

Software Requirements Specification

Version 1.0
10.4.2012

Political Mood Gauge Based on Social Media

Team Vikings:

Gresham Schlect
Serendel Macpherson
Ryan Sacksteder

Prepared for CS383
University of Idaho - Dr. Paul Oman
Fall 2012

Table of Contents

1.0 Introduction

- 1.1 Purpose
- 1.2 Scope
- 1.3 Definitions, Acronyms, and Abbreviations
- 1.4 References
- 1.5 Overview

2.0 Overall Description

- 2.1 Product Perspective
- 2.2 Product Functions
- 2.3 User Characteristics
- 2.4 Constraints
 - 2.4.1 Personal Data
 - 2.4.2 Twitter Data
- 2.5 Assumptions and Dependencies
 - 2.5.1 Twitter Population
 - 2.5.2 Sentiment Analysis
 - 2.5.3 Political Parties

3.0 Specific Requirements

- 3.1 External Interface Requirements
 - 3.1.1 Hardware Interfaces
 - 3.1.1.1 Single-User System
 - 3.1.1.2 Multi-User System (Optional)
 - 3.1.2 Software Interfaces
 - 3.1.2.1 Inputs
 - 3.1.2.2 Outputs
 - 3.1.3 User Interfaces
 - 3.1.3.1 Graphic 1: Mood of the Nation Gauge (Mandatory)
 - 3.1.3.2 Graphic 2: Mood of the Nation Time-Lapse (Optional)
 - 3.1.3.3 Graphic 3: Mood of the Nation by State (Optional)
- 3.2 Functional Requirements
 - 3.2.1 Real-Time Processing
 - 3.2.2 Retrieving Social Media
 - 3.2.3 Filtering Tweets
 - 3.2.4 Parsing Tweets
 - 3.2.5 Weighting Tweets
 - 3.2.6 Sentiment Analysis
 - 3.2.7 Aggregation
 - 3.2.8 Output
- 3.3 Performance Requirements
 - 3.3.1 Real-Time
 - 3.3.2 Tweet Throughput
- 3.4 Use Cases

3.5 Design Constraints

3.6 Software System Attributes

3.6.1 Reliability

3.6.2 Availability

3.6.3 Security

3.6.4 Maintainability

3.6.5 Portability

1.0 Introduction

1.1 Purpose

The purpose of this document is to outline the requirements which the social media political mood gauge software must meet. The audience of this document includes: Project developers and overseers, and users who wish to view the project requirements and specifications.

1.2 Scope

This purpose of this project is to develop an automated way of gauging the political mood of the country based on social media, specifically Twitter. The software should determine whether the country is leaning toward more conservative or more liberal politics, without looking at traditional news forums. The product name will be Twitter Political Mood Gauge. The project is part of the coursework for CS383 at the University of Idaho. It is meant to provide users with a quick snapshot of the nation's political mood.

1.3 Definitions, Acronyms, and Abbreviations

1.3.1 Social Media - technologies that create online communities/networking capabilities, e.g. Twitter, Facebook, MySpace and/or internet blogs

1.3.2 Twitter - a social media website that allows users to publicly share short sentences, or Tweets

1.3.3 Tweet - a text string posted by a Twitter user (140 characters or fewer)

1.3.4 Retweet - the process of reposting someone else's tweet on one's own profile

1.3.5 Follower - one user on Twitter can 'follow another in order to directly receive their Tweets

1.3.6 Sentiment Analysis - the process of analyzing a statement and determining its degree of positivity or negativity (also known as sentiment)

1.4 References

Twitter: <https://twitter.com/about>

Twitter Political Index: <https://election.twitter.com/>

SRS Template: www.jsu.edu/mcis/docs/SRSTemplate.doc (Jacksonville State University MCIS Dept.)

1.5 Overview

The remaining sections of this document will cover an overall description of this software, as well as requirements and specifications for the software. The overall description of the software is a reference for users, and the detailed specifications and requirements should guide development. Section 2 (Overall Description) will cover the system environment, functional requirement specifications, and user interface. Section 3 (Specific Requirements) will cover external interface requirements and any functional/non-functional requirements.

2.0 Overall Description

2.1 Product Perspective

This product is similar to the Twitter Political Index in that it uses Twitter as a form of social media in order to gauge of how Twitter users are discussing politics. It is different however, because it does not limit the analysis to just the Presidential Candidates. This application will rather examine the mood with which the Liberal and Conservative political parties, each as a whole, are discussed on Twitter. It will also provide more meaningful and more easily understood output than the Twitter Political Index. Depending on time constraints, the software will also display additional forms of output beyond a simple gauge comparison of the two main political parties.

2.2 Product Functions

The application will provide the user with a snapshot of the nation's current political mood. Twitter will be used for social media retrieval, returning tweets that contain specific keywords. The information returned will be filtered, sent through a sentiment analysis tool, weighted according to possible influence, and sent to the application's output. The user will have minimized interaction with the application, and will only be allowed to view the output and switch between forms of output. For example, a simple gauge will allow the user to identify the most current political mood of the nation, while another graph will provide a time-lapse view of recent trends in the nation's political mood.

