I081_Aniruddh_Kulkarni_NLP_Exp8

May 28, 2023

1 Name: Aniruddh Kulkarni

2 Roll no: I081

3 Stream: CS (AI)

4 Division: I

5 Semester: 5th Semester

6 Batch: I-3

7 Subject: NLP

8 Assignment-8

[1]:
```
import nltk
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     /Users/pushpakulkarni/nltk_data…
[nltk_data]   Package stopwords is already up-to-date!
```

[1]: True

[2]:
```
import nltk
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     /Users/pushpakulkarni/nltk_data…
[nltk_data]   Package punkt is already up-to-date!
```

[2]: True

[3]:
```
nltk.download('averaged_perceptron_tagger')
```

```
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     /Users/pushpakulkarni/nltk_data…
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]       date!
```

[3]: True

[4]:
```
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize, sent_tokenize
stop_words = set(stopwords.words('english'))

# Dummy text
txt = "Sukanya, Rajib and Naba are my good friends. " \
        "Sukanya is getting married next year. " \
        "Marriage is a big step in one's life." \
        "It is both exciting and frightening. " \
        "But friendship is a sacred bond between people." \
        "It is a special kind of love between us. " \
        "Many of you must have tried searching for a friend "\
        "but never found the right one."

# sent_tokenize is one of instances of
# PunktSentenceTokenizer from the nltk.tokenize.punkt module

tokenized = sent_tokenize(txt)
for i in tokenized:

        # Word tokenizers is used to find the words
        # and punctuation in a string
        wordsList = nltk.word_tokenize(i)

        # removing stop words from wordList
        wordsList = [w for w in wordsList if not w in stop_words]

        # Using a Tagger. Which is part-of-speech
        # tagger or POS-tagger.
        tagged = nltk.pos_tag(wordsList)

        print(tagged)
```

```
[('Sukanya', 'NNP'), (',', ','), ('Rajib', 'NNP'), ('Naba', 'NNP'), ('good',
'JJ'), ('friends', 'NNS'), ('.', '.')]
[('Sukanya', 'NNP'), ('getting', 'VBG'), ('married', 'VBN'), ('next', 'JJ'),
('year', 'NN'), ('.', '.')]
[('Marriage', 'NN'), ('big', 'JJ'), ('step', 'NN'), ('one', 'CD'), ('’', 'NN'),
('life.It', 'NN'), ('exciting', 'VBG'), ('frightening', 'NN'), ('.', '.')]
[('But', 'CC'), ('friendship', 'NN'), ('sacred', 'VBD'), ('bond', 'NN'),
```

```
('people.It', 'NN'), ('special', 'JJ'), ('kind', 'NN'), ('love', 'VB'), ('us',
'PRP'), ('.', '.')]
[('Many', 'JJ'), ('must', 'MD'), ('tried', 'VB'), ('searching', 'VBG'),
('friend', 'NN'), ('never', 'RB'), ('found', 'VBD'), ('right', 'JJ'), ('one',
'CD'), ('.', '.')]
```

NER using Spacy

[5]:
```
import spacy
from spacy import displacy

NER = spacy.load("en_core_web_sm")
```

[6]:
```
raw_text="The Indian Space Research Organisation or is the national space␣
 ↪agency of India, headquartered in Bengaluru. It operates under Department of␣
 ↪Space which is directly overseen by the Prime Minister of India while␣
 ↪Chairman of ISRO acts as executive of DOS as well."
```

[7]:
```
text1= NER(raw_text)
```

[8]:
```
for word in text1.ents:
    print(word.text,word.label_)
```

```
The Indian Space Research Organisation ORG
India GPE
Bengaluru GPE
Department of Space ORG
India GPE
ISRO ORG
DOS ORG
```

[9]:
```
spacy.explain("ORG")
```

[9]:
```
'Companies, agencies, institutions, etc.'
```

[10]:
```
spacy.explain("GPE")
```

[10]:
```
'Countries, cities, states'
```

[11]:
```
displacy.render(text1,style="ent",jupyter=True)
```

```
<IPython.core.display.HTML object>
```

[12]:

```
text3='The brand new Apple Yas Mall opened this Thursday, February 3, at the␣
 ↪bustling shopping destination in Abu Dhabi. Nearly doubling the size of the␣
 ↪original location that opened in 2015, the store serves as a reimagined␣
 ↪space for customers to browse Apple's latest products and services, receive␣
 ↪the best support from knowledgeable team members, and participate in free␣
 ↪Today at Apple sessions.With the opening of the newly expanded Apple Yas␣
 ↪Mall, our team is ready to welcome even more of Abu Dhabi's incredibly␣
 ↪diverse and innovative community to this beautiful new space,said Deirdre␣
 ↪O'Brien, Apple's senior vice president of Retail + People. "We look forward␣
 ↪to continuing to bring the best of Apple to the UAE, and building on our␣
 ↪history in the region.Situated in a prime corner location at the mall's town␣
 ↪square, the store features a stunning curved glass exterior and integrates␣
 ↪over 150 feet of glass throughout the storefront. Bianco Cristal floors and␣
 ↪wood ceilings are used throughout the space, resembling similar materials␣
 ↪found in other Apple Store locations around the world. Natural light easily␣
 ↪flows inside with two large skylights sitting directly above six Ficus␣
 ↪Nitida trees. Visitors will find the freestanding video wall and Forum␣
 ↪positioned at the center of the store, which is home to free Today at Apple␣
 ↪sessions. Led by Apple Creative Pros, these daily sessions provide creative␣
 ↪inspiration, teach practical skills, and help participants go further with␣
 ↪their products.Customers can discover Apple's products and services at␣
 ↪surrounding tables and avenues; learn more about Apple's Trade In program␣
 ↪across iPhone, iPad, Mac, and Apple Watch; and get shopping support from␣
 ↪Apple Specialists. Apple Yas Mall includes 100 highly trained team members␣
 ↪who collectively speak 33 languages and represent 32 nationalities. The␣
 ↪diverse team has nearly doubled since the store originally opened in 2015,␣
 ↪and more than half remains part of the team that will welcome customers to␣
 ↪the new Apple Yas Mall location.Apple has been operating in the region for␣
 ↪over 10 years, and has more than 600 team members across the UAE. Since␣
 ↪Apple opened its first stores in the UAE in 2015, they've welcomed nearly 30␣
 ↪million visitors. Apple continues its strong commitment to local users by␣
 ↪providing an incredible experience across products, software, and services.␣
 ↪Customers can enjoy software and apps in Arabic, tailor-made Arabic content␣
 ↪across Apple Music, and a localised App Store available in 14 countries␣
 ↪across the Middle East and North Africa.'
```

