

exp3

October 31, 2021

Neha Kulkarni

A060 BTech. IT SEM 7

Computational Linguistics and Natural Language Processing Lab Experiment 3

```
[1]: import nltk
     from nltk.tokenize import sent_tokenize, word_tokenize
```

```
[4]: text = "Welcome to Mukesh Patel School of Technology Management and Engineering.
     ↪ Have a good day"
     tokenized = sent_tokenize(text)
```

```
[5]: print(tokenized)
```

```
['Welcome to Mukesh Patel School of Technology Management and Engineering.',
 'Have a good day']
```

```
[6]: print(word_tokenize(text))
```

```
['Welcome', 'to', 'Mukesh', 'Patel', 'School', 'of', 'Technology', 'Management',
 'and', 'Engineering', '.', 'Have', 'a', 'good', 'day']
```

```
[7]: from nltk.corpus import stopwords
```

```
[13]: no_stopwords = [word for word in word_tokenize(text) if word not in stopwords.
     ↪ words()]
     no_stopwords
```

```
[13]: ['Welcome',
      'Mukesh',
      'Patel',
      'School',
      'Technology',
      'Management',
      'Engineering',
      '.',
      'Have',
      'good',
      'day']
```

```
[14]: from nltk import pos_tag
```

```
[15]: print(pos_tag(word_tokenize(text)))
```

```
[('Welcome', 'VB'), ('to', 'TO'), ('Mukesh', 'NNP'), ('Patel', 'NNP'),  
('School', 'NNP'), ('of', 'IN'), ('Technology', 'NNP'), ('Management', 'NNP'),  
('and', 'CC'), ('Engineering', 'NNP'), ('.', '.'), ('Have', 'VBP'), ('a', 'DT'),  
('good', 'JJ'), ('day', 'NN')]
```

```
[33]: from nltk.tokenize import PunktSentenceTokenizer
```

```
article = """In a ‘small but significant step’ forward for Delhi Metro, the 850.8-metre stretch extension in Pink Line at East Delhi’s Trilokpuri, which had proved to be a bottleneck for the DMRC authorities for a long time on account of land acquisition issues, was inaugurated jointly on August 6 by union minister of housing and urban affairs Hardeep Singh Puri and Delhi chief minister Arvind Kejriwal.
```

```
The two leaders jointly inaugurated the section between Mayur Vihar Pocket 1 and Trilokpuri Sanjay Lake stations, via video-conferencing.
```

```
With this, the 59-km-long Majlis Park-Shiv Vihar corridor nicknamed the Ring Road Line spanning 38 stations, is fully operational.
```

```
"The inauguration today of an 850.8 metres stretch extension might be small from the point of view of network length, but from the perspective of providing enhanced connectivity to people, this connection will play an extremely crucial role. The missing link in Trilokpuri is now being bridged to allow seamless and end to end travel through Delhi Metro’s longest corridor of 59 kilometers between Majlis Park and Shiv Vihar. Important transport hubs, markets, residential areas across the city will now get connected benefitting lakhs of passengers," Puri said at the inauguration.
```

```
"The last two years have been extremely difficult for all of us. The Covid - 19 pandemic has led to disruptions in work from time to time. Yet, Delhi Metro Rail Corporation is today ready with a new vital connection, thus further consolidating the footprint of the Metro in the national capital," he said.
```

```
[34]: data = [nltk.pos_tag(nltk.word_tokenize(sent)) for sent in nltk.  
sent_tokenize(article)]  
nnp = [[word for word in data1 if word[1] == "NNP"] for data1 in data]  
nnp
```

```
[34]: [(('', 'NNP'),  
('Delhi', 'NNP'),  
('Metro', 'NNP'),
```

```

('Pink', 'NNP'),
('Line', 'NNP'),
('East', 'NNP'),
('Delhi', 'NNP'),
(' ', 'NNP'),
('Trilokpuri', 'NNP'),
('DMRC', 'NNP'),
('August', 'NNP'),
('Hardeep', 'NNP'),
('Singh', 'NNP'),
('Puri', 'NNP'),
('Delhi', 'NNP'),
('Arvind', 'NNP'),
('Kejriwal', 'NNP')],
[('Mayur', 'NNP'),
('Vihar', 'NNP'),
('Pocket', 'NNP'),
('Trilokpuri', 'NNP'),
('Sanjay', 'NNP'),
('Lake', 'NNP')],
[('Majlis', 'NNP'),
('Park-Shiv', 'NNP'),
('Vihar', 'NNP'),
('Ring', 'NNP'),
('Road', 'NNP'),
('Line', 'NNP')],
[],
[('Trilokpuri', 'NNP'),
('Delhi', 'NNP'),
('Metro', 'NNP'),
(' ', 'NNP'),
('Majlis', 'NNP'),
('Park', 'NNP'),
('Shiv', 'NNP'),
('Vihar', 'NNP')],
[(' ', 'NNP'), ('Puri', 'NNP')],
[],
[('Covid', 'NNP'), ('-', 'NNP')],
[('Delhi', 'NNP'),
('Metro', 'NNP'),
('Rail', 'NNP'),
('Corporation', 'NNP'),
('Metro', 'NNP')]]

