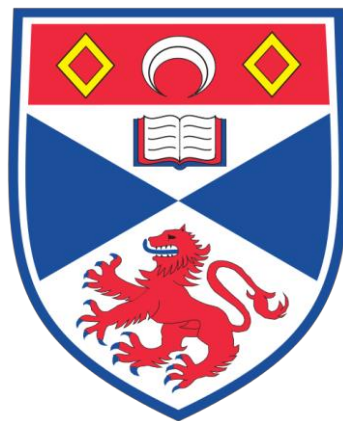# Curated Real-Time Visualisation of Twitter Data

Faridoon Noori

Matriculation Number: 120023515

School of Computer Science

University of St Andrews

August 2013

Submitted in partial fulfilment of the requirements for the degree of

MSc in Management and Information Technology

Declaration

I hereby certify that this dissertation, which is approximately 14,920 words in length, has been composed by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree. This project was conducted by me at the University of St Andrews from June 2013 to August 2013 towards fulfillment of the requirements of the University of St Andrews for the degree of MSc in Management and Information Technology under the supervision of Dr Per Ola Kristensson.

By submitting this project report to the University of St Andrews, I give permission for it to be made available for use in accordance with the regulations of the university. I also give permission to the University of St Andrews to publish the title and abstract of this dissertation on the World Wide Web.

Credit is explicitly given to others by citation or acknowledgment, and I retain the copyright of this work.

Faridoon Noori

August 2013

Acknowledgements

I would like to express sincere gratitude to my supervisor Dr Per Ola Kristensson for his guidance, patience and motivation for this dissertation. The invaluable support and guidance I received from my supervisor throughout this dissertation made it possible for me to learn a new programming language, to develop the application, and to write my dissertation successfully.

I would also like to thank my friends, especially Xinxin Li, for their motivation and support throughout my work for this dissertation.

I am grateful to the members of Processing.org discussion forum, who helped me in learning the programming language and debugging the code throughout my work.

Finally, I would like to thank all the participants of the application's usability study and qualitative investigation. Their feedback and opinions are valuable addition to my work.

Abstract

People generate millions of tweets on the social networking site Twitter per day [1] to share their thoughts, opinions, information, and news online. Understanding this enormous corpus of textual data by reading all or a collection of the tweets is tedious and time-consuming. In order to exploit the human's visual perception and identify easier ways of exploring textual data, information visualisation tools are created, which enable users to understand visually data from Twitter. While most of the existing tools are aimed at targeting a niche topic or a particular user group, there is lack of tools that are aimed at any user who is interested in learning about what is happening across the world by visualising data from Twitter.

This project presents an interactive application called "Curated Real-Time Visualisation of Twitter Data". It has been developed in Processing and it runs on Mac, Windows and Linux operating systems. It enables users to visualise the publicly available data of Twitter in order to understand what is trending in the world. This application visualises real-time geo-tagged tweets geographically, it explores trends in countries, where available, and it enables users to gain insights into these trends by exploring related tweets, images, and tag clouds. It also provides visualisation of real-time tweets from any geographical location, and it enables users to highlight these tweets based on the keywords of their choices. In addition, it provides visual insights into the platforms and languages of real-time tweets.

To leverage the power of information visualisation, a thorough literature review of existing work was carried out to determine best approaches in this domain, and to build this application on the idea of putting the user in centre of the design and implementation process. As a result, the application is very easy to learn and to use. To verify this, a usability study was conducted to collect opinions from first-time users of the application on how useful and effective the application was for them and for which purposes they wanted to use it. More than eighty-seven per cent of the participants found this application useful for various purposes like getting to know what is happening in the world based on trends, gaining insights into real-time tweets, identifying popular real-time hashtags, learning about popular languages on Twitter, and identifying tweets of their interests.

Vision

For anyone who needs to understand what people talk about on Twitter in various parts of the world, *Curated Real-Time Visualisation of Twitter Data* is a visualisation tool that will provide a visual presentation of Twitter's public data. The application will visualise real-time geo-tagged tweets on a world map, reveal which platforms users use to tweet from, reveal the languages of these tweets, show popular hashtags, and provide users with insights into available trending topics for various countries. Unlike Twitter.com and other related prior Twitter front-ends, this system will provide a visualisation of Twitter's data, which will focus on principles of storytelling to make it a novel, simple and enjoyable tool to use.

Terminology

Application: The artefact of this project that is the produced Processing desktop application.

Processing: A programming language based on Java, which is used for creating visualisations and animations.

Tweet: 140 characters long message that Twitter allows its users to share on the website. It also refers to the act of posting the message.

Hashtag: a word or phrase (without spaces) preceded by the # symbol used by Twitter users to tag or define the topics of their tweets.

Platforms: devices that are used by Twitter users to tweet from.

Trend: a hot or trending topic that is immediately popular on Twitter.

API: it stands for Application Programming Interface, and it is used by an application to interact with other applications.

Contents

List of Figures

List of Tables

1 Introduction

Social networking websites, such as Twitter, Facebook, LinkedIn, and Flickr, have led people to come together by building virtual communities. And these websites gained tremendous popularity across the world. Social networking sites enable people to interact and communicate with each other in online communities. Consequently, users generate enormous amounts of 'big data' daily. However, it is difficult to learn and analyse the data and their relationships by reading massive amounts of textual data.

As a result, information visualisation has emerged as a research field that attempts to design visual presentations that illuminate aspects of abstract data. In particular, social media visualisation tackles the problem of gaining insights into large corpora produced by people using online social networking sites.

Information visualisation researchers have focused on representing data in visual formats so that users can gain insights into large datasets visually. Visualising information helps users to perceive meaningful insights and identify intriguing patterns in large datasets. The interactivity aspect of the tools used to visualise information is adding more simplicity and control for the users in terms of exploring and understanding the data.

The focus of this project is a novel visualisation of Twitter data. Twitter.com[1] is a free online social networking and microblogging platform that allows users to share text-based messages of up to 140 characters, known as tweets. Users can also embed images, links to videos and other websites in tweets. It is estimated that users share over 400 million tweets per day [1]. Twitter is one of the most popular social networking sites. According to Alexa's top ranked sites, it is the top fifth website in its social networking sites category[2], and it is ranked the top eleventh most visited website in the globe[3] in August 2013.

This project provides geo-based information visualisation of real-time tweets, trends, and platforms from which people tweet, and the languages used in real-time tweets. In addition, it visualises text of tweets in forms of tag cloud for a simpler understanding of the text in tweets. In summary, this project presents a visual gateway to Twitter's streaming API (Application Programming Interface), and visualises trending topics in various countries across the world based on Twitter data.

1.1 Statement of the Problem

Users generate enormous amounts of 'big data' on Twitter. Understanding this data by reading tweets, and gaining insight into what people are talking about in different countries is tedious and time-consuming. Although research has been carried out in developing tools for exploring Twitter data, they usually target particular niche topics such as news, politics, academics, or particular user groups such as journalists, market researchers, and business users. There is a lack of a general Twitter

---

[1] http://www.twitter.com

[2] http://www.alexa.com/topsites/category/Computers/Internet/On_the_Web/Online_Commu nities/Social_Networking

[3] http://www.alexa.com/topsites/global

visualisation tool that enables any user group to understand visually the real-time data from Twitter.

1.2 Aim

The main scope of this dissertation is to create "Curated Real-Time Visualisation of Twitter Data" – a desktop application developed in Processing that runs on Mac OSX, Windows, and Linux to help people understand visually what is happening around the world based on Twitter's trends and real-time geo-tagged tweets.

1.3 Objectives

To achieve the scope of the project, the following objectives are to be accomplished:

- ☉ Retrieve and visualise real-time geo-tagged tweets from Twitter's streaming API.
- ☉ Retrieve and visualise current trending topics from Twitter's trends API.
- ☉ Allow users to explore languages and hashtags from the real-time tweets, and to visualise platforms from which people tweet.
- ☉ Develop a simple, attractive and interactive user interface.

1.4 Dissertation Outline

The remainder of this dissertation is structured as follows: Chapter 2 provides a review of the literature in the fields of information visualisation, social media analysis, and Twitter data visualisation. It also presents what tools researchers and businesses have developed for visualising Twitter data. Chapter 3 describes detailed requirements specification. Chapter 4 covers project methodology. In Chapter 5, the system design process is described. Chapter 6 shows implementation of this application. Chapter 7 presents four types of testing and evaluation methods that were used to test this application. Lastly, chapter 8 provides conclusions to this dissertation, which also contains an overview of the limitations and further work for this application.

This chapter provided motivation, problem statement, aim, objectives and structure of this dissertation. The following chapter reviews literature about information visualisation, social media analysis, Twitter data visualisation, and related technologies.

2   Context Survey

2.1   Information Visualisation

It is said that a picture is worth a thousand words. This implies the fact that humans can perceive visual presentation of data, such as map or photograph, more easily than textual or spoken presentation in some tasks [2]. Information visualisation can be used as the means for discovering patterns, trends, clusters, outliers and gaps in data [2].

Before discussing social media analysis and visualising Twitter data in particular, the fundamental theory of visualisation data types and user tasks is presented in the following parts.

Shneiderman and Plaisant [2] provide a list of these seven visualisation data types:

2.1.1   Visualisation Data Types

1) **1D linear data:** this data type is one-dimensional and it includes types of data that can be organized sequentially. The data included in this type may be source code, text, dictionary, and alphabetical lists of names. The common user interface design tasks involved in this data type are concerned with choosing the right colours, sizes, and layout. Besides, interactive elements such as overview of data, scrolling, or selection methods can be provided for exploring such data [2].

2) **2D map data:** this type of data can be used to visualise geographic maps, floor plans and newspaper layouts [2]. Every item in this type of data may cover part of the total area, and it may or may not have a rectangular shape. Users can use such visualisations to finding adjacent items, regions that contain similar objects, and paths between these items [2].

3) **3D world data:** this type of data is used to represent three-dimensional objects such as molecules, human body, and even buildings [2]. Advances in computer-aided designs have facilitated the creation of such complex three-dimensional relationships.

4) **Multidimensional data:** in this type of data, items with n attributes become points in an n-dimensional space [2]. In order to visualise this type of data, 2D scattergrams may be used, and sliders may be used to control additional dimensions. Users can explore patterns, clusters, gaps and outliers in visualisations of this data type [2].

5) **Temporal data:** in this type of data, items are having start and finish time, and these items may overlap [2]. Examples of tasks included in visualising such data may be identifying events based on time, or comparing periodical events.

6) **Tree data:** this type of data is also called hierarchy or tree structure. A tree structure is a collection of items in which each item is linked to its parent item [2]. A company's organizational chart is a good example of tree data.

7) **Network data:** in this type of data, relationships among items in the data are determined by connecting them to each other in a network. Network visualisation is an old art, and Shneiderman and Plaisant [2] call it imperfect because of the complexity of relationships in a network and user tasks in such visualisations.

The seven data types in data visualisation were presented. In order to understand the basic tasks involved in interacting with information visualisations, Shneiderman and Plaisant [2] provided a framework that consists of seven basic tasks that users perform on visualisations:

2.1.2 Basic Tasks

1) **Overview:** using overview task on data visualisations, user can gain an overview of the whole dataset. Collections of data are represented in a zoomed-out layout [2] for an overall understanding. Hornbæk et al. [3] describe that it is difficult for many query languages to present overview of a collection, it is important to apply sufficient overview strategies, and use these as a measure to evaluate such interfaces.

2) **Zoom:** this enables users to zoom in on items that are of interest to them. Users have interests in selected parts of the data, and they need tools that allow them to zoom in for more detail in that particular portion of the collection. Zooming tasks are mainly essential in applications of small displays [2].

3) **Filter:** this task enables users to focus on exploring data that is interesting to them by leaving out the uninteresting parts of the data. Filtering tasks are used to control the contents of display, and to allow users to conceal unwanted pieces of information [2].

4) **Details-on-demand:** in this task, users can explore a particular item or group of items for more details by making selections [2]. When users filter a large collection of items, they are left with the data of their interest. Consequently, by allowing them to click on these items for more information, it improves their understanding of the data.

5) **Relate:** in information visualisation, relationships among items of a collection can be represented by distance, by containment, by connected lines or by color-coding [2]. This makes it easier for users to understand and learn about the collection or understand how data items are connected to each other.

6) **History:** in this task, users are enabled to keep a history of the tasks they perform on the data. While performing tasks on data, users do not always get the desired outcomes, and therefore, it is important to allow them to step backward or review the previous tasks performed on data.

7) **Extract:** users can extract a subcollection of data. They are allowed to obtain the item or group of items that they are interested in. It is also important to enable users to save, print, or share the extracted item [2].

2.2 Social Media Analysis

Card et al. [4] define social media as a group of Internet-based applications that are built on Web 2.0, and that enable users to create and share User Generated Content. The concepts of Web 2.0 [5] and User Generated Content [6] have been used together in understanding what social media is. Web 2.0 contributes to the creation of virtual communities on the web that allow users to interact and collaborate with each other, and to generate content [5]. The idea of social media is about two decades old. Internet-based applications that emerged in late 1990's for enabling users to interact and share information with each other is believed to be the origin of social media. SixDegrees.com [7] is given an example of the pioneers in social media in [8]. It enabled users to create their profiles, and to create lists of friends [8]. According to Kaplan and Haenlein [4], the era of social media probably began back in 1998 when

Bruce and Susan Abelson created "Open Diary"[4] that enabled diary writers to form an online community.

Due to the widespread Internet usage and improvements in its speed, social networking and blogging platforms received a significant attention among users of the Internet. The creation of MySpace[5] in 2003 and Facebook[6] in 2004 led to a prominent existence of social media [4]. Most recently, micro-blogging platforms like Twitter[7] which allow creating and sharing real-time updates have emerged [8].

As the content generated and shared in various social media platforms increases, its quality varies significantly [9]. The content may be classified as useful, abuse and spam. It is not only the content that has generated great interest for research, but it is also the non-content information such as links among the items shared by users [9]. Therefore, the task of analysing this content to extract meaningful and high quality information is increasingly becoming important.

Asur and Huberman [10] found that an analysis of the content in Twitter gave more accurate results in making quantitative predictions than those calculated by artificial markets in predicting box-office revenues. The enormous data with a high diversity of opinions shared by users of a community creates opportunities for making specific predictions in the future [10]. In addition, the content of social media can be exploited to help design marketing and advertising campaigns [11], [12].

2.3 First Story Detection

Since social media sites like Twitter enable users to update real-time messages, their content can be analysed for first story detection, also known as new event detection. When a major event takes place, Twitter users inform about the events. Due to its nature of real-time sharing, sometimes the tweets can spread news before any traditional newswire identifies the story [13]. In addition, users share their opinions about a particular event, and this creates interesting information on how a story spreads and what people say about it. For example, Twitter was used to detect the story about protests in Iranian presidential elections in 2009. In another example, US Centre for disease control used Twitter to inform people about the swine flu outbreak [13].

The high volume of data and its noise can add problems to detecting new stories in Twitter [13], [14]. On the other hand, Twitter has the benefit over traditional newswire in terms of adding a social component in detecting these stories [13]. The social aspect in detecting a story helps understand the impact of an event and people's reaction to it.

In addition, research efforts have also focused on detecting events of a specific type such as news events [15] and earthquakes [16]. Other works include the detection of first message in particular events [13], and identifying real-time events by grouping similar tweets based on topics [14].

---

[4] http://www.opendiary.com

[5] http://www.myspace.com

[6] http://www.facebook.com

[7] http://www.twitter.com

2.4 Information Visualisation in Social Networks

Information visualisation is the application of representing abstract and non-spatial data in visual forms. Visualising information can help readers understand the relationships and structure in data [17]. As social media systems enable a collective creation of content, tools can be designed to exploit the way communication flows and to understand connections among content creators [18]. As users create and share content with other users, ties are generated not only among the users but also within the content. By creating visual representation of these ties, it provides interesting and useful insights into the data. Visualising, navigating and finding patterns in networks is not easy, and thus making it difficult to understand networks [19]. Therefore, analysing the relationships among elements of communication and considering its social part can contribute to gaining insights because humans' perceptual abilities can be greatly influenced by visual representations [19].

In the past, various research efforts have focused on presenting interactive and exploratory tools for analysing social networks [20], [21]. Visualisation tools like NodeXL [18] can be used as an add-in for Microsoft Excel 2007 to analyse social media content and create network graphs that visualise connections among various data points. Another tool for visualising social networks is NodeTrix [22] that represents elements of a parent element based on showing node-link diagrams and communities. Likewise, researchers in [23] designed a visualisation tool that analysed patterns of communication in email traffic. The tool offered recommendations for improving productivity within groups in the workplace.

In particular, visualising social media content on the Web has gained increasing attention. The Web offers a great platform of connecting applications, data, information and users. As a result, web-based applications like Google Charts[8], twitInfo[9], Many Eyes[10], and Friend Wheel[11] have emerged for processing and representing data in visual forms.

2.5 Twitter Visualisation

Twitter gained a widespread interest among ordinary people, politicians, celebrities, academics and other public figures. In recent years, research efforts have been carried out in an attempt to understand the large corpus of tweets. Academics, companies and individuals have developed applications for understanding different aspects of the Twitter data. In order to understand the data retrieved from Twitter, tweets are analysed and visualised for detecting events, identifying trends, retrieving news, analysing sentiments, and various other purposes. Companies use Twitter to understand an overall sentiment of customers towards their products and news agencies use it to engage audience with their stories by providing real-time tweets about latest news [24].

Twitter provides APIs to allow access to real-time data that is generated by users. Twitter's widespread use, its provision of large corpuses of user-generated content and the diversity of this content has attracted attention in the field of data mining

---

[8] https://developers.google.com/chart
[9] http://twitinfo.csail.mit.edu
[10] http://www-958.ibm.com/software/data/cognos/manyeyes
[11] http://friend-wheel.com

and information visualisation. In addition, the quality and quantity of metadata [24] that Twitter provides for each tweet help researchers create more interesting and meaningful visualisations. For example, tweets that are paired with geo-location data can be used to create clustered visualisations on a map based on their relevant geographical locations.

A significant amount of work has been done in the context of analysing Twitter data. However, there is only a handful of work relevant to visualising real-time data on Twitter activities. The following section presents works that are closely related to this project.

Truthy [25] collects and analyses tweets, and it clusters the tweets into "memes" that represent clusters of similar messages. This system considers hashtags, mentions, hyperlinks and phrases of tweets as criteria for clustering tweets into memes. Truthy focuses on tweets that relate to themes like politics, social movements and news. Each theme represents a collection of related memes or topics. In Figure 1, Truthy [26] has gathered tweets from the past nine months. And as a result, it provides analysis and visualisation on historical data from Twitter.
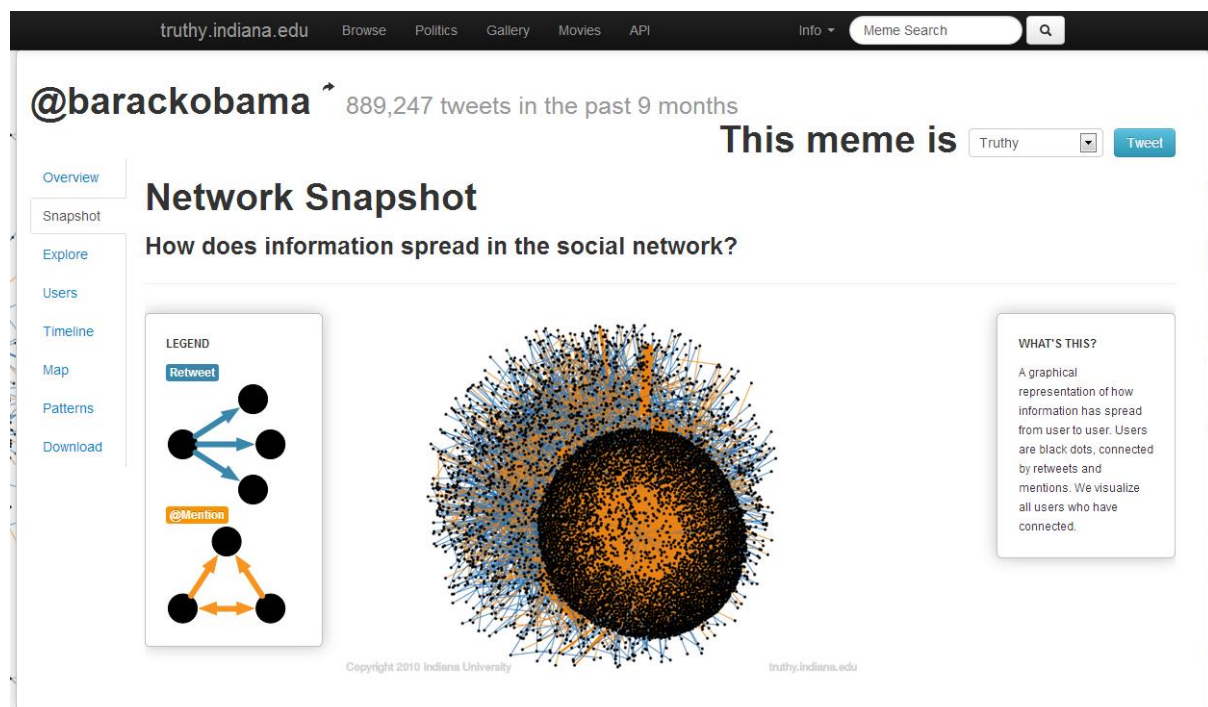
Figure 1: Visualising the @barackobama topic using Truthy.

In an experimental study, Michelson and Macskassy [27] take a different approach of discovering users' topics of interest on Twitter. In their approach, topics of interest are discovered for a particular user. Thus, they create and cluster user profiles based on the identified topics of interest. They discover a user's profile related to the entities about which the user tweets. These entities are then classified under high-level categories of interest. Wikipedia's folksonomy is used for determining categories of interest for the disambiguated entities in given tweets. Similarly, TwitterRank [28] focuses on identifying influential microbloggers in Twitter. It considers similarity in topics between users and the link structure while measuring the influence of Twitter users [28]. Such experimental tools may not attract a wide variety of users. While

journalists, for example, may be interested in identifying top contributors in a particular area of interest in Twitter, these systems may hardly be used for exploring real-time trends happening around a user.

A research paper by Marcus et al. [29] discuss how their system, Twitinfo[12] (Figure 2), helps users to identify events and related sub events by tracking particular keywords. The system enables a user to track an event by logging a number of keywords into Twitinfo. It then tracks these events by serializing tweets related to the given keywords to a database [29]. As a result, a timeline of sub events that determines peaks in tweets frequency per minute is given along with relevant tweets displayed on a map. In addition, tweets are categorized into positive and negative sentiments. While its identification of peaks in events is useful for highly popular topics such as football games, earthquakes or elections, its usefulness can be argued in day-to-day events that happen across many countries. During the evaluation of the system in detecting subevents for three soccer games, it did well in identifying peaks in game start, half time, penalty shots, wrong, and game end events except missing Yellow Card events [29]. However, Twitinfo may not be used in identifying events or trends based on particular locations.
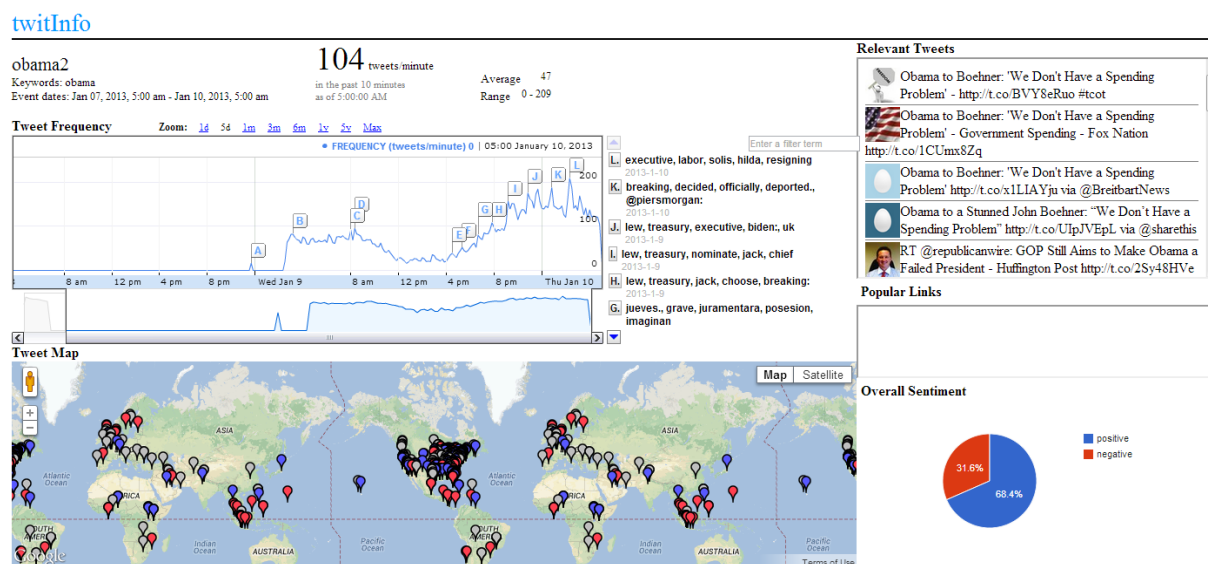
Figure 2: Result of tracking obama keyword for 3 days in TwitInfo.

A challenge with real-time trend detection from tweets is finding the accuracy of results over a short span of collecting real-time tweets [30]. For example, TwitterMonitor claims that it gathers streaming data from Twitter's API and parses the tweets into a database for further detection of bursty keywords [30]. A keyword is said to be bursty if it is received by the system at an unusual high rate. For example, the keyword "Liverpool" may normally be mentioned in 10 tweets per minute, but it suddenly receives a frequency of 100 mentions in tweets per minute.

Another challenge with such systems is their robustness (almost real-time) and scalability (ability to handle large corpus of tweets) for detecting real-time trends). Fang et al. [31] argue that the short paper written on TwitterMonitor lacks detail. And provided that the two-step process of identifying the large amount of bursty

---

[12] http://twitinfo.csail.mit.edu/

keywords and then comparing them to the history for grouping keywords together can be an extremely expensive process. However, the mechanism of identifying trending topics by Fang et al. [31] is similar to that of [30] with the only significant difference of enabling a user to specify a time interval for identifying trending topics. Fang et al. [31] seem to have missed discussing about a comprehensive related work. Instead, they have included only two similar research works by Mathioudakis and Koudas [30] and M. Cataldi et al. [32].

As interest grows in understanding Twitter data, clustering tweets into meaningful groups is becoming a popular research area. For example, TwitterStand[13] (Figure 3) is a news processing system that aggregates user-generated tweets into different categories, such as General, Business, SciTech, Entertainment, Health and Sports. This system captures tweets that correspond to breaking news. These tweets are then plotted on a map and they are differentiated by their corresponding categories' icons [15]. TwitterStand focuses on collecting news from 200 handpicked users known as seeders [15] who frequently publish news via tweets. While this system may be a good tool to use with other news aggregators in order to explore breaking news in depth, its ability to provide a general insight into different types of events and trends across Twitter may be questioned. Enkh-Amgalan Baatarjav and Meesumrarn [33] argue that separating news and noise in tweets that are only 140 characters long can be difficult due to the limitation of message size, and the spelling and grammatical mistakes that people make in these tweets. More importantly, it is difficult to identify trustable sources that report news on Twitter [33]. As no one audits tweets, some tweets are rumours that result in generating false news report [33].
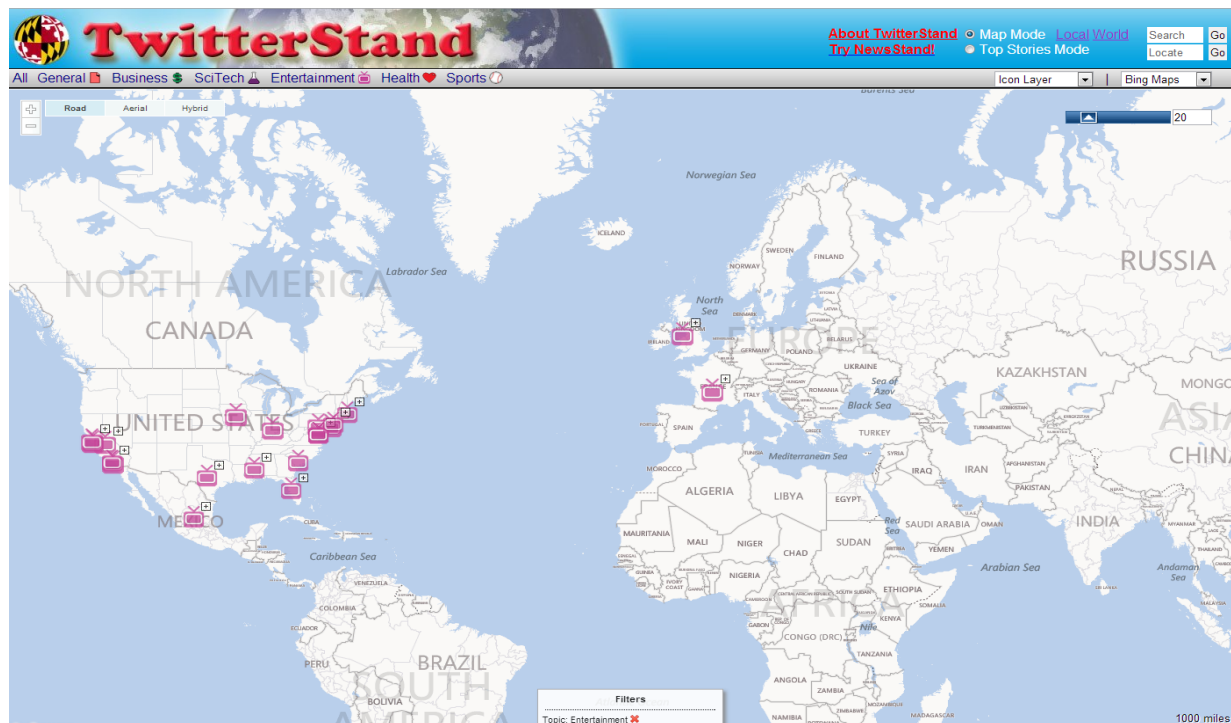
Figure 3: Visualising entertainment-related news in TwitterStand. The TV icons on the map present news reports related to entertainment.

