

**DEPARTMENT OF COMPUTER ENGINEERING**

## CSL804 Computational Lab II

**Eighth Semester, 2021-2022 (Even Semester)**

**Name of Student :** Saurav Kumar

## Roll No. 23

**Division :** BE – CMPN

**Day/ Session :** Monday/Afternoon

**Venue :** SLRTCE Lab 305

## Experiment No. 4

**Title of Experiment :** To perform POS Tagging for English and any Indian language using a tagger.

## Date of Conduction :

**Date of Submission :**

|  |  |  |
| --- | --- | --- |
| **Particulars Max. Marks Marks Obtained** | | |
| Preparedness and Efforts(PE) | **3** |  |
| Knowledge of tools(KT) | **3** |  |
| Debugging and results(DR) | **3** |  |
| Documentation(DN) | **3** |  |
| Punctuality & Lab Ethics(PL) | **3** |  |
| **Total** | **15** |  |

**Grades – Meet Expectations (3 Marks), Moderate Expectations (2 Marks), Below Expectations (1 Mark)**

**Checked and Verified by Name of Faculty :** Prof. Neelam Kulkarni

## Signature :

**Date :**

EXPERIMENT NO: 4

POS TAGGING

**AIM:** To perform POS Tagging for English and any Indian language using a tagger.

**SOFTWARE:** Python, NLTK, CLTK

# THEORY:

The POS tagging process is the process of finding the sequence of tags which is most likely to have generated a given word sequence. We can model this POS process by using a Hidden Markov Model (HMM), where tags are the hidden states that produced the observable output, i.e., the words.

# IMPLEMENTATION:

N-GRAM CODE

import nltk

from nltk.tokenize import word\_tokenize

text = word\_tokenize("And now for something completely different") print(nltk.pos\_tag(text))

nltk.help.upenn\_tagset('.\*')

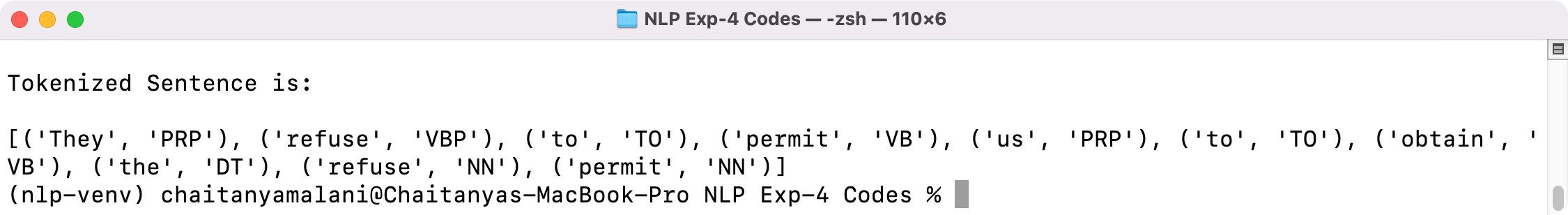
text = word\_tokenize("They refuse to permit us to obtain the refuse permit")

print("\nTokenized Sentence is:\n") print(nltk.pos\_tag(text))

OUTPUT







# CONCLUSION:

Thus we have performed POS Tagging for English and any Indian language using a tagger.