Software Requirements Specification

For

Sentiment Analysis of Geo-Tagged Tweets

**Prepared by:**

**Chinmay Parikh (A059)**

**Qais Makani (A029)**

**Yash Rathod (A042)**

**Date: 04-10-2014**

**Table of Contents**

**Table of Contents**

**Revision History**

**1. Introduction**

1.1 Purpose

1.2 Document Conventions

1.3 Intended Audience and Reading Suggestions

1.4 Product Scope

1.5 References

**2. Overall Description**

2.1 Product Perspective

2.2 Product Functions

2.3 User Classes and Characteristics

2.4 Operating Environment

2.5 Design and Implementation Constraints

2.6 Assumptions and Dependencies

**3. External Interface Requirements**

3.1 User Interfaces

3.2 Hardware Interfaces

3.3 Software Interfaces

3.4 Communications Interfaces

**4. System Features**

4.1 System Feature 1

4.2 System Feature 2

**5. Sentiment Analysis of Tweets**

5.1 Stimulus/Response Sequences

5.2 Functional Requirements

**6. Graphical Representation**

6.1 Stimulus/Response Sequences

6.2 Functional Requirements

**7. Other Nonfunctional Requirements**

7.1 Performance Requirements

7.2 Safety and Security Requirements

7.3 Software Quality Attributes

7.4 Business Rules

1. **Introduction**

**1.1 Purpose**

This document aims to provide literary context to the Sentiment Analysis of

**1.2 Document Conventions**

Main Section Titles

* Font: Times New Roman
* Face: Bold
* Size: 16

Sub Section Titles

* Font: Times New Roman
* Face: Bold
* Size: 14

Other Text Explanations

* Font: Times New Roman
* Face: Normal
* Size: 12

**1.3 Intended Audience and Reading Suggestions**

System users: People using the system to analyze the tweets.

System Developer: For further enhancements to the system.

System Tester: For testing the system to ensure it meets the requirements and

**1.4 Product Scope**

This document aims to provide literary context to the sentiment analysis system which gathers and analyses the tweets from twitter and classifies them as being positive or negative. The analysis is done using techniques like Natural Language Processing and Machine Learning.

**1.5 References**

**Sentiment Analysis**

The process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral.

**Tweet**

A tweet is 140 character long text posted on the social networking website call twitter.

**Machine Learning**

It is a subfield of computer science and statistics that deals with the construction and study of systems that can learn from data, rather than follow only explicitly programmed instructions.

**Natural Language Processing**

It is a field of computer science, artificial intelligence, and linguistics concerned with the interactions between computers and human (**natural**) languages. As such, NLP is related to the area of human–computer interaction.

1. **Overall Description**

**2.1 Product Perspective**

Sentiment analysis involves using techniques like NLP and Machine Learning to make a computer able to understand whether the tweet has negative or positive sentiments. From a user’s perspective, this system will be useful to finding out how people reacted to a certain topic.

**2.2 Product Functions**

* Gather Tweets
* Parse and pre-processing tweets
* Storing tweets in a database
* Analysis of tweets
* Graphical representation of the results

**2.3 User Classes and Characteristics**

Businesses or distinguished individuals like politicians who are interested in knowing their or their product.

**2.4 Operating Environment**

It works on all systems since it is written in python. It only requires python 2.7.x to be installed on the system along with the necessary libraries.

**2.5 Design and Implementation Constraints**

The main design constraint is that the system needs to be trained for a specific domain, like understanding sarcasm.

* 1. **Assumptions and Dependencies**

The main dependency in Sentiment Analysis is the Twitter API and performing Sentiment Analysis. Well we aren’t sure about the results, from where the tweets has come from because people also put a wrong address in their address field while registering.

1. **External Interface Requirements**

**3.1 User Interfaces**

A user interface will be provided to accept the topics the user wants to analyze. It will also accept location based filtering, in case the user is interested in tweets coming from a specific location. The results will be color coded (red for negative and green for positive) and displayed on a 3D globe.

**3.2 Hardware Interfaces**

* The processing power required depends on the amount of data to be gathered, stored, and analyzed.
* A graphics processor will be required to display the 3D globe.
* The computer must be connected to the internet with access to twitter and must have a hard disk drive to store the tweets.
* It must also have a keyboard and a mouse to accept user input
* A screen to display output.

**3.3 Software Interfaces**

* An OS with python 2.7.x and other required libraries.
* SQLlite
* Web Browser

**3.4 Communications Interfaces**

Protocol used is TCP. It can use any port.

1. **System Features**

**4.1 Tweet Gathering**

**4.1.1 Stimulus/Response Sequences**

1. Accepting the filters (topic and location).

2. Streaming the tweets that pass the filter

3. Storing in JSON file

**4.1.2 Functional Requirements**

1. Machine Learning

2. SQLite

**4.2 Storing the tweets in database**

**4.2.1 Stimulus/Response Sequences**

1. Pre-processing the data

2. Parsing the data using JSON parsers

3. Create an SQL table with columns corresponding to attributes of tweets required

4. Store the tweets in the database

**4.2.2 Functional Requirements**

1. Database Functions

2. Block management

**5. Sentiment Analysis of tweets**

**5.1 Stimulus/Response Sequences**

1. Query table to fetch the tweets
2. Analyze tweets using an algorithm like Naïve Bayesian
3. Write the results to another table

**5.2 Functional requirements**

1. Accuracy
2. Flexibility
3. Time to execute
4. **Graphical Representation**

**6.1 Stimulus/Response Sequences**

1. Read the results from the table
2. Display them on a globe according to location
3. Negative tweets will be displayed in red, positive in green.

**6.2 Functional requirements**

1. 3D globe should be interactive
2. Zooming ability
3. Tweets should be displayed correctly at their respective locations
4. **Other Nonfunctional Requirements**

**7.1 Performance**

The performance of the system is a major concern since it is highly likely that it will work on extremely large data sets.

**7.2 Safety and Security Requirements**

The system doesn’t contain any confidential data since it is available on twitter’s public stream. Therefore, it doesn’t have any specific security requirements. The only security requirement would be to hide the content of tweets since twitter doesn’t allow the content to be displayed.

**7.3 Software Quality Attributes**

The source code of the product is going to be open as this is going to be open source software. It will be free for further modifications and improvements.

**7.4 Business Rules**

There cannot be any business rules, as the system is open source.