[13]:
```
text3
```

[13]:
```
'The brand new Apple Yas Mall opened this Thursday, February 3, at the bustling
shopping destination in Abu Dhabi. Nearly doubling the size of the original
location that opened in 2015, the store serves as a reimagined space for
customers to browse Apple's latest products and services, receive the best
support from knowledgeable team members, and participate in free Today at Apple
sessions.With the opening of the newly expanded Apple Yas Mall, our team is
ready to welcome even more of Abu Dhabi's incredibly diverse and innovative
community to this beautiful new space,said Deirdre O'Brien, Apple's senior vice
president of Retail + People. "We look forward to continuing to bring the best
```

of Apple to the UAE, and building on our history in the region.Situated in a prime corner location at the mall's town square, the store features a stunning curved glass exterior and integrates over 150 feet of glass throughout the storefront. Bianco Cristal floors and wood ceilings are used throughout the space, resembling similar materials found in other Apple Store locations around the world. Natural light easily flows inside with two large skylights sitting directly above six Ficus Nitida trees. Visitors will find the freestanding video wall and Forum positioned at the center of the store, which is home to free Today at Apple sessions. Led by Apple Creative Pros, these daily sessions provide creative inspiration, teach practical skills, and help participants go further with their products.Customers can discover Apple's products and services at surrounding tables and avenues; learn more about Apple's Trade In program across iPhone, iPad, Mac, and Apple Watch; and get shopping support from Apple Specialists. Apple Yas Mall includes 100 highly trained team members who collectively speak 33 languages and represent 32 nationalities. The diverse team has nearly doubled since the store originally opened in 2015, and more than half remains part of the team that will welcome customers to the new Apple Yas Mall location.Apple has been operating in the region for over 10 years, and has more than 600 team members across the UAE. Since Apple opened its first stores in the UAE in 2015, they've welcomed nearly 30 million visitors. Apple continues its strong commitment to local users by providing an incredible experience across products, software, and services. Customers can enjoy software and apps in Arabic, tailor-made Arabic content across Apple Music, and a localised App Store available in 14 countries across the Middle East and North Africa.'

[14]:
```
text_1= NER(text3)
```

[15]:
```
for word in text_1.ents:
    print(word.text,word.label_)
```

```
Apple Yas Mall ORG
Thursday DATE
February 3 DATE
Abu Dhabi GPE
2015 DATE
Apple ORG
Today DATE
Apple ORG
Apple Yas Mall ORG
Abu Dhabi's ORG
Deirdre O'Brien PERSON
Apple ORG
Retail + People FAC
Apple ORG
UAE ORG
over 150 feet QUANTITY
Bianco Cristal ORG
```

```
Apple Store ORG
two CARDINAL
six CARDINAL
Ficus Nitida ORG
Today DATE
Apple ORG
Apple Creative Pros ORG
daily DATE
Apple ORG
Apple's Trade ORG
iPhone ORG
iPad ORG
Mac PERSON
Apple Watch ORG
Apple Specialists ORG
Apple Yas Mall PERSON
100 CARDINAL
33 CARDINAL
32 CARDINAL
2015 DATE
more than half CARDINAL
Apple Yas Mall PERSON
Apple ORG
over 10 years DATE
more than 600 CARDINAL
UAE ORG
Apple ORG
first ORDINAL
UAE ORG
2015 DATE
nearly 30 million CARDINAL
Apple ORG
Arabic LANGUAGE
Arabic NORP
Apple Music ORG
App Store PERSON
14 CARDINAL
the Middle East LOC
North Africa GPE
```

[16]:
```
displacy.render(text_1,style="ent",jupyter=True)
```

```
<IPython.core.display.HTML object>
```

[ ]:
```

```

[17]:
```
from nltk.tag import StanfordNERTagger
from nltk.tokenize import word_tokenize
import os
```

[18]:
```
from nltk.tag.stanford import StanfordNERTagger
from nltk.tokenize import word_tokenize
import nltk

%pip install wget
!wget 'https://nlp.stanford.edu/software/stanford-ner-2018-10-16.zip'
!unzip stanford-ner-2018-10-16.zip

nltk.download('punkt')

st = StanfordNERTagger('stanford-ner-2018-10-16/classifiers/english.all.3class.
 ↪distsim.crf.ser.gz',
                       'stanford-ner-2018-10-16/stanford-ner.jar',
                       encoding='utf-8')
```

```
Requirement already satisfied: wget in
/Users/pushpakulkarni/miniconda3/envs/tensorflow/lib/python3.10/site-packages
(3.2)
Note: you may need to restart the kernel to use updated packages.
zsh:1: command not found: wget
unzip:  cannot find or open stanford-ner-2018-10-16.zip, stanford-
ner-2018-10-16.zip.zip or stanford-ner-2018-10-16.zip.ZIP.
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     /Users/pushpakulkarni/nltk_data…
[nltk_data]   Package punkt is already up-to-date!
```

[19]:
```
text2= 'On day 20 of Russia invasion of Ukraine, residents in the capital Kyiv␣
 ↪were placed under a 35-hour curfew - but that did not stop the prime␣
 ↪ministers of Poland, Slovenia and the Czech Republic from travelling there␣
 ↪by train.The trip was a Polish idea, after the EU warned of potential␣
 ↪security risks.The leaders decided to go by train because flying by Polish␣
 ↪military jet could have been viewed by Russia as dangerously provocative,␣
 ↪BBC Europe editor Katya Adler reported. It was not immediately clear when␣
 ↪their train would make the return trip to Warsaw.Polands Mateusz Morawiecki␣
 ↪said history was being made in Ukraine.'
text2
```

[19]:
```
'On day 20 of Russia invasion of Ukraine, residents in the capital Kyiv were
placed under a 35-hour curfew - but that did not stop the prime ministers of
Poland, Slovenia and the Czech Republic from travelling there by train.The trip
was a Polish idea, after the EU warned of potential security risks.The leaders
decided to go by train because flying by Polish military jet could have been
```

```
viewed by Russia as dangerously provocative, BBC Europe editor Katya Adler
reported. It was not immediately clear when their train would make the return
trip to Warsaw.Polands Mateusz Morawiecki said history was being made in
Ukraine.'
```

