

Tagging PoS for SRUniversity text

```
SRUniversity="""The SR University campus is located in Ananthasagar village of Hasanparthy Mandal in Warangal, Telangana, Ind
It is in 150 acres, with both separate hostel facilities for boys and girls.
There is a huge central library along with Indias largest Technology Business Incubator (TBI) in tier 2 cities."""
```

```
import nltk
nltk.download('punkt')
nltk.download('punkt_tab')
nltk.download('averaged_perceptron_tagger_eng') # Changed to download 'averaged_perceptron_tagger_eng'
from nltk.tokenize import word_tokenize
words = word_tokenize(SRUniversity)
nltk.pos_tag(words)
```

```
[nltk_data] /root/nltk_data...
[nltk_data] Unzipping taggers/averaged_perceptron_tagger_eng.zip.
[('The', 'DT'),
 ('SR', 'NNP'),
 ('University', 'NNP'),
 ('campus', 'NN'),
 ('is', 'VBZ'),
 ('located', 'VBN'),
 ('in', 'IN'),
 ('Ananthasagar', 'NNP'),
 ('village', 'NN'),
 ('of', 'IN'),
 ('Hasanparthy', 'NNP'),
 ('Mandal', 'NNP'),
 ('in', 'IN'),
 ('Warangal', 'NNP'),
 (',', ','),
 ('Telangana', 'NNP'),
 (',', ','),
 ('India', 'NNP'),
 (',', ','),
 ('It', 'PRP'),
 ('is', 'VBZ'),
 ('in', 'IN'),
 ('150', 'CD'),
 ('acres', 'NNS'),
 (',', ','),
 ('with', 'IN'),
 ('both', 'DT'),
 ('separate', 'JJ'),
 ('hostel', 'NN'),
 ('facilities', 'NNS'),
 ('for', 'IN'),
 ('boys', 'NNS'),
 ('and', 'CC'),
 ('girls', 'NNS'),
 (',', ','),
 ('There', 'EX'),
 ('is', 'VBZ'),
 ('a', 'DT'),
 ('huge', 'JJ'),
 ('central', 'JJ'),
 ('library', 'NN'),
 ('along', 'IN'),
 ('with', 'IN'),
 ('Indias', 'NNP'),
 ('largest', 'JJS'),
 ('Technology', 'NN'),
 ('Business', 'NNP'),
 ('Incubator', 'NNP'),
 ('(', '('),
 ('TBI', 'NNP'),
 (')', ')'),
 ('in', 'IN'),
 ('tier', '$'),
 ('2', 'CD'),
 ('cities', 'NNS'),
 (',', ',')]
```

```
import spacy
nlp = spacy.load("en_core_web_sm")
doc = nlp("""The SR University campus is located in Ananthasagar village of Hasanparthy Mandal in Warangal, Telangana, India.
It is in 150 acres, with both separate hostel facilities for boys and girls.
There is a huge central library along with Indias largest Technology Business Incubator (TBI) in tier 2 cities.""")
```

```
for token in doc:
    print(token.text, token.pos_)
```

```
University PROPN
campus NOUN
is AUX
located VERB
in ADP
Ananthasagar PROPN
village NOUN
of ADP
Hasanparthy PROPN
Mandal PROPN
in ADP
Warangal PROPN
, PUNCT
Telangana PROPN
, PUNCT
India PROPN
. PUNCT
```

```
SPACE
It PRON
is AUX
in ADP
150 NUM
acres NOUN
, PUNCT
with ADP
both DET
separate ADJ
hostel NOUN
facilities NOUN
for ADP
boys NOUN
and CCONJ
girls NOUN
. PUNCT
```

```
SPACE
There PRON
is VERB
a DET
huge ADJ
central ADJ
library NOUN
along ADP
with ADP
Indias PROPN
largest ADJ
Technology PROPN
Business PROPN
Incubator PROPN
( PUNCT
TBI PROPN
) PUNCT
in ADP
tier NOUN
2 NUM
cities NOUN
. PUNCT
```

```
import spacy
from collections import Counter
nlp = spacy.load("en_core_web_sm")
text = ""The SR University campus is located in Ananthasagar village of Hasanparthy Mandal in Warangal, Telangana, India.
It is in 150 acres, with both separate hostel facilities for boys and girls.
There is a huge central library along with Indias largest Technology Business Incubator (TBI) in tier 2 cities.""
doc = nlp(text)
nouns = []
verbs = []
for token in doc:
    if token.pos_ in ["NOUN", "PROPN"]:
        nouns.append(token.text)
    elif token.pos_ == "VERB":
        verbs.append(token.text)
noun_freq = Counter(nouns)
verb_freq = Counter(verbs)
print("Noun Frequency:", noun_freq)
print("Verb Frequency:", verb_freq)
```

```
Noun Frequency: Counter({'SR': 1, 'University': 1, 'campus': 1, 'Ananthasagar': 1, 'village': 1, 'Hasanparthy': 1, 'Mandal': 1})
Verb Frequency: Counter({'located': 1, 'is': 1})
```

PoS Tagging for whatsapp file

```
import nltk
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger_eng')
sentence = "Students are learning Natural Language Processing"
tokens = nltk.word_tokenize(sentence)
pos_tags = nltk.pos_tag(tokens)
print(pos_tags)
```

```
[('Students', 'NNS'), ('are', 'VBP'), ('learning', 'VBG'), ('Natural', 'NNP'), ('Language', 'NNP'), ('Processing', 'NNP')]
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger_eng to
[nltk_data] /root/nltk_data...
[nltk_data] Package averaged_perceptron_tagger_eng is already up-to-
[nltk_data] date!
```

```
import spacy
nlp = spacy.load("en_core_web_sm")
doc = nlp("Students are learning Natural Language Processing")
for token in doc:
    print(token.text, token.pos_)
```

```
Students NOUN
are AUX
learning VERB
Natural PROPN
Language PROPN
Processing NOUN
```

```
import spacy
from collections import Counter
nlp = spacy.load("en_core_web_sm")
text = "Students are learning Natural Language Processing"
doc = nlp(text)
nouns = []
verbs = []
for token in doc:
    if token.pos_ in ["NOUN", "PROPN"]:
        nouns.append(token.text)
    elif token.pos_ == "VERB":
        verbs.append(token.text)
noun_freq = Counter(nouns)
verb_freq = Counter(verbs)
print("Noun Frequency:", noun_freq)
print("Verb Frequency:", verb_freq)
```

```
Noun Frequency: Counter({'Students': 1, 'Natural': 1, 'Language': 1, 'Processing': 1})
Verb Frequency: Counter({'learning': 1})
```

Assignment 3.2

```
from google.colab import files
files.upload()
```


<https://colab.research.google.com/drive/1pMEfci3U6NAiU281n33u0bJI6eDAs6B#scrollTo=bHoYcXIoNdvR&printMode=true>

<https://colab.research.google.com/drive/1pMEfci3U6NAiU281n33u0bJl6eDAs6B#scrollTo=bHoYcXloNdvR&printMode=true> 6/6