

Web mining lab -31/07/2018, time: 5:30pm to 7 pm ~Roy

1. Write a program (using **nltk** toolkit in python environment) to tokenize

- a) Sentence
- b) Multiple sentences
- c) A paragraph
- d) Information of a complete web page

2. Write a program to do stop word removal and stemming from a paragraph. Prepare a table (write in a excel file) that includes “Word” and “frequency (2 columns). Find the frequency of words (terms) starting with A,B,S,D,E(You can use **nltk** tool). Find maximum frequency terms/term.