[20]:
```
import pandas as pd
```

[22]:
```
tokenized_text = nltk.word_tokenize(raw_text)
classified_text = st.tag(tokenized_text)

classified_text_df = pd.DataFrame(classified_text)

classified_text_df.drop_duplicates(keep='first', inplace=True)
classified_text_df.reset_index(drop=True, inplace=True)
classified_text_df.columns = ["Entities", "Labels"]
classified_text_df
```

[22]:

	Entities	Labels
0	The	0
1	Indian	ORGANIZATION
2	Space	ORGANIZATION
3	Research	ORGANIZATION
4	Organisation	ORGANIZATION
5	or	0
6	is	0
7	the	0
8	national	0
9	space	0
10	agency	0
11	of	0
12	India	LOCATION
13	,	0
14	headquartered	0
15	in	0
16	Bengaluru	LOCATION
17	.	0
18	It	0
19	operates	0
20	under	0
21	Department	ORGANIZATION
22	of	ORGANIZATION
23	which	0
24	directly	0
25	overseen	0
26	by	0
27	Prime	0
28	Minister	0

```
29          while              O
30       Chairman              O
31           ISRO  ORGANIZATION
32           acts              O
33             as              O
34      executive              O
35            DOS              O
36           well              O
```

9 Exercises

10 Create a corpus on your own(paragraph 15-20 lines). Perform POS tagging and NER

11 Perform NER using NLTK and Stanford NLP

[24]:
```
new = "World War II or the Second World War, often abbreviated as WWII or WW2,␣
 ↪was a global war that lasted from 1939 to 1945."\
"It involved the vast majority of the world's countries-including all of the␣
 ↪great powers-forming two opposing military alliances: the Allies and the␣
 ↪Axis powers." \
"World War II was a total war that directly involved more than 100 million␣
 ↪personnel from more than 30 countries."\
"The major participants in the war threw their entire economic, industrial, and␣
 ↪scientific capabilities behind the war effort, blurring the distinction␣
 ↪between civilian and military resources." \
"Aircraft played a major role in the conflict, enabling the strategic bombing␣
 ↪of population centres and deploying the only two nuclear weapons ever used␣
 ↪in war." \
"World War II was by far the deadliest conflict in human history; it resulted␣
 ↪in 70 to 85 million fatalities, mostly among civilians."\
"Tens of millions died due to genocides (including the Holocaust), starvation,␣
 ↪massacres, and disease. In the wake of the Axis defeat, Germany and Japan␣
 ↪were occupied, and war crimes tribunals were conducted against German and␣
 ↪Japanese leaders."\
"The causes of World War II are debated, but contributing factors included the␣
 ↪Second Italo-Ethiopian War, the Spanish Civil War, the Second Sino-Japanese␣
 ↪War, the Soviet-Japanese border conflicts, the rise of fascism in Europe and␣
 ↪rising European tensions since World War I."\
"World War II is generally considered to have begun on 1 September 1939, when␣
 ↪Nazi Germany, under Adolf Hitler, invaded Poland. The United Kingdom and␣
 ↪France subsequently declared war on Germany on 3 September."\
"Under the Molotov-Ribbentrop Pact of August 1939, Germany and the Soviet Union␣
 ↪had partitioned Poland and marked out their spheres of influence across␣
 ↪Finland, Estonia, Latvia, Lithuania and Romania."\
```

```
"From late 1939 to early 1941, in a series of campaigns and treaties, Germany␣
 ↪conquered or controlled much of continental Europe, and formed the Axis␣
 ↪alliance with Italy and Japan (with other countries later)." \
"Following the onset of campaigns in North Africa and East Africa, and the fall␣
 ↪of France in mid-1940, the war continued primarily between the European Axis␣
 ↪powers and the British Empire, with war in the Balkans, the aerial Battle of␣
 ↪Britain, the Blitz of the United Kingdom, and the Battle of the Atlantic."
"On 22 June 1941, Germany led the European Axis powers in an invasion of the␣
 ↪Soviet Union, opening the Eastern Front, the largest land theatre of war in␣
 ↪history."
```

[24]:
```
'On 22 June 1941, Germany led the European Axis powers in an invasion of the
Soviet Union, opening the Eastern Front, the largest land theatre of war in
history.'
```

[25]:
```
tokenizedi = sent_tokenize(new)
for i in tokenizedi:

        # Word tokenizers is used to find the words
        # and punctuation in a string
        wordsList = nltk.word_tokenize(i)

        # removing stop words from wordList
        wordsList = [w for w in wordsList if not w in stop_words]

        # Using a Tagger. Which is part-of-speech
        # tagger or POS-tagger.
        tagged = nltk.pos_tag(wordsList)

        print(tagged)
```

```
[('World', 'NNP'), ('War', 'NNP'), ('II', 'NNP'), ('Second', 'NNP'), ('World',
'NNP'), ('War', 'NNP'), (',', ','), ('often', 'RB'), ('abbreviated', 'VBN'),
('WWII', 'NNP'), ('WW2', 'NNP'), (',', ','), ('global', 'JJ'), ('war', 'NN'),
('lasted', 'VBD'), ('1939', 'CD'), ('1945.It', 'CD'), ('involved', 'VBN'),
('vast', 'JJ'), ('majority', 'NN'), ('world', 'NN'), ("'s", 'POS'),
('countries-including', 'VBG'), ('great', 'JJ'), ('powers-forming', 'VBG'),
('two', 'CD'), ('opposing', 'VBG'), ('military', 'JJ'), ('alliances', 'NNS'),
(':', ':'), ('Allies', 'NNS'), ('Axis', 'NNP'), ('powers.World', 'NNP'), ('War',
'NNP'), ('II', 'NNP'), ('total', 'JJ'), ('war', 'NN'), ('directly', 'RB'),
('involved', 'VBD'), ('100', 'CD'), ('million', 'CD'), ('personnel', 'NNS'),
('30', 'CD'), ('countries.The', 'NNS'), ('major', 'JJ'), ('participants',
'NNS'), ('war', 'NN'), ('threw', 'VBD'), ('entire', 'JJ'), ('economic', 'JJ'),
(',', ','), ('industrial', 'JJ'), (',', ','), ('scientific', 'JJ'),
('capabilities', 'NNS'), ('behind', 'IN'), ('war', 'NN'), ('effort', 'NN'),
(',', ','), ('blurring', 'VBG'), ('distinction', 'NN'), ('civilian', 'JJ'),
('military', 'JJ'), ('resources.Aircraft', 'NN'), ('played', 'VBD'), ('major',
```

