Experiment No.6
Perform POS tagging on the given English and Indian Language Text
Date of Performance:
Date of Submission:

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:

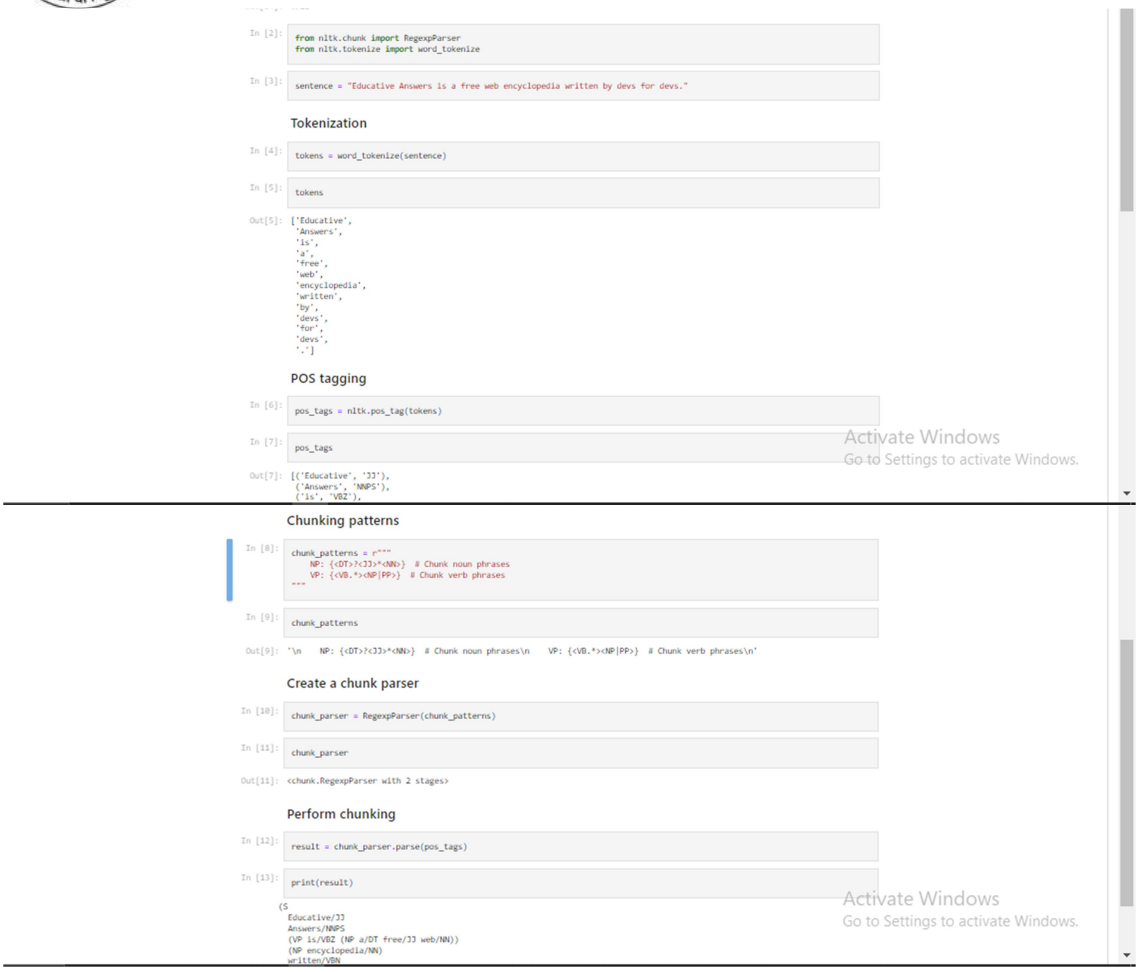
The primary target of Part-of-Speech (POS) tagging is to identify the grammatical group of a given word. Whether it is a NOUN, PRONOUN, ADJECTIVE, VERB, ADVERBS, etc. based on the context. POS Tagging looks for relationships within the sentence and assigns a corresponding tag to the word.

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)

Output:



```
In [2]: from nltk.chunk import RegexpParser
        from nltk.tokenize import word_tokenize

In [3]: sentence = "Educative Answers is a free web encyclopedia written by devs for devs."
```

Tokenization

```
In [4]: tokens = word_tokenize(sentence)

In [5]: tokens

Out[5]: ['Educative',
         'Answers',
         'is',
         'a',
         'free',
         'web',
         'encyclopedia',
         'written',
         'by',
         'devs',
         'for',
         'devs',
         '.']
```

POS tagging

```
In [6]: pos_tags = nltk.pos_tag(tokens)

In [7]: pos_tags

Out[7]: [('Educative', 'JJ'),
         ('Answers', 'NNPS'),
         ('is', 'VBZ'),
```


Chunking patterns

```
In [8]: chunk_patterns = r"""
            NP: {<DT>?<JJ>*<NN>}  # Chunk noun phrases
            VP: {<VB.*><NP|PP>}  # Chunk verb phrases
        """

In [9]: chunk_patterns

Out[9]: '\n    NP: {<DT>?<JJ>*<NN>}  # Chunk noun phrases\n    VP: {<VB.*><NP|PP>}  # Chunk verb phrases\n'
```

Create a chunk parser

```
In [10]: chunk_parser = RegexpParser(chunk_patterns)

In [11]: chunk_parser

Out[11]: <chunk.RegexpParser with 2 stages>
```

Perform chunking

```
In [12]: result = chunk_parser.parse(pos_tags)

In [13]: print(result)

(S
  Educative/JJ
  Answers/NNPS
  (VP is/VBZ (NP a/DT free/JJ web/NN))
  (NP encyclopedia/NN)
  written/VBN
```


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.