
SENTIMENT ANALYSIS OF SOCIAL MEDIA WITH
DATA VISUALISATION

BY

JOHN KELLY

SOFTWARE DEVELOPMENT

FUNCTIONAL SPECIFICATION

5TH APRIL 2017

TABLE OF CONTENTS

Abstract	2
1. Introduction	3
2. Target Market	4
3. Metrics	6
4. Project Difference	7
5. System Architecture	8
6. Supplementary Specification	9
6.1 Functionality	9
6.2 Usability	9
6.3 Reliability	9
6.4 Performance	9
6.5 Supportability	10
6.6 Security	10
7. Iteration Plan	11
7.1 Iteration 1	11
7.1.1 Schedule Of Work	11
7.1.2 Changes in requirement	11
7.2 Iteration 2	12
7.2.1 Schedule Of Work	12
7.3 Iteration 3	13
7.3.1 Schedule Of Work	13
7.3.2 Change in Requirements	13
8. Bibliography	14

ABSTRACT

The purpose of this document is to outline the functional and non-functional requirements for this project. It will state who is the target market of this application and why they might use it. This document also states the work that is scheduled for each of the iterations.

1. INTRODUCTION

This project is a sentiment analysis tool that can be used to mine the public's opinion to calculate the sentiment of a certain brand. This brand can be a product, person, sports team, etc. In this project, the word brand will be replaced with the word term. This term will then be used to retrieve tweets from Twitter using the Twitter Search API or an alternative social media.

This project will allow the user to search for a term, and enable them to see how the users of Twitter feel about that term. This will be done by first querying the database to check if it contains any relevant data associated with the searched term and if no data is found, start querying Twitter's Search API to reveal results. This will facilitate the ability to giving the user the option to view the sentiment of the chosen topic in the present as well as giving the option to show the trends in the past.

The user will be able to have several views of the outcome of the data returned from the server. They will be presented on several different graphs. One graph will present the average sentiment over a set interval of time, this interval will depend on range of results returned by Twitter and the date they were published. Another graph will present the entire sentiment over all the results in the database. A map will be produced giving a rough idea where the tweets have been posted from. From this, the user will be able to see where in the world people have been discussing the current term. Finally, the tweets will be revealed to allow the user to view what people are saying about the chosen term.

Any terminology issues can be referred to the glossary section of the research document. This is where all the terms used to describe any part of this project will be defined.

2. TARGET MARKET

This project is a tool that is created for developers to aid them in the development of their analysis. However, it is implemented using a sentiment analysis on Twitter. There are several aspects of this project that will exist to help developers to achieve their goal. The first aspect is that the project isn't platform dependent. This means that although the project is developed using Twitter, if a developer wishes to implement Facebook or Movie Reviews, they will be able to with very little difficulty.

The next aspect is the category classifications. As project is implemented as a sentiment analysis tool, its implemented using the positive and negative sentiments. However, if the developer wishes to add neutral to the classifications, they will be able to with minimal effort. This same feature also works if the developer wishes to make another tool other than sentiment analysis, for example a spam filter. This is easily done by replacing positive and negative with spam and not spam.

Another feature to add to its generalisation is the ability to analyse different formats other than just text, e.g. images. This feature would allow developers to analyse posts off Instagram for example or other image based platforms.

The classifiers used can also be easily changed and adjusted to suit the needs of the developer. Depending on the above choices, the API created to for this project will adjust to suit the requirements of the above circumstances. This is a useful feature as it allows the developer test different methods of classifying and adjust easily. For example, for the sentiment analysis project, it is currently implemented with positive and negative. However, if the developer wanted to test the output with the neutral sentiment, the API will automatically adjust to suit the new circumstances.

Sentiment Analysis of Social Media – Functional Specification

As for the sentiment analysis part of this project, this tool is an invaluable asset for companies as it allows them to track the sentiment of different topics about their business. The map feature allows them to track where people are posting about their topic. This tool would also prove helpful for tracking how advertisements are doing and how the public is responding to it. Also, as this application will use the top trends from Google and Twitter in America, Ireland and England it will help companies stay up to date with current events.