'JJ'), ('role', 'NN'), ('conflict', 'NN'), (',', ','), ('enabling', 'VBG'),
('strategic', 'JJ'), ('bombing', 'VBG'), ('population', 'NN'), ('centres',
'NNS'), ('deploying', 'VBG'), ('two', 'CD'), ('nuclear', 'JJ'), ('weapons',
'NNS'), ('ever', 'RB'), ('used', 'VBD'), ('war.World', 'NNP'), ('War', 'NNP'),
('II', 'NNP'), ('far', 'RB'), ('deadliest', 'JJS'), ('conflict', 'JJ'),
('human', 'JJ'), ('history', 'NN'), (';', ':'), ('resulted', 'VBD'), ('70',
'CD'), ('85', 'CD'), ('million', 'CD'), ('fatalities', 'NNS'), (',', ','),
('mostly', 'RB'), ('among', 'IN'), ('civilians.Tens', 'NNS'), ('millions',
'NNS'), ('died', 'VBD'), ('due', 'JJ'), ('genocides', 'NNS'), ('(', '('),
('including', 'VBG'), ('Holocaust', 'NNP'), (')', ')'), (',', ','),
('starvation', 'NN'), (',', ','), ('massacres', 'NNS'), (',', ','), ('disease',
'NN'), ('.', '.')]
[('In', 'IN'), ('wake', 'NN'), ('Axis', 'NNP'), ('defeat', 'NN'), (',', ','),
('Germany', 'NNP'), ('Japan', 'NNP'), ('occupied', 'VBD'), (',', ','), ('war',
'NN'), ('crimes', 'NNS'), ('tribunals', 'NNS'), ('conducted', 'JJ'), ('German',
'JJ'), ('Japanese', 'JJ'), ('leaders.The', 'NN'), ('causes', 'NNS'), ('World',
'NNP'), ('War', 'NNP'), ('II', 'NNP'), ('debated', 'VBD'), (',', ','),
('contributing', 'VBG'), ('factors', 'NNS'), ('included', 'VBD'), ('Second',
'JJ'), ('Italo-Ethiopian', 'JJ'), ('War', 'NNP'), (',', ','), ('Spanish',
'NNP'), ('Civil', 'NNP'), ('War', 'NNP'), (',', ','), ('Second', 'NNP'), ('Sino-
Japanese', 'NNP'), ('War', 'NNP'), (',', ','), ('Soviet-Japanese', 'JJ'),
('border', 'NN'), ('conflicts', 'NNS'), (',', ','), ('rise', 'NN'), ('fascism',
'NN'), ('Europe', 'NNP'), ('rising', 'VBG'), ('European', 'JJ'), ('tensions',
'NNS'), ('since', 'IN'), ('World', 'NNP'), ('War', 'NNP'), ('I.World', 'NNP'),
('War', 'NNP'), ('II', 'NNP'), ('generally', 'RB'), ('considered', 'VBD'),
('begun', 'VBN'), ('1', 'CD'), ('September', 'NNP'), ('1939', 'CD'), (',', ','),
('Nazi', 'NNP'), ('Germany', 'NNP'), (',', ','), ('Adolf', 'NNP'), ('Hitler',
'NNP'), (',', ','), ('invaded', 'VBD'), ('Poland', 'NNP'), ('.', '.')]
[('The', 'DT'), ('United', 'NNP'), ('Kingdom', 'NNP'), ('France', 'NNP'),
('subsequently', 'RB'), ('declared', 'VBD'), ('war', 'NN'), ('Germany', 'NNP'),
('3', 'CD'), ('September.Under', 'NNP'), ('Molotov-Ribbentrop', 'NNP'), ('Pact',
'NNP'), ('August', 'NNP'), ('1939', 'CD'), (',', ','), ('Germany', 'NNP'),
('Soviet', 'NNP'), ('Union', 'NNP'), ('partitioned', 'VBD'), ('Poland', 'NNP'),
('marked', 'VBD'), ('spheres', 'NNS'), ('influence', 'NN'), ('across', 'IN'),
('Finland', 'NNP'), (',', ','), ('Estonia', 'NNP'), (',', ','), ('Latvia',
'NNP'), (',', ','), ('Lithuania', 'NNP'), ('Romania.From', 'NNP'), ('late',
'JJ'), ('1939', 'CD'), ('early', 'JJ'), ('1941', 'CD'), (',', ','), ('series',
'NN'), ('campaigns', 'NNS'), ('treaties', 'NNS'), (',', ','), ('Germany',
'NNP'), ('conquered', 'VBD'), ('controlled', 'VBN'), ('much', 'JJ'),
('continental', 'NN'), ('Europe', 'NNP'), (',', ','), ('formed', 'VBD'),
('Axis', 'NNP'), ('alliance', 'NN'), ('Italy', 'NNP'), ('Japan', 'NNP'), ('(',
'('), ('countries', 'NNS'), ('later', 'RB'), (')', ')'), ('.Following', 'VBG'),
('onset', 'JJ'), ('campaigns', 'NNS'), ('North', 'NNP'), ('Africa', 'NNP'),
('East', 'NNP'), ('Africa', 'NNP'), (',', ','), ('fall', 'NN'), ('France',
'NNP'), ('mid-1940', 'NN'), (',', ','), ('war', 'NN'), ('continued', 'VBD'),
('primarily', 'RB'), ('European', 'JJ'), ('Axis', 'NNP'), ('powers', 'NNS'),
('British', 'NNP'), ('Empire', 'NNP'), (',', ','), ('war', 'NN'), ('Balkans',
'NNP'), (',', ','), ('aerial', 'JJ'), ('Battle', 'NNP'), ('Britain', 'NNP'),

```
(',', ','), ('Blitz', 'NNP'), ('United', 'NNP'), ('Kingdom', 'NNP'), (',', ','),
('Battle', 'NNP'), ('Atlantic', 'NNP'), ('.', '.')]
```

