



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)



# Vidyavardhini's College of Engineering & Technology

## Department of Computer Science and Engineering (Data Science)

### Output:

```
29_Dinesh Madhav_Expt 06_NLP.ipynb
File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text
RAM
Disk

[5] import nltk
    nltk.download('punkt')
    nltk.download('averaged_perceptron_tagger')

[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 to
[nltk_data] /root/nltk_data...
[nltk_data] Package averaged_perceptron_tagger is already up-to-
[nltk_data] date!
[nltk_data] True

[2] from nltk.chunk import RegexpParser
    from nltk.tokenize import word_tokenize

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

▼ Tokenization

[4] tokens = word_tokenize(sentence)

[6] tokens

['Educative',
```

```
29_Dinesh Madhav_Expt 06_NLP.ipynb
File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text
RAM
Disk

[6] 'written',
    'by',
    'devs',
    'for',
    'devs',
    '.']

▼ POS tagging

[7] pos_tags = nltk.pos_tag(tokens)

pos_tags

[('Educative', 'JJ'),
 ('Answers', 'NNPS'),
 ('is', 'VBZ'),
 ('a', 'DT'),
 ('free', 'JJ'),
 ('web', 'NN'),
 ('encyclopedia', 'NN'),
 ('written', 'VBN'),
 ('by', 'IN'),
 ('devs', 'NN'),
 ('for', 'IN'),
 ('devs', 'NN'),
 ('.', '.')]

0s completed at 2:56 PM
```



# Vidyavardhini's College of Engineering & Technology

## Department of Computer Science and Engineering (Data Science)

```
29_Dinesh Madhav_Expt 06_NLP.ipynb
File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text
[8] ('for', 'IN'),
    ('devs', 'NN'),
    ('.', '.')
```

### Chunking patterns

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

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

### Create a chunk parser

```
[11] chunk_parser = RegexpParser(chunk_patterns)

[12] chunk_parser
<chunk.RegexpParser with 2 stages>
```

0s completed at 2:56 PM

```
29_Dinesh Madhav_Expt 06_NLP.ipynb
File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text
Perform chunking

[13] result = chunk_parser.parse(pos_tags)

print(result)
```

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

0s completed at 2:56 PM

### 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.