Report on

SENTIMENTAL ANALYSIS USING NLTK

BACHELOR OF SCIENCE
IN

COMPUTER SCIENCE ENGINEERING

REPORT SUBMITTED BY

SIDDHARTH DALAL 17070122063
URJIT TEMBHURNIKAR 17070122069
VARUN SAHA 17070122073
GUNVANT RANE 18070122512

SYMBIOSIS

Under the Guidance of

PROF. POOJA V. KAMAT

SYMBIOSIS INSTITUTE OF TECHNOLOGY

(A CONSTITUTENT OF SYMBIOSIS INTERNATIONAL UNIVERSITY)

Pune - 412115

Software Requirement Specification

Table of Contents

1. INTRODUCTION

1.1 Purpose

The purpose of this document is to give a brief idea regarding the project. Our project is about analyzing the twitter opinions of the people on the movies. This project works on the API provided by the twitter which is called tweepy. All the functions and the working idea behind the project is introduced in the document.

1.2 Feasibility Study

There are some software available in the market that uses tweepy API to predict the sentimental analysis but they just state the readings based on the perticular time when it was executed. We are planning to record the data for a week and plot the graph using the gatherd data.This will help us for analysing the statical analysis that how the movie has impact on the audience during week, From that analysis we can predict that what is future results and statics will reach in future.

1.3 Intended Audience

Intended audience for using this software are statistical analysers, students and also the public that want to analyse the review of any movie that is recently released of already available. By analysing using this software it will be easy to decide weather to watch that movie or not.

1.4 Product Scope

The functionality of this project is based upon the twitter comments. Based on the comments made by people the tweepy API analyse the sentiments of people on the basis of the keywords in the comments.So this limitations of this project is that it is limited to twitter for now.In future we will try to extend its scope so it can be used over different applications running on the net.

1.5 References

1) Documents available on Data camp
2) YouTube tutorials.
3) Tweepy Documentation (http://docs.tweepy.org/en/latest/)
4) TextBlob Documentation (https://textblob.readthedocs.io/en/dev/quickstart.html)

2. Overall Description

2.1 Product Perspective

The product name is self describing. The Sentimental analyzer is designed to analyse tremendous number of tweets on twitter and give us the total percentage of positive as well as negative tweets . The input format can be a tweet #'__' and output will be the percentage of both positive and negative tweets .

2.2 Product Functions

2.2.1 Analyzer :

1) Taking the input from the user

2) Taking the input from the user and search all the related . on twitter as well as analyzes them.

2.2.2 Graph plotter :

1) Translating the data into statistical format that is graphs. (inputted by the user).

2) also storing the data (positive/negative pole)

2.3 User Classes and Characteristics

This part is to identify various user classes that will use the web application. User classes will be differentiated based on the use, product functions and features, technical expertise, security and privilege levels and educational level. No special knowledge or skills should be assumed for the part of the regular users. Users are not expected to learn or remember a set of commands in order to start using the application.
The following clearly describes the role of each participant.

- Users: users with no particular knowledge needed, users who are interested to use the tool looking for knowing people's thoughts about a desired topic

- System Administrators:
    1. Develop and maintain installation and configuration procedures and operational requirements
    2. Perform weekly/monthly backup operations, ensuring all required files and data are successfully backed up
    3. Repair and recover from hardware or software failures

2.4 Operating Environment

Windows OS has been used for testing purposes. Hence, the software is made to work in Windows at present.

1)Access to internet

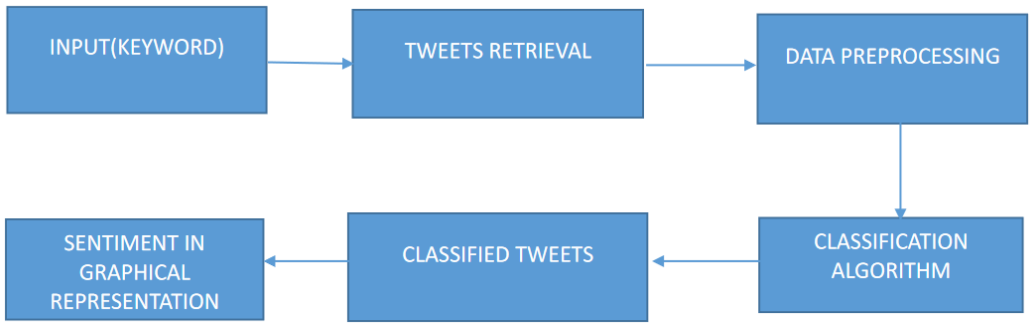
2)Browser

3)Twitter account

4)python version 3.7.2

2.5 Design and Implementation Constraints

Design:



Constraints:

- Cannot work offline.
- The software is designed to go through only Twitter opinions.

2.6 Process Model

We are planning to use the waterfall model.

Reasons to choose waterfall model:
The primary reason of choosing this model is that all the requirements are known, fixed and clear prior to implementing the project. There are sufficient resources and time available.

2.7 Assumptions and Dependencies

Assumptions:

- The user should have an active internet connection.
- The input should be valid.
- It is assumed that the software designed will work correctly.

Dependencies:

- Python 3.0+
- Tweepy (to access Twitter API)
- TextBlob (to perform sentiment analysis)
- Matplotlib (for visual representation in graphical form)

Dependencies can be installed using ‘pip’.

3. Functional Requirements

3.1 Requirement 1

Purpose : Login/Sign up Function
Input : User ID and Password
Output : Display successful authentication and allow user to use the desktop application

3.2 Requirement 2

Purpose : Search movie name followed by #
Input : Movie name
Output : Sentimental Analysis Report(poles)

3.3 Requirement 3

Purpose : Statistical Report in tabular form
Input : Backend from database
Output : Table

3.4 Requirement 4

Purpose : Statistical Report in Graphical form
Input : Backend from database
Output : Graphs/pie chart

3.5 Requirement 5

Purpose: See top 10 retweeted tweets
Input: backend from database
Output :top 10 tweets

3.6 Requirement 6

Purpose: Search tweets by username
Input: Twitter username
Output : Show tweets by the user