2.3 User Characteristics

The intended user will be a member of the general public who is interested in the current political mood of the nation. Users are not expected to have a very high level of technical expertise, but should have a general understanding of how to run the application given the installation instructions provided with the software.

2.4 Constraints

2.4.1 Personal Data

Twitter does not return any information about the user which that user has not made public. Any personal information collected from Twitter will not be stored or used in anyway, except to anonymously categorize the tweet.

2.4.2 Twitter Data

The program must comply with Twitter's terms of service. All information returned from Twitter is to be considered Twitter's data and must be treated accordingly.

2.5 Assumptions and Dependencies

2.5.1 Twitter Population

It will be assumed, for the purposes of this application, that the Twitter population is an accurate representation of the entire nation.

2.5.2 Sentiment Analysis

It will be assumed that the sentiment will be returned accurately for 140 characters.

2.5.3 Political Parties

It will be assumed, for the purposes of this application, that a user within the Twitter population will be either a republican or a democrat; i.e. negative sentiment toward one of the parties is equivalent to positive sentiment toward the other.

3.0 Specific Requirements

3.1 External Interface Requirements

3.1.1 Hardware Interfaces

The application is intended to be a stand-alone, single-user system with the capability of being deployed as a multi-user system over a server in the future. The following are short descriptions of each of these systems.

3.1.1.1 Single-User System

In the single-user system environment, the application will run on the user's local machine. No further hardware devices or interfaces will be required.

3.1.1.2 Multi-User System (Optional)

In the event that the multi-user system environment is implemented, the application will be deployed from a server, allowing multiple users to access the output of the application via a project website on their local machine.

3.1.2 Software Interfaces

3.1.2.1 Inputs

The software will take social media as input, specifically current tweets containing predetermined keywords corresponding to the liberal and conservative political parties. Each tweet is expected to be retrieved in JSON format.

3.1.2.2 Outputs

The output will portray the current political mood of the nation (as determined by the application) in the form of a simple gauge. Time permitting, the output will also include a graphical time-lapse view of recent political mood trends seen in the simple mood gauge over the past few hours, as well as a state-by-state map gauge of current political moods. A web browser will be necessary for viewing output.

3.1.3 User Interfaces

The interface will meet the following requirements to conform to the users' needs. It will be simple and easy to understand. Controls which allow the user to interact with the application will be clear and imply their functionality within the application. The interface will consist of a number of graphics. These graphics will all be displayed on a single webpage, or otherwise be separated in a way in which the user can specify which graphics they would like to see. The graphics displayed to the user will be the main interface between the user and the application, providing the user with a graphical representation of the output produced.

3.1.3.1 Graphic 1: Mood of the Nation Gauge (Mandatory)

This graphic is guaranteed to be in the final product of the application. It will provide a simple gauge which shows the current political mood of the nation. This will be done by giving the percentage of the nation who are currently pro-conservative versus a percentage of the nation who are pro-liberal. The gauge will use the color red to represent pro-conservative, and blue to represent pro-liberal. It will also display the total number of tweets which have been processed

in order to provide this output.

3.1.3.2 Graphic 2: Mood of the Nation Time-Lapse (Optional)

This graphic is intended to be in the final product; however implementation is dependent on time constraints. If available, this graphic will show the most recent trends in the nation's political moods. These trends will be constructed by periodically taking snapshots of the gauge described in section 2.3.3.1. These snapshots will be displayed in a clear and meaningful way in order to provide the user with a time-lapse view of the nation's political mood. The user may interact with the time-lapse graph by hovering or clicking on a particular point on the line graph to see the specific numeric gauge value at that point in time.

3.1.3.3 Graphic 3: Mood of the Nation by State (Optional)

This graphic is also not mandatory but is desired. It would provide a breakdown of the political mood by state. Using the Twitter API and a map service API, coordinates for particular tweets will be used to calculate the origin of the tweet. Results can then be broken down to a per-state basis. Color variations will be used to express the degree of political stance a particular state has. For example, a dark red would suggest a strong conservative stance, while a light blue would suggest a slightly liberal stance. A neutral color (such as grey or black) would be used when necessary to depict a neutral political stance. Similar to the other graphics, the state-by-state breakdown would be updated real-time. The user may interact with this map by hovering or clicking on a state to see the specific numeric gauge value for that state.

3.2 Functional Requirements

3.2.1 Real-Time Processing

The software will take input, process data, and display output real-time. This will enforce that the snapshot provided by the simple gauge, is a current view of the nation's political mood.

3.2.2 Retrieving Social Media (Input)

The software will input social media (specifically tweets) in an easily parsable format (most likely JSON). Along with the text for each tweet, additional information should be retrieved, such as number of retweets for the tweet, the number of followers for the given Twitter user, and any other necessary data for processing. The Twitter API is expected to be the primary means of retrieving tweets. Only tweets with predetermined keywords are intended to be retrieved.

