Name: Prathamesh Arvind Jadhav
Roll No: 4059

Experiment No:2

Experiment Title:

Implement Tokenization by Word and Sentence

Aim:

To implement POS tagging techniques in NLP to identify grammatical categories of words in a sentence using NLTK and spaCy libraries..

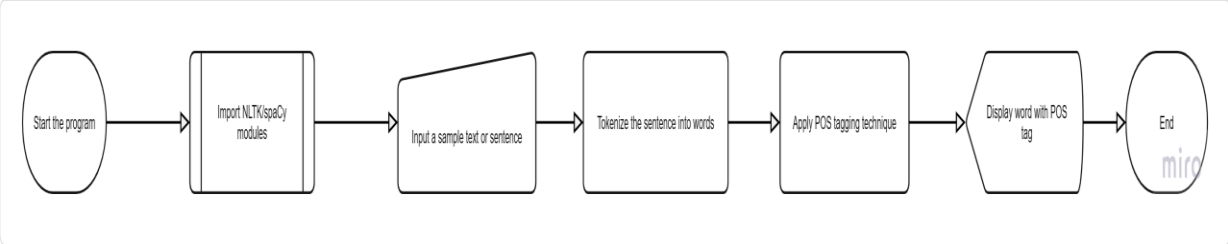
Objective:

- Understand the concept of Parts of Speech (POS) in natural language.
- Perform POS tagging on a sample sentence.
- Implement POS tagging using Python with NLTK and spaCy libraries.

Procedure & Flowchart:

- ## Procedure:
- Install required NLP libraries like NLTK or spaCy.
- Import necessary modules for POS tagging.
- Define a sample input sentence.
- Tokenize the input text into words.
- Apply POS tagging using the selected NLP library.
- Print each word along with its POS tag.

- **Flowchart:**



Code / Implementation:

```
import nltk
from nltk.tokenize import word_tokenize
from nltk import pos_tag

# Download required resources
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')
nltk.download('averaged_perceptron_tagger_eng')

# Sample text
text = "I am Prathamesh Jadhav And Currently I am Performing POS Tagging Experiment."

# Step 1: Tokenization
tokens = word_tokenize(text)

# Step 2: POS Tagging
tagged_words = pos_tag(tokens)

# Output result
print("Tokenized Words:", tokens)
print("POS Tagged Words:")
for word, tag in tagged_words:
    print(f"{word} -> {tag}")
```

```
Tokenized Words: ['I', 'am', 'Prathamesh', 'Jadhav', 'And', 'Currently', 'I', 'am', 'Performing', 'POS', 'Tagging', 'Experiment', '.']
POS Tagged Words:
I -> PRP
am -> VBP
Prathamesh -> JJ
Jadhav -> NNP
And -> CC
Currently -> NNP
I -> PRP
am -> VBP
Performing -> VBG
POS -> NNP
Tagging -> NNP
Experiment -> 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 to
[nltk_data]     /root/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]       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!
```

Student Activity - Code & Output

Student Task:

1. Prepare a tokenization program using a different paragraph related to your subject.
2. Visualize the POS tagging

Sample Code for Student Activity:

Student Activity with Visualization

```
import nltk
from tabulate import tabulate
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')

# Sample Text
text = "Data Science combines statistics, computer science, and domain knowledge to extract insights."

# Tokenize and tag
tokens = nltk.word_tokenize(text)
tags = nltk.pos_tag(tokens)

# Display as table
print(tabulate(tags, headers=["Word", "POS Tag"], tablefmt="grid"))
```

```
+------------+-----------+
| Word       | POS Tag   |
+============+===========+
| Data       | NNP       |
+------------+-----------+
| Science    | NNP       |
+------------+-----------+
| combines   | NNS       |
+------------+-----------+
| statistics | NNS       |
+------------+-----------+
| ,          | ,         |
+------------+-----------+
| computer   | NN        |
+------------+-----------+
| science    | NN        |
+------------+-----------+
| ,          | ,         |
+------------+-----------+
| and        | CC        |
+------------+-----------+
| domain     | VB        |
+------------+-----------+
| knowledge  | NN        |
+------------+-----------+
| to         | TO        |
+------------+-----------+
| extract    | VB        |
+------------+-----------+
| insights   | NNS       |
+------------+-----------+
| .          | .         |
+------------+-----------+
[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!
```

Questions & Answers:

1. What is POS tagging in NLP? (CO1)

Answer:
POS tagging is the process of assigning a part of speech (such as noun, verb, adjective) to each word in a sentence based on its meaning and context.

---

2. Differentiate between NLTK and spaCy for POS tagging. (CO1)

Answer:

- **NLTK:** Rule-based with statistical models; useful for academic exploration and prototyping.
- **spaCy:** Industrial-strength NLP library with faster and more context-aware tagging.

---

3. Why is POS tagging important in NLP tasks? (CO3)

Answer:
POS tagging helps in understanding the grammatical structure of sentences, enabling downstream tasks like parsing, information extraction, machine translation, and sentiment analysis.

Conclusion:

In this experiment, we successfully implemented POS tagging using Python's NLTK and spaCy libraries. POS tagging provides crucial grammatical information that helps in deeper language analysis and forms the base for many advanced NLP applications.