EEP702-Software Lab
Assignment7 : Sentiment Analysis in Python

Harshit Kumar Gupta

2013EET2369

Computer Technology

Department Of Electrical Engineering

IIT DELHI

February 20, 2014

Contents

Chapter 1

PROBLEM STATEMENT

Write a program in Python as per the given statement.

1. You are given a text file which contains random facebook status.

2. You have to do sentiment analysis of those posts on the basis of positive,negative and neutral feelings.

3. To differentiate between feelings, create a dictionary having various positive and negative words.

4. Match whether a post has any of those words and if it has, it gets counted into the respective category.Also Include emoticons (for ex. :) for positive and :(for sad).

5. Consider a post having neither positive nor negative feelings as neutral.

6. TASKS

   (a) Count the number of posts with each kind of 'feeling' for a given hour (use a command line tool).

   (b) Make a table with entries "feeling" and "its count in terms of posts in a given hour"

   (c) For each hour,normalize this counted data on the scale of [-1,0,1] i.e. assign weight of -1 to negative feeling, +1 to positive feeling , 0 to neutral feeling and adding all, divide result by total number of posts in that hour.From this calculated data, plot a graph with hour as X-axis and normalized feeling value as Y-axis.

7. Find the respective hours in which most number of posts arrived for each category of feeling.

8. Given any two "geographically separate" places, compare the number of posts in those places containing different category of feelings.

9. Extract the location of the places in the post and give a graphical representation with the place as X-axis and the normalized mood value for the whole file on Y-axis.

Chapter 2

ABSTRACT

The Intention of the Python Code is to practice and get familiar with the python programming. Sentiment analysis or opinion mining is the computational study of people's opinions, sentiments, attitudes, and emotions expressed in written language. It is one of the most active research areas in natural language processing and text mining in recent years. Its popularity is mainly due to two reasons. First, it has a wide range of applications because opinions are central to almost all human activities and are key influencers of our behaviors. Whenever we need to make a decision, we want to hear others' opinions. Second, it presents many challenging research problems, which had never been attempted before the year 2000.

Chapter 3

[INTRODUCTION]{.underline}

Python is a widely used general-purpose, high-level programming language. Its design philosophy emphasizes code readability, and its syntax allows programmers to express concepts in fewer lines of code than would be possible in languages such as C. The language provides constructs intended to enable clear programs on both a small and large scale

Python supports multiple programming paradigms, including object-oriented, imperative and functional programming or procedural styles. It features a dynamic type system and automatic memory management and has a large and comprehensive standard library.

Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials.

Generally speaking, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation (see appraisal theory), affective state (that is to say, the emotional state of the author when writing), or the intended emotional communication (that is to say, the emotional effect the author wishes to have on the reader).

Chapter 4

SPECIFICATIONS AND ASSUMPTIONS

Specifications

1. Different Functions have been used for different objectives.

2. Show function is used to display the answers.

Assumptions

1. There is new line and time stamp between two comments.

2. The time is input as hh:mmam/pm.

3. The file is in utf-8 format.

Chapter 5

LOGIC USED/METHODOLOGY

The methodology that is used for developing the program is defined below:

1. We intiate our code in Python by importing the required libraries.

2. We input the data.txt file and extract the post, time and place and store all the data.

3. Now we create a dictionary with positive and negative words.

4. Next we compare the post with the dictionary and calculate the positive and negative words in every post

5. We normalize the positive and negative words data and plotted graph between hour and normalize value.

6. The values are recorded into a file system.

7. further Searching of the values done from the file named data.txt

8. Records may be deleted added as per specification.

Chapter 6
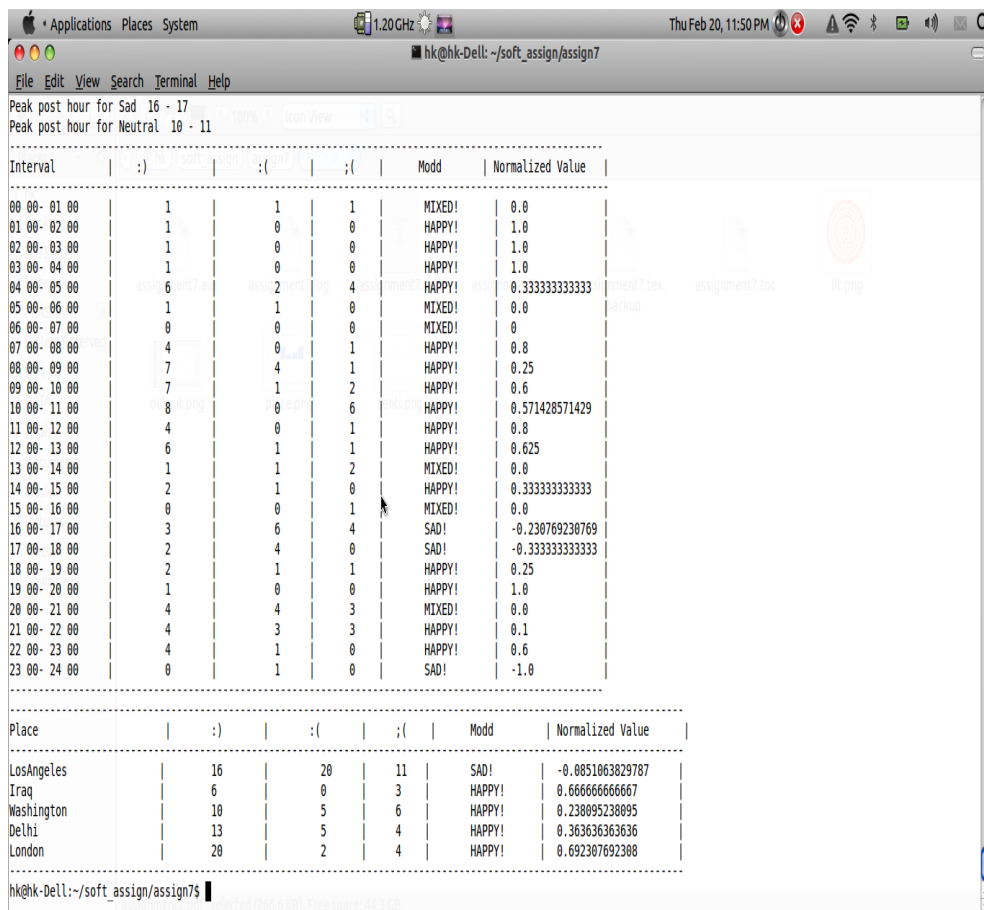
EXECUTION INSTRUCTIONS

1. For program following instructions are used.

   (a) python sentiment.py

Chapter 7

RESULTS AND CONCLUSIONS

For the given set of inputs the program exactly displays the Library management Database.

sentiment analysis per hour

sentiment analysis per city