3.2.3 Parsing Tweets

The software will parse the social media data into a plaintext string (to be later sent off for sentiment analysis). Any other information which is needed (e.g. number of retweets, number of followers, etc.) will be parsed and stored for processing. To maintain speed and performance, a 3rd party parser will likely be used.

3.2.4 Filtering Tweets

The software will send parsed tweets through a filtering algorithm to discard any 'flagged' tweets. Flagged tweets include tweets which have already been seen and processed by the software, tweets which contain text written in a language not handled by the software, tweets which are outside the boundaries of the United States, and tweets which do not contain any of the predetermined keywords.

3.2.5 Weighting Tweets

The software will use a weighting algorithm to establish the 'weight' of a given tweet. Weight will likely be dependent on an estimate of the overall exposure a particular tweet has. For example, number of followers, as well as number of times a post has been retweeted are all likely considerations for the weight assigned to a specific tweet. The greater the weighting of a tweet, the more of an effect the tweet will have on the overall end results after processing.

3.2.6 Sentiment Analysis

The software will need to perform sentiment analysis on each tweet it processes. Given the keywords in a particular tweets, the sentiment analysis of that tweet will help to determine the mood of the tweet towards those keywords. The sentiment analysis will provide a negative, neutral, or positive sentiment which can then be converted to a numeric value for processing. Because of time constraints and the complexity of sentiment analysis, the software will likely use a 3rd party tool for analysis. The 3rd party tool will also likely help in maintaining the speed and performance necessary for the software to maintain real-time processing.

3.2.7 Aggregation

The software will aggregate a set amount (probably between 100 and 200) of the most recent tweets in order to maintain real-time results. The keywords, sentiment analysis, and weight of each tweet will be used to calculate the political mood (pro-liberal or pro-conservative) of the tweet. A summation of all the tweets and their political moods will be written to an output file in an easily interpreted form. As more tweets are processed and enter the aggregation process, older tweets should be discarded to keep output results current. The most recent results will then overwrite the previous data stored in the output file.

3.2.8 Output

The software will output results from processing in the form of a web page. Given the hardware interface in section 3.1.1.1 of this document, the web page will be viewed locally on the user's machine. If time constraints allow for the hardware implementation discussed in section 3.1.1.2, the web page will be viewable on the internet to multiple users at once. The web page will likely be implemented with HTML/CSS, and will also use JavaScript. The output file mentioned in section 3.2.7 of this document will be read by the web page, parsed, and used to populate one or more gauges displayed on the web page. These gauges will show the current political mood as determined by the processing algorithms. The web page will need to continuously poll the output file in order to present real-time data.

3.3 Performance Requirements

3.3.1 Real-Time

The software will provide up-to-date information, limited only by social media input. The gauge output should incorporate the latest results at all times, and if it lags behind, the user should be notified as part of the output.

3.3.2 Tweet Throughput

The software should process tweets at a rate to match the input. The input may be limited by any number of external factors, including internet connection speed and social media input speed. The

software will be unusable if fewer than 100 tweets can be processed to create an aggregate result at any given time, excluding the period when fewer than 99 tweets have been delivered to the software as the application starts.

3.4 Use Cases

The only foreseeable use case is that the software will serve as a tool of interest, providing intrigued parties with the current political mood of the nation according to the Twitter population. Over time, the software could be used for research and statistical data mining related to the nation's current political stance.

3.5 Design Constraints

The software should be designed to run on either a server (multi-user system) or an individual's computer (single-user system). This means that the software should run smoothly on the hardware of an average user's system.

In addition to the resources required to operate a browser, suggested single-user hardware is as follows:

- 5Mb free disc space
- 2Gb of RAM
- Internet Access

3.6 Software System Attributes

3.6.1 Reliability

The software will meet all of the functional requirements without any unexpected behavior; at no time should the gauge output display incorrect or outdated information without alerting the user to potential errors.

3.6.2 Availability

The software should be available at all times on a single-user system. The functionality of the software may be dependent on any external services that are used; if those services are unavailable, the user should be cleanly alerted. On a multi-user system, the availability will also be dependent on the server and website hosting solution.

3.6.3 Security

The software should never disclose any personal information of Twitter users, and should collect no personal information from its own users. In a multi-user system the output of the application may be publicly available.

3.6.4 Maintainability

The software should be written clearly and concisely. The code will be well commented and documented. Particular care will be taken to design the software modularly to ensure that maintenance is easy.

3.6.5 Portability

The software process that gathers and generates the political mood gauge data will be designed to run only on a Linux operating system. Support and instructions will be specific to a debian distro. For a multi-user system the data crunching runs on a server (Linux), and the output can be viewed on other machines. Any machine running an operating system that supports modern web browsing should be capable of viewing the output.