Fr. C. Rodrigues Institute of Technology, Vashi
Department of Computer Engineering

Amroz Siddiqui

Course Code: CSDL7013	Academic Year: Second Half - 2022
Course Name: Natural Language Processing Lab	Branch/Semester: Computer Engineering / VII

Name: ______________________________

Roll No.: ____________

Lab.: 06

1. **Title:** Parts of Speech Tagging
2. **Objective/Aim:**
   - To implement viterbi algorithm for HMM
   - To illustrate parts of speech tagging and exploring tagged corpora using nltk and supporting libraries
3. **Tools/Techniques/Technology Used:** python
4. **Due Date:** Friday September 16, 2022
5. **Lab Instructors:**
   - Amroz Siddiqui
   - Ms. Padmashree

PART 1

- **Precursor to PART 1**
  1. Consider the sentence
     "I closed the last bag".
     Identify the tags for each word of this sentence. Use Universal tagset.
  2. Build transition probability matrix (A) and emission probability matrix (B) for the tags and words from the previous exercise. Use "mystery" category of Brown corpus. Use Universal tagset.
     NOTE: Incorporate the start probability matrix (π) in the first row of matrix A.
- **Solve the following exercise:**
  [05 Marks] Implement the viterbi algorithm using the A and B matrices from the precursor exercises.

PART 2

- *[05 Marks]* **This lab is for exploring NLTK for POST**
  1. Get acquainted with various tags in different corpus. Write explanation about two tags from each corpora. Check other tags, namely, JJ, RB, VB etc. How do you list all tags in a corpora?
  2. Read the given file **hitchhikersguidetogalaxy.txt**. Tokenize and tag each word.
  3. Write code to find most frequent, say, Noun tags, Adjective tags, Verb tags, Adverb tags
  4. (a) Find the next word of a given word.
     (b) Find part of speech tag of the next word of a given word.
     Use different categories of brown words (mystery, news, learned) Use Universal tag.
     Given Words = { often, will, would, chance, likely, what, give, try, early }
  5. Perform Unigram Tagging