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.: 05

1. **Title:** Minimum Edit Distance
2. **Objective/Aim:** To implement Minimum Edit Distance Algorithm
3. **Tools/Techniques/Technology Used:** python
4. **Due Date:** Friday September 09, 2022
5. **Lab Instructors:**
   - Amroz Siddiqui
   - Ms. Padmashree

- There are seven text files given.
- **bagofwords.txt** and **errordocument.txt** to be used for exercise 2.
- The files, **newtonlaws.txt** and **answers1.txt** to **answers4.txt** are for exercise 3.

Solve the given exercise

1. *[03 Marks]* Implement the minimum edit distance algorithm in Python, and find out the distances between following pair of words:
   (apple,mango), (there,their), (laughter,daughter), (rain,reign), (right,write)
2. *[03 Marks]* Words with correct spellings are given in the file **bagofwords.txt**, and in the **errordocument.txt**, each line contains a sentence with few words misspelled. Find the correct word from the bag of words and replace it and write the corrected sentence in **correcteddocument.txt**
3. *[04 Marks]* From the file **newtonlaws.txt**, ignoring typical stopwords, generate the bag of words. From the files **answers1.txt** to **answers4.txt**, find out the number of words misspelled and the degree of wrong spelling, and assign a score to that file. (Think of some relevant metric.)

Checklist for exercises on **Minimum Edit Distance**
The exercises have been designed with specific learning objectives. The following self-assessment list should give you a fair idea of the extent to which you have learnt the expected material. Please respond to the statements after doing the exercises with ratings between 1 and 10. (1 being the lowest)

1. I know how to implement minimum edit distance algorithm in python: _____
2. I know how to build bag of words from given data set in Python: _____
3. I know how to replace misspelled words in a given input document in Python: _____