**SVKM’s NMIMS**

**Mukesh Patel School of Technology Management & Engineering**

**Department of Artificial Intelligence**

Program: B. Tech/MBA Tech Semester: V

**Course: Natural Language Processing**

**Faculty Name: Ami Munshi**

**Experiment No.08**

PART A

(PART A: TO BE REFFERED BY STUDENTS)

**A.1 Aim:** Explore POS tagger and Chunker in Virtual Labs

1. To explore POS Tagger using Hidden Markov Model in Virtual Lab for NLP
2. To explore Chunker in Virtual Lab for NLP

**A.2 Prerequisite: -** Python

**A.3 Outcome:**

**After successful completion of this experiment students will be able to**

1. Understand the usage of POS tagger and chunker

**A.4 Theory:**

POS tagging or part-of-speech tagging is the procedure of assigning a grammatical category like noun, verb, adjective etc. to a word. In this process both the lexical information and the context play an important role as the same lexical form can behave differently in a different context.

For example the word "Park" can have two different lexical categories based on the context.

The boy is playing in the park. ('Park' is Noun)

Park the car. ('Park' is Verb)

* Assigning part of speech to words by hand is a common exercise one can find in an elementary grammar class. But here we wish to build an automated tool which can assign the appropriate part-of-speech tag to the words of a given sentence. One can think of creating hand crafted rules by observing patterns in the language, but this would limit the system's performance to the quality and number of patterns identified by the rule crafter. Thus, this approach is not practically adopted for building POS Tagger. Instead, a large corpus annotated with correct POS tags for each word is given to the computer and algorithms then learn the patterns automatically from the data and store them in form of a trained model. Later this model can be used to POS tag new sentences.

**A.5 Task to be completed in PART B**

**A.5.1. Task**

1. There are three corpus given in the simulation tab in the link below. Select any one corpus and obtain emission and transition probability tables

2. Observe the working chunker

**Ref:**

https://nlp-iiith.vlabs.ac.in/exp/markov-model/index.html

**PART B**

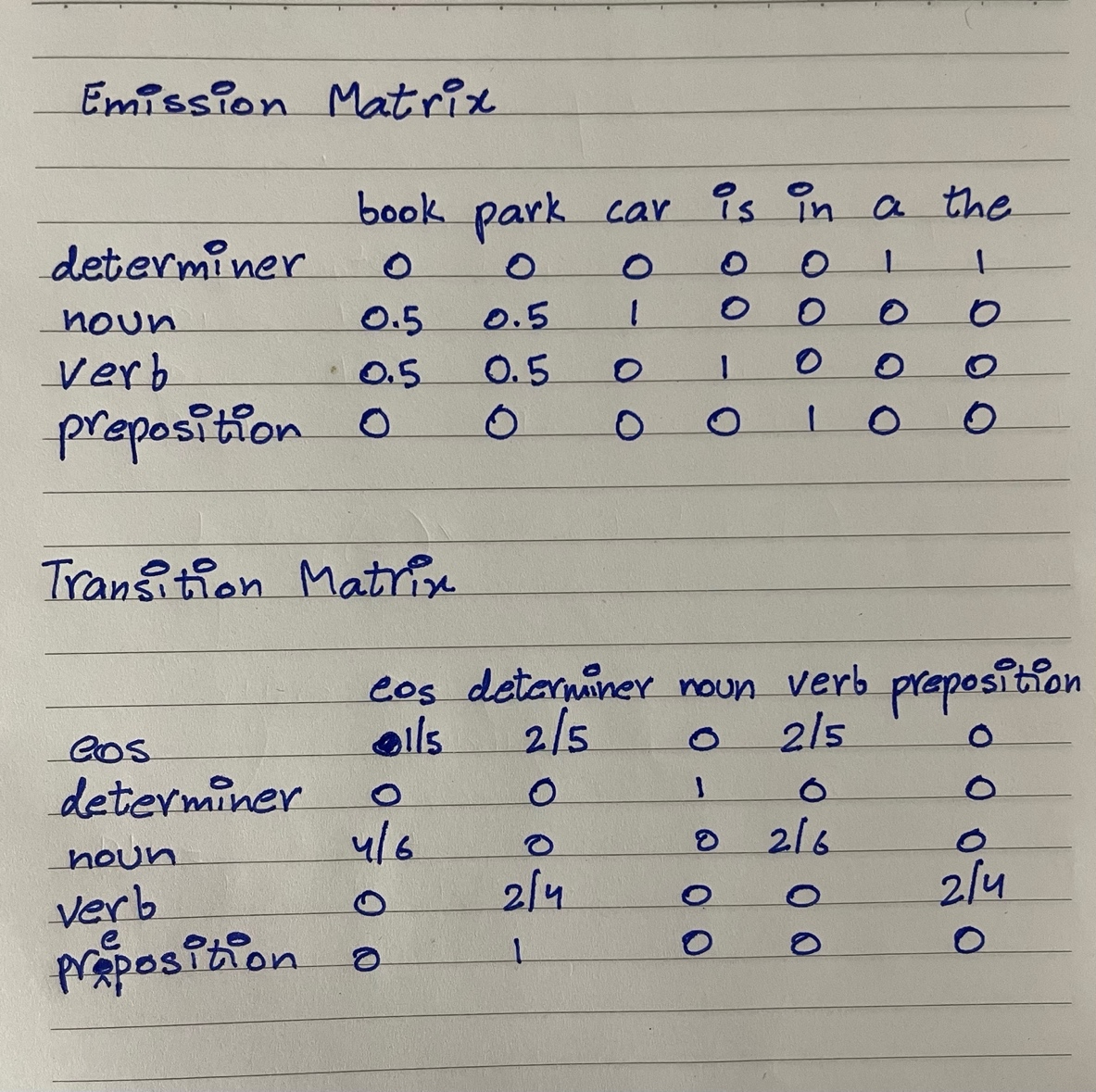
(PART B: TO BE COMPLETED BY STUDENTS)

**(Students must submit the soft copy as per following segments within two hours of the practical. The soft copy must be uploaded on the Portal/Teams or emailed to the concerned lab in charge faculties at the end of the practical in case the there is no portal access available)**

|  |  |
| --- | --- |
| Roll No. : I028 | Name: Aditya Kothari |
| Program : BTech AI | Division: |
| Batch: B1 | Date of Experiment: 11/10/2023 |
| Date of Submission: 11/10/2023 | Grade : |

**B.1 Tasks given in PART A to be completed here**

*(****Students must write the answers of the task(s) given in the PART A )***

****

**B.2 Observations and Learning:**

**1.** How is Hidden Markov Model different from Markov Model?

A Markov Model deals with observable states, where each state depends only on the previous state. In contrast, a Hidden Markov Model (HMM) has both observable and hidden states. The hidden states are not directly observed and are used to model situations with uncertainty, like speech recognition or part-of-speech tagging.

1. How does the corpus size effect the transition and emission matrix?

A larger corpus generally leads to more reliable transition and emission matrices in an HMM, while a smaller corpus can result in less accurate estimates.

**B.3 Conclusion:**

Explored POS Tagger using Hidden Markov Model in Virtual Lab for NLP and made transition and emission matrix.

*(****Students must write the conclusive statements as per the attainment of individual outcomes listed above and learning/observation noted in section B.2)***

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*