[26]:
```
newtext=NER(new)
```

[27]:
```
for word in newtext.ents:
    print(word.text,word.label_)
```

```
World War II EVENT
the Second World War EVENT
WWII EVENT
from 1939 DATE
two CARDINAL
World War II EVENT
more than 100 million CARDINAL
more than 30 CARDINAL
only two CARDINAL
World War II EVENT
70 to 85 million MONEY
Tens of millions CARDINAL
Holocaust EVENT
Germany GPE
Japan GPE
German NORP
Japanese NORP
World War II EVENT
Second ORDINAL
the Spanish Civil War EVENT
Second ORDINAL
Sino-Japanese NORP
Soviet NORP
Japanese NORP
Europe LOC
European NORP
World War I.World War II EVENT
1 September 1939 DATE
Nazi Germany GPE
Adolf Hitler PERSON
Poland GPE
The United Kingdom GPE
France GPE
Germany GPE
3 September DATE
August 1939 DATE
Germany GPE
the Soviet Union GPE
Poland GPE
Finland GPE
```

```
Estonia GPE
Latvia GPE
Lithuania GPE
Romania GPE
late 1939 to early 1941 DATE
Germany GPE
Europe LOC
Italy GPE
Japan GPE
North Africa GPE
East Africa GPE
France GPE
mid-1940 DATE
European NORP
the British Empire GPE
Balkans LOC
Battle of Britain GPE
the Blitz of the United Kingdom ORG
the Battle of the Atlantic LOC
```

[28]:
```
spacy.explain("NORP")
```

[28]:
```
'Nationalities or religious or political groups'
```

[29]:
```
spacy.explain("CARDINAL")
```

[29]:
```
'Numerals that do not fall under another type'
```

[30]:
```
spacy.explain("WORK_OF_ART")
```

[30]:
```
'Titles of books, songs, etc.'
```

[31]:
```
displacy.render(newtext,style="ent",jupyter=True)
```

```
<IPython.core.display.HTML object>
```

[32]:
```
tokenized_text = nltk.word_tokenize(new)
classified_text = st.tag(tokenized_text)

classified_text_df = pd.DataFrame(classified_text)

classified_text_df.drop_duplicates(keep='first', inplace=True)
classified_text_df.reset_index(drop=True, inplace=True)
classified_text_df.columns = ["Entities", "Labels"]
pd.set_option('display.max_rows', None)
classified_text_df
```

[32]:

	Entities	Labels
0	World	O
1	War	O
2	II	O
3	or	O
4	the	O
5	Second	O
6	,	O
7	often	O
8	abbreviated	O
9	as	O
10	WWII	O
11	WW2	O
12	was	O
13	a	O
14	global	O
15	war	O
16	that	O
17	lasted	O
18	from	O
19	1939	O
20	to	O
21	1945.It	O
22	involved	O
23	vast	O
24	majority	O
25	of	O
26	world	O
27	's	O
28	countries-including	O
29	all	O
30	great	O
31	powers-forming	O
32	two	O
33	opposing	O
34	military	O
35	alliances	O
36	:	O
37	Allies	O
38	and	O
39	Axis	O
40	powers.World	O
41	total	O
42	directly	O
43	more	O
44	than	O
45	100	O

46	million	O
47	personnel	O
48	30	O
49	countries.The	O
50	major	O
51	participants	O
52	in	O
53	threw	O
54	their	O
55	entire	O
56	economic	O
57	industrial	O
58	scientific	O
59	capabilities	O
60	behind	O
61	effort	O
62	blurring	O
63	distinction	O
64	between	O
65	civilian	O
66	resources.Aircraft	O
67	played	O
68	role	O
69	conflict	O
70	enabling	O
71	strategic	O
72	bombing	O
73	population	O
74	centres	O
75	deploying	O
76	only	O
77	nuclear	O
78	weapons	O
79	ever	O
80	used	O
81	war.World	O
82	by	O
83	far	O
84	deadliest	O
85	human	O
86	history	O
87	;	O
88	it	O
89	resulted	O
90	70	O
91	85	O
92	fatalities	O

93	mostly	O
94	among	O
95	civilians.Tens	O
96	millions	O
97	died	O
98	due	O
99	genocides	O
100	(	O
101	including	O
102	Holocaust	O
103	)	O
104	starvation	O
105	massacres	O
106	disease	O
107	.	O
108	In	O
109	wake	O
110	defeat	O
111	Germany	LOCATION
112	Japan	LOCATION
113	were	O
114	occupied	O
115	crimes	O
116	tribunals	O
117	conducted	O
118	against	O
119	German	O
120	Japanese	O
121	leaders.The	O
122	causes	O
123	are	O
124	debated	O
125	but	O
126	contributing	O
127	factors	O
128	included	O
129	Italo-Ethiopian	O
130	Spanish	O
131	Civil	O
132	Sino-Japanese	O
133	Soviet-Japanese	O
134	border	O
135	conflicts	O
136	rise	O
137	fascism	O
138	Europe	LOCATION
139	rising	O

140	European	O
141	tensions	O
142	since	O
143	I.World	O
144	is	O
145	generally	O
146	considered	O
147	have	O
148	begun	O
149	on	O
150	1	O
151	September	O
152	when	O
153	Nazi	ORGANIZATION
154	under	O
155	Adolf	PERSON
156	Hitler	PERSON
157	invaded	O
158	Poland	LOCATION
159	The	O
160	United	LOCATION
161	Kingdom	LOCATION
162	France	LOCATION
163	subsequently	O
164	declared	O
165	3	O
166	September.Under	O
167	Molotov-Ribbentrop	O
168	Pact	O
169	August	O
170	Soviet	LOCATION
171	Union	LOCATION
172	had	O
173	partitioned	O
174	marked	O
175	out	O
176	spheres	O
177	influence	O
178	across	O
179	Finland	LOCATION
180	Estonia	LOCATION
181	Latvia	LOCATION
182	Lithuania	LOCATION
183	Romania.From	O
184	late	O
185	early	O
186	1941	O

187	series	O
188	campaigns	O
189	treaties	O
190	conquered	O
191	controlled	O
192	much	O
193	continental	O
194	formed	O
195	alliance	O
196	with	O
197	Italy	LOCATION
198	other	O
199	countries	O
200	later	O
201	.Following	O
202	onset	O
203	North	LOCATION
204	Africa	LOCATION
205	East	LOCATION
206	fall	O
207	mid-1940	O
208	continued	O
209	primarily	O
210	powers	O
211	British	O
212	Empire	O
213	Balkans	LOCATION
214	aerial	O
215	Battle	O
216	Britain	LOCATION
217	Blitz	O
218	Atlantic	LOCATION