```

```

[35]: jj = [[word for word in data1 if word[1] == "JJ"] for data1 in data]
      jj

```

```
[35]: [[('', 'JJ'),
      ('small', 'JJ'),
      ('significant', 'JJ'),
      ('850.8-metre', 'JJ'),
      ('long', 'JJ'),
      ('urban', 'JJ'),
      ('chief', 'JJ')],
      [],
      [('59-km-long', 'JJ'), ('operational', 'JJ')],
      [('stretch', 'JJ'), ('small', 'JJ'), ('enhanced', 'JJ'), ('crucial', 'JJ')],
      [],
      [('Important', 'JJ'), ('residential', 'JJ')],
      [('last', 'JJ'), ('difficult', 'JJ')],
      [],
      [('ready', 'JJ'), ('new', 'JJ'), ('vital', 'JJ'), ('national', 'JJ')]]
```

```
[36]: vbd = [[word for word in data1 if word[1] == "VBD"] for data1 in data]
vbd
```

```
[36]: [['had', 'VBD'), ('was', 'VBD')],
      [('inaugurated', 'VBD')],
      [('nicknamed', 'VBD')],
      [],
      [('s', 'VBD')],
      [('said', 'VBD')],
      [],
      [],
      [('said', 'VBD')]]
```

```
[42]: from nltk.stem import PorterStemmer
ps = PorterStemmer()
stem = [ps.stem(w) for w in ['Playing', 'Played', 'Player']]
print(stem)

['play', 'play', 'player']
```

```
[46]: from nltk.stem import WordNetLemmatizer
lem = WordNetLemmatizer()
print(lem.lemmatize(article))
```

In a ‘small but significant step’ forward for Delhi Metro, the 850.8-metre stretch extension in Pink Line at East Delhi’s Trilokpuri, which had proved to be a bottleneck for the DMRC authorities for a long time on account of land acquisition issues, was inaugurated jointly on August 6 by union minister of housing and urban affairs Hardeep Singh Puri and Delhi chief minister Arvind Kejriwal.

The two leaders jointly inaugurated the section between Mayur Vihar Pocket 1 and Trilokpuri Sanjay Lake stations, via video-conferencing.

With this, the 59-km-long Majlis Park-Shiv Vihar corridor nicknamed the Ring Road Line spanning 38 stations, is fully operational.

"The inauguration today of an 850.8 metres stretch extension might be small from the point of view of network length, but from the perspective of providing enhanced connectivity to people, this connection will play an extremely crucial role. The missing link in Trilokpuri is now being bridged to allow seamless and end to end travel through Delhi Metro's longest corridor of 59 kilometers between Majlis Park and Shiv Vihar. Important transport hubs, markets, residential areas across the city will now get connected benefitting lakhs of passengers," Puri said at the inauguration.

"The last two years have been extremely difficult for all of us. The Covid - 19 pandemic has led to disruptions in work from time to time. Yet, Delhi Metro Rail Corporation is today ready with a new vital connection, thus further consolidating the footprint of the Metro in the national capital," he said.

```
[3]: from nltk.corpus import wordnet
for word in ["Hi", "Hello", "Bye"]:
    print(f"Word: {word}")
    obj = wordnet.synsets(word)
    print(f"Definition: {obj[0].definition()}")
    print(f"Examples: {obj[0].examples()}\n")
```

Word: Hi

Definition: an expression of greeting

Examples: ['every morning they exchanged polite hellos']

Word: Hello

Definition: an expression of greeting

Examples: ['every morning they exchanged polite hellos']

Word: Bye

Definition: you advance to the next round in a tournament without playing an opponent

Examples: ['he had a bye in the first round']

```
[66]: for syn in wordnet.synsets('has'):
    for lemma in syn.lemmas():
        print(lemma.name())
    if lemma.antonyms():
        print(lemma.antonyms()[0].name())
```

hour_angle

HA

have

have_got

hold
have
feature
experience
receive
have
get
own
have
possess
get
let
have
consume
ingest
take_in
take
have
have
hold
throw
have
make
give
have
have
have
experience
have
induce
stimulate
cause
have
get
make
accept
take
have
receive
have
suffer
sustain
have
get
have
get
make
give_birth

deliver
bear
birth
have
take
have

```
[2]: syn = wordnet.synsets('hello')[0]
      print ("Syn tag : ", syn.pos())
      syn = wordnet.synsets('doing')[0]
      print ("Syn tag : ", syn.pos())
      syn = wordnet.synsets('beautiful')[0]
      print ("Syn tag : ", syn.pos())
      syn = wordnet.synsets('quickly')[0]
      print ("Syntag : ", syn.pos())
```

Syn tag : n
Syn tag : v
Syn tag : a
Syntag : r

```
[ ]:
```