# Sentiment Analysis

## Rishi Dua, 2010EE50557

February 20, 2014

## 1 BASH PROGRAM

### 1.1 PROBLEM STATEMENT

You are given a text file which contains random facebook status. You have to do sentiment analysis of those posts on the basis of positive,negative and neutral feelings. To differentiate between feelings, create (hardcode) a dictionary having various positive and negative words. 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). Consider a post having neither positive nor negative feelings as neutral.
Tasks:

1. Count the number of posts with each kind of feeling for a given hour

2. Make a table with entries feeling and its count in terms of posts in a given hour

3. 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.

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

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

6. 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.

7. Plot the overall feelings of a location normalized against all other locations

## 1.2 ABSTRACT

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.

Millions of people use Facebook everyday to keep up with friends, upload an unlimited number of photos, share links and videos, and learn more about the people they meet. Comment Policy: We love your comments, but please be respectful of others. Processing this is an important part of Sentiment Analysis.

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, affective state , or the intended emotional communication.

## 1.3 SPECIFICATION AND ASSUMPTIONS

**Tool Specifications:**
Language used: Python 2.7.3 (default, Sep 26 2013, 20:03:06) [GCC 4.6.3] on linux2
Platform: Ubuntu 12.04
Additional tools used: gnuplot
Bash Version: GNU bash, version 4.2.25(1)-release (x86_64-pc-linux-gnu)

**Problem specifications**: Assume the following utilities are present on the system
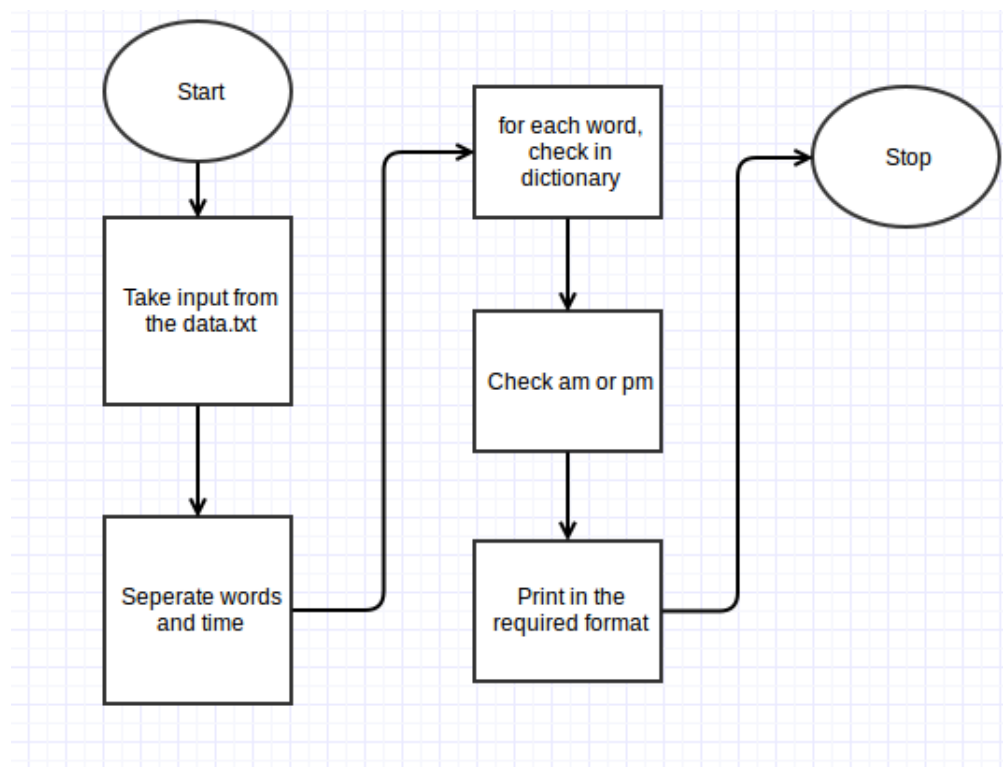gnuplot
gnuplot python libraries

**Assumptions**
User has read and write prividlges to run the program, create files and directories
If a status has both positive and negative sentiment, it is counted in both

## 1.4 FLOW CHART

The flowcharts is as follows

## 1.5  LOGIC IMPLEMENTATION

1. Take input from the data.txt

2. Separate words and time

3. for each word, check in dictionary

4. check am or pm

5. print in the required format

6. Plot the normalized value using gnu

## 1.6  EXECUTION DIRECTIVE
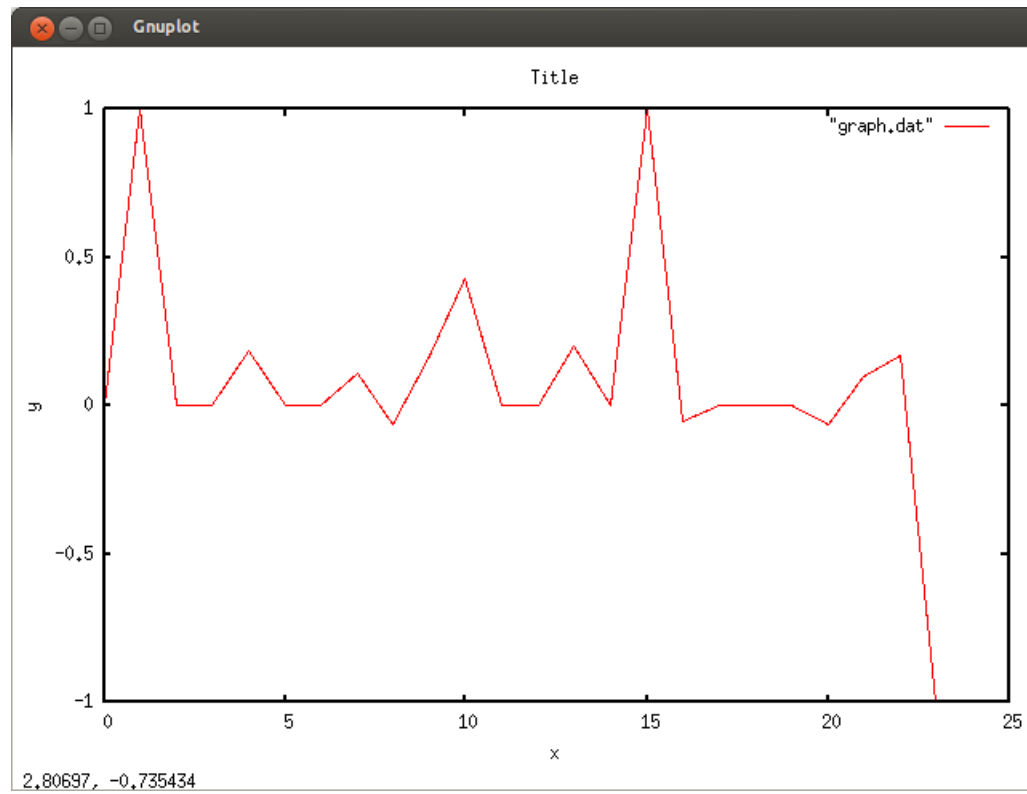
No compilation required. Directly run by typing

python code.py

## 1.7 Output Of The Program

```
rishi@rishi: /media/teknovates/iit/sem8/eep702/ass7/code

rishi@rishi: /media/teknovates/iit/sem8/css310/...    ✖    rishi@rishi: /media/teknovates/iit/sem8/eep702...    ✖


rishi@rishi:/media/teknovates/iit/sem8/eep702/ass7/code$ python code.py
Interval            :)        :(        ;)        Comment

_____          _____     _____     _____     _____

0:00-1:00           2         2         30        SAD!
1:00-2:00           1         0         0         HAPPY!
2:00-3:00           0         0         1         NONE!
3:00-4:00           0         0         1         NONE!
4:00-5:00           8         5         3         HAPPY!
5:00-6:00           1         1         1         SAD!
6:00-7:00           0         0         0         NONE!
7:00-8:00           5         4         0         HAPPY!
8:00-9:00           4         5         6         SAD!
9:00-10:00          5         3         4         HAPPY!
10:00-11:00         7         1         6         HAPPY!
11:00-12:00         4         4         1         SAD!
12:00-13:00         0         0         2         NONE!
13:00-14:00         2         1         2         HAPPY!
14:00-15:00         0         0         3         NONE!
15:00-16:00         1         0         0         HAPPY!
16:00-17:00         8         9         1         SAD!
17:00-18:00         4         4         2         SAD!
18:00-19:00         2         2         2         SAD!
19:00-20:00         1         1         0         SAD!
20:00-21:00         5         6         5         SAD!
21:00-22:00         3         2         5         HAPPY!
22:00-23:00         2         1         3         HAPPY!
23:00-24:00         0         1         0         SAD!

_____          _____     _____     _____     _____

gnuplot> set terminal x11
gnuplot> clear
gnuplot> set title "Title"
gnuplot> set xlabel "x"
gnuplot> set ylabel "y"
gnuplot> plot "graph.dat" with lines
```
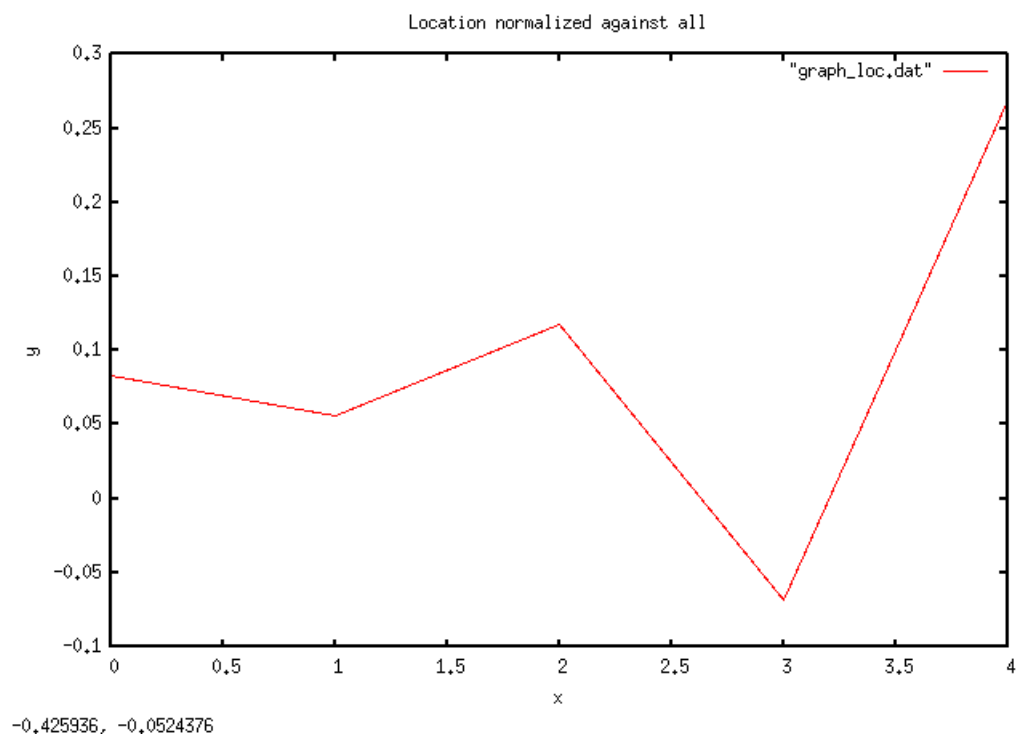
```
_____        _____    _____    _____    _____

Enter place 1 (eg: LosAngeles): LosAngeles
Enter place 2 (eg: Iraq): Iraq
LOCATION 1:     LosAngeles

Interval         :)       :(       ;)       Comment

0:00-1:00        0        0        2        NONE!
1:00-2:00        0        0        0        NONE!
2:00-3:00        0        0        0        NONE!
3:00-4:00        0        0        0        NONE!
4:00-5:00        1        0        1        HAPPY!
5:00-6:00        0        0        1        NONE!
6:00-7:00        0        0        0        NONE!
7:00-8:00        1        0        0        HAPPY!
8:00-9:00        0        0        2        NONE!
9:00-10:00       0        0        1        NONE!
10:00-11:00      0        0        0        NONE!
11:00-12:00      1        1        0        SAD!
12:00-13:00      0        0        1        NONE!
13:00-14:00      1        0        1        HAPPY!
14:00-15:00      0        0        0        NONE!
15:00-16:00      0        0        0        NONE!
16:00-17:00      7        8        0        SAD!
17:00-18:00      4        4        0        SAD!
18:00-19:00      0        0        0        NONE!
19:00-20:00      0        0        0        NONE!
20:00-21:00      4        5        4        SAD!
21:00-22:00      2        0        4        HAPPY!
22:00-23:00      1        0        3        HAPPY!
23:00-24:00      0        1        0        SAD!
LOCATION 2:     Iraq

Interval         :)       :(       ;)       Comment
_____        _____    _____    _____    _____

0:00-1:00        1        1        0        SAD!
1:00-2:00        0        0        0        NONE!
2:00-3:00        0        0        0        NONE!
3:00-4:00        0        0        0        NONE!
4:00-5:00        1        1        0        SAD!
5:00-6:00        0        0        0        NONE!
6:00-7:00        0        0        0        NONE!
7:00-8:00        0        0        0        NONE!
8:00-9:00        0        0        0        NONE!
9:00-10:00       2        2        1        SAD!
10:00-11:00      1        0        2        HAPPY!
11:00-12:00      1        1        0        SAD!
12:00-13:00      0        0        0        NONE!
13:00-14:00      0        0        0        NONE!
14:00-15:00      0        0        0        NONE!
15:00-16:00      0        0        0        NONE!
16:00-17:00      0        0        0        NONE!
17:00-18:00      0        0        0        NONE!
18:00-19:00      0        0        0        NONE!
19:00-20:00      0        0        0        NONE!
20:00-21:00      0        0        0        NONE!
```
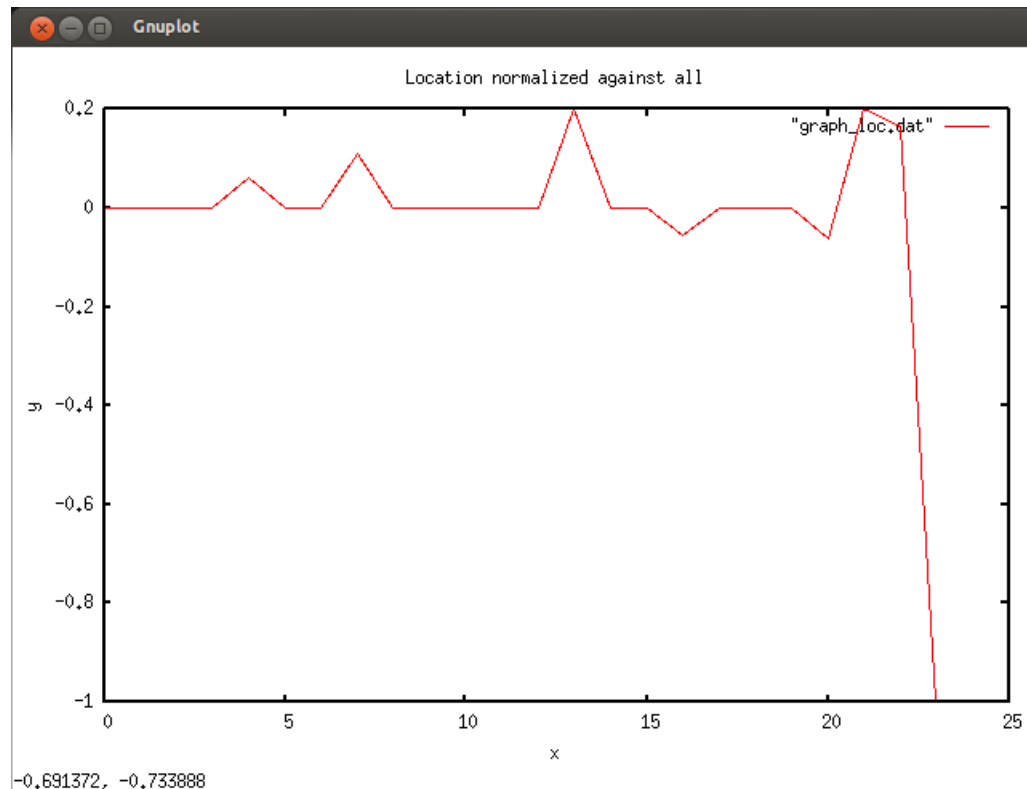
7

Location normalized against all



-0.425936, -0.0524376

## 1.8 RESULT

The program successfully does the sentiment analysis of given Facebook data

**Problems encountered:**

1. Write permissions error

2. Execute rights error

**Solution**: check and set permissions using chmod

In Unix-like operating systems, chmod is the name of a Unix shell command and a system call, which both change the access permissions to file system objects (including files and directories), as well as specifying special flags.

The data file given for this assignment has the following issues:

1. No proper format. For seniment analysis, the raw file generated by crawler is parsed and used. However, the file given is generated by manually copying the text, hence has a lot of errors

2. The file has non readable characters (unicode characters inconsistent with the file encoding) and required a cleanup resulting in some loss of data.

## 1.9 CONCLUSION

Learnt about Python programming

Python is a multi-paradigm programming language: object-oriented programming and structured programming are fully supported, and there are a number of language features which support functional programming and aspect-oriented programming (including by metaprogramming and by magic methods). Many other paradigms are supported using extensions, including design by contract and logic programming.