S. B. JAIN INSTITUTE OF TECHNOLOGY, MANAGEMENT & RESEARCH, NAGPUR.

Practical No. 6

Aim: Perform data filtration using POS tagging.

Name of Student: **Shrutika Pradeep Bagdi**

Roll No.: **CS22130**

Semester/Year: **IV/VII**

Academic Session: 2025-2026

Date of Performance: _____

Date of Submission: _____

AIM: Perform data filtration using POS tagging

OBJECTIVE/EXPECTED LEARNING OUTCOME:

- Understanding POS tagging
- Understanding data filtration

HARDWARE AND SOFTWARE REQUIRMENTS:

Hardware Requirement:

Software Requirement:

THEORY:

Part-of-speech (POS) tagging is a fundamental task in natural language processing (NLP). It involves the process of assigning a specific grammatical category (such as noun, verb, adjective, adverb, pronoun, preposition, conjunction, etc.) to each word in a text or a sentence. POS tagging is essential for various NLP applications, including syntactic analysis, information retrieval, machine translation, and text-to-speech synthesis.

Purpose:

- **Linguistic Analysis:** POS tagging helps in understanding the grammatical structure of a sentence, which is essential for parsing and syntactic analysis.
- **Disambiguation:** Many words can have multiple meanings and can function as different parts of speech depending on the context. POS tagging disambiguates these cases.

Applications:

- **Information Retrieval**: POS tagging can help improve search and retrieval systems by considering the grammatical structure of queries and documents.
- **Machine Translation:** Accurate POS tagging is crucial for translating sentences from one language to another.
- **Text-to-Speech Synthesis**: Proper pronunciation and intonation often depend on the POS of words in a sentence.
- **Named Entity Recognition (NER):** POS tags can be used as features for identifying named entities in text.

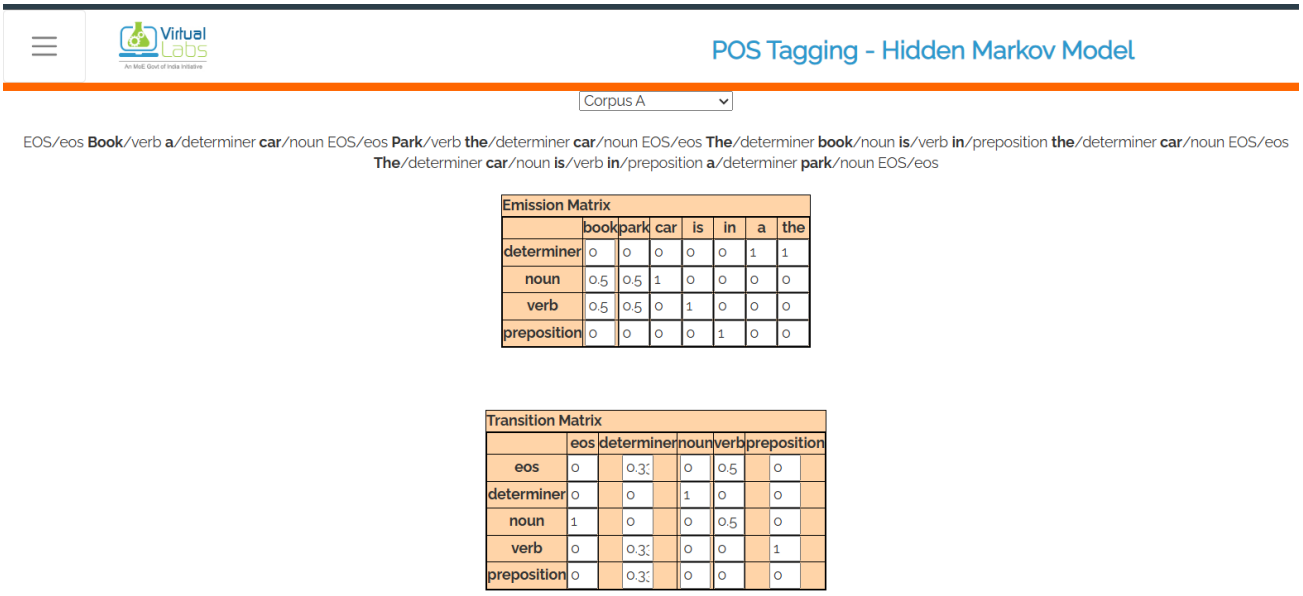
CODE:

```
import nltk
# Download necessary NLTK resources
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')
# Take sentence input from the user
sentence = input("Enter a sentence: ")
# Tokenize and tag parts of speech
words = nltk.word_tokenize(sentence)
tags = nltk.pos_tag(words)
# Filter and print by part of speech
print("\nFiltered Words:")
for word, tag in tags:
    if tag.startswith('DT'):   # Determiner
        print(f"Determiner: {word}")
    elif tag.startswith('NN'): # Noun
        print(f"Noun: {word}")
    elif tag.startswith('VB'): # Verb
        print(f"Verb: {word}")
    elif tag == 'IN':          # Preposition
        print(f"Preposition: {word}")
```

OUTPUT (SCREENSHOT):

```
[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!
Enter a sentence: Book a car Park the car The book is in the car The car is in a park

Filtered Words:
Verb: Book
Determiner: a
Noun: car
Noun: Park
Determiner: the
Noun: car
Determiner: The
Noun: book
Verb: is
Preposition: in
Determiner: the
Noun: car
Determiner: The
Noun: car
Verb: is
Preposition: in
Determiner: a
Noun: park
```



CONCLUSION:

DISCUSSION AND VIVA VOCE:

- What is POS tagging?
- What are methods of implementing POS tagging?
- How it is implemented through code?

REFERENCE:

- *www.w3schools.com*
- *www.tutorialsmade.com*
- *www.towardsdatascience.com*