31_Bhavik Maru
Experiment No.6
Perform POS tagging on the given English and Indian Language Text

Aim: Perform POS tagging on the given English and Indian Language Text

Objective: To study POS Tagging and tag the part of speech for given input in english and an Indian Language.

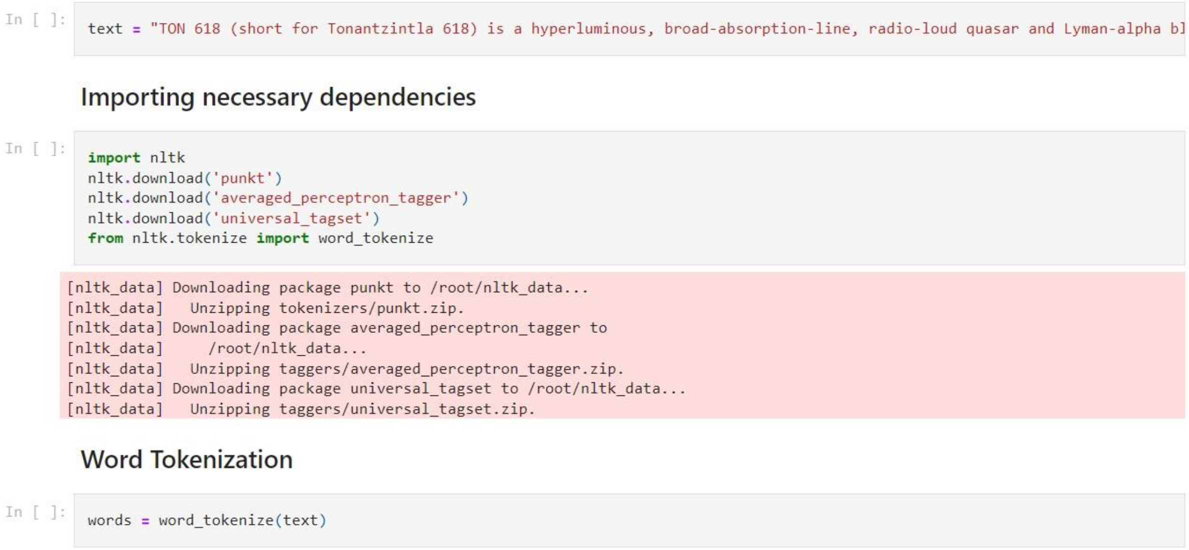
Theory:

The primary target of Part-of-Speech (POS) tagging is to identify the grammatical group of a given word. Whether it is a NOUN, PRONOUN, ADJECTIVE, VERB, ADVERBS, etc. based on the context. POS Tagging looks for relationships within the sentence and assigns a corresponding tag to the word.

POS Tagging (Parts of Speech Tagging) is a process to mark up the words in text format for a particular part of a speech based on its definition and context. It is responsible for text reading in a language and assigning some specific token (Parts of Speech) to each word. It is also called grammatical tagging.

Steps Involved in the POS tagging example:

- Tokenize text (word_tokenize)
- apply pos_tag to above step that is nltk.pos_tag(tokenize_text)



In [ ]:
```
text = "TON 618 (short for Tonantzintla 618) is a hyperluminous, broad-absorption-line, radio-loud quasar and Lyman-alpha bl
```