3. METRICS

As this project is a sentiment analysis tool which measures public opinions from Twitter, there will be several metrics used in assessing the success of this application. The first being whether the specifications have been completed. The ability to allow a user to search a term that they want to find the sentiment of, successfully retrieving tweets back from Twitter, processing the text, discovering the sentiment of the tweets and finally displaying in the results on a webpage in an intuitive way for the user.

The classifiers are a key aspect to this project as the actual sentiment analysis is going to be done by them. The accuracy of these are going to be vital. This will be measured using cross validation, this method is used to see how the algorithms perform on unseen data. The dataset is split into several sets, one set is used as training and the another is used to test the algorithm. The number of times the algorithms have correctly predicted the correct classification is then shown as a percentage and the standard deviation is then calculated and shown as a percentage as well. This will provide a number that represents how correct the algorithms are. (Schneider, 1997)

4. PROJECT DIFFERENCE

There are many different types sentiment analysis applications. The main difference between all of them is the frontend. Most of them show charts of the sentiment and others show a range of emotions. The application being developed in this project will be created with several graphs including the sentiment throughout time and a map to show where users of Twitter are posting about the searched term. This is different from the from other applications they don't show a map that shows this information.

This application differs from its competitors as a framework is also being developed. This allows developers to change the analysis type, classifications, platforms and classifiers as well as providing an API that will adjust to the choices made in those categories. This enhances the reusability of the code and allows unlimited uses for this product. It's not bound by just sentiment analysis. Other uses include spam filters and social media monitoring where an analysis could check posts and comments for bullying messages for example. However, this software isn't limited to just text classification, image classifications could be implemented by a developer.

5. SYSTEM ARCHITECTURE

This will be a web and phone based application. Figure 1 shows the layout that this project will adapt.

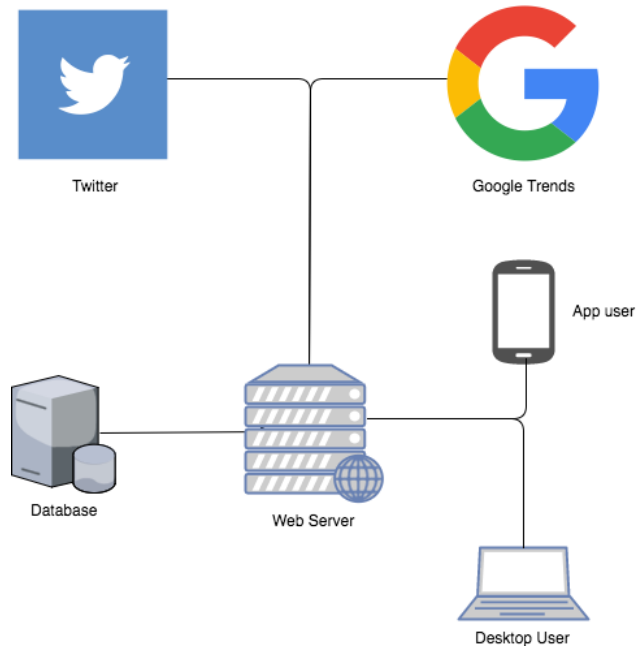


Figure 1. Design Layout

The webserver will retrieve trends from Google and Twitter which will be used to query Twitter for tweets about those topics. Once the tweets have successfully been retrieved, the tweets are classified and stored. The end user, using either the phone or desktop application will be able to search a term which will be sent to the webserver via an Application Programming Interface (API). Once the webserver receives the term, it will check the database for occurrences and if none are found, it will query Twitter for tweets about that term. The results are then returned to the user.

