EEP773

Telecom Software Lab

Assignment 9

September 24, 2014

Abhishek Madan

2014JTM2866

Indian Institute of Technology, Delhi

Contents

1 Problem Statement

In this assignment we have to analyze a data file on the basis of different emoticons and words that reflect emotions. A dictionary is given beforehand relation each emoticon to its correpsonding emotion. In this assignment we have to count the percentage of occurence of each emoticon and also find out the mood of each user on the basis of the emoticons he has used the most in his/her conversation. The chat file and the dicitonary have to be pulled from a github account. The output file containing the percentage of usage of each emoticon and user mood has to be pushed as well on the github account.

2 Assumptions

1) Only one emoticon is present in each line.
2) The content.txt file is present in the same directory as the python script file.

3 Explanation of Logic

1) A 7x7 matrix was created to hold the info of 7 emoticons of 7 users.

2) All 14 emoticons were predefined.

3) File to be read and file to be written were opened.

4) Each line of content.txt file was read individually. The first word of each line was the name of the user who was sending the message, which was identified using if-elif.

5) Using another if-elif, it was found that which emoticon was present in wach line. FOr this purpose, the method to search a substring in a given string was used: if "substring" in "string". If yes, then for that particular user, the number of that particular emoticon was incremented by 1.

6) Then it was found out that a particular user used which emoticon the maximum and according to this, his/her mood was written in the output file. This process was repeated for each of the 7 users.

7) Then the total occurence of each emoticon was found out along with the total number of emoticons used. Using this information, percentage of occurence of each emoticon was obtained and was written to tha ouptut file.