Importing necessary dependencies

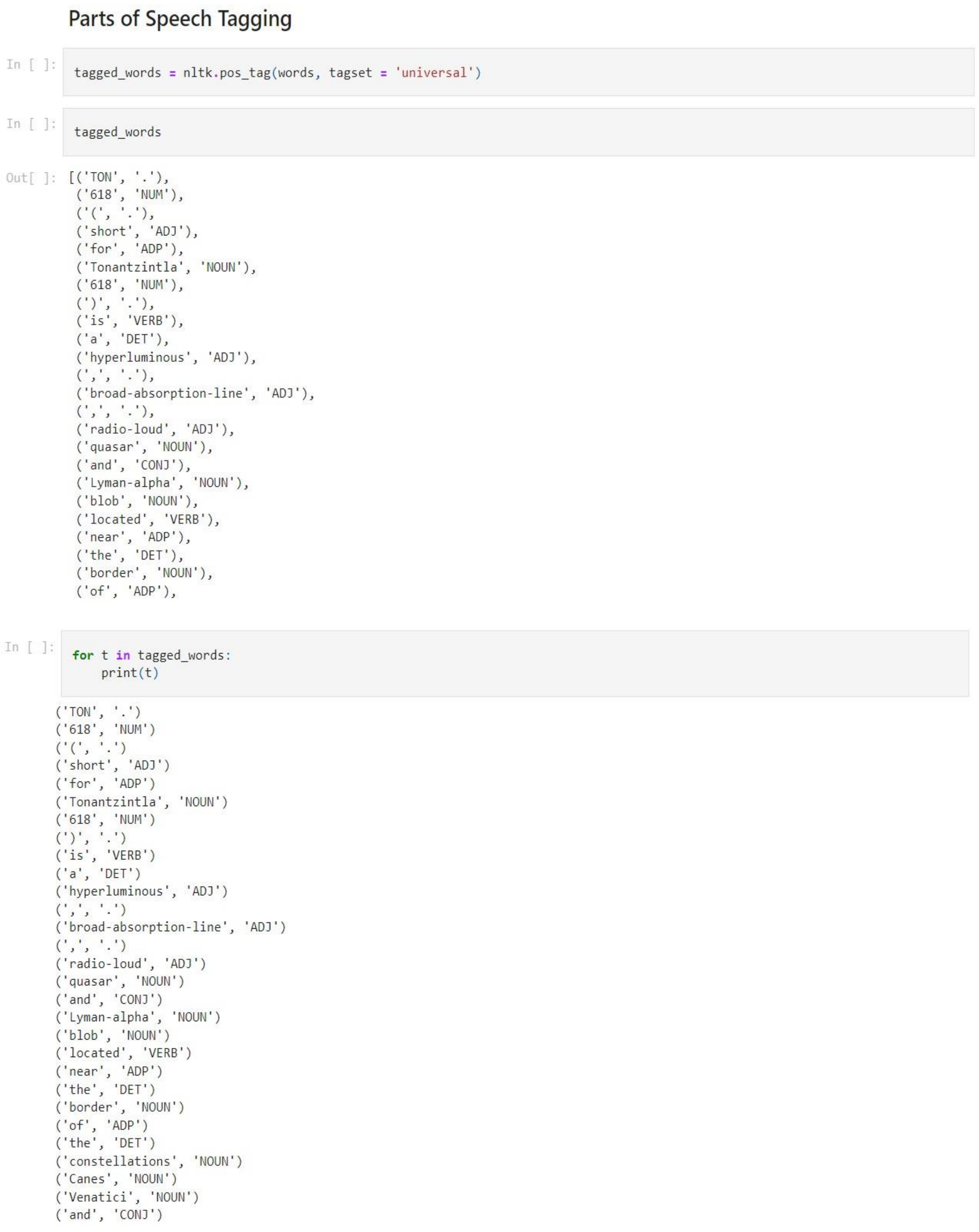
In [ ]:
```
import nltk
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')
nltk.download('universal_tagset')
from nltk.tokenize import word_tokenize
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     /root/nltk_data...
[nltk_data]   Unzipping taggers/averaged_perceptron_tagger.zip.
[nltk_data] Downloading package universal_tagset to /root/nltk_data...
[nltk_data]   Unzipping taggers/universal_tagset.zip.
```

Word Tokenization

In [ ]:
```
words = word_tokenize(text)
```

Output:



Parts of Speech Tagging

```
In [ ]: tagged_words = nltk.pos_tag(words, tagset = 'universal')
```

```
In [ ]: tagged_words
```

```
Out[ ]: [('TON', '.'),
         ('618', 'NUM'),
         ('(', '.'),
         ('short', 'ADJ'),
         ('for', 'ADP'),
         ('Tonantzintla', 'NOUN'),
         ('618', 'NUM'),
         (')', '.'),
         ('is', 'VERB'),
         ('a', 'DET'),
         ('hyperluminous', 'ADJ'),
         (',', '.'),
         ('broad-absorption-line', 'ADJ'),
         (',', '.'),
         ('radio-loud', 'ADJ'),
         ('quasar', 'NOUN'),
         ('and', 'CONJ'),
         ('Lyman-alpha', 'NOUN'),
         ('blob', 'NOUN'),
         ('located', 'VERB'),
         ('near', 'ADP'),
         ('the', 'DET'),
         ('border', 'NOUN'),
         ('of', 'ADP'),
```

```
In [ ]: for t in tagged_words:
            print(t)
```

```
('TON', '.')
('618', 'NUM')
('(', '.')
('short', 'ADJ')
('for', 'ADP')
('Tonantzintla', 'NOUN')
('618', 'NUM')
(')', '.')
('is', 'VERB')
('a', 'DET')
('hyperluminous', 'ADJ')
(',', '.')
('broad-absorption-line', 'ADJ')
(',', '.')
('radio-loud', 'ADJ')
('quasar', 'NOUN')
('and', 'CONJ')
('Lyman-alpha', 'NOUN')
('blob', 'NOUN')
('located', 'VERB')
('near', 'ADP')
('the', 'DET')
('border', 'NOUN')
('of', 'ADP')
('the', 'DET')
('constellations', 'NOUN')
('Canes', 'NOUN')
('Venatici', 'NOUN')
('and', 'CONJ')
```

Conclusion:

POS tagging is the process of assigning a part-of-speech tag to each word in a sentence. Part-of-speech tags are labels that indicate the grammatical function of a word in a sentence, such as noun, verb, adjective, adverb, etc. The result of POS tagging is a sequence of part-of-speech tags, one for each word in the sentence. For example, the POS tagging for the sentence "The cat sat on the mat" would be: DET NN VBD IN DET NN There are two main types of POS tagging techniques: rule-based and statistical.