[33]:

```
meds = "This study aimed to compare the outcomes of patients who underwent␣
 ↪laparoscopic and open resections for colorectal cancer. Comparison of␣
 ↪colectomy in 2 consecutive periods (period 1: January 1996-May 2000; period␣
 ↪2: June 2000-December 2004), with laparoscopic surgery being a surgical␣
 ↪option in period 2, was also performed. During period 2, the operative␣
 ↪mortality rates of patients with laparoscopic (n = 401) and open resection␣
 ↪(n = 255) were 0.8% and 3.7%, respectively (P = 0.022), and the morbidity␣
 ↪rates were 21.7% and 15.7%, respectively (P = 0.068). The patients who␣
 ↪underwent laparoscopic resection had significantly earlier return of bowel␣
 ↪function, earlier resumption of diet, and shorter hospital stay. The 3-year␣
 ↪overall survivals in those with nondisseminated disease were 74.4% and 78.8%␣
 ↪for open and laparoscopic resection, respectively (P = 0.046). The operative␣
 ↪morality rates were 4.4% and 2.6% in period 1 and period 2, respectively (P␣
 ↪= 0.132). The 3-year overall survivals for patients with nondisseminated␣
 ↪disease were 69.7% and 76.1% for period 1 and period 2, respectively (P = 0.
 ↪019). The overall survivals in patients who underwent open resection in the␣
 ↪2 periods were similar (P = 0.284). Preoperative workup included blood␣
 ↪tests, chest x-rays, and serum carcinoembryonic antigen. CT scan was not a␣
 ↪routine and depended on the availability of the test, especially in the␣
 ↪early part of the study. During the latter part, more patients had␣
 ↪preoperative CT scan. The surgical approach was decided with the consent of␣
 ↪the patients, after thorough discussion on the pros and cons of the approach.
 ↪ The decision also depended on the availability of operating time and␣
 ↪laparoscopic surgeons. Patients with large, fixed tumors with invasion to␣
 ↪other organs were advised against laparoscopic resection. The patient␣
 ↪received mechanical bowel preparation with polyethylene glycol electrolytes␣
 ↪solution the day before surgery and prophylactic intravenous antibiotics␣
 ↪were administered at the induction of anesthesia. A urinary catheter was␣
 ↪inserted after the patient was put under general anesthesia. Nasogastric␣
 ↪tube was not used as a routine. Open resections were performed through a␣
 ↪midline incision. The extent of resection was determined by the site of the␣
 ↪tumor and the method of anastomosis was decided by the surgeon. In surgery␣
 ↪for upper rectal cancer, the rectum was mobilized by sharp perimesorectal␣
 ↪dissection to keep the visceral pelvic fascia, which enveloped the␣
 ↪mesorectum, intact. Total mesorectal excision was not performed for upper␣
 ↪rectal cancer. Instead, the rectum and mesorectum was transected 4 to 5 cm␣
 ↪below the distal extent of the tumor."
```