---

[13] http://twitterstand.umiacs.umd.edu/news/

Mentionmapp[14] (Figure 4) enables a user to explore a user's most active followers and hashtags on Twitter. It retrieves all tweets of the user and then identifies which followers and hashtags have been used most frequently.
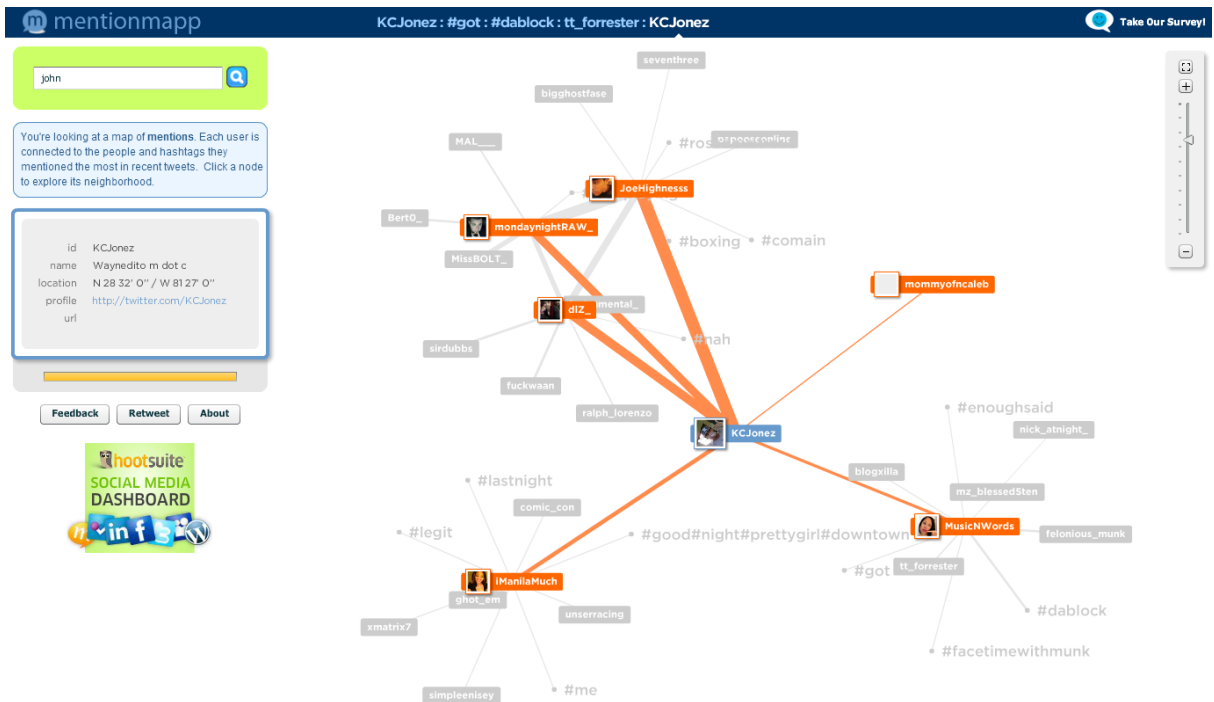
Figure 4: Exploring KCJonez user profile using mentionmapp.

TwitterCounter [15] claims that it is the number 1 stats site for Twitter users. TwitterCounter (Figure 5) tracks Twitter users and Twitter's usage. It offers a number of widgets that users can add to their websites or blogs for showing their recent Twitter visitors and number of followers.

---

[14] http://mentionmapp.com/
[15] http://twittercounter.com/

Figure 5: Twitter statistics for BarackObama user profile using TwitterCounter.

Tagdef[16] allows users to look for definitions of hashtags from Twitter. If a hashtag is an acronym or a unique term, it becomes difficult to understand what its relevant tweets are discussing about. By using tagdef.com (Figure 6), one can find definitions of these hashtags and can even add new definitions to the hashtags.
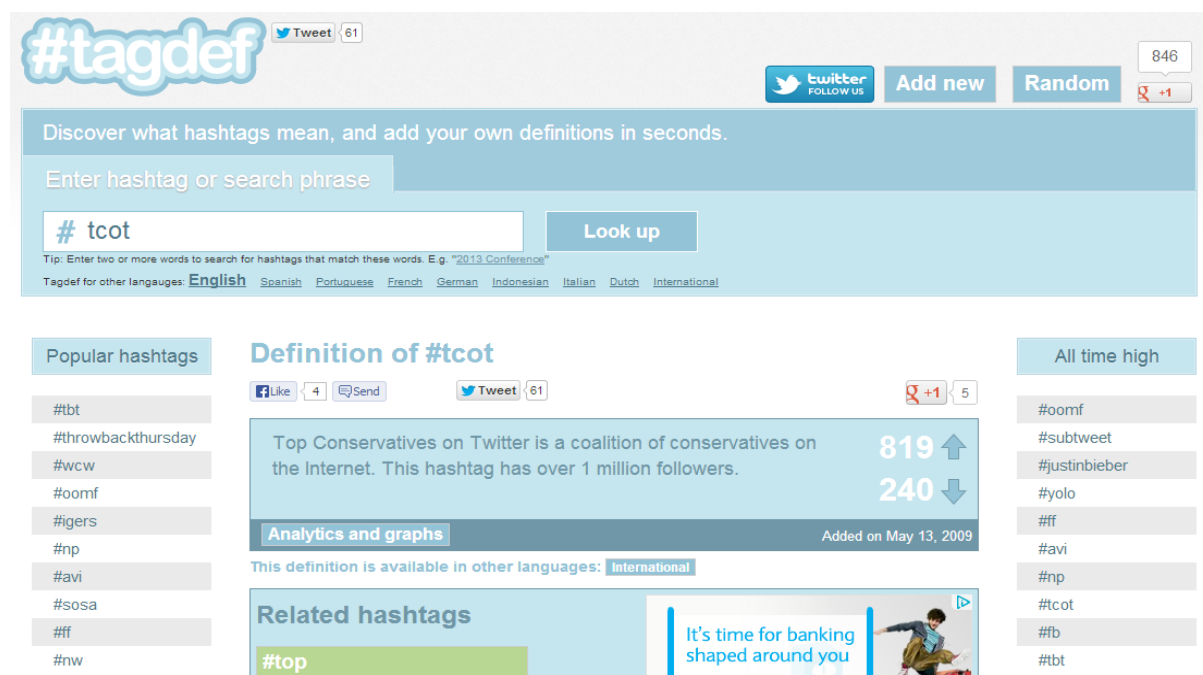
Figure 6: Definition of #tcot hashtag provided by tagdef.

---

[16] http://tagdef.com/

What the trend[17] (Figure 7), similar to tagdef.com, it is a web-based service that offers user-generated explanations of Twitter's trending topics. The accuracy of definitions is determined by the tool's voting and flagging mechanism. It allows users to choose their most favourite explanation for a trend by voting.
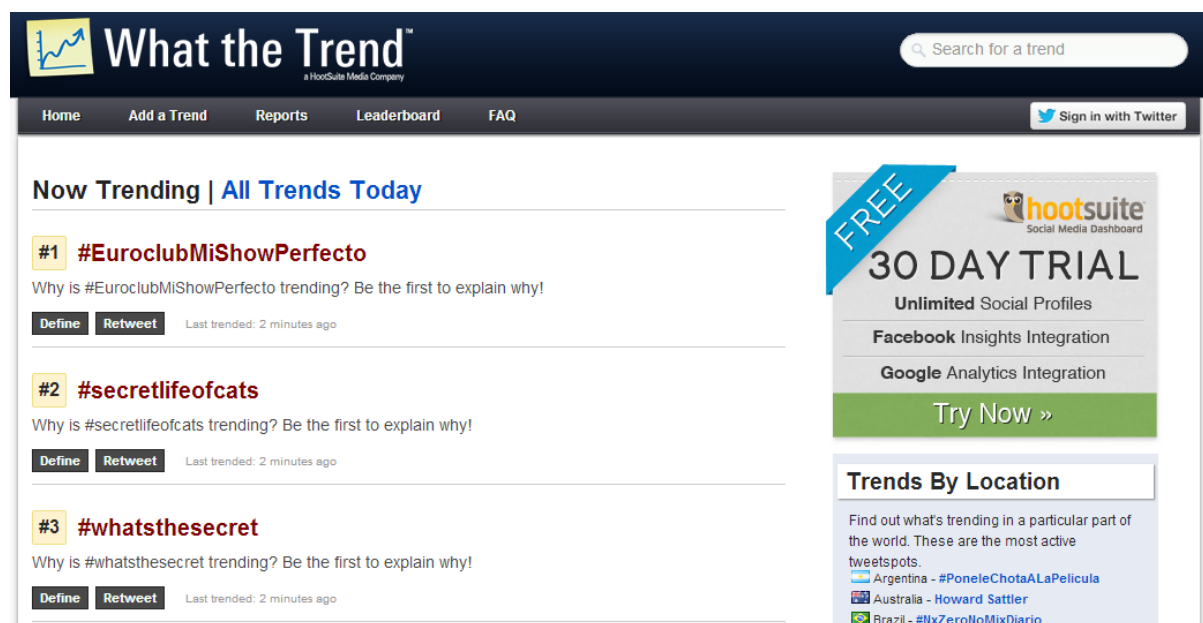
Figure 7: Trending topics and their definitions from What the trend.

Trendsmap[18] processes 80 million tweets daily. The application, as shown in Figure 8, tracks most popular trends received in tweets. It offers a geo-tagged visualisation of the trends for only one week. However, if the trend is less popular, it may be removed from the map. Further exploration of the topics is offered in a list of most recent tweets.
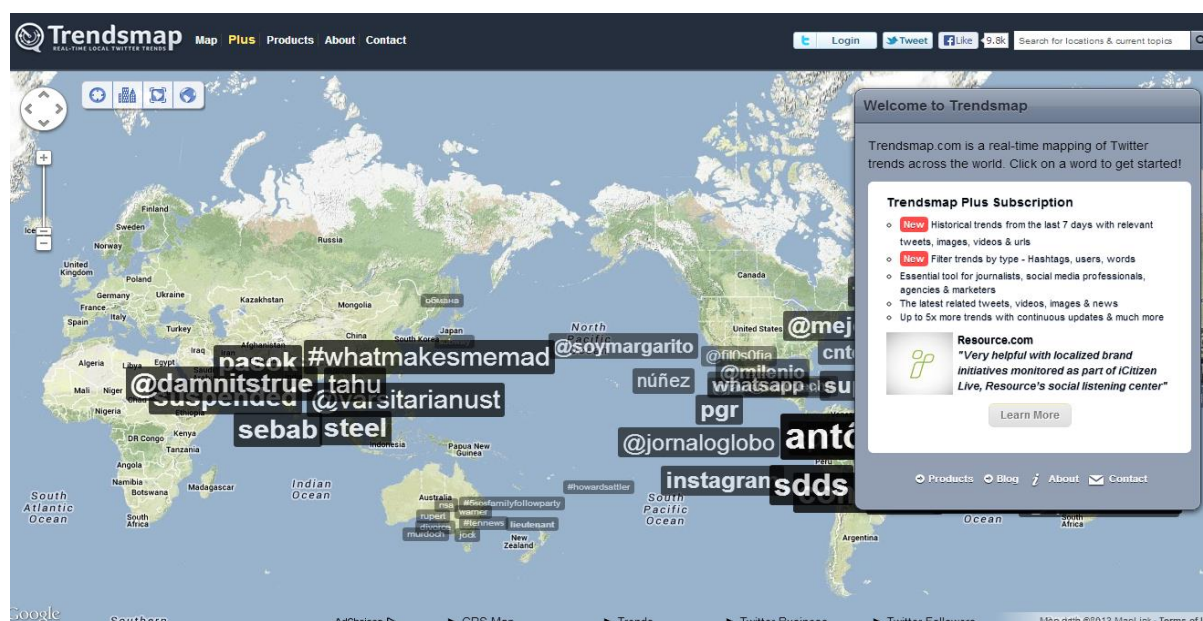
Figure 8: Popular trends provided by Trendsmap.

---

[17] http://whatthetrend.com/
[18] http://trendsmap.com

12

Twendr [19] (Figure 9) retrieves trending topics from Twitter and presents them in country-based categories. Each category contains a list of trending topics for that particular country. These topics are linked to their corresponding Twitter pages where relevant tweets are displayed.
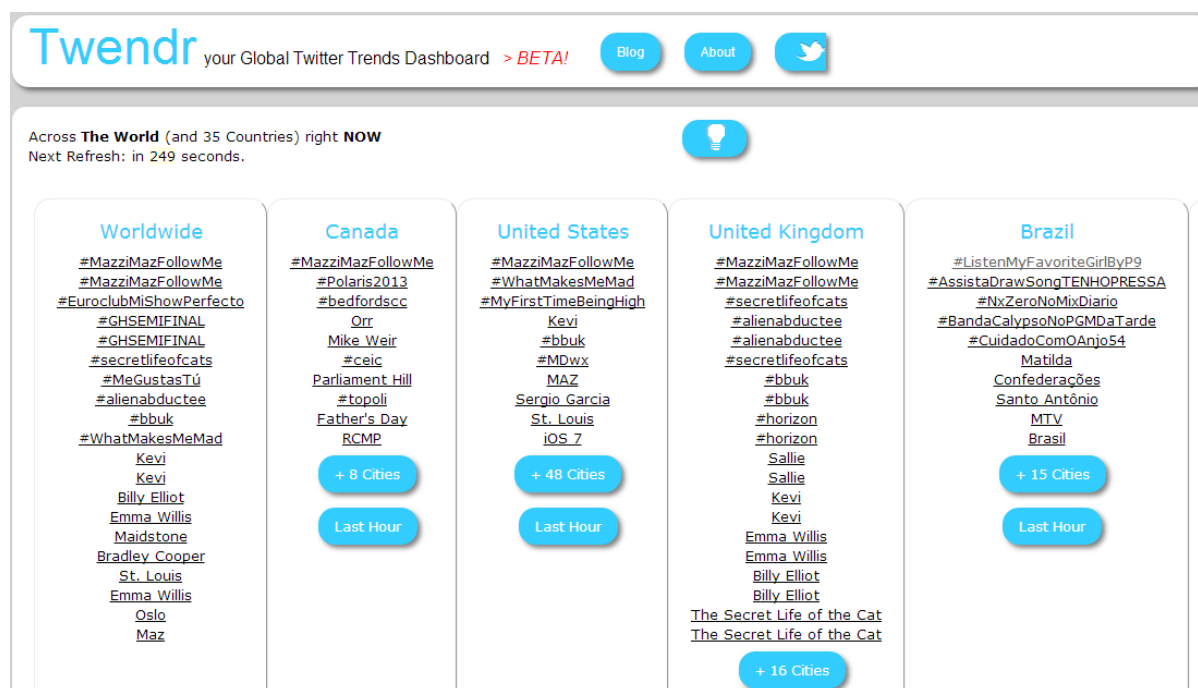
Figure 9: Trending topics from Twitter are displayed by Twendr.

2.6 Summary

Past research has mostly focused on how to analyse tweets in order to identify topics or latest trends. These papers have been discussed in the above section. This work mainly focuses on offering a simple and understandable visualisation of the trending topics, trend-related and real-time tweets. This work aims at providing a usable and intuitive, yet versatile, interface for an easy exploration of what is happening around the world by taking into account the trends provided by Twitter.

Twitter offers a list of trending topics called "trends" that are tailored to a user's location and who the user follows. These trends can be customized according to the various locations in which trending topics are available. The algorithm [34] used by Twitter provides trends that are immediately popular. This application uses Twitter-provided trends in order to help users discover the most popular topics that are being discussed on Twitter and enable them to explore further by visualising the related tweets.

2.7 Related Technologies

This section provides an overview of all the technologies and tools that were used in developing this application.

---

[19] http://twendr.com

2.7.1 Processing

Processing[20], created in 2001, is a simple and open source programming language and integrated development environment (IDE) that is used for creating drawings, animation, and interactive graphics [35]. Processing is built on the Java language, but its syntax and graphics-programming model are simplified.

2.7.2 Twitter API[21]

Twitter provides access to its public data via its APIs. Each API is used to represent a facet of Twitter. This application uses the following APIs:

1) **Search API** enables a user to query for Twitter content. It can be used to find tweets by a set of keywords, tweets by a particular user, or tweets referencing a user. The Search API can also be used to access trends on Twitter. However, this is not suitable for applications that require extreme velocities and hit rate limits. Instead, Twitter provides a Streaming API that can be used to address the limitation of Search API. Only tweets from the past week are included in the search.

2) **REST API v1.1** enables access to the core primitives of Twitter, such as timeline, status updates and information for a particular user. This API is used for leveraging core Twitter objects. It can be useful for building profile of a user, which contains user's name, Twitter handle, profile avatar, and a graph of people the user is following. In addition, this API can be used to create and update tweets, reply to tweets, add tweets as favourite, and retweet other tweets.

   Rate limits in REST API v1.1 are divided into fifteen-minute intervals. If the rate limit for a particular API call is fifteen per window, it means that only fifteen calls can be made in fifteen minutes.

3) **Streaming API** provides a real-time sample of Twitter Firehose. It is particularly useful for data intensive needs such as data mining and analytics. The Streaming API enables users to specify and track large quantities of keywords, retrieve geo-tagged tweets for a particular location, and retrieve public statuses of a set of users. However, it requires establishing long-lived HTTP (Hyptertext Transfer Protocol) connection and maintaining that connection.

   The Streaming APIs offer various streaming endpoints that can be customized to different use cases:
   a. *Public streams:* endpoints like filter, sample and Firehose can be used to access samples of the public data from Twitter. Once an application establishes connection to a streaming endpoint, Twitter delivers a feed of tweets to the application.
   b. *POST statuses/filter:* this returns public statuses based on one or more filter predicates. Both GET and POST requests can be used, but GET request that uses too many parameters may cause rejection of the request due to excessive URL length. Thus, using POST request method solves this problem. This endpoint allows up to 400 track keywords and 5,000 follow userids.

---

[20] http://processing.org/
[21] https://dev.twitter.com/docs

c. *GET statuses/sample:* this returns a small random sample of all public tweets. It counts for 1% of the Firehose (all public statuses).

2.7.3   TileMill[22]

It is a design environment that is used for creating interactive and static maps. This tool was used for creating the two world maps in the application.

The following libraries were used in the development of this application:

2.7.4   gifAnimation[23]

This library is used for drawing gif animations in Processing sketches. This library was used for creating a gif animation that is used while a tag cloud is being generated.

2.7.5   Twitter4J[24]

An alternative to accessing Twitter API is the Java's unofficial library, Twitter4. The application uses this library to connect to Twitter and retrieve data from its different APIs.

2.7.6   G4P[25]

This Processing library is an easy tool for adding 2D Graphical user interfaces in Processing sketches. This library was used in this application for designing the GUI (Graphical User Interface).

2.7.7   JLangDetect[26]

This is a lightweight library for Java, which is used to detect language of text. This library was used for detecting languages of real-time tweets.

2.7.8   WordCram[27]

This is a Processing library that is used for creating tag clouds. It was used to visualise tag clouds from the tweets in our application.

2.7.9   Toxiclibs[28]

This library is built for Processing, which is used for general data visualisation purposes. This useful library was used for visualising histogram and bar charts.

This chapter presented the context survey, overview of related prior Twitter front-ends, and related technologies that were used in this project. The next chapter will discuss requirements of the application.

---

[22] http://www.mapbox.com/tilemill/

[23] http://www.extrapixel.ch/processing/gifAnimation/

[24] http://twitter4j.org/en/index.html

[25] http://www.lagers.org.uk/g4p/

[26] http://www.jroller.com/melix/entry/jlangdetect_0_2_released

[27] http://wordcram.org/

[28] http://toxiclibs.org/

3   Requirements Specification

Requirements of a system represent the features that it must have. Functional requirements specify the functions that a system should provide. On the other hand, non-functional requirements specify constraints on the operation of the system [36]. The following section of the dissertation will cover functional and non-functional requirements of the Curated Real-Time Visualisation of Twitter Data application.

3.1   Functional Requirements:

The following are functional requirements of the application. These describe *what* functions the application provides. Functional requirements are divided into three sets i.e. Primary, Secondary, and Tertiary.

3.1.1   Primary:

The primary functional requirements represent the basic features of the application that must be implemented.

1) The application shall retrieve and visualise real-time geo-tagged tweets from Twitter's Streaming API on a World map.
2) The application shall retrieve real-time trends from Twitter's trends API in all countries where available.
3) The application shall retrieve and visualise tweets from Twitter's Search API for each trend.
4) The application shall classify and visualise platforms that users tweet from in real-time on the World map.
5) The application shall classify and visualise real-time tweets' languages.
6) The application shall retrieve and visualise hashtags from real-time tweets.
7) The application shall allow user to toggle between two types of World maps i.e. Map with coloured countries, and Map with single-coloured countries
8) The application shall allow user to select tweets from a location of choice on the map and then visualise these tweets.
9) The application shall provide help to the user on how to use its various functions.
10) The application shall allow user to highlight real-time tweets based on two keywords, which are retrieved from a particular area.

3.1.2   Secondary:

The secondary functional requirements represent a set of features that came up during the development of the application. In order to enrich the application with more than the basic requirements, these features were required to be implemented.

1) The application shall generate a tag cloud by analysing retrieved tweets for a particular trend.
2) The application shall generate a tag cloud by analysing all the tweets that user wishes to select from an area on the World map.
3) The application shall show relationships of replies to tweets in the retrieved tweets for a particular trend.
4) The application shall classify retrieved tweets from most popular to less popular for a particular trend.
5) The application shall classify real-time tweets in an area based on their users from most followers to least followers.

6) The application shall retrieve and display images if available from tweets for a particular trend.

7) The application shall classify tweets that contain images for a particular trend.

8) The application shall allow user to explore each tweet when retrieved for a trend or in a selected area on the Map.

9) The application shall allow user to view a particular tweet in the web browser, which is retrieved for a trend or in a selected area.

3.1.3 Tertiary:

The tertiary set of functional requirements may be implemented if there is enough time. These may also be considered for future development of the application.

1) The application shall be deployed on the web and it shall work using Web standards.

3.2 Non-Functional Requirements:

The following are non-functional requirements of the application. These describe *how* our application achieves the required functionality.

3.2.1 Design constraints:

This section outlines a set of design constraints of the application. In particular, it determines how the system design should be restricted to the specifications.

1) The application shall be implemented in Processing.
2) The application shall be implemented as a stand-alone desktop application.
3) The application shall be supported by OSX, Windows and Linux operating systems.
4) The application shall use temporary memory to store incoming data from Twitter APIs.

3.2.2 Quality attributes:
1) The application shall be simple and easy to learn and use.
2) The application shall have consistent user interface that is usable and friendly.
3) The application shall not require username and password of user's Twitter account.

3.3 Use Case Diagram

Use case diagrams are used to represent the functionality of the system [36]. These diagrams are used to explain how user interacts with the system, and how the system produces a visible result. In this application, as shown in Figure 10, there is a single user that interacts with the application in order to view real-time tweets on a World map, view languages that are used in these tweets, and reveal platforms used as source of the tweets. In addition, user also interacts with the system in order to retrieve trends for a country or read tweets in a particular area on the map. When the user makes a request, the application interacts with Twitter via its streaming or search API and responds to user's request.
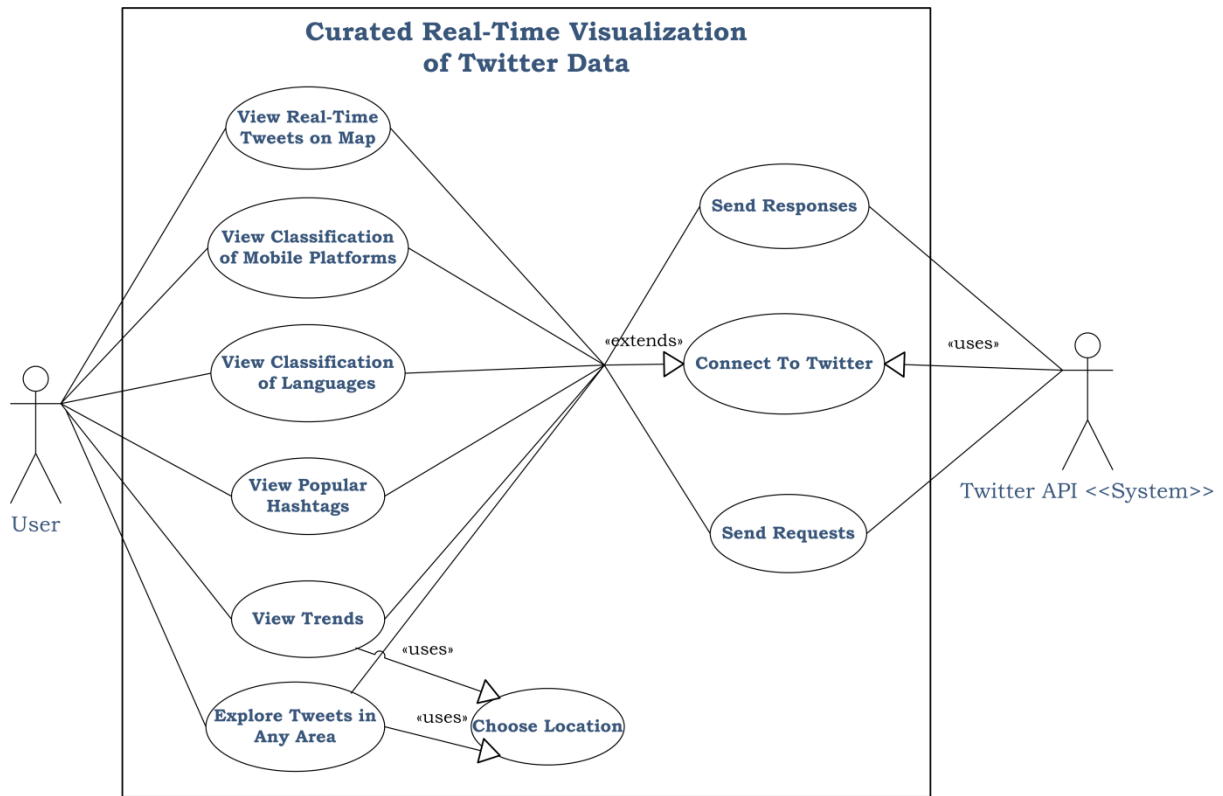
Figure 10: Use case diagram presents how user interacts with our application, and how the application interacts with Twitter APIs.

This chapter discussed in detail the requirements of this project. It included explanation of the functional and non-functional requirements of the application. The next chapter will discuss the methodology that was undertaken while implementing this project.

4   Project Methodology

This chapter provides an overview of the software development methodology that was used to build this application.

Sommerville [37] explains that projects that insist on specifying thorough requirements, and then move on to design, implementation and testing result in slow software development. As a result of new requirements or problems during the development of a system, the traditional waterfall process must start over and over again. Thus, it results in delaying the end product. Likewise, due to time constraints on the project, it was wise to focus on rapid software development and delivery.