6. SUPPLEMENTARY SPECIFICATION

6.1 FUNCTIONALITY

- The application will send a slack message to developers for all runtime errors (i.e. crashes)
- The application will need to interact with Twitter and Google to collect data
- Search terms and number of results will be logged for improvements later. This will enable the developer to enhance the downloading feature.
- The classifiers accurate as a human (70% - 79%) (BRNRD, 2013)

6.2 USABILITY

- Once the user searches a term, all other options to search a term will be locked and disable the user to change their decision.
- A user should be able to see the analysis of their chosen term within one click.
- The user must be kept up to date with the current activity on the website.

6.3 RELIABILITY

- If the server crashes, the system will send a report to the developer including the error via slack.
- For accurate information, a script must be running continuously on the server to retrieve new opinions on recent topics

6.4 PERFORMANCE

- The must be sustainable with multiple users

6.5 SUPPORTABILITY

- The classifier must not be accessible to the user; however, the admin should be able to access it whenever necessary.
- As this is a webpage it must be functional with all modern browsers

6.6 SECURITY

- This application uses OAuth authentication to access the Twitter API, so user login information won't be sent.
- All tweets being retrieved from Twitter are public tweets so no security is required for storing them.

7. ITERATION PLAN

This project adapts the agile approach towards the development of this application and therefore, there will be multiple iterations. At the end of each iteration the author will present the application to interested parties and feedback will be given. The application will then be improved upon before the end of the next iteration.

7.1 ITERATION 1

This iteration is four weeks long and will start on the 21st November 2016 and will finish on the 15th December 2016.

7.1.1 SCHEDULE OF WORK

- The collection of trends will be completed. This will use the top trends on Twitter and Google in Ireland, America and England. This will enable the application to stay up to date with current topics of conversation. These will aid the user to in choosing what to search for.
- The database will be created.
- The ability to download tweets will be created. This will be using the Twitters Search API as it allows 100 tweets to be downloaded per request and 450 requests to be made every 15 minutes.
- The development of the text classifier will be started.
- Search for an accurate dataset to be found to be used to train the classifiers.

7.1.2 CHANGES IN REQUIREMENT

- Change from MySQL to MongoDB as the data received by Twitter is in JSON format and not structured. Some fields required by this application may not be in the data received for example, Coordinates, this is an element that is used by the application for the map feature however, only a small number of tweets contain this information.

7.2 ITERATION 2

This iteration is seven weeks long and will start on the 9th January 2017 and will finish on the 24th February 2017.

7.2.1 SCHEDULE OF WORK

- Complete the development on the twitter download of tweets.
- Continue with the development of the classifiers.
- Change from MySQL to MongoDB
- Set up testing environment for AngularJS for both functional and unit tests and Python for unit tests
- Create a graphical user interface for the application.

7.3 ITERATION 3

This iteration is six weeks long and will start on the 27th February 2017 and will finish on the 5th April 2017.

7.3.1 SCHEDULE OF WORK

- Finish the development of the classifiers.
- Refactor code and generalise parts of the code to be used in any circumstances. This includes changing the dependency on Twitter and the classifiers, so both aspects can easily be changed. Generalise the sentiments expressed to allow a developer to add different selection of sentiments. Also, separate the text aspect of the classification into its own module to allow developers to use a wider range of classifications.
- Create a mobile application. As the frontend of this application is using the technology AngularJS and styling is implemented using the Sass technology. This creates a perfect fit to use on a mobile application.
- Test software

7.3.2 CHANGE IN REQUIREMENTS

- Creating a mobile application was not in the original spec, however, due to the chosen technologies, it would enable the developer to transpose the application onto a phone platform.
- Generalisation allows this tool to have a wider range of applications than just sentiment analysis.

8. BIBLIOGRAPHY

BRNRD, 2013. *On Social Sentiment and Sentiment Analysis*. [Online]

Available at: <http://brnr.me/social-sentiment-sentiment-analysis/>

[Accessed 28 March 2017].

Schneider, J., 1997. *Cross Validation*. [Online]

Available at: <https://www.cs.cmu.edu/~schneide/tut5/node42.html>

[Accessed 28 March 2017].