[34]:
```
tokenized = sent_tokenize(meds)
for i in tokenized:

        # Word tokenizers is used to find the words
        # and punctuation in a string
        wordsList = nltk.word_tokenize(i)

        # removing stop words from wordList
        wordsList = [w for w in wordsList if not w in stop_words]
```

```
        # Using a Tagger. Which is part-of-speech
        # tagger or POS-tagger.
        taggedii = nltk.pos_tag(wordsList)

        print(taggedii)
```

```
[('This', 'DT'), ('study', 'NN'), ('aimed', 'VBD'), ('compare', 'JJ'),
('outcomes', 'NNS'), ('patients', 'NNS'), ('underwent', 'JJ'), ('laparoscopic',
'JJ'), ('open', 'JJ'), ('resections', 'NNS'), ('colorectal', 'JJ'), ('cancer',
'NN'), ('.', '.')]
[('Comparison', 'NNP'), ('colectomy', 'VBD'), ('2', 'CD'), ('consecutive',
'JJ'), ('periods', 'NNS'), ('(', '('), ('period', 'NN'), ('1', 'CD'), (':',
':'), ('January', 'NNP'), ('1996-May', 'CD'), ('2000', 'CD'), (';', ':'),
('period', 'NN'), ('2', 'CD'), (':', ':'), ('June', 'NNP'), ('2000-December',
'CD'), ('2004', 'CD'), (')', ')'), (',', ','), ('laparoscopic', 'JJ'),
('surgery', 'NN'), ('surgical', 'JJ'), ('option', 'NN'), ('period', 'NN'), ('2',
'CD'), (',', ','), ('also', 'RB'), ('performed', 'VBN'), ('.', '.')]
[('During', 'IN'), ('period', 'NN'), ('2', 'CD'), (',', ','), ('operative',
'JJ'), ('mortality', 'NN'), ('rates', 'NNS'), ('patients', 'NNS'),
('laparoscopic', 'VBP'), ('(', '('), ('n', 'JJ'), ('=', 'NNP'), ('401', 'CD'),
(')', ')'), ('open', 'JJ'), ('resection', 'NN'), ('(', '('), ('n', 'JJ'), ('=',
'NNP'), ('255', 'CD'), (')', ')'), ('0.8', 'CD'), ('%', 'NN'), ('3.7', 'CD'),
('%', 'NN'), (',', ','), ('respectively', 'RB'), ('(', '('), ('P', 'NNP'), ('=',
'NNP'), ('0.022', 'CD'), (')', ')'), (',', ','), ('morbidity', 'NN'), ('rates',
'NNS'), ('21.7', 'CD'), ('%', 'NN'), ('15.7', 'CD'), ('%', 'NN'), (',', ','),
('respectively', 'RB'), ('(', '('), ('P', 'NNP'), ('=', 'NNP'), ('0.068', 'CD'),
(')', ')'), ('.', '.')]
[('The', 'DT'), ('patients', 'NNS'), ('underwent', 'JJ'), ('laparoscopic',
'JJ'), ('resection', 'NN'), ('significantly', 'RB'), ('earlier', 'RBR'),
('return', 'JJ'), ('bowel', 'NN'), ('function', 'NN'), (',', ','), ('earlier',
'JJR'), ('resumption', 'NN'), ('diet', 'NN'), (',', ','), ('shorter', 'JJR'),
('hospital', 'NN'), ('stay', 'NN'), ('.', '.')]
[('The', 'DT'), ('3-year', 'JJ'), ('overall', 'JJ'), ('survivals', 'NNS'),
('nondisseminated', 'VBD'), ('disease', 'JJ'), ('74.4', 'CD'), ('%', 'NN'),
('78.8', 'CD'), ('%', 'NN'), ('open', 'JJ'), ('laparoscopic', 'NN'),
('resection', 'NN'), (',', ','), ('respectively', 'RB'), ('(', '('), ('P',
'NNP'), ('=', 'NNP'), ('0.046', 'CD'), (')', ')'), ('.', '.')]
[('The', 'DT'), ('operative', 'JJ'), ('morality', 'NN'), ('rates', 'NNS'),
('4.4', 'CD'), ('%', 'NN'), ('2.6', 'CD'), ('%', 'NN'), ('period', 'NN'), ('1',
'CD'), ('period', 'NN'), ('2', 'CD'), (',', ','), ('respectively', 'RB'), ('(',
'('), ('P', 'NNP'), ('=', 'NNP'), ('0.132', 'CD'), (')', ')'), ('.', '.')]
[('The', 'DT'), ('3-year', 'JJ'), ('overall', 'JJ'), ('survivals', 'NNS'),
('patients', 'NNS'), ('nondisseminated', 'JJ'), ('disease', 'JJ'), ('69.7',
'CD'), ('%', 'NN'), ('76.1', 'CD'), ('%', 'NN'), ('period', 'NN'), ('1', 'CD'),
('period', 'NN'), ('2', 'CD'), (',', ','), ('respectively', 'RB'), ('(', '('),
('P', 'NNP'), ('=', 'NNP'), ('0.019', 'CD'), (')', ')'), ('.', '.')]
```

```
[('The', 'DT'), ('overall', 'JJ'), ('survivals', 'NNS'), ('patients', 'NNS'),
('underwent', 'JJ'), ('open', 'JJ'), ('resection', 'NN'), ('2', 'CD'),
('periods', 'NNS'), ('similar', 'JJ'), ('(', '('), ('P', 'NNP'), ('=', 'NNP'),
('0.284', 'CD'), (')', ')'), ('.', '.')]
[('Preoperative', 'NNP'), ('workup', 'NN'), ('included', 'VBD'), ('blood',
'NN'), ('tests', 'NNS'), (',', ','), ('chest', 'JJS'), ('x-rays', 'NNS'), (',',
','), ('serum', 'NN'), ('carcinoembryonic', 'JJ'), ('antigen', 'NN'), ('.',
'.')]
[('CT', 'NNP'), ('scan', 'JJ'), ('routine', 'NN'), ('depended', 'VBD'),
('availability', 'NN'), ('test', 'NN'), (',', ','), ('especially', 'RB'),
('early', 'JJ'), ('part', 'NN'), ('study', 'NN'), ('.', '.')]
[('During', 'IN'), ('latter', 'JJ'), ('part', 'NN'), (',', ','), ('patients',
'NNS'), ('preoperative', 'VBP'), ('CT', 'NNP'), ('scan', 'NN'), ('.', '.')]
[('The', 'DT'), ('surgical', 'JJ'), ('approach', 'NN'), ('decided', 'VBD'),
('consent', 'NN'), ('patients', 'NNS'), (',', ','), ('thorough', 'JJ'),
('discussion', 'NN'), ('pros', 'NNS'), ('cons', 'NNS'), ('approach', 'VBP'),
('.', '.')]
[('The', 'DT'), ('decision', 'NN'), ('also', 'RB'), ('depended', 'VBD'),
('availability', 'NN'), ('operating', 'NN'), ('time', 'NN'), ('laparoscopic',
'JJ'), ('surgeons', 'NNS'), ('.', '.')]
[('Patients', 'NNS'), ('large', 'JJ'), (',', ','), ('fixed', 'JJ'), ('tumors',
'NNS'), ('invasion', 'VBP'), ('organs', 'NNS'), ('advised', 'VBD'),
('laparoscopic', 'JJ'), ('resection', 'NN'), ('.', '.')]
[('The', 'DT'), ('patient', 'NN'), ('received', 'VBD'), ('mechanical', 'JJ'),
('bowel', 'NN'), ('preparation', 'NN'), ('polyethylene', 'NN'), ('glycol',
'NN'), ('electrolytes', 'VBZ'), ('solution', 'JJ'), ('day', 'NN'), ('surgery',
'NN'), ('prophylactic', 'JJ'), ('intravenous', 'JJ'), ('antibiotics', 'NNS'),
('administered', 'VBN'), ('induction', 'NN'), ('anesthesia', 'NN'), ('.', '.')]
[('A', 'DT'), ('urinary', 'JJ'), ('catheter', 'NN'), ('inserted', 'VBN'),
('patient', 'NN'), ('put', 'VBD'), ('general', 'JJ'), ('anesthesia', 'NN'),
('.', '.')]
[('Nasogastric', 'NNP'), ('tube', 'NN'), ('used', 'VBN'), ('routine', 'NN'),
('.', '.')]
[('Open', 'JJ'), ('resections', 'NNS'), ('performed', 'VBD'), ('midline', 'JJ'),
('incision', 'NN'), ('.', '.')]
[('The', 'DT'), ('extent', 'NN'), ('resection', 'NN'), ('determined', 'VBD'),
('site', 'NN'), ('tumor', 'NN'), ('method', 'NN'), ('anastomosis', 'NN'),
('decided', 'VBD'), ('surgeon', 'NN'), ('.', '.')]
[('In', 'IN'), ('surgery', 'NN'), ('upper', 'JJ'), ('rectal', 'NN'), ('cancer',
'NN'), (',', ','), ('rectum', 'VB'), ('mobilized', 'VBN'), ('sharp', 'JJ'),
('perimesorectal', 'JJ'), ('dissection', 'NN'), ('keep', 'VB'), ('visceral',
'JJ'), ('pelvic', 'JJ'), ('fascia', 'NN'), (',', ','), ('enveloped', 'VBD'),
('mesorectum', 'NN'), (',', ','), ('intact', 'JJ'), ('.', '.')]
[('Total', 'JJ'), ('mesorectal', 'JJ'), ('excision', 'NN'), ('performed',
'VBD'), ('upper', 'JJ'), ('rectal', 'JJ'), ('cancer', 'NN'), ('.', '.')]
[('Instead', 'RB'), (',', ','), ('rectum', 'JJ'), ('mesorectum', 'NN'),
('transected', 'VBD'), ('4', 'CD'), ('5', 'CD'), ('cm', 'NN'), ('distal', 'JJ'),
('extent', 'NN'), ('tumor', 'NN'), ('.', '.')]
```