The rapid software development processes are designed to develop the software as a series of increments where each increment represents a new system functionality [37]. Agile software development methods were used to build our application. This approach facilitates rapid development of the system and adds new functions to the application during the course of development lifecycle.

4.1   Agile Software Development

Agile methods include incremental development methods in which each increment represents new functionality of the system [37]. Figure 11 describes the process. Every new release of the system provides a new function or solves a problem that resulted in previous releases.
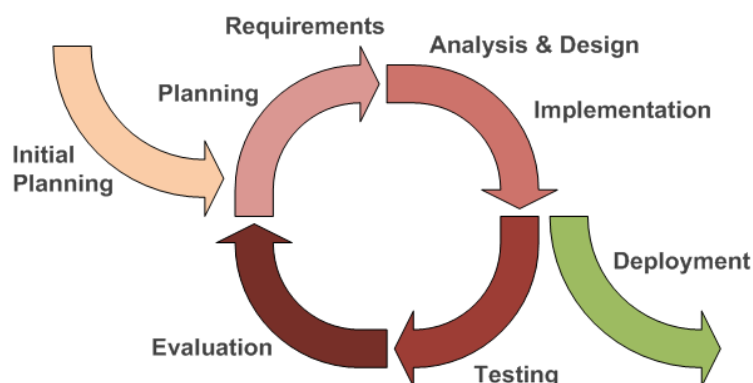
Figure 11: Agile Development Process[29]

The following sections explain all the agile methods that were used in development of our application:

4.1.1   Plan-driven Approach

In plan-driven agile development approach, separate stages of the software and their outputs are identified and implemented [37]. During the development lifecycle, the outputs of each stage were used to determine what had to be achieved in the next stage. Thus, planning for the next milestones was facilitated as a result of weekly meetings with the supervisor.

---

[29] Source: http://agile-development-tools.com

4.1.2 Extreme Programming

Extreme Programming (XP) is one of the most widely used agile development methods. In this method, requirements are specified to express scenarios, also known as user stories. Such scenarios are implemented as a series of tasks [37]. It is used to improve the quality of the software and to respond to changing or new requirements of the customer. Adapting extreme programming when developing this application facilitated implementing new functions in incremental steps. Each new release of the application included these new functions, which expressed scenarios. For example, in the first stage of the application, it was required to retrieve streaming tweets from Twitter and plot these on a world map. This scenario was implemented as a series of tasks that included implementing Twitter4J library, authenticating to Twitter's Streaming API, retrieving and parsing tweets, and plotting these tweets on a map using Mercator projection.

This chapter included an overview of the methodology and software development lifecycle throughout this project. The next chapter will discuss the system design and the major design decisions that were taken for the project.

5 System Design

5.1 Architecture Design

The architectural design describes how a system should be organized and it is used to design the general structure of the system [37]. Creating and documenting software architecture is an important step in developing software systems. It is the first step in software design process. Specifying the architectural design of a system builds a bridge between the design and requirements engineering [37].

During this stage, design goals of the project were defined and the system was decomposed to subsystems that had to be implemented in different stages of the development lifecycle. The following figure shows a model of the architecture of this application. It provides an overview of what components have to be developed.
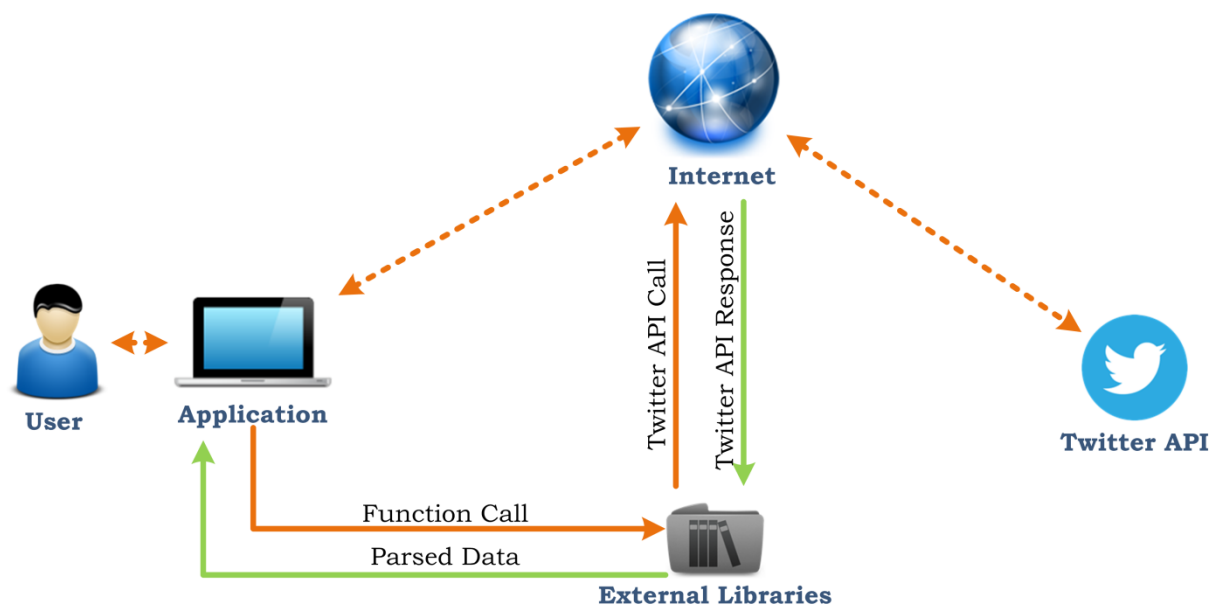
Figure 12: System architecture diagram of our application.

As seen in the above system architectural diagram, the user interacts with the application by making a request to view tweets or trends. The application uses Twitter4J external library to make a call to Twitter API over the Internet. Twitter API responds to the call by sending data in JSON format. Again, Twitter4J library parses the data according to the first request that was initiated by the user and sends the data to our application. The application then visualises the data. In addition, languages of tweets are classified using JLangDetect external library. The text of each incoming tweet is analysed and its result is returned to the application in String that determines its language. The application then plots the percentage of these various languages on a bar chart.

5.2 User Interface Design

According to Lauesen [38], the user interface is the part of a system that user sees, hears and feels. In fact, other parts of the system are hidden to the user. Most of the functionality in our application provides a user interface that depicts a gateway to Twitter's streaming API. It is easy to understand what is happening in various parts of the world by interacting with the application. As it visualises real-time Trending topics and tweets, it allows user to see what is important in the interface while

21

ignoring the hidden components that interact with Twitter's API and other libraries to retrieve, analyse and visualise the data. Once the application is opened, it establishes a persistent connection to Twitter API in order to listen to real-time tweets.

This application's user interface implements the following usability factors [38]:

1) Functionality: the application enables user to accomplish tasks laid out in the requirements specification.
2) Ease of Use: the application provides a simple interface that is easy to learn and use. The design of this application emphasizes on using conventional user interface elements that every user is familiar with. In addition, easy to understand guidelines are given in every part of the application. The application provides a help that explains how to achieve a particular task. This makes it easy for a user to understand how his requests are executed.
3) Task efficiency: our application is efficient at responding to user requests.
4) Ease of remembering: the application's user interface is minimal and conventional. This enables user to remember how to use the application even if it is used occasionally.
5) Subjective satisfaction: based on the evaluation of the system, users have responded with great satisfaction of achieving the required tasks.

After gathering requirements of the application, the next step was creating a prototype design of the user interface. During the design process of application's user interface, an HCI Classic approach of iterative design was taken. It facilitated learning the user and tasks that were required to be accomplished. As a result, prototype was created in the early design stage and it was reviewed with the supervisor. After the feedback, design was corrected to remove usability problems. The diagram below (Figure 13) depicts the approach towards designing user interface:
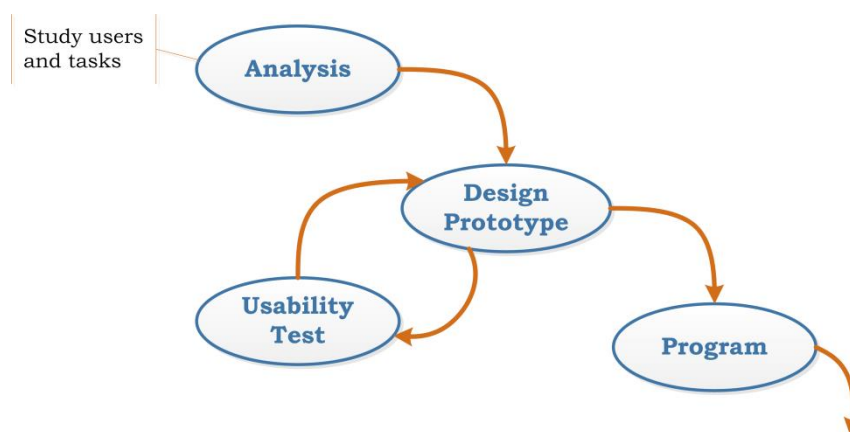
Figure 13: HCI Classic - iterative design[30] approach that was used in designing the user interface of our application.

---

[30] Derived from [38]   S. Lauesen, *User Interface Design: A Software Engineering Perspective*: Pearson Education Limited, 2005.

5.2.1 User Interface Mock-ups

The following section presents a set of user interface mock-ups of the application. These show how the user interface changed from the initial stages of design to the final.

It can be observed from the following mock-ups that the user interface of this application changed from time to time. These changes have been made on the grounds of best design practices in the field.
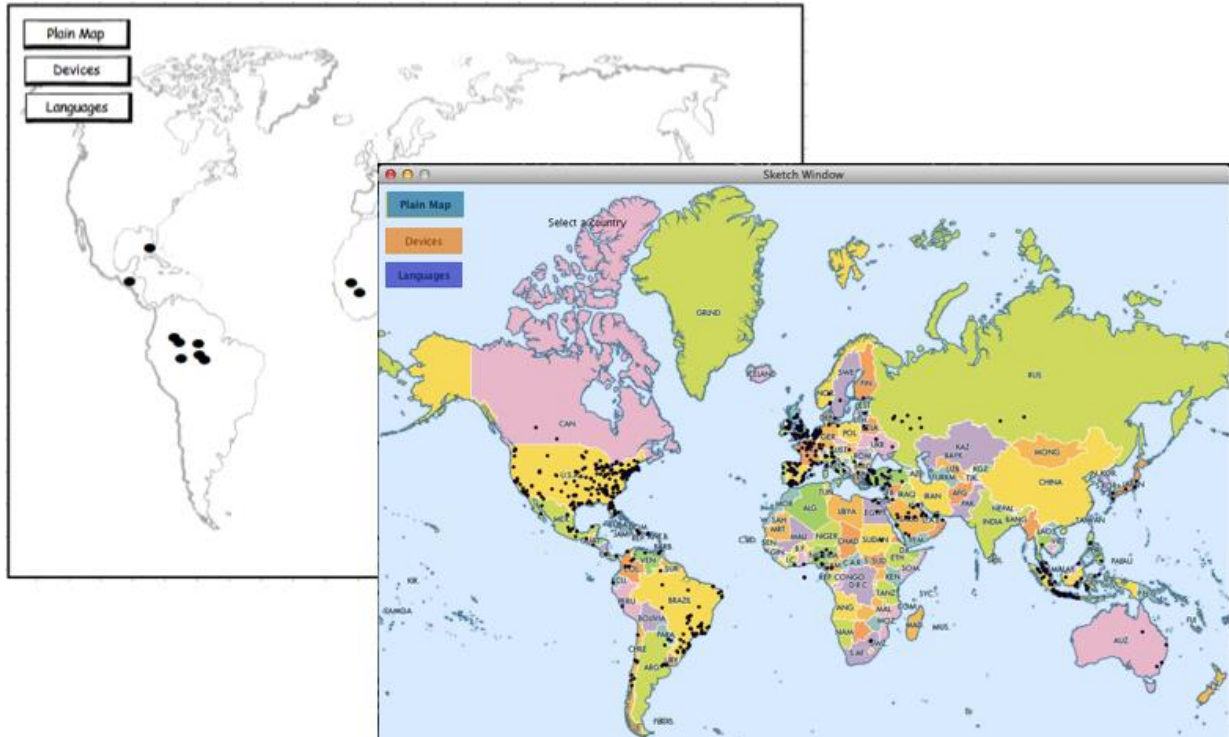
Figure 14: First mock-up and design of the home screen. The wireframe in the back shows how our initial idea of design looked like, and the screenshot in the front shows how the wireframe was converted into a working prototype.
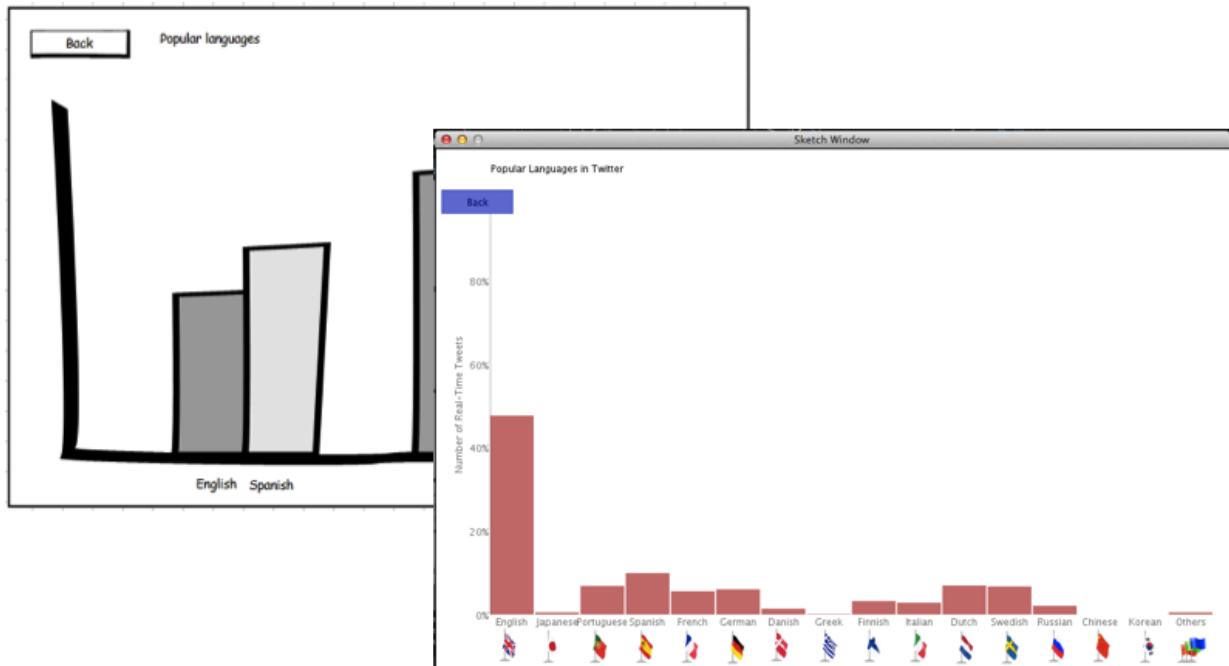
Figure 15: First mock-up and design of the languages screen. The wireframe in the back shows our initial idea of using bar charts to visualize languages of real-time tweets. The actual design in the front shows the working prototype of languages screen of the application.
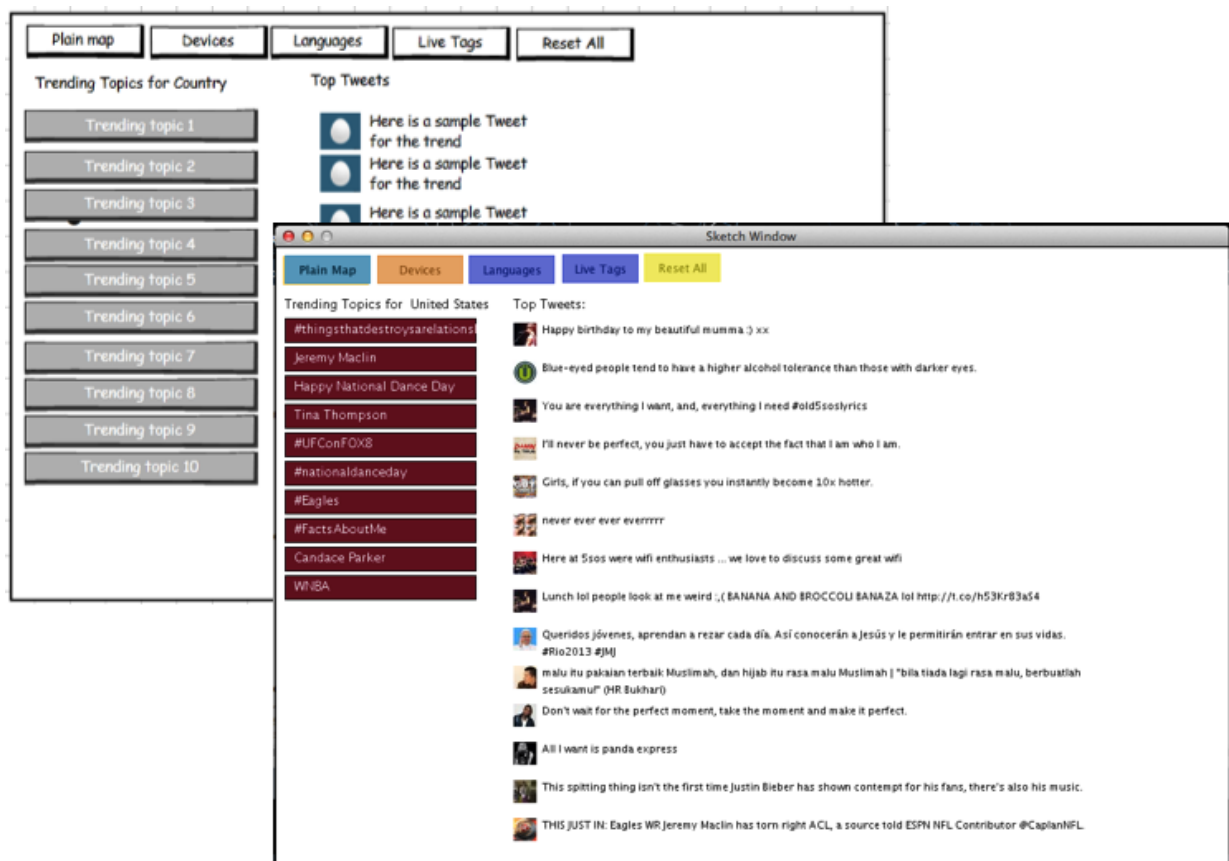
Figure 16: First mock-up and design of the trends screen. Initially, tweets were visualized as a list of messages along with the profile pictures of the users who posted them. However, this design was totally changed to present more information in this small space more effectively. As a result, tag cloud, images and tweet card were added to the design.
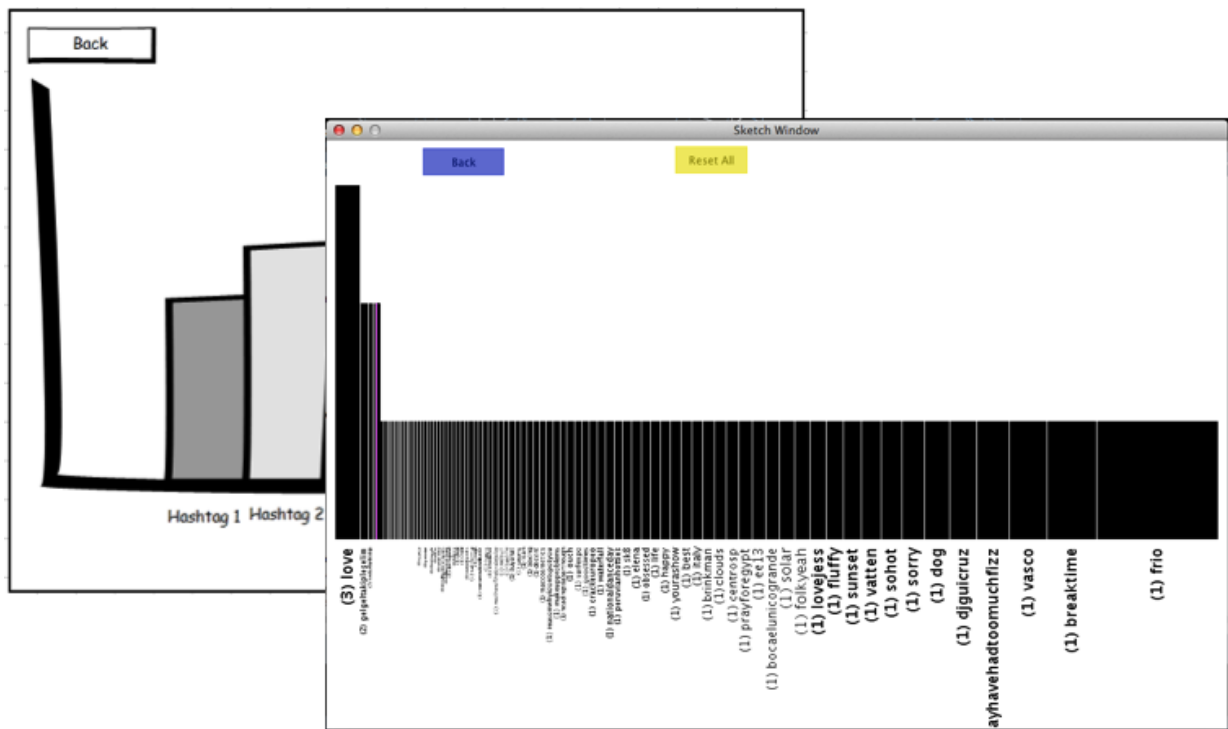
Figure 17: First mock-up and design of the Live Tags (hashtags) screen. This wireframe proposed simple bar chart to visualize popular hashtags. The screenshot on the front shows the first working prototype of languages screen.
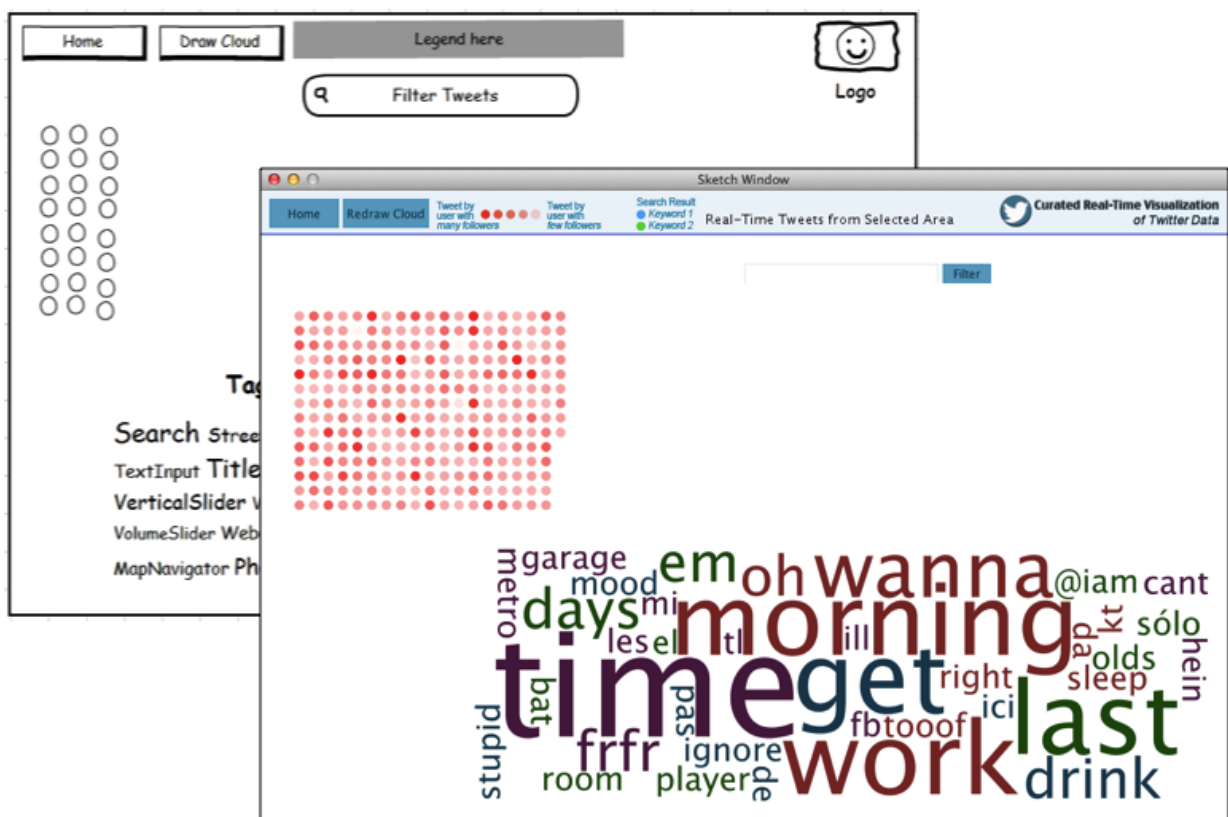
Figure 18: First mock-up and design of the screen showing selected tweets in an area on the map.

While creating the mock-ups and prototype designs, the main principle of user interface design [39] was applied in every stage of the user interface:

"Focus on the users and their tasks, not on the technology."

To keep this principle in mind, the design was simplified, unnecessary controls were eliminated, and appropriate colour combinations were used consistently throughout the entire application.

In order to justify the design decisions, focusing on users and their tasks influenced application's user interface in the following ways:

- The application is designed for an ordinary user. Any user with general computer savvy can use it. Complex tasks and interactions have been simplified to enable user to learn and remember how to use the application.
- Corresponding to the requirements, the application enables users to accomplish number tasks, such as exploring trending topics in a country, analysing tweets in a particular geographical area, viewing real-time classification of mobile platforms and languages and more.
- Easy to understand tips and instructions are provided throughout the application. Each screen is supplemented with a Help screen that explains how the functions of that particular screen can be accessed and used.
- Users are not required to perform any type of configuration-related tasks in order to start the application. It works well on any of the three platforms – Mac OS X, Windows and Linux, as long as they support Java.

In order to get user attention to the more important elements on a screen, the following guidelines were taken into consideration [40]:

Items such as legends, tweet cards and control elements have been highlighted or placed in a box and isolated from the rest of the elements in order to improve their visibility to the user and simplify navigation. The mouse pointer changes on clickable elements to indicate that these controls trigger actions. Moreover, the mouse pointer changes to a cross when it hovers over the map in order to allow user to choose a location more precisely.

The application uses the "direct manipulation" and "menu selection" interaction patterns [40]. The former emphasizes on creating visual representations of the tasks that should be carried out by the application. For example, a user can click on an ellipse to see a tweet information or user can click somewhere on the map to explore more. The latter is used to allow users to manipulate actions by selecting the appropriate menu item. For instance, a user can click on "Languages" button to see what languages are used in real-time tweets, and likewise, click on the "Home" button to go back to the first screen of the application. More examples of the "Menu Selection" can be seen in Figure 19. Direct manipulation facilitates easy learning, easy remembering, encourages exploration and helps user to avoid errors [40]. Example of "Direct Manipulation" is given in Figure 20. Menu selection interaction style reduces keystrokes, structures decision making and shortens learning [40].

Figure 19: Main navigation remains in the header of every screen in the application. It contains main controls that the user can use to switch between various screens.

Figure 20: Mouse pointer changes when hovered over the map to allow user to select a location more precisely.

Figure 21: Legend of mobile platforms is placed in a box on the map to follow conventional design principles when it comes to showing maps and their legends.
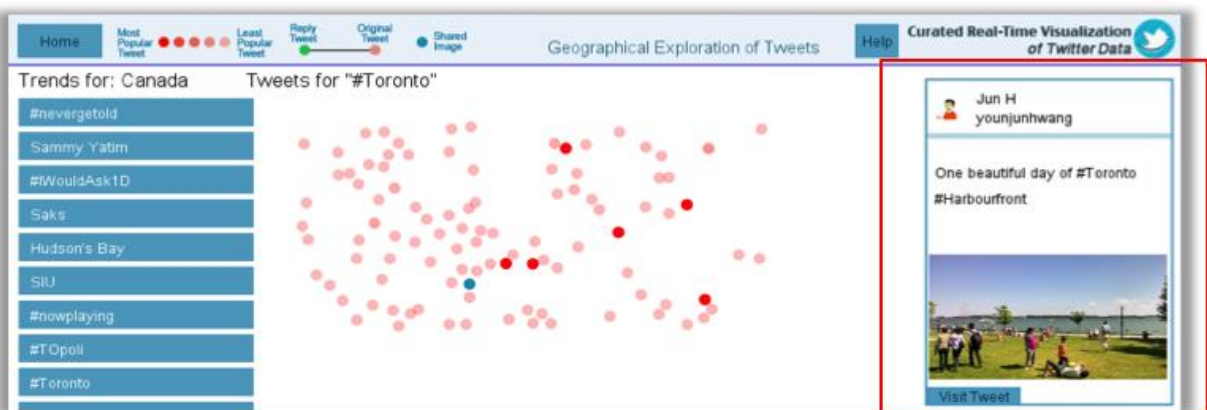
Figure 22: A tweet card on the right side enables user to read more about the selected tweet or visit it on a web page. It is separated from the rest of the page and it opens only when user demands for one by clicking on a tweet (the ellipse).

The following are final screenshots of the user interface for various screens of the application:
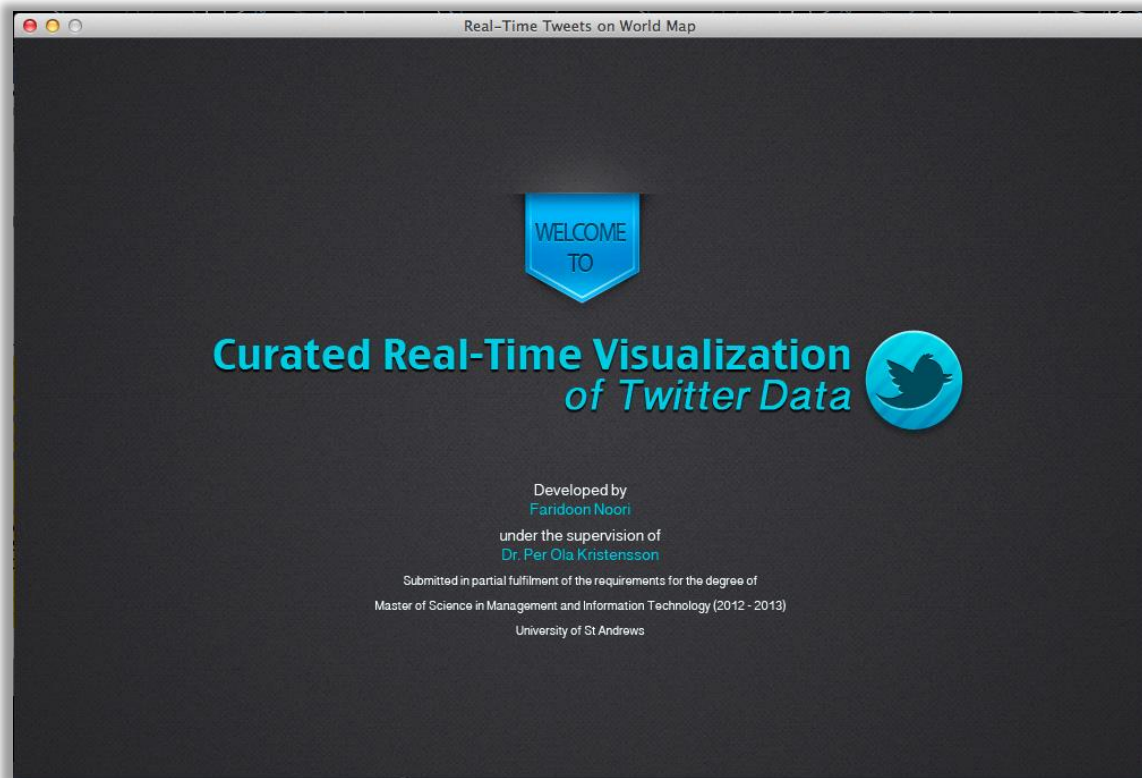
Figure 23: Splash screen appears for three seconds once the application starts.
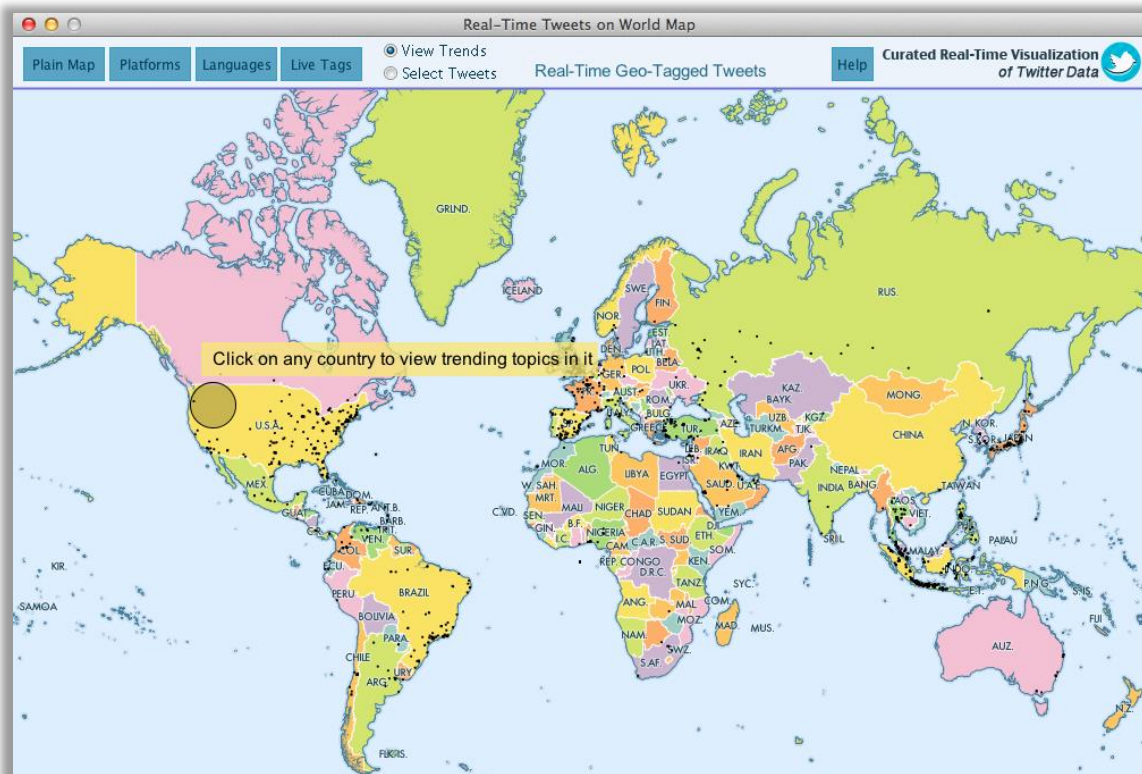
Figure 24: Home screen displays tweets from Streaming API on the world map. The tweets are presented as small black ellipses.
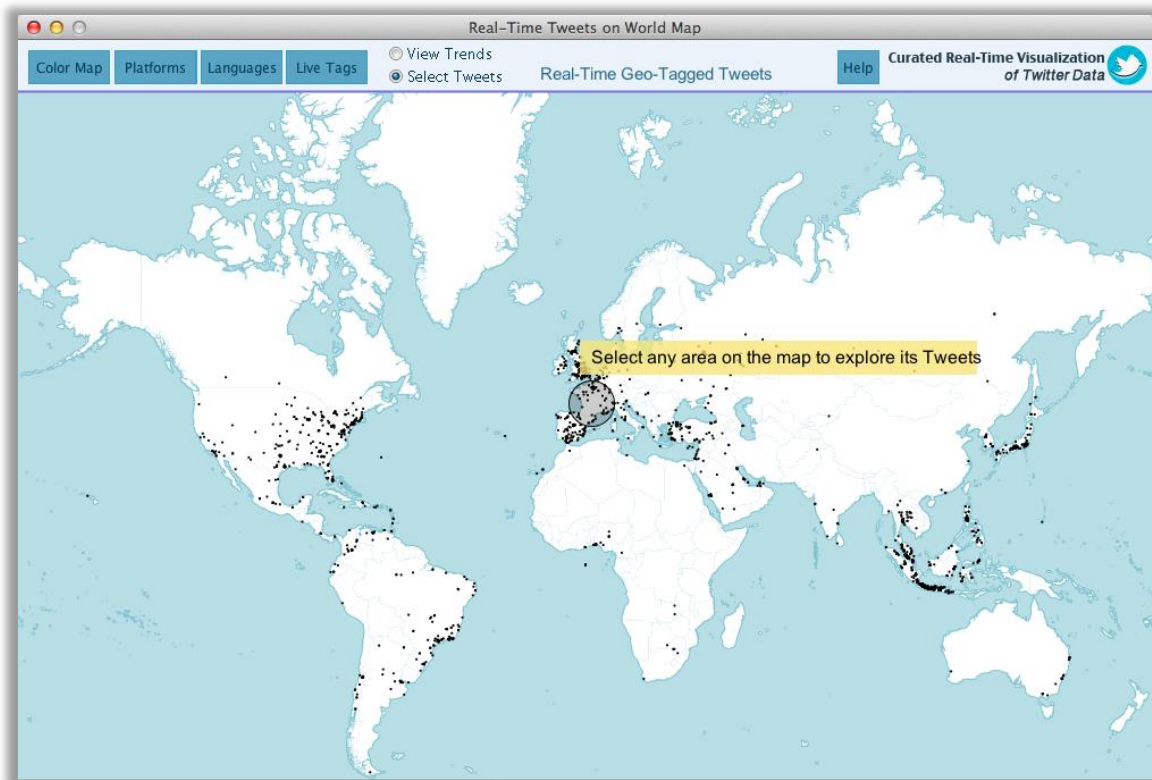
Figure 25: A variation of home screen map is the plain map where countries are not coloured. As shown in this screenshot, tweets are easy to see on the white background of countries.
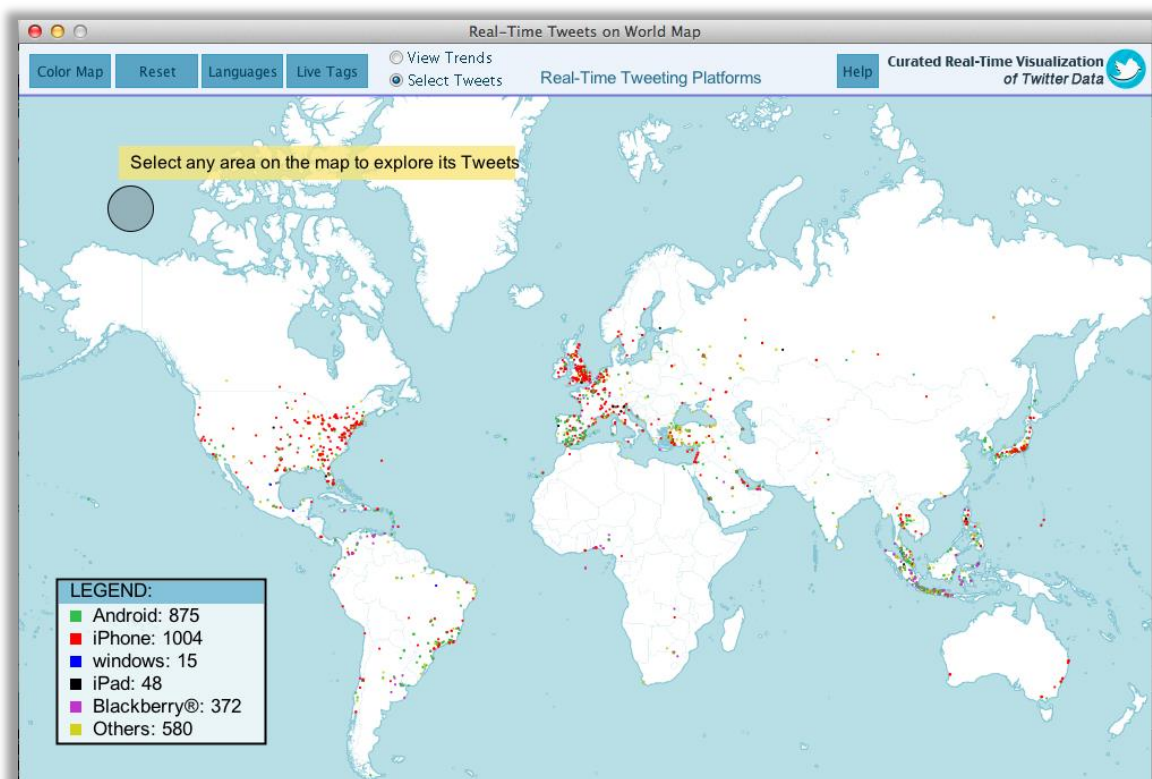
Figure 26: Platforms screen displays various mobile platforms used to tweet from. The legend box on the left bottom corner shows the classification of these platforms by presenting tweets in different colours on the map. The legend also provides a real-time number of tweets updated via these devices.
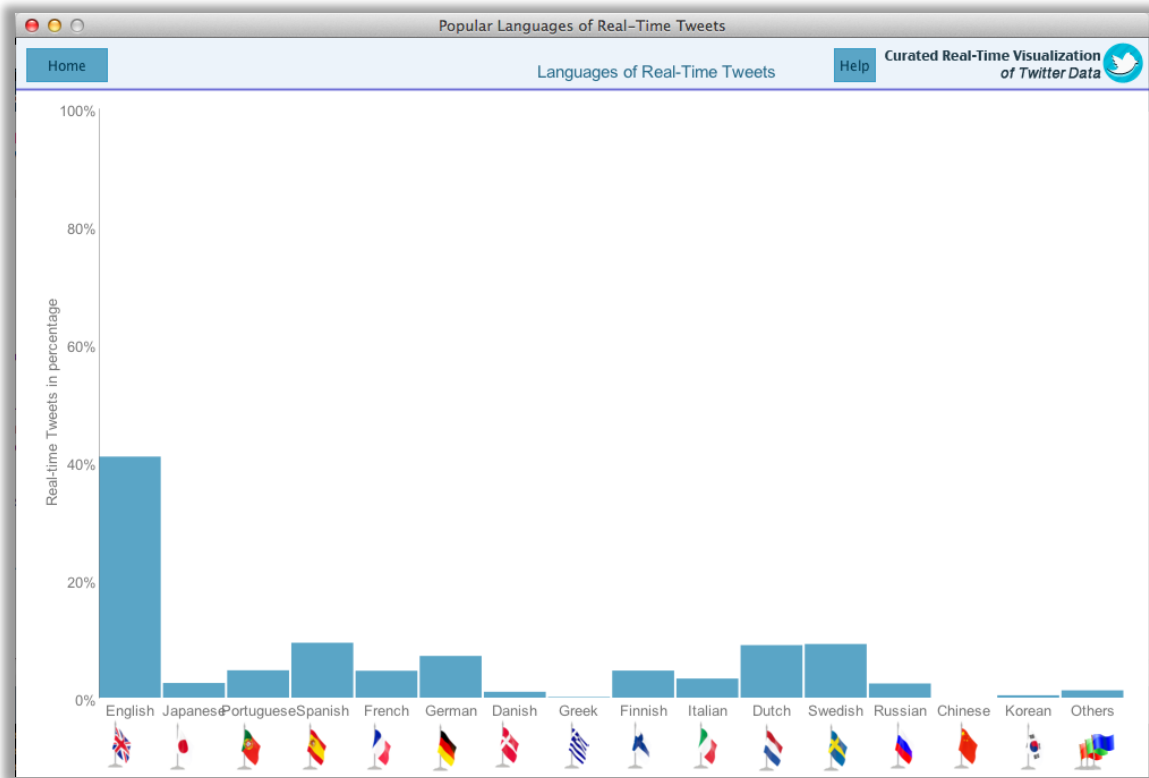
Figure 27: Languages screen displays real-time languages in tweets from Streaming API. This dynamic bar chart visualizes the amount of tweets for the specified languages in percentage.
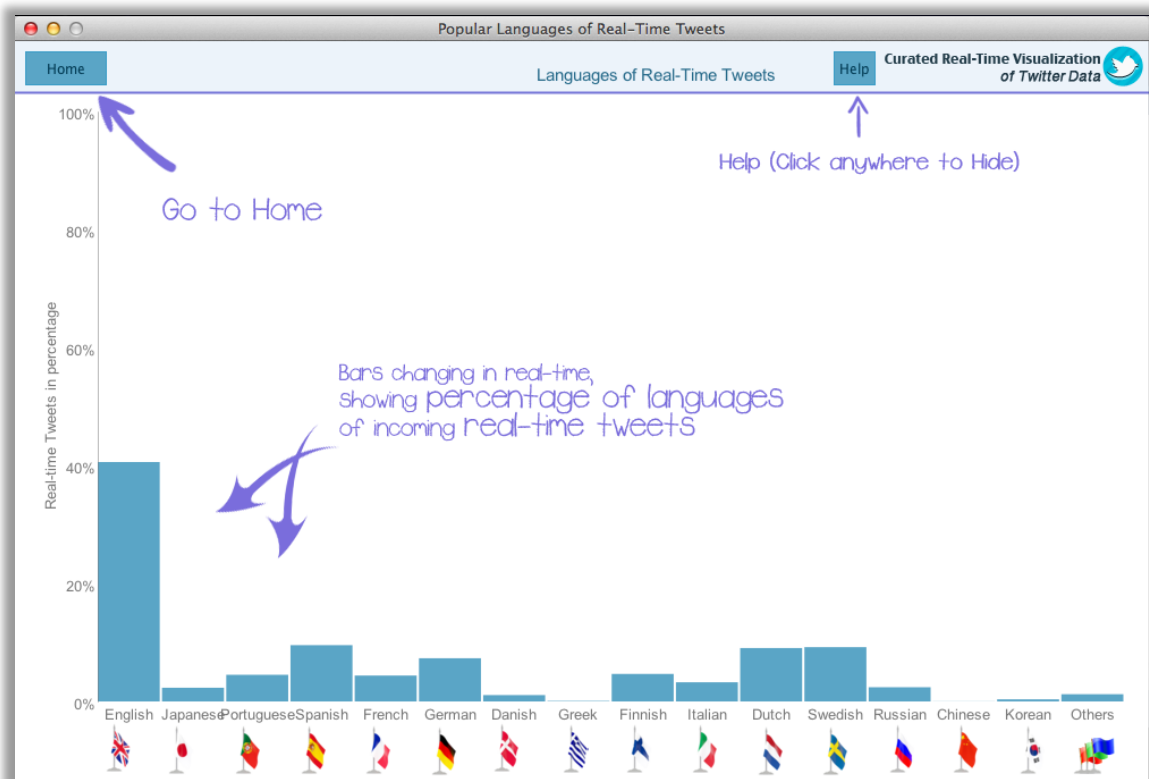
Figure 28: Languages help displays simple tips on how to use the different available controls. It also explains to user the different parts of the screen that are important for understanding the visualization.

Figure 29: Live tags screen displays popular hashtags from tweets in real-time. This dynamic histogram enables users to see the usage frequency of all hashtags that are extracted from real-time tweets since the application is started.

Figure 30: Popup warns user in case no trends are available in a country. It gives three options to the user so that it keeps the user engaged and improves interaction.

Figure 31: Trends screen displays top ten trends in the selected country with tweet relationships, images, tags cloud and a tweet card. The tweets are retrieved only for the selected trend. Opacity in the colour of ellipses that represent tweets determines their popularity or number of retweets. In addition, colour coding of the ellipses helps user to understand whether a tweet is reply to another tweet, or it shared the image that is on display.
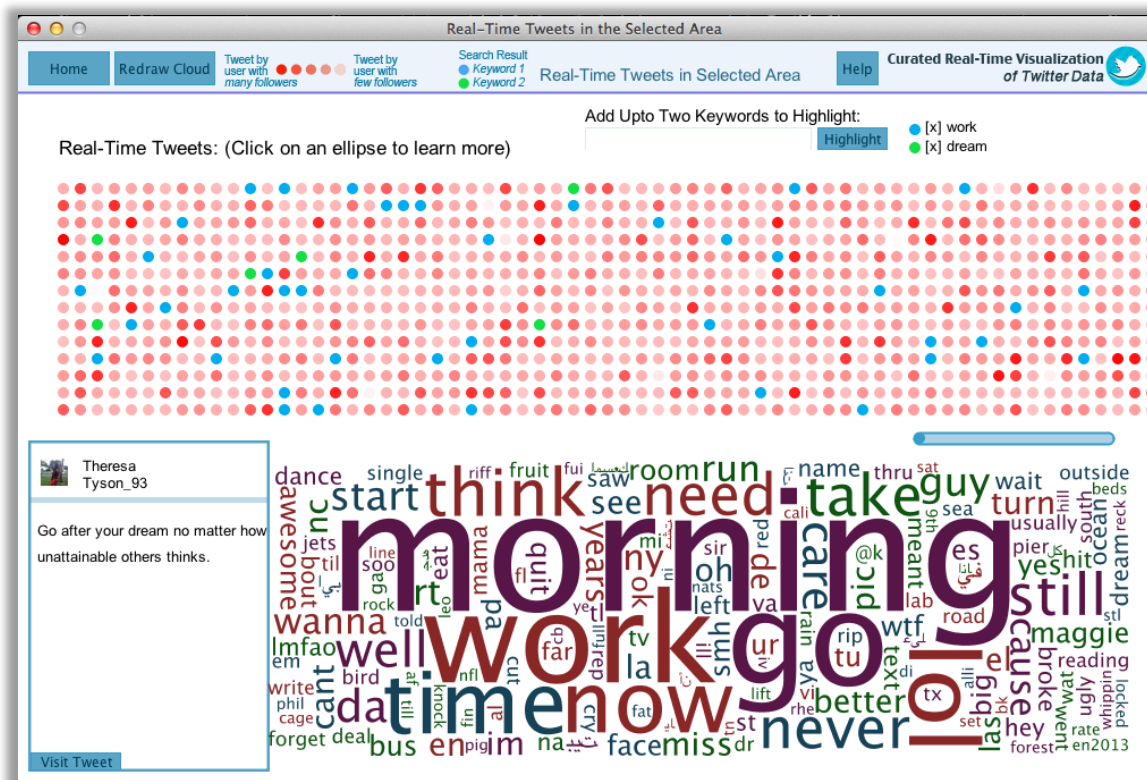
Figure 32: Selected tweets are incoming real-time tweets in a selected geographical location. This screen also shows tweet card, tags cloud and allows user to add two keywords to highlight the tweets. After adding the keywords, all retrieved tweets and further incoming real-time tweets are colour coded to disginguish the tweets that contain the given keywords.

This chapter discussed in detail the design decisions and the entire user interface design of the application. The next chapter will discuss the implementation of this project. It will provide overview of how all the requirements were turned into a real application during this phase.

6   Implementation

This chapter describes the implementation process of this application – the Curated Real-Time Visualisation of Twitter Data. The goal of this project is to visualise real-time data from Twitter in such a way that user can quickly get insight into what is happening in various countries by visualising tweets and trends geographically. In addition, users can gain more interesting insights from the streaming tweets in terms of exploring the popular languages, hashtags, and platforms used by users of Twitter.

After designing the system architecture and user interface, the next stage of development was the implementation phase. Based on the requirements specification, the functions of the system were incrementally implemented. The following sections describe how each requirement was turned into a functionality of the application and what major design and implementation decisions were made.

6.1   Connecting to Twitter API

The application uses OAuth[31] to connect to Twitter API. In this project, Application-only authentication model of API v1.1 is used. This enables the application to make requests to Twitter API on its own behalf without requiring user-level authentication. This takes place in the following three steps:

1) Encoding application's consumer key and secret into a set of encoded credentials.
2) Making a request to the POST oauth2/token[32] endpoint which will return a bearer token.
3) Using the bearer token to access the API.

6.2   Retrieving Tweets

The primary data that the application visualises is tweets. These tweets are consumed from Twitter's Streaming and Search APIs. This application uses Twitter4J library to simplify the process of parsing tweets that are retrieved in JSON format from Twitter.

The application retrieves data from the following three APIs:

6.2.1   Streaming API

As the application has to plot tweets on a map, it needs to retrieve tweets that contain geo-location data. Therefore, "Filter" endpoint of Streaming API is used in this application in order to pass location parameters. Twitter then sends only those tweets that contain latitude and longitude values.

In addition, it is important to note that a persistent HTTP connection is maintained to Twitter API. The application listens to a feed of tweets and stores the tweet objects in a list. This list of tweets is then used throughout various functions of the application. These objects are reused for the purposes of classifying languages, analysing hashtags, and classifying platforms that are used as sources of tweets by the users. This process is visualised in the following figure:

---

[31] http://oauth.net
[32] https://dev.twitter.com/docs/api/1.1/post/oauth2/token
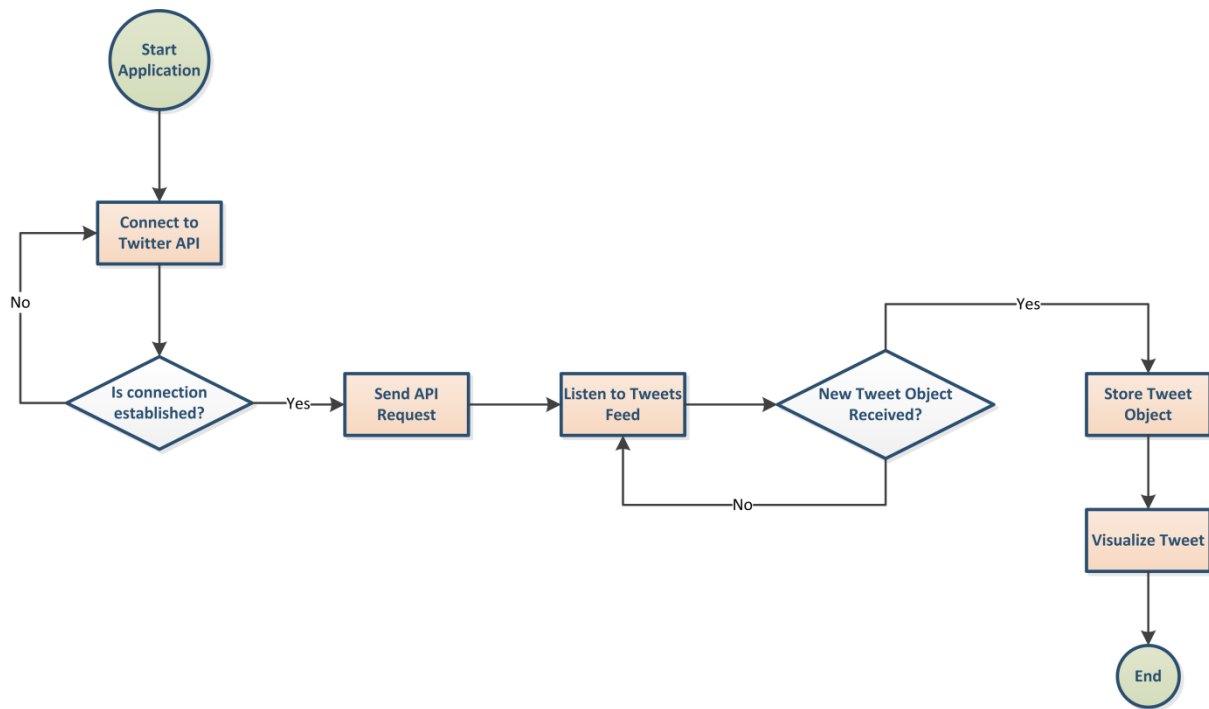
Figure 33: Process Flow Diagram for retrieving tweets from Streaming API.

6.2.2   Search API

The Search API is used in retrieving tweets for trends. Once the application retrieves top ten trends for a particular country, and then it makes a request to Search API to retrieve a hundred relevant tweets for that particular trend. Using the Search API in this context is used to pass a parameter that defines the trend. The result is a set of tweet objects that are stored in a list. These are then further exploited for determining relationships among the tweets by visualising replies and classifying popularity of each tweet by determining the retweet count. This process is explained visually in Figure 34 below.
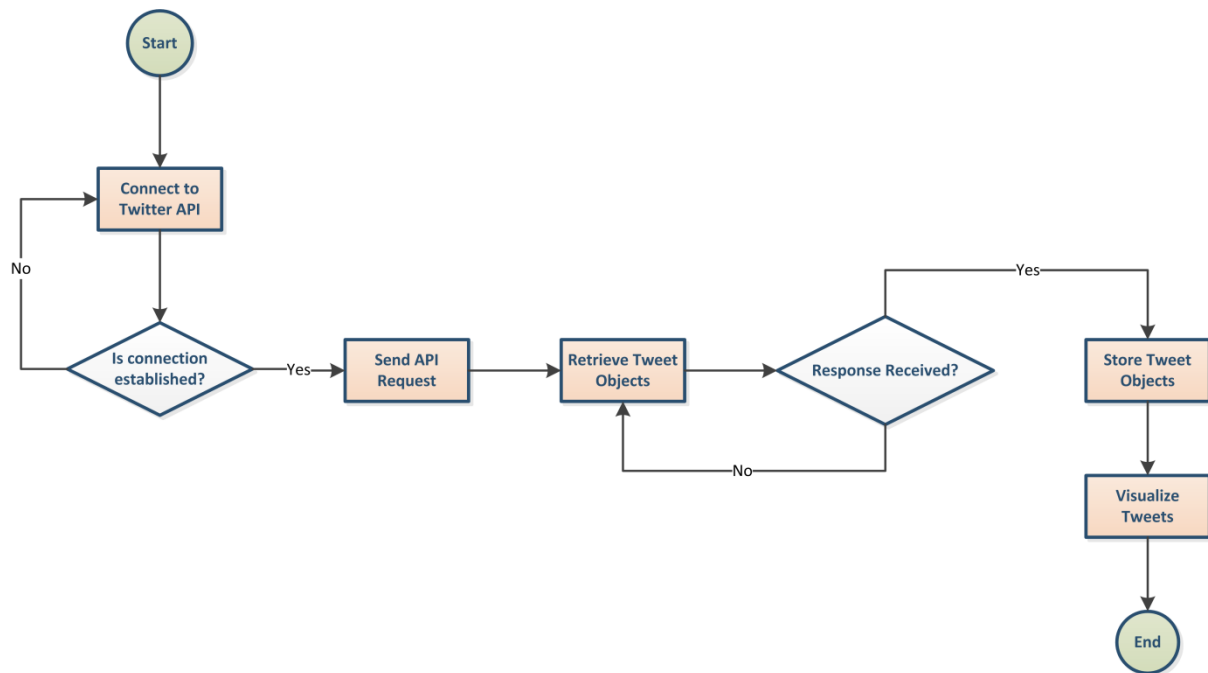
Figure 34: Process flow diagram for retrieving tweets from Search API.

6.2.3   Trends API

This application first determines all the locations where Twitter provides trends. The result of this step is stored in two lists of WOEIDs (Yahoo! Where On Earth ID)[33] and corresponding country names. By clicking on any location on the map, the application recognizes name of the country and then passes its corresponding WOEID as a parameter to the trends API. Twitter then returns ten trending topics for this given country. If trends are not found for any given location, the application warns a user about this and provides three options:

1) Continue viewing top ten trends in the world: this will retrieve trending topics in the world. Their relevant tweets, tag cloud and images are visualised subsequently.
2) Search for tweets in the selected region: if user chooses to see tweets from the area where the map was clicked, tweets from the Streaming API are retrieved and visualised subsequently.
3) Go home: the user is simply taken back to the home screen of the application.

The following figures explains visually how trends are retrieved in our application:

---

[33] http://developer.yahoo.com/geo/geoplanet
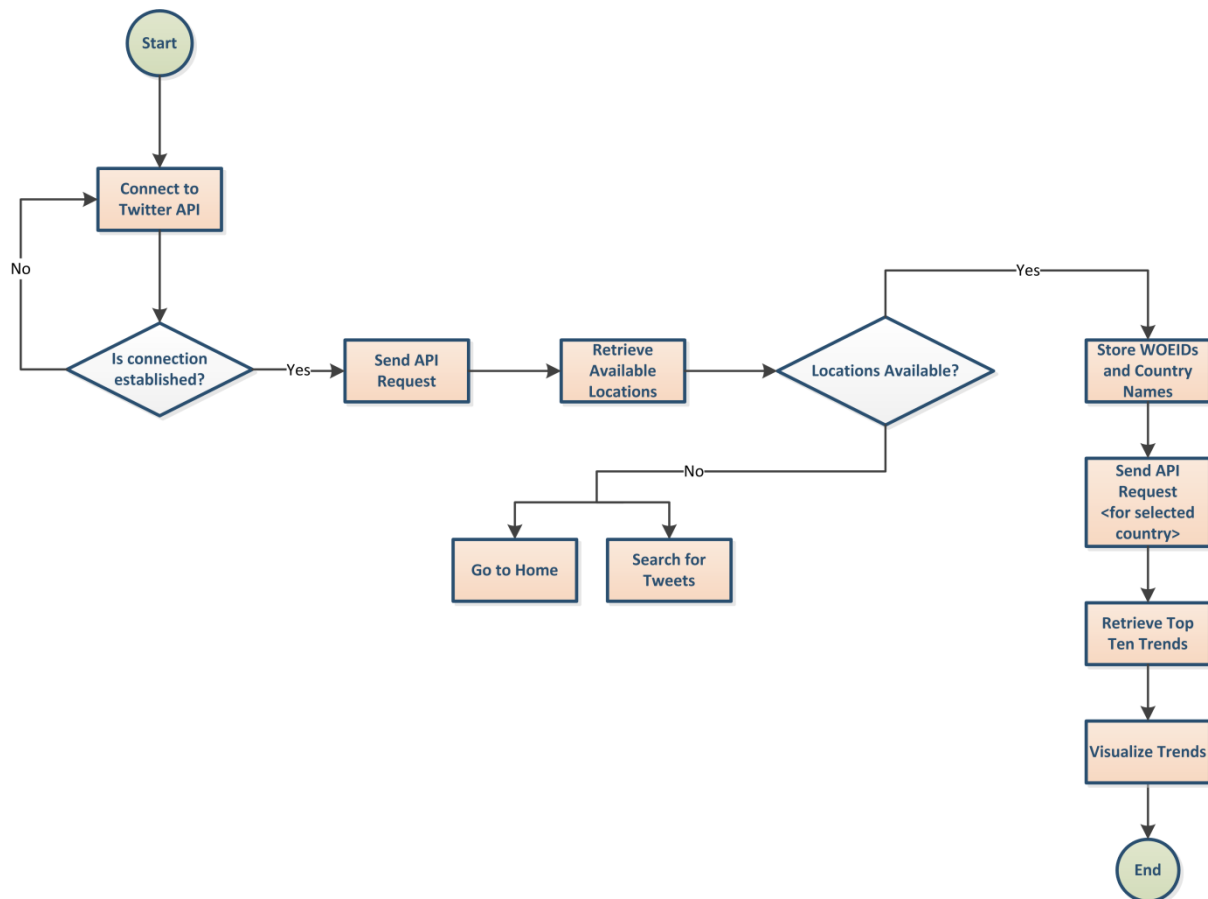
Figure 35: Process flow diagram for retrieving trends.

6.3   Visualising Tweets on Map

Plotting real-time tweets on a world map makes the application more interactive and interesting to users. As tweets are retrieved with their geo-tags (latitude and longitude coordinates), these are plotted on the map using Mercator[34] projection. This type of map projection facilitated creating and deploying custom maps using TileMill[35]. The two map variations, a) with coloured countries, b) with plain colour, makes the application more customizable for the user. A user can decide to choose either of the maps for more personalized visualisation.

6.4   Visualising Platforms of Tweets

Tweets that are fed by the Streaming API to this application are stored in lists. These objects are parsed for determining their sources in this component of the application. The tweets are represented by ellipses of different colours that distinguish the classification of platforms in real-time on the map. Users are also allowed to switch between the two types of maps for customizing the visualisation. A legend is provided on the map that contains real-time numbers of the different platforms of tweets.

---

[34] https://en.wikipedia.org/wiki/Mercator_projection
[35] http://www.mapbox.com/tilemill

6.5   Visualising Languages of Tweets

The application uses an external Java library – JlangDetect[36] to classify languages of real-time tweets. For the time being, the library can detect fifteen languages (English, Japanese, Portuguese, Spanish, French, German, Danish, Greek, Finnish, Italian, Dutch, Swedish, Russian, Chinese and Korean). The rest of the tweets are categorized as "Others" in the languages classification. The bar chart that visualises these languages is dynamic which changes its values based on the percentage of tweets received for each language from Streaming API.

6.6   Visualising Hashtags in Tweets

People use the hashtag symbol (#) with keywords or phrases (without any spaces) to categorize their tweets. As a result, hashtags become topics of interest across Twitter. A hashtag can describe its relevant tweets and the included subject of discussion. This application extracts hashtags from real-time tweets and appends a frequency number to each of them. Subsequently, these hashtags are visualised in a real-time bar chart that sorts the elements from most popular to least popular. This enables a user to understand what people are discussing in that particular time in the world based on real-time tweets as opposed to the top ten trends that are determined by Twitter for a given country or the entire world.

6.7   Visualising Trends

When top ten trends are retrieved for a particular country or the world from the trends API, the application displays gives a list of ten buttons that represent these trends.

In order to allow users to view trends of a country by clicking on the map, pixels of another map that contains unique red colours for each country are loaded to the home screen. As a result, the colour of pixels beneath the mouse pointer is picked up and its corresponding country name is stored as a string. This name is checked against all the available trending locations. If a match is found in the list of locations, it means that Twitter provided trends for the given country. Subsequently, a request is made to retrieve the top ten trends for this country from trends API. This process is visualised in the following figure:

---

[36] http://www.jroller.com/melix/entry/jlangdetect_0_2_released
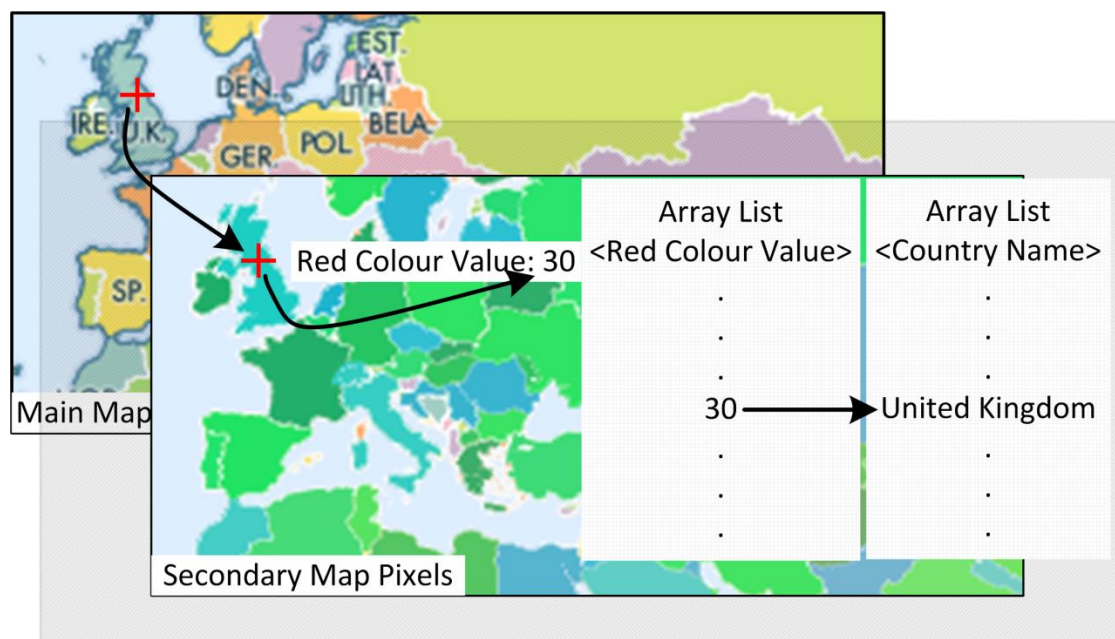
Figure 36: This figure demonstrates how a country is selected on the map in order to retrieve its top ten trends.

6.7.1   Visualising Related Tweets

A user can explore a particular trend by clicking on it, and subsequently the application performs a Search for the trend and retrieves relevant hundred tweets. These tweets are visualised in shapes of ellipses that are classified according to their popularity. The popularity of each tweet is determined by the number of times it has been retweeted by others on Twitter. Fill colour opacity is used to visualise and distinguish the level popularity for the tweets. In addition, relationships are determined among tweets that are replies to each other. Again, different colours are used to determine the original and reply tweets.

6.7.2   Retrieving Related Embedded Images

Twitter provides embedded links of images in tweets where people shared images. By extracting only those links that contain photos from tweet objects, the application shows a slideshow of these images for a selected trend. Tweets that have embedded a particular image are given a different colour so that users can easily explore more about this image or the tweet by clicking on the corresponding tweet ellipse. A constraint that retrieves only unique images in a given trend has been implemented in the application. Usually, Twitter users may embed the same image in their tweets. To avoid showing the same image repeatedly, the application checks for unique embedded URLs and then parses the relevant images.

6.8   Generating Tag Cloud from Tweets

For visual presentation of the textual content of tweets, the application generates tag clouds with the help of WordCram[37] library for Processing. A tag cloud can help user to quickly understand what people talk about in a particular set of tweets. These tag clouds are generated for all the trends as well as selected real-time tweets from a geographical area.

---

[37] http://wordcram.org

The application retrieves text of all tweets and stores it in a temporary external file. The tag is generated as a thread to prevent interruption and waiting times in carrying out other tasks in the application. This process creates a temporary image of the tag cloud and places it on the screen.

While generating tag cloud for a trend, the trend name is passed as a stop word to the cloud. This prevents this particular keyword from taking a bigger space on the tag cloud. As a result, it allows the other relevant keywords to be more emphasized.

6.9   Reading a Tweet

To allow users to read a tweet from a trend or selected area on the map, the user can click on an ellipse and a box that displays username, profile image, status message and embedded image (if available) will appear. This is helpful for those who would like to read the actual text and gain detailed insights into what is happening around a particular trend or in a geographical area.

In addition, the application provides a link to the original tweet. By clicking on this button, the user is taken to the tweet's webpage in the default web browser as seen in the following screenshot:

Figure 37: A tweet card retrieved from #Melbourne trend in Australia.

6.10 Visualising Tweets from a Geographical Area

In addition to the major component of this application that visualises trending topics in countries around the world, another component that explores real-time tweets that fall in a particular geographical area has also been implemented. The idea behind visualising a set of tweets from the Streaming API is to allow users to gain more customized insights into what is happening around the world. By selecting an area on the map, the user is taken to a screen where real-time tweets are visualised in ellipses that categorize tweets by users with least and most followers. This enables users to determine whether these tweets are coming from a popular Twitter user. In addition, tag cloud is generated for the incoming tweets in order to allow users to gain insights into a summary of what these tweets are talking about. The tag cloud is

generated after thirty tweets are collected in an area, and later, a user can redraw the cloud as many times as desired.

There could be thousands of tweets coming from the selected area, and to allow easy exploration of these tweets, the application visualises all of these in rows and columns of ellipses. If the columns exceed the width of screen, the application is then displays a horizontal scrollbar to allow user to scroll through all the tweets back and forth. This process is visualised in the following diagram:
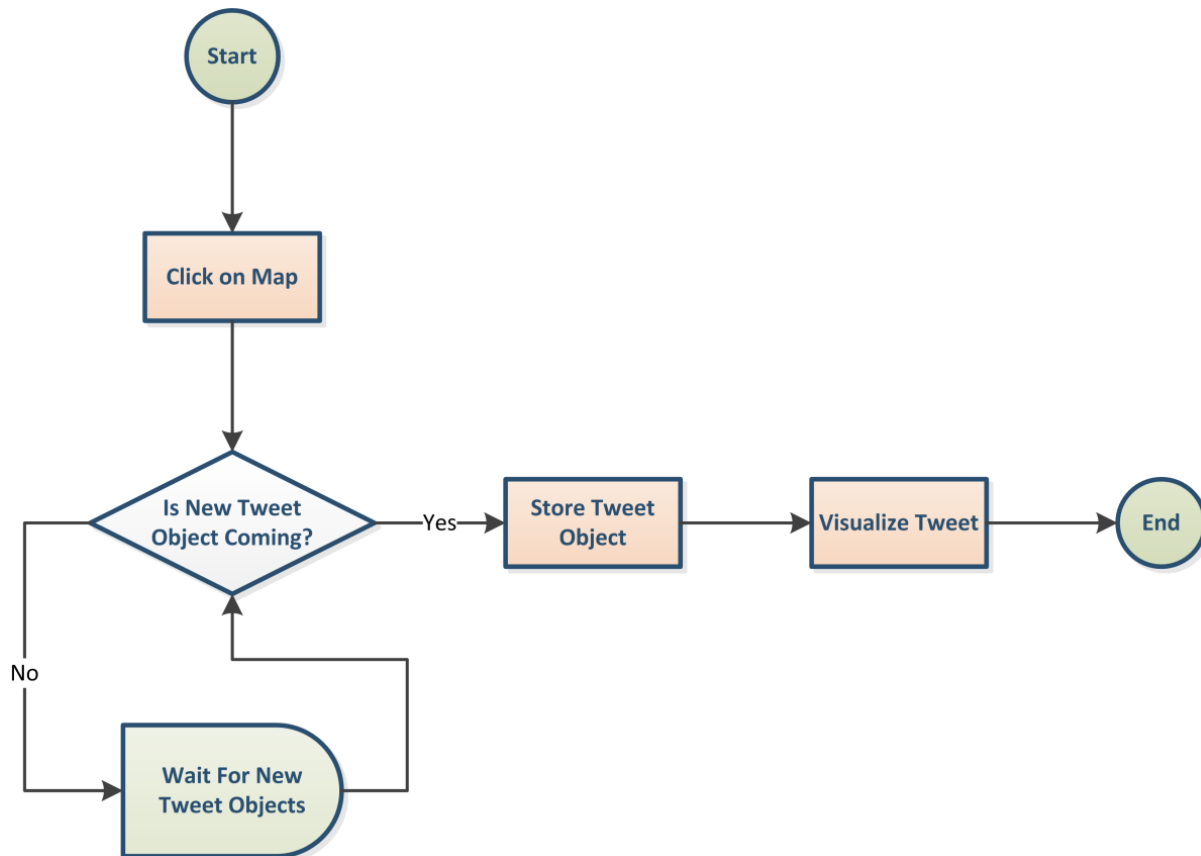
Figure 38: Process flow diagram for visualising real-time tweets from a selected area on the map.

6.10.1 Highlighting tweets

When tweets are visualised in a selected area from the map, these could provide intriguing patterns if user is enabled to highlight them based on keywords. This application allows users to add up to two keywords that would eventually highlight the received and new tweets in two different colours. This can be used to compare two trends or terms and track them in real-time.

Users can add a keyword in a text box and a legend that corresponds to highlighted tweets is created. Any keyword can be removed and the tweets classification can be reset to original colours. A relevant example is given in Figure 39.
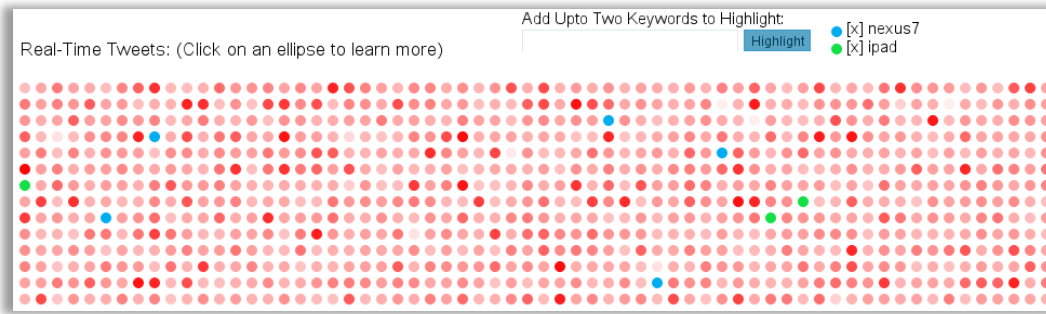
Figure 39: Real-time tweets retrieved from an area around Texas, US during the release of new Nexus 7 Tablet in late July, 2013. It shows how the application highlighted the tweets that mentioned any of the two given keywords.

6.11 Providing Help

The application provides simple guidelines on how to use the various functions. In every different screen, the help is customized and provides visual tips that can be easily understood and remembered. Below is a screenshot of the Help for application's home screen:
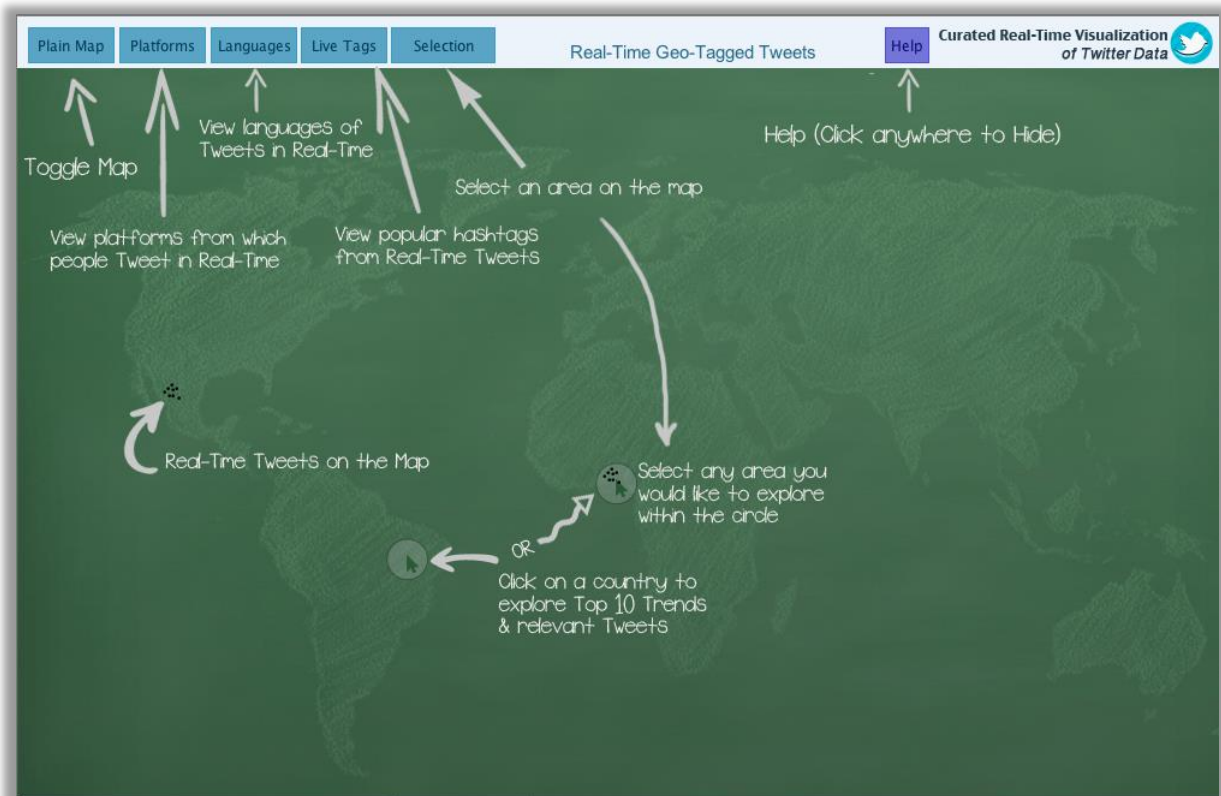
Figure 40: Help for home screen of the application gives visual tips and instructions on how to perform the various possible tasks.

6.12 Data Visualisation Methods

A *taxonomy,* as defined by Kirk [41], refers to the organization of members that share the same characteristics into groups. An example of such taxonomy in data visualisation can be chart types that contain data of similar characteristics.

42

Finding the most appropriate visualisation method depends on the following questions [41]:

- ⊙ What is to be said?
- ⊙ How should it be showed?

In this project, it is clear that users are presented with data from Twitter, which should help them understand what is happening in the world and to provide a gateway to the Streaming API of Twitter.

To visualise this data, the following methods have been used:

6.12.1 Comparing categories

The purpose of this method is to allow users to compare categorical values relatively [41]. An example of this method is the use of bar charts in comparing languages and hashtags of real-time tweets in the application.

6.12.2 Showing changes over time

The purpose of showing changes over time in data visualisation is to exploit temporal data [41]. As real-time data is retrieved from Twitter APIs, it is important to show changes in the data visualisation continuously. This method is used in plotting tweets on the map, classifying mobile platforms, classifying languages of streaming tweets, calculating frequency of hashtags and visualising them in bar charts, and retrieving real-time tweets from a selected area on the map.

6.12.3 Plotting connections and relationships

The purpose of plotting connections and relationships is to visualise associations and patterns among datasets [41]. This gives users the ability to gain insights into how data elements are connected to each other or what communication patterns exist. This method is used to display relationships between original tweets and their replies for a given trend.

6.12.4 Mapping geo-spatial data

The purpose of this method is to plot data based on its geo-location data on a map [41]. Real-time tweets are visualised on a world map, which allows users to understand where on earth people tweet.

6.13 Elements of Design

The application uses the following elements of design to represent and visualise dimensions of data points. The data has been encoded by colour, position, shape and size. The following section explains each technique and justifies the decisions that were taken to pursue a particular technique in the data visualisation:

6.13.1 Colour

Colour can be an excellent encoding tool for labelling categorized data [42]. It helps users to differentiate between different types of data. For example, this application applies colour encoding to classifying sources of tweets of different categories. As tweets of different colours that represent mobile platforms are plotted on a map (Figure 41). The real-time nature of data and classification makes it more appealing and interesting for users to understand what kind of device is used the most for Tweeting in various locations across the world. Likewise, the popularity of tweets is also represented by the opacity of red colour that has been applied to ellipses

representing tweets. These help users understand which tweets are more popular than the others. Another example is given in Figure 42 where real-time tweets have been highlighted by two keywords. The colour encoding helps users to visually compare the volume of tweets for both keywords.
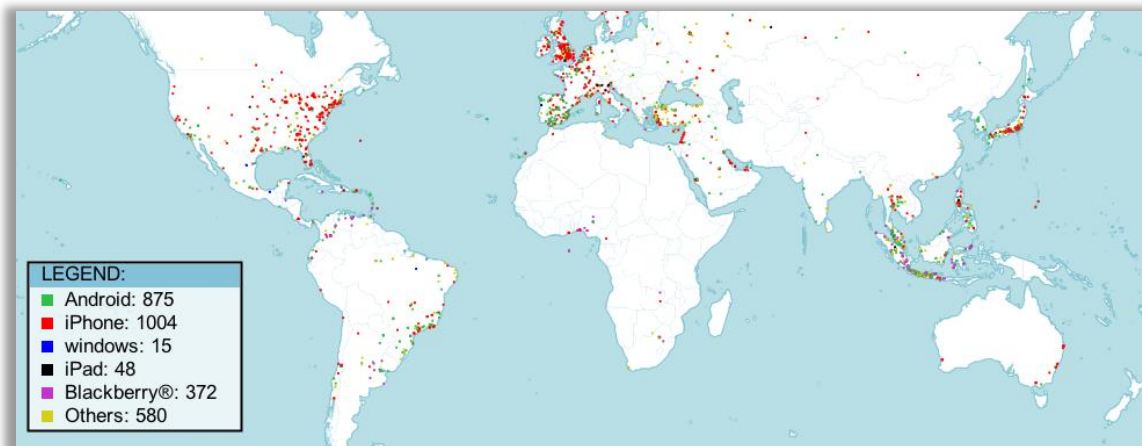
Figure 41: Colour encoding used in visualising classification of platforms of real-time tweets.
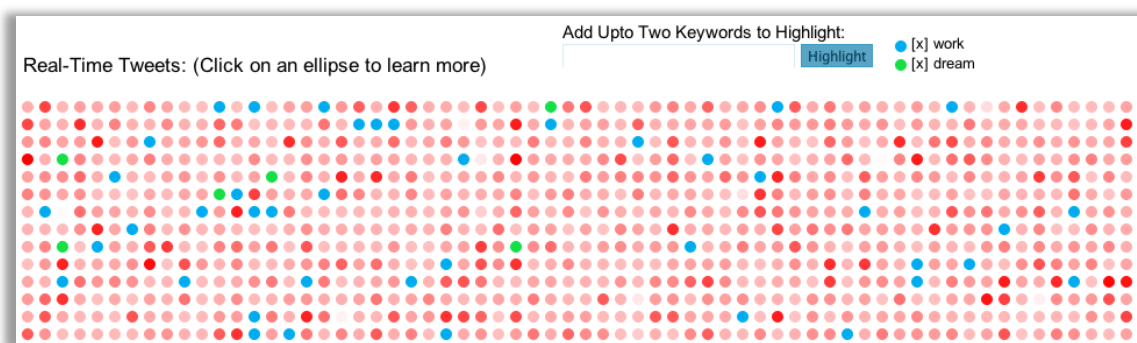
Figure 42: Colour encoding used in classifying and highlighting real-time tweets from a selected area on the map.

6.13.2 Position
tweets are presented visually on a world map. Thus, it is important to plot each tweet based on its latitude and longitude coordinates on the map. This enables users to easily understand where on earth people discuss in Twitter. Besides, it helps users to choose any location without having to search by typing in a text box.

6.13.3 Shape
Throughout the application, tweets are represented consistently by ellipses, and their relationships are represented by lines. This helps users remember the association of data objects with the shapes easily. Subsequently, learning how to use the application becomes easy and enjoyable.

6.13.4 Size
Size can be an influential encoding technique in visualising data. It can be used to highlight the importance of entities in order to make them more eye-catching [42]. Different sizes of bar charts have been used to visualise real-time hashtags (Figure 43) based on their popularity of use, and to visualise classification of tweets'

languages in real-time. This helps users understand what languages are used on Twitter. It is not possible to gain such insights by visiting Twitter.com.
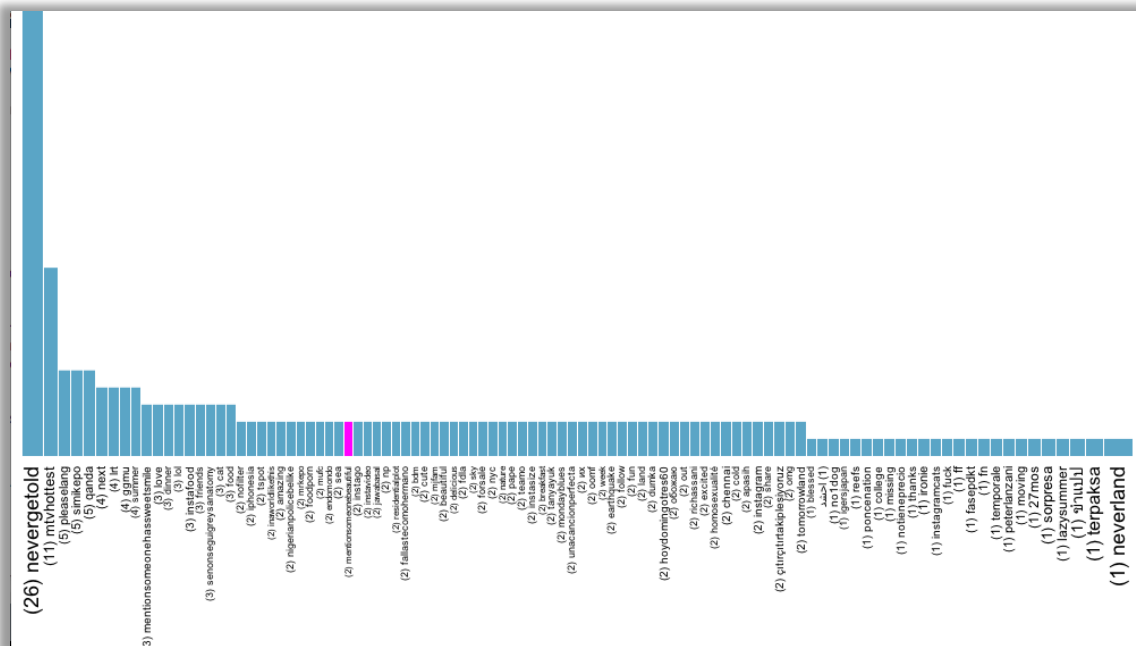
Figure 43: Encoding by size in visualising hashtags from real-time tweets.

6.13.5 Text and Typography

It is important to use appropriate text and numbers as titles, axis labels or other kinds of labels or indicators to create effective visualisations [42]. The design of this application emphasizes on providing meaningful and minimalist textual and numerical labels, and legend with each various entities such as languages, platforms, and hashtags. Titles are used to describe each screen of the application.

6.13.6 Tag Cloud

Tag clouds are used to show the frequency of words in a set of text [41]. Tags or words of larger sizes depict their higher frequency and therefore making it easy for users to understand a general theme of the entire dataset. The application is generating tag clouds for tweets of a trend (Figure 44), and for real-time tweets that are retrieved from a particular geographical area on the map.

Figure 44: Tag cloud generated from NSA trend in United States in late July.

This chapter provided a complete overview of how the system was implemented. The following chapter will provide an overview of the functional testing of the application, usability study, and qualitative investigation.

7 Testing

Testing is an important phase of software development lifecycle, which can be used to determine whether the system does what it is intended to do. Furthermore, it can be used to discover flaws, and use its outcome to improve the system.

This chapter explains how the application was tested from its functionality point of view, and presents the results of this test. It also discusses the usability study that was carried out for evaluating the usability of this application, and a qualitative investigation that was undertaken to gather feedback from participants of semi-structured interviews.

The testing process has two objectives [37]:

1) To demonstrate whether the software meets the requirements.
2) To determine situations where the software behaves abnormally or does not match its specification.

To ensure that this application meets the requirements and it operates without any faults, the following testing strategies were used:

7.1 Unit Testing

When performing unit testing, the whole system was isolated into several smaller units. This facilitated testing of each unit of the application and determined whether improvement was required. The units that were tested in this stage were individual methods. For each method, a test was carried out for all possible ranges of inputs. Once this stage was completed successfully, and the resulted bugs were fixed, the next testing strategy was executed.

7.2 Integration Testing

In this stage of testing, the smaller units that were tested in the first stage were put together as single unit. The whole unit was tested for different ranges of inputs. This was used to evaluate the interaction between the units that were tested in earlier stage of the testing strategy. Once all the major components were tested successfully, and bugs were fixed, the application was ready for the next level of testing.

7.3 System Testing

System testing falls under the black box testing category. It is used to test a complete system without having to understand the underlying logic or system design. The main rationale for this testing procedure was to discover any possible inconsistencies among the software units, and ensure that the system responds to user interactions and data inputs.

The test cases obtained in system testing are described below:

7.3.1 System Test Cases

Test Case 1: Visualise real-time tweets on World map.

Properties:

Input Source: Geo-tagged tweets from Twitter's Streaming API

47

Expected Result: Visualisation of real-time tweets represented by ellipses on a World map.

Actual Result: As expected.

Test Case 2: Visualise classification of platforms from which Twitter users tweet in real-time.

Properties:

Input Source: Geo-tagged tweets from Twitter's Streaming API

Expected Result: Visualisation of classification of devices from which people tweet in real-time.

Actual Result: As expected.

Test Case 3: Visualise classification of languages of real-time tweets.

Properties:

Input Source: Geo-tagged tweets from Twitter's Streaming API

Expected Result: Visualisation of classification of languages of these real-time tweets.

Actual Result: As expected.

Test Case 4: Visualise top trends and their related tweets in a particular country by selecting it from the map.

Properties:

Input Source: trends from Twitter's trend API, and tweets from Twitter's Search API

Expected Result: Visualisation of top ten trends in the given country, relevant tweets from Search API for a given trend in this country, slideshow of retrieved images from the tweets (if available) and a tag cloud that summarizes content of these tweets.

Actual Result: As expected.

Test Case 5: Visualise real-time tweets in a particular area on the map, generate a tag cloud based on these tweets, and highlight them based on two keywords.

Properties:

Input Source: Geo-tagged tweets from Twitter's Streaming API, and two keywords to highlight tweets

Expected Result: Visualisation of real-time tweets, tag cloud generated from these tweets, highlighting the tweets based on two keywords from user input, and showing a tweet card by clicking on any of the tweets.

Actual Result: As expected.

7.3.2 Graphical User Interface Testing

Testing the Graphical User Interface (GUI) of this application was an important step of the testing process. In this step, all the GUI controls used in the application were verified to generate the anticipated actions. It helped in identifying some lags in the GUI, and these were fixed.

7.4 Usability Study

A usability study was carried out to evaluate the application and to understand the complexity of this application from the perspectives of first-time users. This study also helped to discover new ways of use for the application as participants of the study were allowed to express their thoughts on where and how the application can be used in new ways. Furthermore, it was used to determine the need for future improvements.

7.4.1 Participants

In order to receive unbiased feedback, participants with various levels of computer and Twitter savvy were recruited. This ensured that the application was targeting users from any background and level of knowledge in computing and using Twitter as social media platform. A total number of eight participants were recruited from the University of St Andrews based on personal connections and word-of-mouth. In return of participation in the study, no one was offered any type of compensation.

7.4.2 Procedure

The study concentrated on a 30-minute usability testing. It consisted of the following tasks: A summary of the usability study procedure for this application.

Table I: A summary of the usability study procedure.

Steps	Task Description	Duration (minutes)	Outcome
1	Watch an introductory video about this application.	5	User gets familiar with application
2	Download and install the application on user's computer.	5	User can run application on own computer
3	Use the application.	10	User interacts with application
4	Fill out an online questionnaire to collect feedback of the user on this application.	10	User-generated feedback about application

During the study, participants were asked to use the application on their own computers. This ensured that the application was able to run on any of the three operating systems i.e. Windows, Mac, Linux. It was important to give an introduction of the application to the participants in order to familiarize them with the various functions of the application. The users had to get familiar with the available features of the application so that they could use the application thoroughly and give more accurate feedback. This also helped users to interact with the application freely

without any restrictions that would be emphasized in traditional lab testing environments by assigning tasks to participants.

Finally, participants were asked to fill out online questionnaire that collected feedback about their interaction experience with the application. Participants were also asked to provide feedback on whether they liked or disliked the various features this application. This helped in collecting interesting insights into how the application's functions could be leveraged and what improvements could be added in future. Questionnaire is attached in Appendix A.

The following section provides a summary of the study results:

7.4.3   Study Results

After collecting responses from eight participants of this study, the following two types of data were gathered:

1) Quantitative: Participants were asked to rate usefulness of various features of the application, and provide responses on how often they would like to use the application. These types of responses generated a quantitative dataset that can be analysed and presented using charts or graphs. Bar charts have been used to present findings from the quantitative dataset.

2) Qualitative: Participants were asked to justify their responses with explicit textual answers. These responses led to a set of qualitative data that were used to understand what participants thought about each feature of the application, and their overall perception.

For each question, the responses from participants are summarized and presented in the following section:

1. Please rate your experience in learning to use the application.
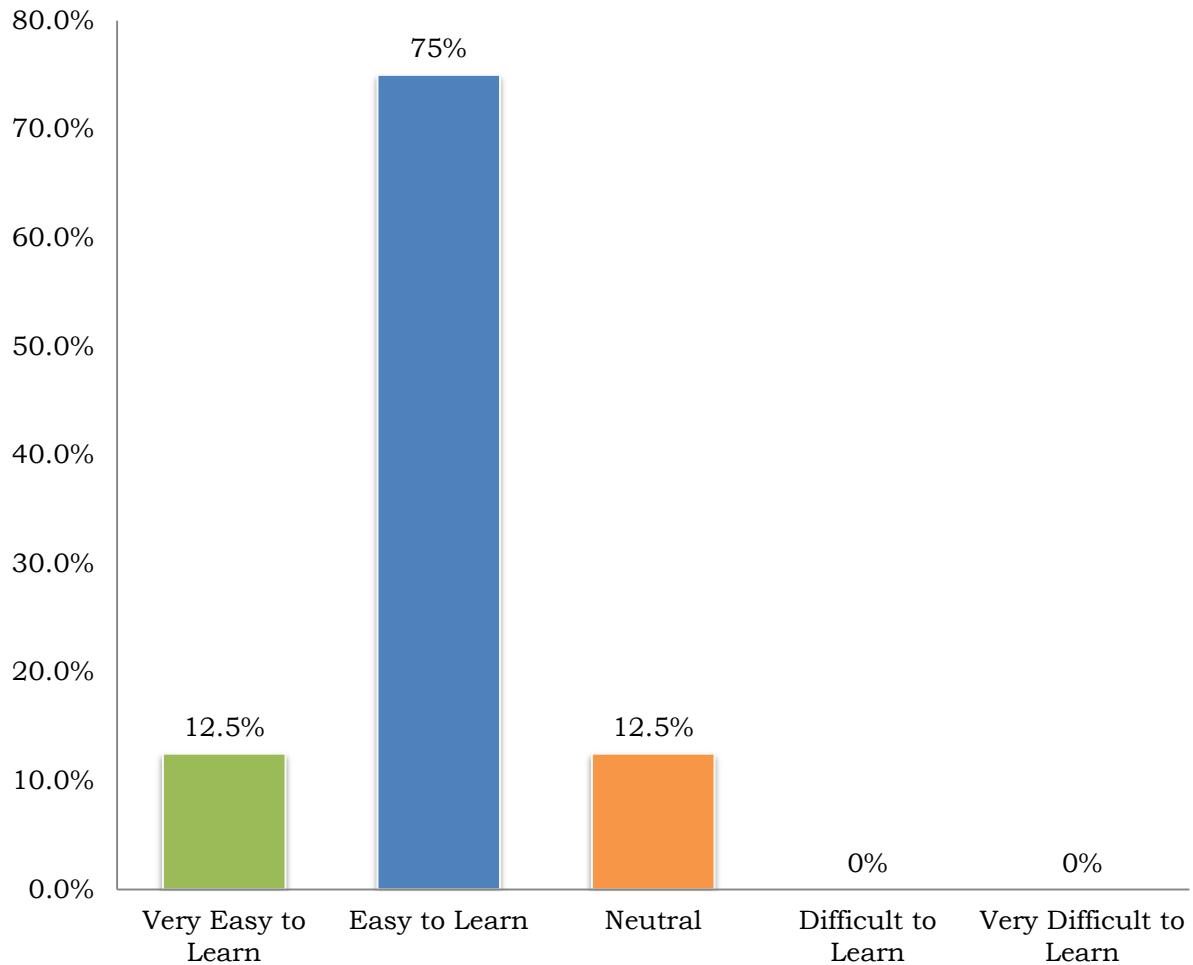
The purpose of this question was to evaluate this application's learnability factor. Participants were asked to watch a short demo on how to use the application, and then they had to use the application on their own. As a result, seventy-five per cent of the participants found this application very easy to learn while none of them felt that the application was difficult to learn how to use.

The main reasons given by the participants for easy learnability of the application were:

⊙ Very simple user interface
⊙ Provision of Help screens
⊙ Good navigation (toolbar)
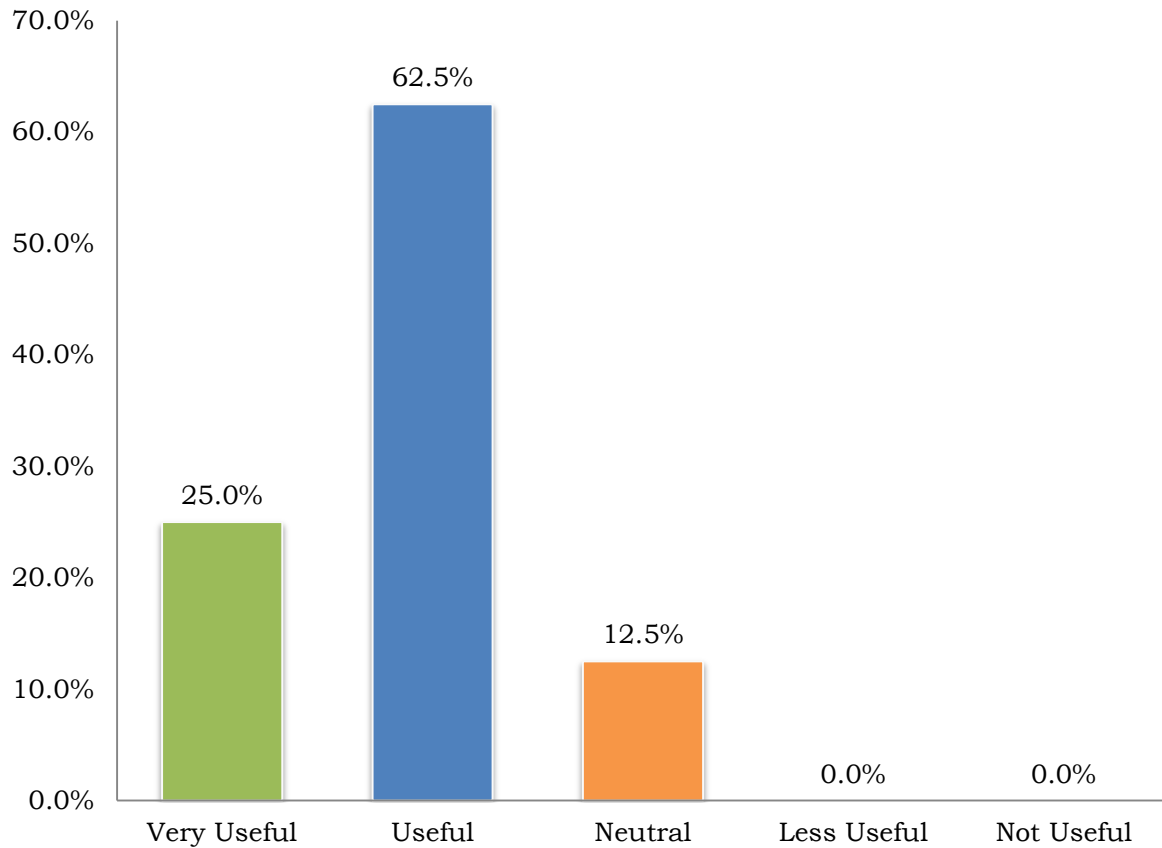
2. Please rate visualising real-time tweets on the map.

The purpose of this question was to rate the usefulness of visualising real-time tweets on a world map. Participants were also asked to provide explicit reasons on why they found the function useful. As it can be seen from the above bar chart, all but one participant found the application useful. The main reasons they mentioned in their responses were:

- ⊙ Reading real-time tweets easily
- ⊙ Toggling between two types of map
- ⊙ Finding real-time tweets distribution in different countries

3. How useful is visualising classification of platforms of real-time tweets on the map?
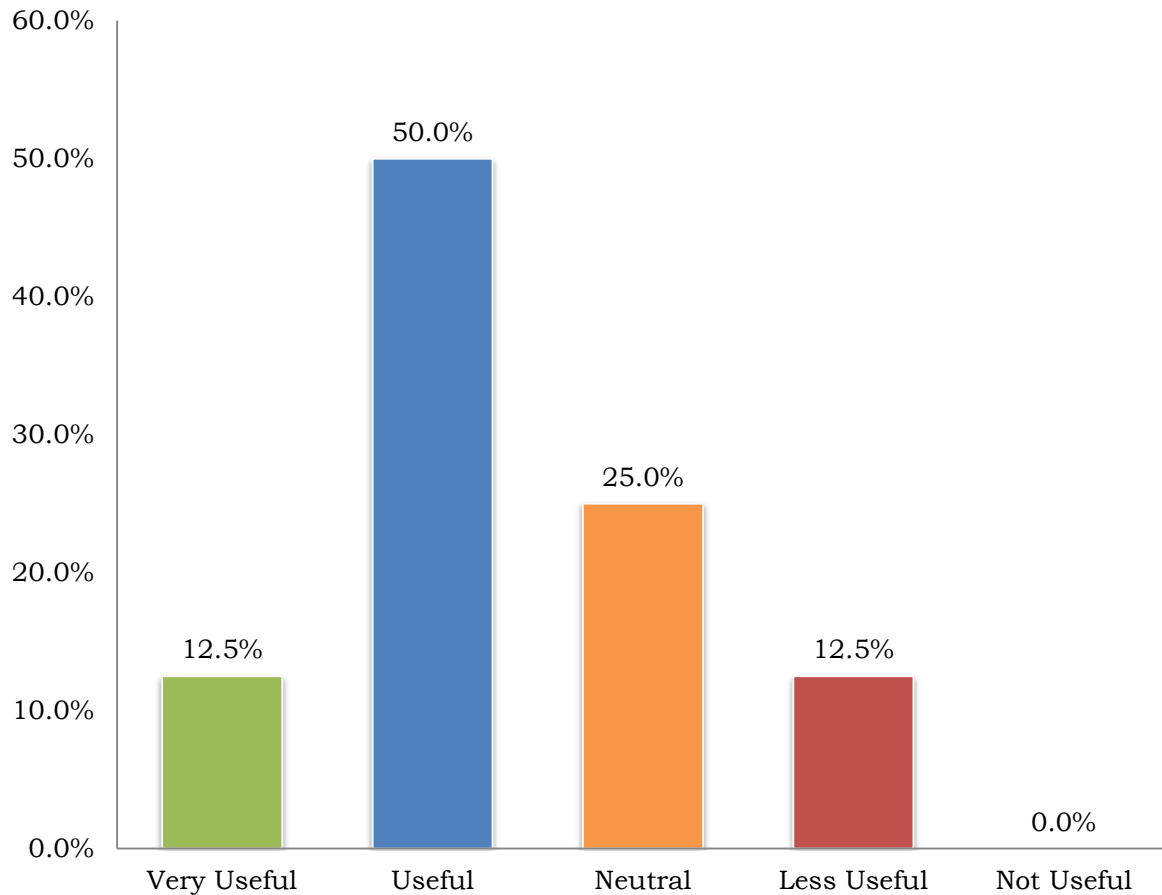
Figure 47: Responses of participants about rating the usefulness of visualization platforms of real-time tweets on a world map in this application.

The purpose of this question was to determine participants' interest in the visualisation of real-time tweets' platforms that this application provides. Half of the participants found this feature of the application useful. The main reasons given by participants for liking this feature were:

⊙ Determining which platform is dominant in which part of the world
⊙ Finding out about trending phones used by Twitter users

4. How useful is classifying languages of real-time tweets?
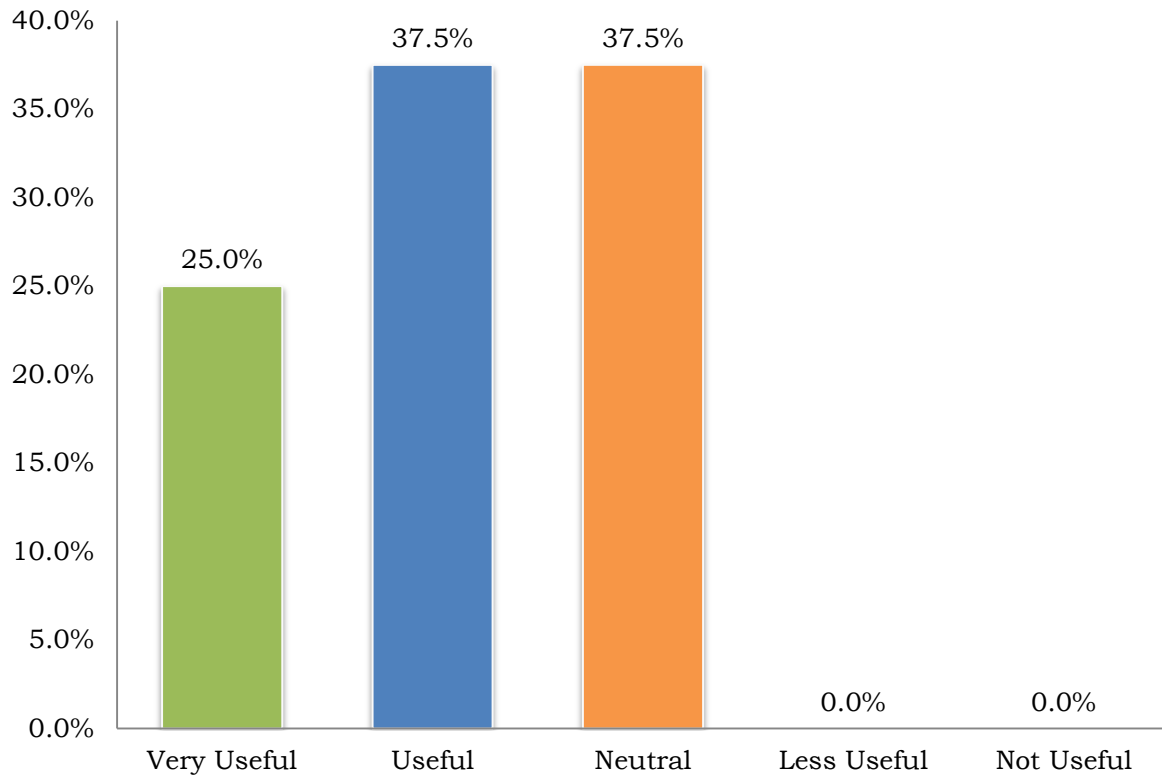
Figure 48: Responses of participants on rating the usefulness of visualising languages of real-time tweets in the application.

The purpose of this question was to evaluate the usefulness of visualising different languages of real-time tweets. Participants were asked to rate its usefulness and provide reasons for justifying their answers. As it can be seen from the above bar chart, more than half of the participants found this feature helpful while 35.5% of the participants were not interested in this visualisation. The main reasons that participants gave for liking this feature were:

- ⊙ Finding what languages people use on Twitter to communicate
- ⊙ Determining popularity of Twitter in different regions of the world

5. How useful is visualising popular hashtags of real-time tweets?
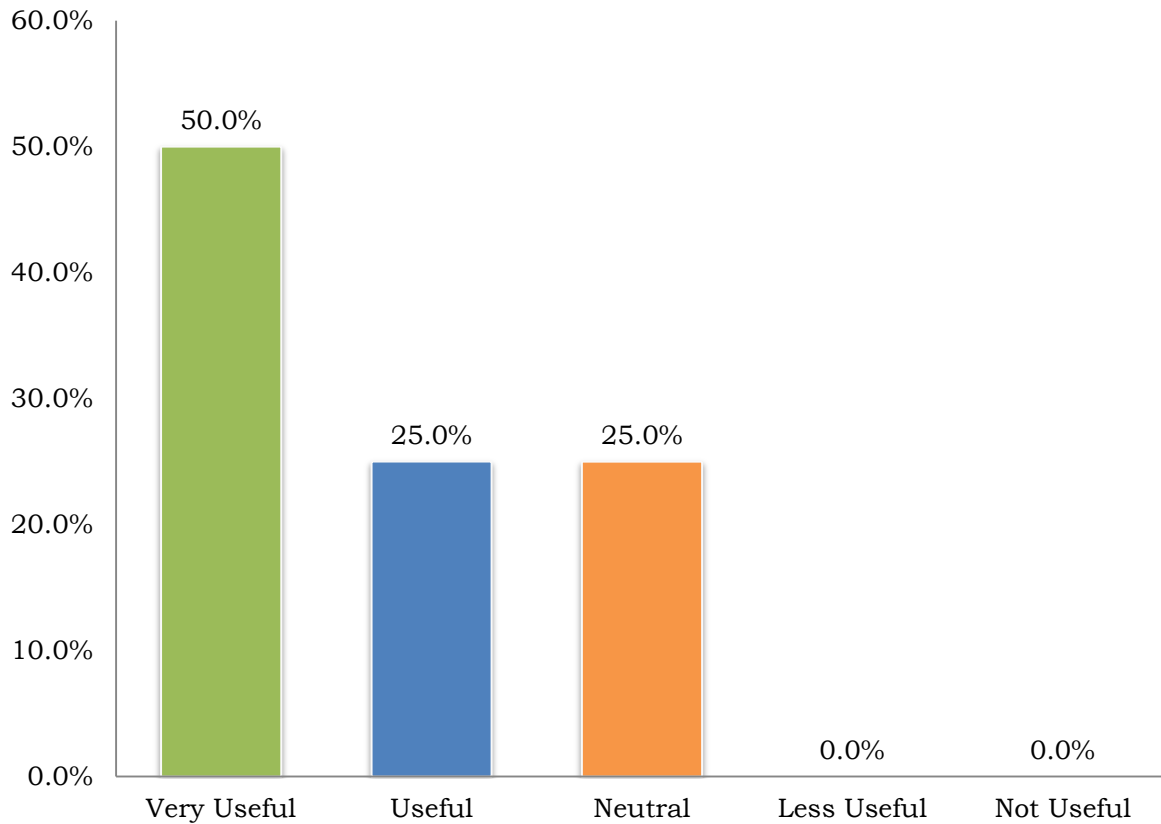
The purpose of this question was to allow participants to evaluate whether visualising real-time hashtags was useful in this application. This feature of the application was appealing to the users – 50% participants said that it was very useful while 25% participants found it useful. However, 25% of participants were not interested in this visualisation. The main reasons of participants who found this function useful were:

- ☉ It is trendy and interesting to see what is popular on Twitter
- ☉ Finding what people are talking about on Twitter.
- ☉ Determining real-time, hottest, and up-to-date topics

6. How useful is visualising top trends in a selected country?
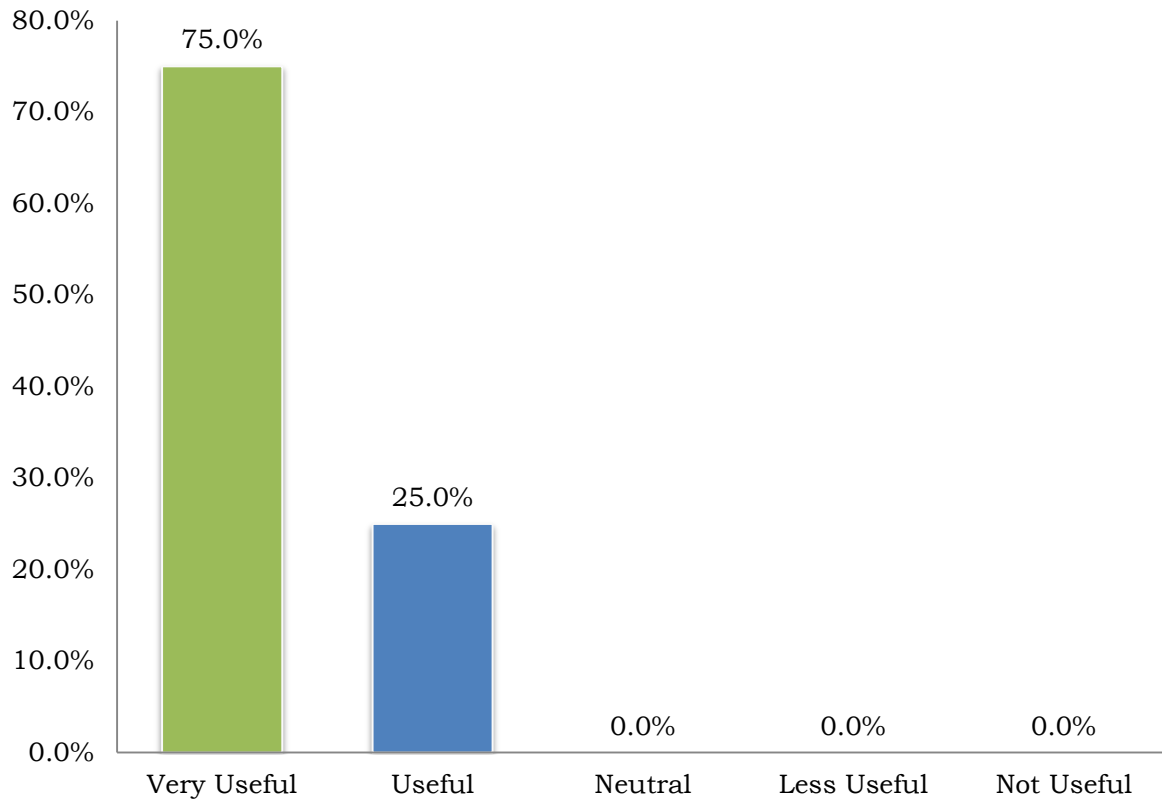
Figure 50: Responses of participants on rating the usefulness of visualizing top ten trends in a country.

This question of the study aimed at determining usefulness of visualising top trends in countries. All the participants found this feature of the application useful for the following reasons:

- ⊙ Very useful for understanding hottest topics in different countries
- ⊙ Viewing related images for a trend is meaningful and interesting
- ⊙ Interface is reasonable, clear and informative.
- ⊙ Images allow users to understand visually what a trend is about

7. How useful is visualising tweets related to a selected trend?
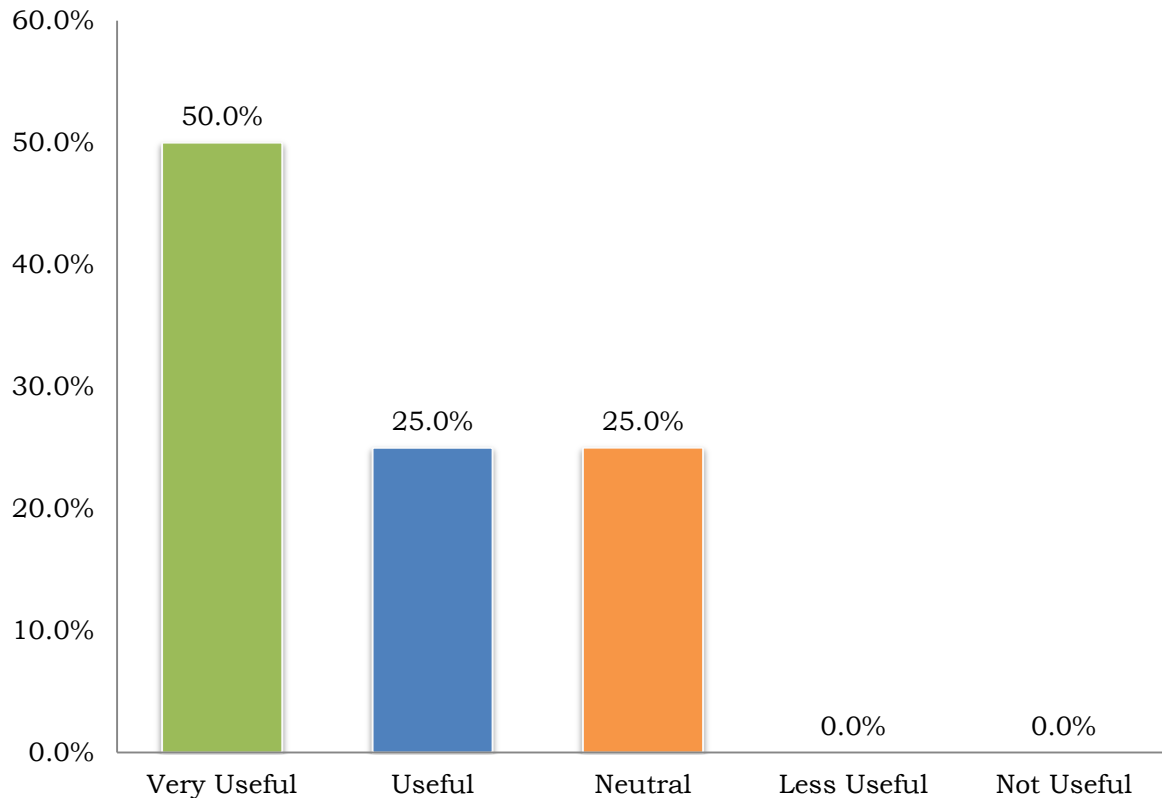
Figure 51: Responses of participants about rating the usefulness of visualising tweets about a particular trend in this application.

The application visualises tweets related to a particular trend. In this question, participants were asked to evaluate whether this feature of the application was helpful to them. As a result, 75% of participants found this visualisation helpful, while 25% were not interested in it. The main reasons that participants gave for justifying why they found it useful were:

- ⊙ Faster to find interesting tweets
- ⊙ Relationships of tweets improves understanding
- ⊙ Classification of tweets from most popular to least popular is helpful
- ⊙ Relationships of tweets (original and reply) is easy to understand

8. How useful is visualising tag cloud based on tweets?
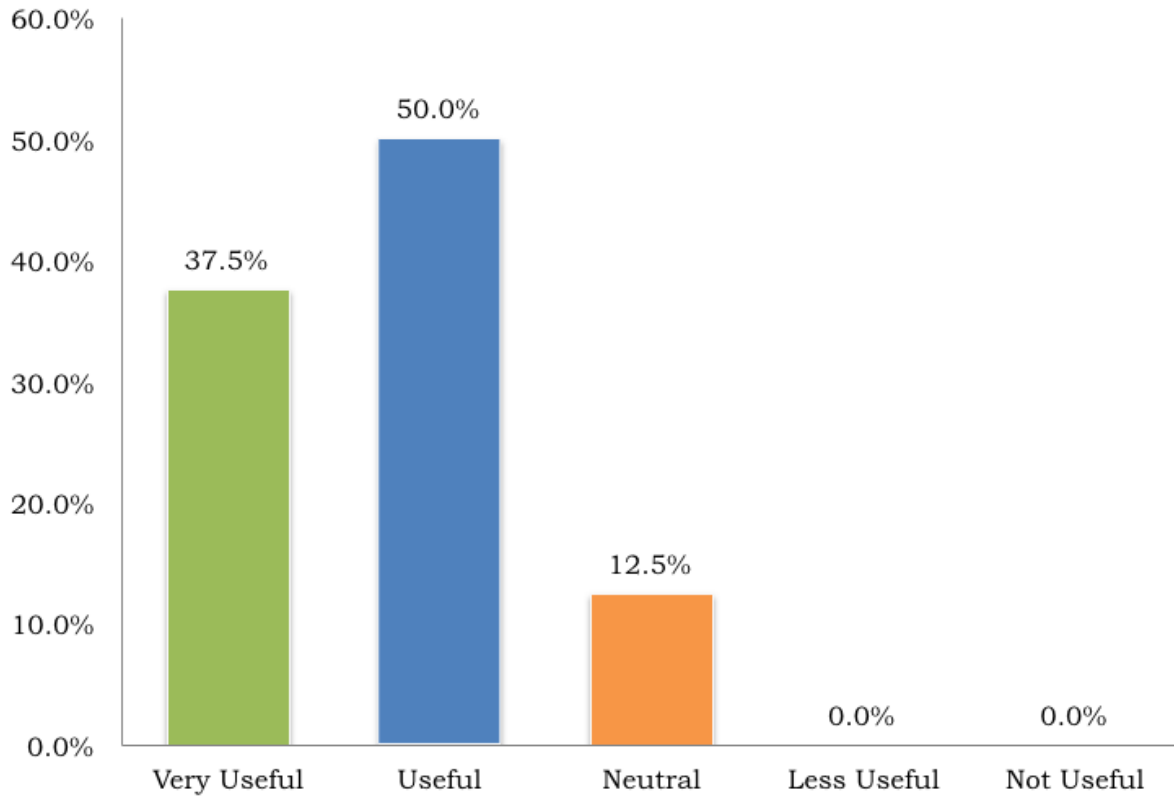
Figure 52: Responses of participants about rating usefulness of visualising tag cloud based on the tweets of a trend or from a selected geographical location on the map in our application.

The purpose of this question was to evaluate how useful it was for the users to view a tag cloud generated from tweets. This feature of the application was very useful to 37.5% of the participants, and useful to 50% of the participants. However, 12.5% of the participants were not interested in this visualisation. The main reasons for finding this feature useful were:

⊙ Convenient and interesting to read content of tweets
⊙ Can explain the trend, and to learn about the topic in a trend

9. How useful is viewing a tweet Card by clicking on a tweet?
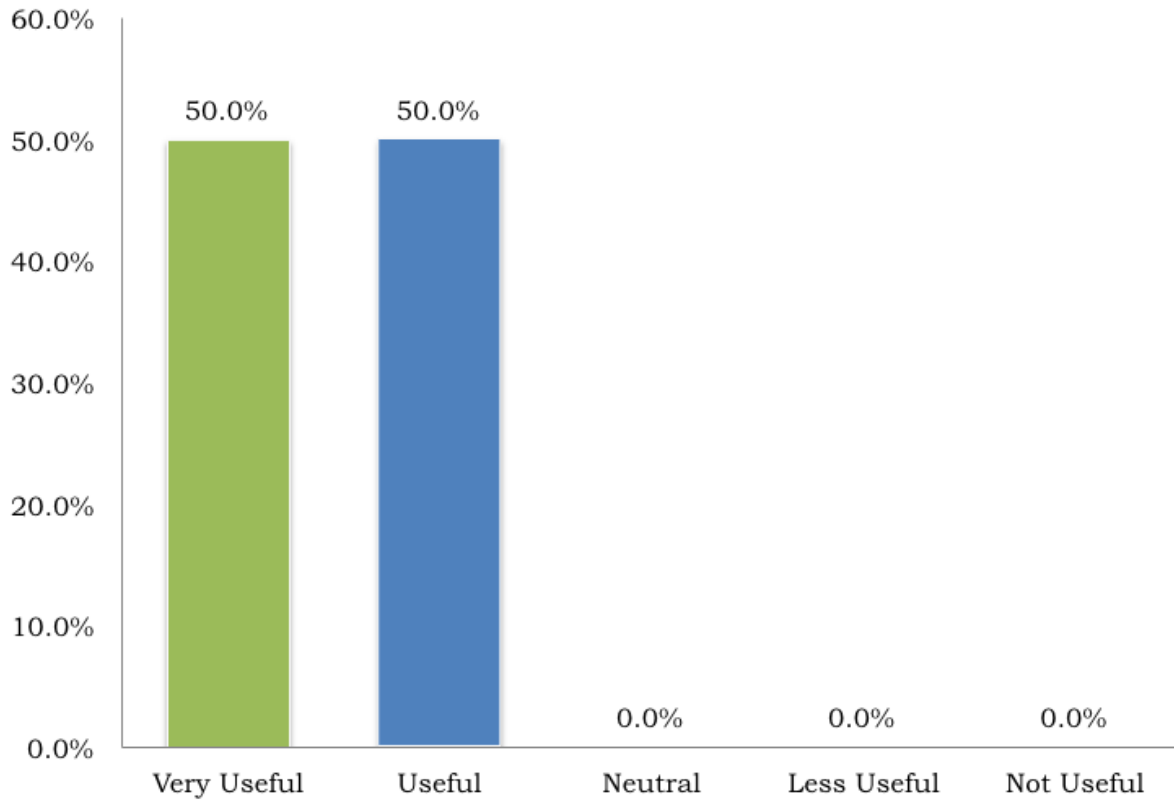
Figure 53: Responses of participants about rating the usefulness of visualizing a particular tweet via a tweet card in our application.

The application allows users to read more about a tweet by clicking on its representative ellipse. This information is presented in form of a tweet Card. Participants were asked to rate the usefulness of this function of the application. All of the participants found this feature useful. The main reasons were:

- ⊙ Accessing a tweet easily without having to open a browser or login to Twitter account
- ⊙ Saving time in reading more about a tweet
- ⊙ Gaining detailed insights into a trend

10. How useful is visualising selected tweets from a particular area on the map?
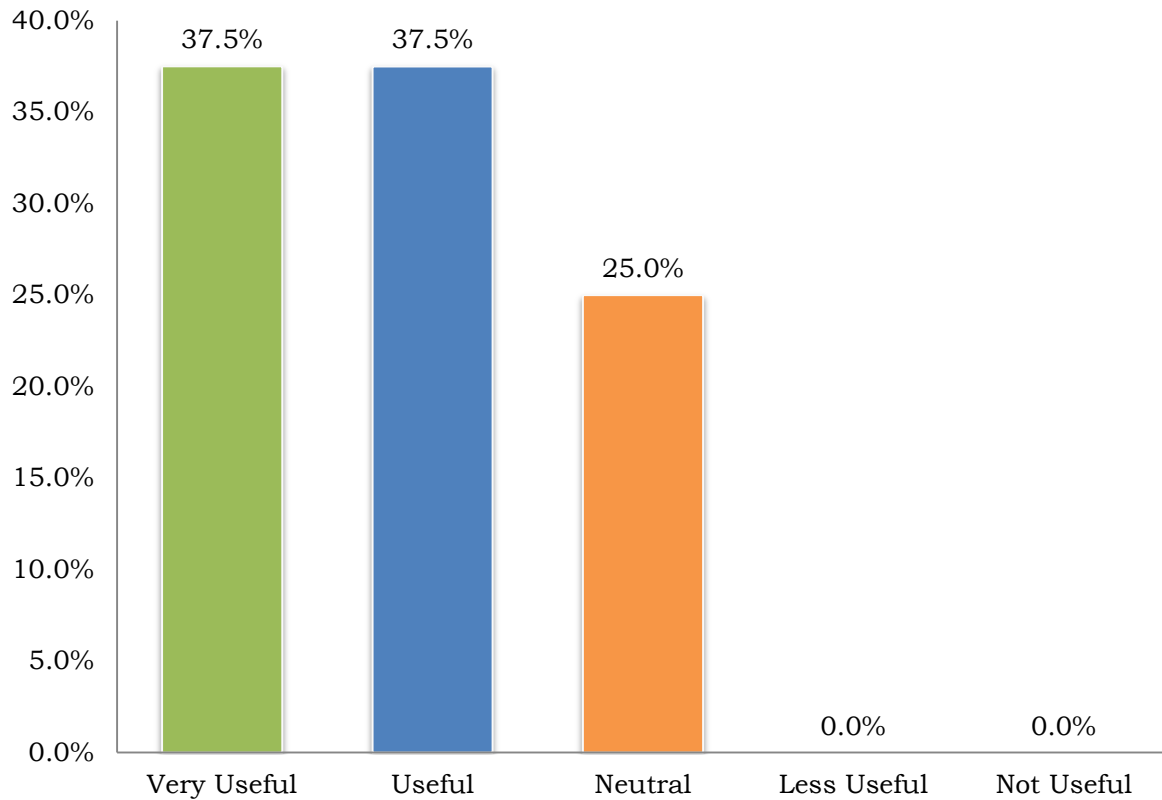
Figure 54: Responses of participants about rating the usefulness of visualising real-time tweets from a selected area on the map in this application.

The purpose of this question was to ask users to rate usefulness and give reasons for evaluating visualisation of real-time tweets from a selected area on the map. As it can be seen from the above bar chart, 75% of participants found this function useful, and the main reasons were:

⊙ Exploring tweets based on given keywords
⊙ Reading tweets from a particular area is flexible
⊙ Tag cloud from the selected tweets is useful
⊙ Good for travellers and businessmen

11. How useful is highlighting real-time tweets based on your own keywords?
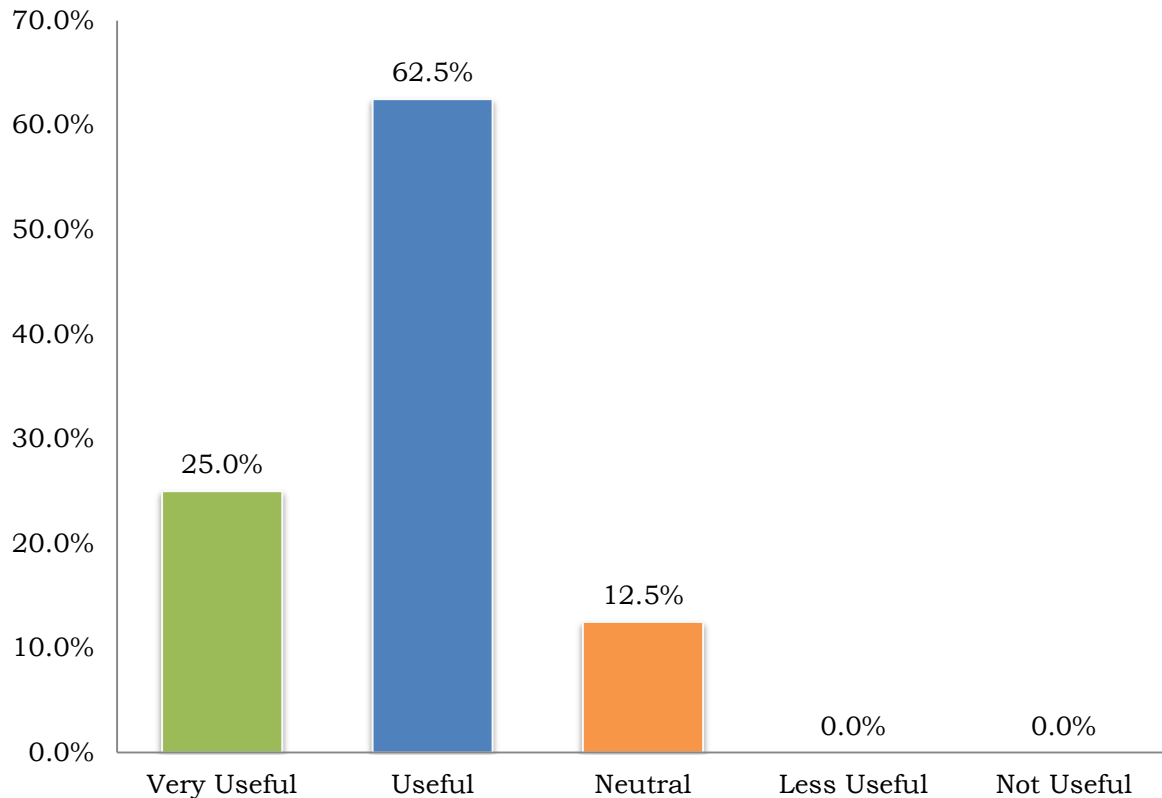
This question helped in understanding the usefulness of highlighting real-time tweets based on keywords of user's choice. This feature was useful to 62.5% of participants, and very useful to 25% of the participants. Only 12.5% participants were not interested in this visualisation. The following reasons were given for liking this feature of the application:

- ⊙ Reading tweets of user's choice quickly
- ⊙ Finding out about tweets that user is interested in

12. Please rate the application overall.
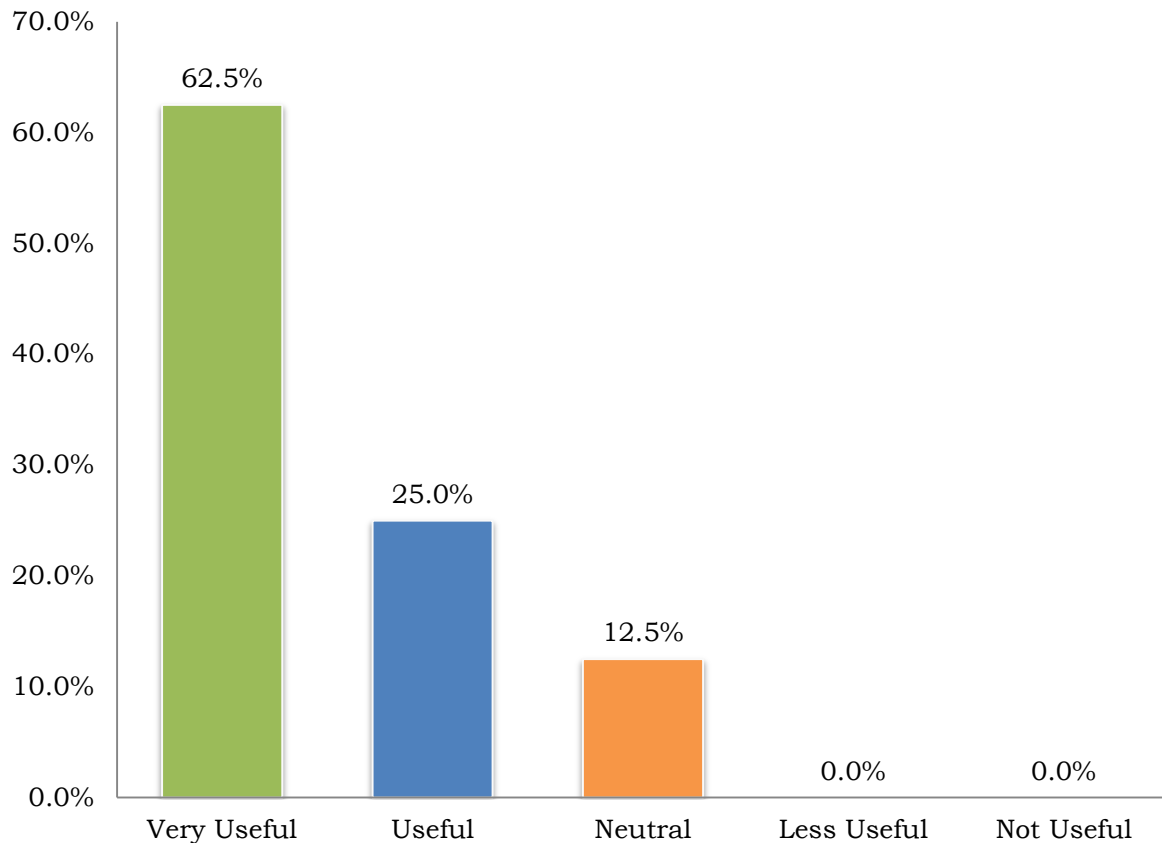
Figure 56: Responses of participants for rating the application overall in this usability study.

At this stage of the study, participants had a good understanding of the application. They evaluated all the features explicitly. In this question, the participants were asked to give an overall rating on how useful the application was to them. As it can be seen in the above chart, 62.5% found the application very useful, 25% found it useful and 12.5% were neutral towards the application. Participants gave the following reasons for finding the application useful:

- ⊙ Reading tweets in a smarter way
- ⊙ Easy to use
- ⊙ Important for understanding what is happening now in the world
- ⊙ Visualisation of thousands of tweets is simple and clear
- ⊙ Delivery of information is well-organized
- ⊙ Visualisation is informative
- ⊙ Handy for analysing tweets in real-time

13. How often will you be willing to use this application?
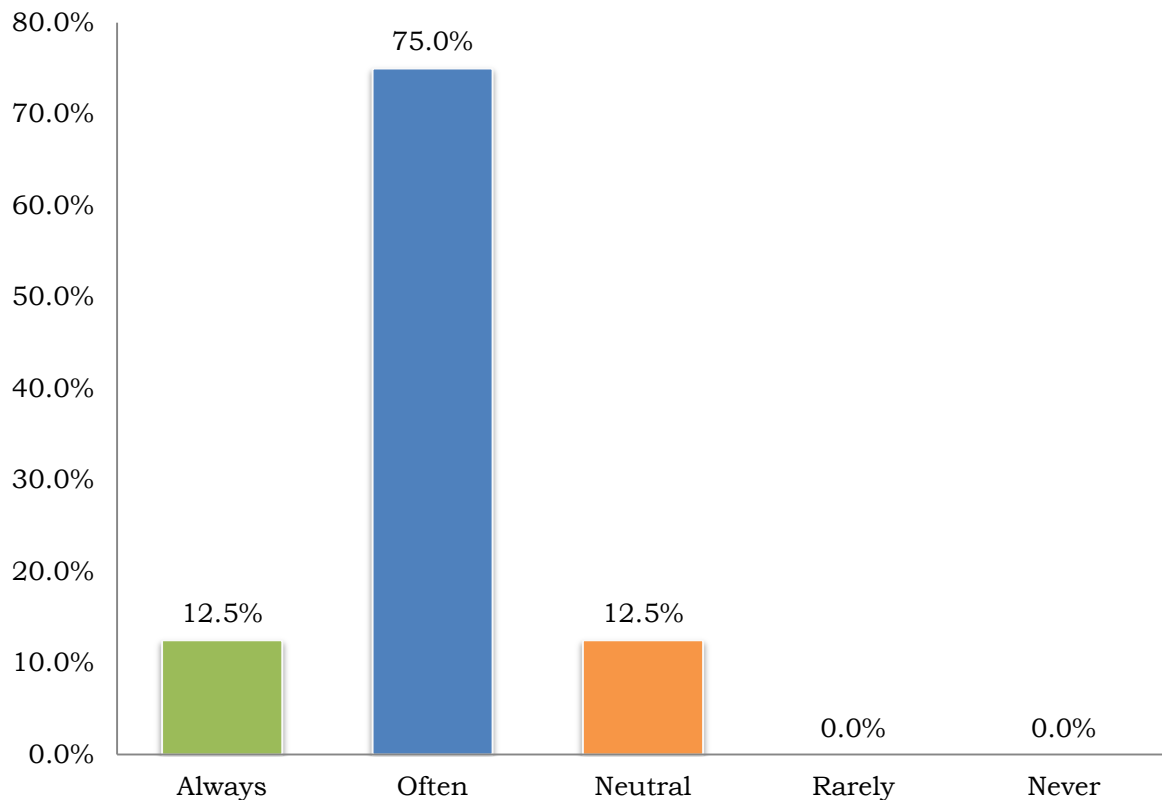
Figure 57: Responses of participants that show how often they wanted to use this application.

Now that the participants rated the overall usefulness of the application, participants were asked if they were willing to use this application, and to determine how often they would like to use it. As a result, 75% of the participants said that they were willing to use the application often, 12.5% responded that they were willing to use it always. However, 12.5% of the participants had a neutral feeling towards using this application.

14. For which purposes will you use this application?

At the end of the study, participants were asked to provide their opinions on how they wanted to use this application. They found the following main purposes of the application:

- Finding most popular trends in a country
- Seeing what languages people use to communicate on Twitter
- Finding out about what is happening now around the world
- Finding popular topics (hashtags) in real-time

Figure 58: Tag cloud generated from participants' responses that revealed their purposes of use of the application.

7.5 Qualitative Investigation

After conducting the usability study to identify how useful the application was for first time users, a qualitative investigation was carried out to gather user's using experience and attitude towards this application.

Two participants (hereafter referred to as participant A and participant B) took part in a thirty-minute semi structured interview. They were asked to use the application and express their attitudes and opinions based on their experience in this application simultaneously. The interviews were designed with varying open-ended questions to stimulate them to express their opinions in an open environment. The following part presents the results of the qualitative investigation.

For participant A, the application's interface was simple yet powerful. "There is a lot of information that one can retrieve from Twitter through this application. On the contrary, Twitter.com doesn't offer such a useful interface," he said. For example, finding latest trends by clicking a country on the map was very easy while finding this information on Twitter.com was time-consuming and rather difficult. Participant A referred to the whole experience of using this application as "new type of social media browsing." According to him, single tweets may not make any sense, but when they are combined in form of a tag cloud; it is easier to understand what these tweets are about.

After seeing classification of platforms, participant A was surprised to see the popularity of platforms (e.g. Android, Windows, Blackberry, etc.) in different countries. For example, by comparing 572 tweets posted by Windows device and 15,643 tweets posted by iPhone, he said that this could give a meaningful message to Microsoft that Windows-based devices were failing to gain popularity in the market. He explained that such data can be obtained from this application free of charge instead of buying it from marketing or research firms.

His final comments were that this application could further improve the usefulness of Twitter. "I never thought I could get such useful information from Twitter before using this app," he added.

Participant B believed that this application could be useful for journalists and reporters who would like to see top trends based on Twitter data. She added that this visualization makes it easier and less time-consuming to understand what is happening in Twitter. Participant B also found it useful to see relationships between original and reply tweets. She said that she did not like the idea of showing threaded replies in tweets on Twitter, but this application made it more interesting for her to see how people replied to the tweets of one another.

After seeing thousands of tweets visualized on the world map in just a few minutes, the participants were amazed that the world is getting crazy about social media. They expressed that it helped them understand visually how people tweeted so extensively in different parts of the world. The participants also found it useful to highlight real-time tweets based on the keywords of their choice. It gave them control to see if people talked about these keywords in real-time. By highlighting tweets, marketing practitioners can track opinions of real people in real-time about their products or services especially when they are recently released.

Both of the participants mentioned that the application gave them a feeling of browsing through a storybook. They found it fascinating to see that navigating information from general (tweets on a world map) to specific (reading each tweet) was very well structured and organized. They found a connection between every stage of the visualization that helped them dig deeper into the information presented from Twitter.

In summary, both of the participants provided positive feedback and expressions towards this application. They believed that this application informative to provide insights into Twitter data; good navigation improved the usability of this application; and this application presented a novel visualization of Twitter data.

This chapter discussed the various types of testing and evaluation methods such as unit testing, integration testing, system testing, usability study, and qualitative investigation. The following chapter will present the conclusions to this project, and will include limitations and future development of this project.

8   Conclusions

The aim of this project was to build a tool that visualises the publicly available data of Twitter via the Twitter Streaming API. This project succeeded in achieving the original objectives, and in implementing all of the primary and secondary requirements. The outcome of this project is this dissertation report and a working product named "Curated Real-Time Visualisation of Twitter Data."

This dissertation summarized all of the work that was carried out towards completing this project. First, the motivation behind this project was introduced, fundamentals of data visualisation were explained, literature was reviewed in this domain, and related work was discussed in the field of visualising Twitter data. Then, the design and implementation processes of this project were presented. Finally, to verify that this application is usable, a user study was conducted. The findings of this usability study were discussed in the previous chapter.

The main scope of this project was to provide a novel interactive visualisation of the real-time data from Twitter. It has been achieved by developing an interactive desktop application that is available for Mac, Windows and Linux operating systems. It visualises real-time tweets and current trends, it reveals classifications of platforms and languages, and it enables users to learn more about a trend in a particular country by exploring its relevant tweets, their relationships, and classification of popularity. By visualising tweets and trends in real-time, the application enables users to gain insights visually about what is happening across the world.

8.1   Limitations

Although the primary and secondary requirements were implemented successfully, there are two major limitations to this work. These limitations were caused by the nature of the Twitter Streaming APIs and the way map projections are implemented in Processing.

First, there are rate limits in accessing data via Twitter APIs. There are different limits for calling the various APIs of Twitter that have been used in this application. For example, one can call any of the resources from REST API v1.1 of Twitter only fifteen times in fifteen minutes. As a result, it imposes a limitation on this application, which is restricting the use of this application when several users use it at the same time.

Second, the application is designed to provide a window of fixed size that is 1000 pixels of width and 652 pixels of height. As tweets are visualised geographically, custom maps of fixed sizes were used to plot the data using its latitude and longitude values. Maps of appropriate sizes had to be deployed so that it could leverage screen real estate of both normal and wide screen displays. However, this may not be a serious limitation because it doesn't cause any problems in usability of the application.

8.2   Future Development

Information visualisation is a rather interesting and challenging area of research. There is a lot of scope in developing creative tools that visualise 'big data' from social networking sites.

Likewise, this application can be further developed to add more features and visualisations of Twitter data. Future development can ensure implementing the tertiary requirement of the application that was not implemented due to time constraints of the project. As a result, the application can be further developed to extend its presence on the World Wide Web. At the moment, programs in Processing that use external libraries like Twitter4J and JLangDetect cannot be exported as a web application. There is scope in using some JavaScript libraries that can substitute the external libraries in this application, which are not supported on the web.

9   References

[1]   H. Tsukayama. (2013, 09/08/2013). *Twitter turns 7: Users send over 400 million tweets per day*. Available: http://articles.washingtonpost.com/2013-03-21/business/37889387_1_tweets-jack-dorsey-twitter

[2]   B. Shneiderman and C. Plaisant, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*: Addison-Wesley/Pearson, 2010.

[3]   K. Hornbæk, B. B. Bederson, and C. Plaisant, "Navigation patterns and usability of zoomable user interfaces with and without an overview," *ACM Transactions on Computer-Human Interaction (TOCHI),* vol. 9, pp. 362-389, 2002.

[4]   A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," *Business horizons,* vol. 53, pp. 59-68, 2010.

[5]   (14/06/2013). *Web 2.0*. Available: http://en.wikipedia.org/wiki/Web_2.0

[6]   (14/06/2013).   *User   Generated   Content.*   Available: http://en.wikipedia.org/wiki/User-generated_content

[7]   (14/06/2013).   *Six   Degrees.*   Available: http://en.wikipedia.org/wiki/SixDegrees.com

[8]   J. H. Kietzmann, K. Hermkens, I. P. McCarthy, and B. S. Silvestre, "Social media? Get serious! Understanding the functional building blocks of social media," *Business horizons,* vol. 54, pp. 241-251, 2011.

[9]   E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne, "Finding high-quality content in social media," presented at the Proceedings of the 2008 International Conference on Web Search and Data Mining, Palo Alto, California, USA, 2008.

[10]   S. Asur and B. A. Huberman, "Predicting the Future with Social Media," in *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on,* 2010, pp. 492-499.

[11]   J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," *ACM Trans. Web,* vol. 1, p. 5, 2007.

[12]   B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitter power: Tweets as electronic word of mouth," *J. Am. Soc. Inf. Sci. Technol.,* vol. 60, pp. 2169-2188, 2009.

[13]   S. Petrović, M. Osborne, and V. Lavrenko, "Streaming first story detection with application to twitter," in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics,* 2010, pp. 181-189.

[14]   H. Becker, M. Naaman, and L. Gravano, "Beyond trending topics: Real-world event identification on twitter," in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM'11),* 2011.

[15]   J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling, "Twitterstand: news in tweets," in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems,* 2009, pp. 42-51.

[16]   T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors," in *Proceedings of the 19th international conference on World wide web,* 2010, pp. 851-860.

[17]   R. M. Rohrer and E. Swing, "Web-based information visualization," *Computer Graphics and Applications, IEEE,* vol. 17, pp. 52-59, 1997.

[18] M. A. Smith, B. Shneiderman, N. Milic-Frayling, E. M. Rodrigues, V. Barash, C. Dunne, T. Capone, A. Perer, and E. Gleave, "Analyzing (social media) networks with NodeXL," presented at the Proceedings of the fourth international conference on Communities and technologies, University Park, PA, USA, 2009.

[19] A. Perer and B. Shneiderman, "Balancing systematic and flexible exploration of social networks," *Visualization and Computer Graphics, IEEE Transactions on,* vol. 12, pp. 693-700, 2006.

[20] J. Heer and D. Boyd, "Vizster: Visualizing online social networks," in *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, 2005, pp. 32-39.

[21] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry, "Task taxonomy for graph visualization," in *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, 2006, pp. 1-5.

[22] N. Henry, J. Fekete, and M. J. McGuffin, "NodeTrix: a Hybrid Visualization of Social Networks," *Visualization and Computer Graphics, IEEE Transactions on,* vol. 13, pp. 1302-1309, 2007.

[23] P. Gloor, R. Laubacher, Y. Zhao, and S. Dynes, "Temporal visualization and analysis of social networks," in *NAACSOS Conference, June*, 2004, pp. 27-29.

[24] B. Schnitzer and R. Williams, "Visualization and Harvesting of Hurricane Sandy Geo-Tagged Tweets."

[25] K. McKelvey, A. Rudnick, M. D. Conover, and F. Menczer, "Visualizing Communication on Social Media: Making Big Data Accessible," *arXiv preprint arXiv:1202.1367,* 2012.

[26] Truthy. (2012, 12/06/2013). *Popular memes in Twitter's past 9 months.* Available: http://truthy.indiana.edu/themedetail?id=5&sort_type=16&filter_type=0

[27] M. Michelson and S. A. Macskassy, "Discovering users' topics of interest on twitter: a first look," presented at the Proceedings of the fourth workshop on Analytics for noisy unstructured text data, Toronto, ON, Canada, 2010.

[28] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in *Proceedings of the third ACM international conference on Web search and data mining*, 2010, pp. 261-270.

[29] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller, "Twitinfo: aggregating and visualizing microblogs for event exploration," presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 2011.

[30] M. Mathioudakis and N. Koudas, "TwitterMonitor: trend detection over the twitter stream," presented at the Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, Indianapolis, Indiana, USA, 2010.

[31] F. Fang, N. Pervin, A. Datta, K. Dutta, and D. VanderMeer, "Detecting Twitter Trends in Real-Time," in *Proceedings of the 21st Workshop on Information Technologies and System (WITS)*, 2011.

[32] M. Cataldi, L. Di Caro, and C. Schifanella, "Emerging topic detection on Twitter based on temporal and social terms evaluation," in *Proceedings of the Tenth International Workshop on Multimedia Data Mining*, 2010, p. 4.

[33] J. C. Enkh-Amgalan Baatarjav and T. Meesumrarn, "A summary and Critique of "TwitterStand: News in Tweets"."

[34]   Twitter-FAQ. (n.d., 12/06/2013). *FAQ about Trends on Twitter*. Available: https://support.twitter.com/articles/101125-faq-about-trends-on-twitter

[35]   C. Reas and B. Fry, *Getting Started with Processing*: Make, 2010.

[36]   B. Bruegge and A. H. Dutoit, *Object-Oriented Software Engineering: Using Uml, Patterns And Java, 2/E*: Pearson Education, 2004.

[37]   I. Sommerville, *Software Engineering*: Pearson/Addison-Wesley, 2011.

[38]   S. Lauesen, *User Interface Design: A Software Engineering Perspective*: Pearson Education Limited, 2005.

[39]   J. Johnson, *GUI Bloopers 2.0: Common User Interface Design Don'ts and Dos*: Elsevier Science, 2007.

[40]   B. Schneiderman and C. Plaisant, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*: Pearson/Addison Wesley, 2005.

[41]   A. Kirk, *Data Visualization: A Successful Design Process*: Packt Publishing, Limited, 2012.

[42]   J. Steele and N. Iliinsky, *Designing Data Visualizations: Representing Informational Relationships*: O'Reilly Media, 2011.

10 Appendices

10.1 Appendix 1: Questionnaire of Usability Testing:

(1) Please rate your experience in learning to use the application.

1. Very Easy to Learn   2. Easy to Learn   3. Neutral   4. Difficult to Learn   5. Very Difficult to Learn

- Please explain why: _____

(2) Please rate visualising real-time tweets on the map?

1. Very Useful   2. Useful   3. Neutral   4. Less Useful   5. Not Useful

- Please explain why: _____

(3) How useful is visualising classification of platforms of real-time tweets on the map?

1. Very Useful   2. Useful   3. Neutral   4. Less Useful   5. Not Useful

- Please explain why: _____

(4) How useful is classifying languages of real-time tweets?

1. Very Useful   2. Useful   3. Neutral   4. Less Useful   5. Not Useful

- Please explain why: _____

(5) How useful is visualising popular hashtags of real-time tweets?

1. Very Useful   2. Useful   3. Neutral   4. Less Useful   5. Not Useful

- Please explain why: _____

(6) How useful is visualising top trends in a selected country?

1. Very Useful   2. Useful   3. Neutral   4. Less Useful   5. Not Useful

- Please explain why: _____

(7) How useful is visualising tweets related to a selected trend?

1. Very Useful   2. Useful   3. Neutral   4. Less Useful   5. Not Useful

- Please explain why: _____

(8) How useful is visualising tag cloud based on tweets?

1. Very Useful   2. Useful   3. Neutral   4. Less Useful   5. Not Useful

- Please explain why: _____

(9) How useful is viewing a tweet Card by clicking on a tweet?

1. Very Useful   2. Useful   3. Neutral   4. Less Useful   5. Not Useful

- Please explain why: _____

(10) How useful is visualising selected tweets from a particular area on the map?

1. Very Useful   2. Useful   3. Neutral   4. Less Useful   5. Not Useful

- Please explain why: _____

(11) How useful is highlighting real-time tweets based on your own keywords?

1. Very Useful   2. Useful   3. Neutral   4. Less Useful   5. Not Useful

- Please explain why: _____

 (12) Please rate the application overall?

1. Very Useful   2. Useful   3. Neutral   4. Less Useful   5. Not Useful

- Please explain why: _____

(13) How often will you be willing to use this application?

1. Always   2. Often   3. Neutral   4. Rarely   5. Never

(14) For which purposes will you use this application?

- Please explain _____

10.2 Appendix 2: Responses of Usability Testing

Participant 1:
Question: 1. Please rate your experience in learning to use the application.

Answer: 2. Easy to Learn

Please explain why:

Answer: It is easy to learn through the video tutorial. The manual is detailed and the comments are very clear.

Question: 2. Please rate visualising real-time tweets on the map.

Answer: 2. Useful

Please explain why:

Answer: If it is possible to zoom in and out the map, to see the city names, it would be better!

Question: 3. How useful is visualising classification of platforms of real-time tweets on the map?

Answer: 3. Neutral

Please explain why:

Answer: I think I didn't concern which platforms. But it gives me an idea of the trendy of using what kinks of phones. This is interesting.

Question: 4. How useful is classifying languages of real-time tweets?

Answer: 2. Useful

Please explain why:

Answer: From the figure of languages of real-time tweets, I can find the tweets information depends on my own country.I can also know which is the most popular tweet country!

Question: 5. How useful is visualising popular hashtags of real-time tweets?

Answer: 2. Useful

Please explain why:

Answer: I can have the idea of what is the hottest topic in the real time. It's cool and I think the information I gathered won't be out of date!

Question: 6. How useful is visualising top trends in a selected country?

Answer: 1. Very Useful

Please explain why:

Answer: If I am a businessman, and I could utilise the trends of tweets based on the country to expand my business.

Question: 7. How useful is visualising tweets related to a selected trend?

Answer: 1. Very Useful

Please explain why:

Answer: It integrates some slash tags and it may save time for me to viewing the trends.

Question: 8. How useful is visualising tag cloud based on tweets?

Answer: 1. Very Useful

Please explain why:

Answer: Based on the visualising tag, I know what is the hottest word just based on the size of font. This is outstanding and striking

Question: 9. How useful is viewing a tweet Card by clicking on a tweet?

Answer: 1. Very Useful

Please explain why:

Answer: With this, I don't need to log into my tweet to see the tweets. The tweet card shows the most essential information. Additional, If I am interested in replying and commenting, I can just simple click the tweet card.

Question: 10. How useful is visualising selected tweets from a particular area on the map?

Answer: 2. Useful

Please explain why:

Answer: I can find what is the hottest topic in a certain country. It is good for travellers and businessmen. However, it can delete some previous records after a period, otherwise, the screen will be too colourful.

Question: 11. How useful is highlighting real-time tweets based on your own keywords?

Answer: 2. Useful

Please explain why:

Answer: The only weakness is that it can only add unto 2 keywords, and I think it is also better to display the numbers of tweets which is highlighted.

Question: 12. Please rate the application overall?

Answer: 1. Very Useful

Please explain why:

> Answer: The application is interesting overall. I can search the hottest topic based on the keywords, view the tweets as well. Maybe I can find some friends with same interests in my country even in other countries!!! It is so nice.

Question: 13. How often will you be willing to use this application?

> Answer: 2. Often

Question: 14. For which purposes will you use this application? Please give reasons.

> Answer: I think the target audience for this application is businessman and educators. E.g. a business man wanna open a new shop in a certain country, so he can investigate what is the most interesting in public. for educators, because the tweet user are almost teenage and students, they can obtain an idea of what are they thinking and take the corresponding measure to educate. however, for other users, they may just play it for fun.

Participant 2:
Question: 1. Please rate your experience in learning to use the application.

Answer: 2. Easy to Learn

Please explain why:

Answer: The video tutorial is very helpful. The functions of the application are intuitive.

Question: 2. Please rate visualising real-time tweets on the map.

Answer: 2. Useful

Please explain why:

Answer: For understanding the trend of different countries.

Question: 3. How useful is visualising classification of platforms of real-time tweets on the map?

Answer: 4. Less Useful

Please explain why:

Answer: Personally I don't care what devices others are using.

Question: 4. How useful is classifying languages of real-time tweets?

Answer: 3. Neutral

Please explain why:

Answer: English would be dominating language without doubt. People in my country don't use Twitter. And the use of other languages makes no difference to me.

Question: 5. How useful is visualising popular hashtags of real-time tweets?

Answer: 1. Very Useful

Please explain why:

Answer: Let me know what people are talking about, what's new.

Question: 6. How useful is visualising top trends in a selected country?

Answer: 1. Very Useful

Please explain why:

Answer: For investigating what's happening in a country.

Question: 7. How useful is visualising tweets related to a selected trend?

Answer: 2. Useful

Please explain why:

Answer: An interesting function.

Question: 8. How useful is visualising tag cloud based on tweets?

Answer: 2. Useful

Please explain why:

Answer: Check what is the hottest word for now. I can directly see the text.

Question: 9. How useful is viewing a tweet Card by clicking on a tweet?

Answer: 1. Very Useful

Please explain why:

Answer: To get the full story rather than just a little piece of it.

Question: 10. How useful is visualising selected tweets from a particular area on the map?

Answer: 3. Neutral

Please explain why:

Question: 11. How useful is highlighting real-time tweets based on your own keywords?

Answer: 1. Very Useful

Please explain why:

Answer: That's what I really care, instead of seeing others' monologue.

Question: 12. Please rate the application overall?

Answer: 1. Very Useful

Please explain why:

Answer: Very handy tool for analyzing tweeds. You made NSA's work much easier :)

Question: 13. How often will you be willing to use this application?

Answer: 2. Often

Question: 26. 14. For which purposes will you use this application? Please give reasons.

Answer: Personal interest +++ trends research ++ Commercial +

Participant 3:
Question: 1. Please rate your experience in learning to use the application.

>   Answer: 2. Easy to Learn

Please explain why:

>   Answer: this application has a good accessibility. It contains a "help" button, allowing me to quickly know the function and meaning of every button. By the way, I like the design of help interface, it is very clear and cute. Additionally, this application also present a good navigation via its tool bar. simple but clear design make it easy to operate for users.

Question: 2. Please rate visualising real-time tweets on the map.

>   Answer: 1. Very Useful

Please explain why:

>   Answer: It is very useful. First, application developer provides choices about types of map (plain map & color map). Plain map enables users to develop a visual impression to distribution of tweets in the whole map. And color map enables to check distribution of tweets in each country. Second, visualising real-time tweets on the map help users to develop the awareness: - how tweets are distributed in the whole of map. -which continents are people tweeting the most from. - which countries are people tweeting the most from. - which cities in one country are people tweeting the most from.

Question: 3. How useful is visualising classification of platforms of real-time tweets on the map?

>   Answer: 2. Useful

Please explain why:

>   Answer: It is also useful. Visualising classification of platforms of real-time tweets in the map enables users to know which platform is the most popular used to tweet by people. and I think that the result is helpful for Twitter company to check using situation of its production in different platforms.

Question: 4. How useful is classifying languages of real-time tweets?

>   Answer: 2. Useful

Please explain why:

>   Answer: this function is basic (less informative) but necessary, allowing users to visualise distribution of tweets from the language's perspective.

Question: 5. How useful is visualising popular hashtags of real-time tweets?

>   Answer: 3. Neutral

Please explain why:

Answer: visualising popular hashtags of real-time tweets is less useful than other functions. the advantage of this function is to visually present popular hashtags of real-time tweets. it is easy for users to capture which hashtag is the most popular and which hashtags are less popular. However, most of these hashtags are difficult to be understood. it is less meaningful to provide extra information to users.

Question: 6. How useful is visualising top trends in a selected country?

Answer: 1. Very Useful

Please explain why:

Answer: visualising top trends in a selected country is very useful. First, the layout of entire interface looks like reasonable and clear. Second, this interface delivers progressive information to users. 1. users can know the top ten trends in the selected country. 2. users can view 100 tweets relevant with anyone of top 10 trends. 3. users can view key words appeared in these 100 tweets via tags cloud. 4. user can visually understand situations of shared image, reply tweet and popular tweet about one topic based on colors. (this design is very good, it visually display the relationship of these tweets. 4. of course, users also can check detailed content of every tweet. Thirdly, this interface is informative.

Question: 7. How useful is visualising tweets related to a selected trend?

Answer: 1. Very Useful

Please explain why:

Answer: It visually display the relationship of these tweets.

Question: 8. How useful is visualising tag cloud based on tweets?

Answer: 2. Useful

Please explain why:

Answer: Tag cloud can be treated as assistant explanation to trend. Because the title of trend is just some key words and it is difficult for users to understand the meaning of this trend, tag cloud enables users to further learn what is the meaning of this trend.

Question: 9. How useful is viewing a tweet Card by clicking on a tweet?

Answer: 2. Useful

Please explain why:

Answer: it is useful. Because most information delivered by the interface of visualising top trends in a selected country belong to summarized style, such as top 10 trends and relationship of tweets, tweet Card provides users an opportunity to check every tweets and see the detailed content. For example, I always check the most popular tweet and tweets with shared image.

Additionally, based on clear color classification, it is easy for users to retrieve specific tweet.

Question: 10. How useful is visualising selected tweets from a particular area on the map?

Answer: 1. Very Useful

Please explain why:

Answer: it is very useful. the design of this application is very flexible. users not only can visualise top trends in a selected country, but also visualise selected tweets from a particular area on the map. then, users is allowed to see the distributed situation of keywords via keyword search. Users can regard tags cloud as reference of keyword search.

Question: 11. How useful is highlighting real-time tweets based on your own keywords?

Answer: 2. Useful

Please explain why:

Answer: It is useful to retrieve specific or required tweets quickly.

Question: 12. Please rate the application overall?

Answer: 1. Very Useful

Please explain why:

Answer: Generally, this application is very useful. I am so surprised that I can visualise thousands of tweets as the simple and clear way. The entire design is structured and clear. Additionally, reasonable classification and planning of interface make delivery of information well-organized and informative.

Question: 13. How often will you be willing to use this application?

Answer: 2. Often

Question: 14. For which purposes will you use this application? Please give reasons.

Answer: - check tweet dynamics of specific country - learn popular topics in twitter

Participant 4:
Question: 1. Please rate your experience in learning to use the application.

Answer: 2. Easy to Learn

Please explain why:

Answer: I haven't not Twitter before.

Question: 2. Please rate visualising real-time tweets on the map.

Answer: 3. Neutral

Please explain why:

Answer: To see which country's using Twitter most, but it's not my interest.

Question: 3. How useful is visualising classification of platforms of real-time tweets on the map?

Answer: 2. Useful

Please explain why:

Answer: It's trendy!

Question: 4. How useful is classifying languages of real-time tweets?

Answer: 3. Neutral

Please explain why:

Answer: n/a

Question: 5. How useful is visualising popular hashtags of real-time tweets?

Answer: 1. Very Useful

Please explain why:

Answer: It's trendy, and interesting.

Question: 6. How useful is visualising top trends in a selected country?

Answer: 1. Very Useful

Please explain why:

Answer: I like the part most as a Twitter user.

Question: 7. How useful is visualising tweets related to a selected trend?

Answer: 3. Neutral

Please explain why:

Answer: n/a

Question: 8. How useful is visualising tag cloud based on tweets?

Answer: 2. Useful

Please explain why:

Answer: It's colourful and quickly gotten.

Question: 9. How useful is viewing a tweet Card by clicking on a tweet?

Answer: 2. Useful

Please explain why:

Answer: n/a

Question: 10. How useful is visualising selected tweets from a particular area on the map?

Answer: 1. Very Useful

Please explain why:

Answer: Can search the highlight.

Question: 11. How useful is highlighting real-time tweets based on your own keywords?

Answer: 2. Useful

Please explain why:

Answer: Can search the highlight.

Question: 12. Please rate the application overall?

Answer: 3. Neutral

Please explain why:

Answer: I don't know whether i need the app as a twitter user.

Question: 13. How often will you be willing to use this application?

Answer: 2. Often

Question: 14. For which purposes will you use this application? Please give reasons.

Answer: To see the trends/ popular hashtags in twitter.

Participant 5:

Question: 1. Please rate your experience in learning to use the application.

Answer: 2. Easy to Learn

Please explain why:

Question: 2. Please rate visualising real-time tweets on the map.

Answer: 2. Useful

Please explain why:

Question: 3. How useful is visualising classification of platforms of real-time tweets on the map?

Answer: 2. Useful

Please explain why:

Question: 4. How useful is classifying languages of real-time tweets?

Answer: 2. Useful

Please explain why:

Question: 5. How useful is visualising popular hashtags of real-time tweets?

Answer: 1. Very Useful

Please explain why:

Question: 6. How useful is visualising top trends in a selected country?

Answer: 2. Useful

Please explain why:

Question: 7. How useful is visualising tweets related to a selected trend?

Answer: 2. Useful

Please explain why:

Question: 8. How useful is visualising tag cloud based on tweets?

Answer: 2. Useful

Please explain why:

Question: 9. How useful is viewing a tweet Card by clicking on a tweet?

Answer: 2. Useful

Please explain why:

Question: 10. How useful is visualising selected tweets from a particular area on the map?

Answer: 3. Neutral

Please explain why:

Question: 11. How useful is highlighting real-time tweets based on your own keywords?

Answer: 2. Useful

Please explain why:

Question: 12. Please rate the application overall?

Answer: 2. Useful

Please explain why:

Answer: It's easy to use and of great importance in catching up with what's happening now in the world

Question: 13. How often will you be willing to use this application?

Answer: 2. Often

Question: 14. For which purposes will you use this application? Please give reasons.

Answer: It's easy to use and of great importance in catching up with what's happening now in the world

Participant 6:
Question: 1. Please rate your experience in learning to use the application.

Answer: 1. Very Easy to Learn

Please explain why:

Answer: The set of controls at the top of the screen are conveniently located and makes it easy for user to use the application

Question: 2. Please rate visualising real-time tweets on the map.

Answer: 1. Very Useful

Please explain why:

Answer: The colored countries map and other color coded aspects of the application makes the application very useful

Question: 3. How useful is visualising classification of platforms of real-time tweets on the map?

Answer: 1. Very Useful

Please explain why:

Question: 4. How useful is classifying languages of real-time tweets?

Answer: 1. Very Useful

Please explain why:

Answer: I guess this is the most interesting and useful feature. Besides what languages people use to communicate on twitter, it also can give people idea of the popularity of twitter in different regions of the world

Question: 5. How useful is visualising popular hashtags of real-time tweets?

Answer: 1. Very Useful

Please explain why:

Answer: Visualising popular hashtags of real-time tweets is very useful. One can get an idea of the most popular hashtags (on the left side, with the tallest bars) by having a glimpse at chart.

Question: 6. How useful is visualising top trends in a selected country?

Answer: 1. Very Useful

Please explain why:

Question: 7. How useful is visualising tweets related to a selected trend?

Answer: 1. Very Useful

Please explain why:

Question: 8. How useful is visualising tag cloud based on tweets?

Answer: 1. Very Useful

Please explain why:

Question: 9. How useful is viewing a tweet Card by clicking on a tweet?

Answer: 1. Very Useful

Please explain why:

Question: 10. How useful is visualising selected tweets from a particular area on the map?

Answer: 1. Very Useful

Please explain why:

Question: 11. How useful is highlighting real-time tweets based on your own keywords?

Answer: 1. Very Useful

Please explain why:

Question: 12. Please rate the application overall?

Answer: 1. Very Useful

Please explain why:

Question: 13. How often will you be willing to use this application?

Answer: 2. Often

Question: 14. For which purposes will you use this application? Please give reasons.

Answer: To know about the most popular tweets in a particular country and see what languages people use to communicate on twitter

Participant 7:

Question: 1. Please rate your experience in learning to use the application.

Answer: 3. Neutral

Please explain why:

Question: 2. Please rate visualising real-time tweets on the map.

Answer: 2. Useful

Please explain why:

Question: 3. How useful is visualising classification of platforms of real-time tweets on the map?

Answer: 3. