# Software Lab Assignment - 7 : PYTHON

## Basic Reading Material

- URL: http://docs.python.org/2/tutorial/
- Plotting software
    - (easy) GNUplot : people.duke.edu/ hpgavin/gnuplot.html
    - (slightly more involved) Graphviz
    - (versatile) Qt (5.1) - this is also next week's tool to learn

## PROBLEM STATEMENT

## Problem 1:(compulsory:100 marks)

- 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 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**:
    - Count the **number** of posts with each kind of 'feeling' for a given hour (use a command line tool).
    - Make a table with entries "feeling" and "its count in terms of posts in a given hour".eg.

| Interval | :) | :( | ;( | Comment |
|---|---|---|---|---|
| 0800-0900 | 2 | 44 | 1 | SAD ! |
| 0900-1000 | 100 | 1 | 1 | HAPPY ! |

    - 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.

## Problem 2:(Difficulty level:**; Bonus: 5 marks)

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

## Problem 3:(Difficulty level:***; Bonus: 5 marks)

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

## Problem 4:(Difficulty level:****; Bonus: 5 marks)

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.