[35]:
```
text3= NER(meds)                          #NER-Spacy
for word in text3.ents:
    print(word.text,word.label_)
```

```
2 CARDINAL
January 1996 DATE
May 2000 DATE
2 CARDINAL
June 2000 DATE
December 2004 DATE
2 CARDINAL
2 CARDINAL
401 CARDINAL
255 CARDINAL
0.8% PERCENT
3.7% PERCENT
0.022 CARDINAL
21.7% PERCENT
15.7% PERCENT
0.068 CARDINAL
3-year DATE
74.4% PERCENT
78.8% PERCENT
0.046 CARDINAL
4.4% PERCENT
2.6% PERCENT
1 CARDINAL
2 CARDINAL
0.132 CARDINAL
3-year DATE
69.7% PERCENT
76.1% PERCENT
1 CARDINAL
2 CARDINAL
0.019 CARDINAL
2 CARDINAL
0.284 CARDINAL
CT ORG
the day DATE
anesthesia GPE
anesthesia GPE
4 CARDINAL
5 cm QUANTITY
```

[36]:
```
displacy.render(text3,style="ent",jupyter=True)
```

```
<IPython.core.display.HTML object>
```

[37]:
```
spacy.explain("GPE")
```

[37]:
```
'Countries, cities, states'
```

[38]:
```
tokenized_text = nltk.word_tokenize(meds)          #NER-Stanford nltk
classified_text = st.tag(tokenized_text)

classified_text_df = pd.DataFrame(classified_text)

classified_text_df.drop_duplicates(keep='first', inplace=True)
classified_text_df.reset_index(drop=True, inplace=True)
classified_text_df.columns = ["Entities", "Labels"]
pd.set_option('display.max_rows', None)
classified_text_df
```

[38]:

	Entities	Labels
0	This	O
1	study	O
2	aimed	O
3	to	O
4	compare	O
5	the	O
6	outcomes	O
7	of	O
8	patients	O
9	who	O
10	underwent	O
11	laparoscopic	O
12	and	O
13	open	O
14	resections	O
15	for	O
16	colorectal	O
17	cancer	O
18	.	O
19	Comparison	O
20	colectomy	O
21	in	O
22	2	O
23	consecutive	O
24	periods	O
25	(	O
26	period	O
27	1	O
28	:	O
29	January	O
30	1996-May	O

31	2000	O
32	;	O
33	June	O
34	2000-December	O
35	2004	O
36	)	O
37	,	O
38	with	O
39	surgery	O
40	being	O
41	a	O
42	surgical	O
43	option	O
44	was	O
45	also	O
46	performed	O
47	During	O
48	operative	O
49	mortality	O
50	rates	O
51	n	O
52	=	O
53	401	O
54	resection	O
55	255	O
56	were	O
57	0.8	O
58	%	O
59	3.7	O
60	respectively	O
61	P	O
62	0.022	O
63	morbidity	O
64	21.7	O
65	15.7	O
66	0.068	O
67	The	O
68	had	O
69	significantly	O
70	earlier	O
71	return	O
72	bowel	O
73	function	O
74	resumption	O
75	diet	O
76	shorter	O
77	hospital	O

78	stay	O
79	3-year	O
80	overall	O
81	survivals	O
82	those	O
83	nondisseminated	O
84	disease	O
85	74.4	O
86	78.8	O
87	0.046	O
88	morality	O
89	4.4	O
90	2.6	O
91	0.132	O
92	69.7	O
93	76.1	O
94	0.019	O
95	similar	O
96	0.284	O
97	Preoperative	O
98	workup	O
99	included	O
100	blood	O
101	tests	O
102	chest	O
103	x-rays	O
104	serum	O
105	carcinoembryonic	O
106	antigen	O
107	CT	O
108	scan	O
109	not	O
110	routine	O
111	depended	O
112	on	O
113	availability	O
114	test	O
115	especially	O
116	early	O
117	part	O
118	latter	O
119	more	O
120	preoperative	O
121	approach	O
122	decided	O
123	consent	O
124	after	O

```
125          thorough      O
126        discussion      O
127              pros      O
128              cons      O
129          decision      O
130         operating      O
131              time      O
132          surgeons      O
133          Patients      O
134             large      O
135             fixed      O
136            tumors      O
137          invasion      O
138             other      O
139            organs      O
140           advised      O
141           against      O
142           patient      O
143          received      O
144        mechanical      O
145       preparation      O
146     polyethylene       O
147           glycol       O
148     electrolytes       O
149          solution      O
150               day      O
151            before      O
152     prophylactic       O
153       intravenous      O
154       antibiotics      O
155     administered       O
156               at       O
157        induction       O
158        anesthesia      O
159                 A      O
160           urinary      O
161          catheter      O
162          inserted      O
163               put      O
164             under      O
165           general      O
166       Nasogastric      O
167              tube      O
168              used      O
169                as      O
170              Open      O
171           through      O
```

172	midline	O
173	incision	O
174	extent	O
175	determined	O
176	by	O
177	site	O
178	tumor	O
179	method	O
180	anastomosis	O
181	surgeon	O
182	In	O
183	upper	O
184	rectal	O
185	rectum	O
186	mobilized	O
187	sharp	O
188	perimesorectal	O
189	dissection	O
190	keep	O
191	visceral	O
192	pelvic	O
193	fascia	O
194	which	O
195	enveloped	O
196	mesorectum	O
197	intact	O
198	Total	O
199	mesorectal	O
200	excision	O
201	Instead	O
202	transected	O
203	4	O
204	5	O
205	cm	O
206	below	O
207	distal	O

12 Conclusion

1) Spacy NER performs better than NLTK and Stanford NLP NER modules. This is because Spacy is trained on a big corpus and has more classes of entities.

2) An alternative to NLTK's named entity recognition (NER) classifier is provided by the Stanford NER tagger. This tagger is largely seen as the standard in named entity recognition, but since it uses an advanced statistical learning algorithm it's more computationally expensive than the option provided by NLTK.

3) Some special domains like Medicine domain might be difficult for NER tasks as the libraries are trained to recognize daily common things/entities and not complex procedure names/medicine names etc. We might have to train our own NER module for these type of specific tasks.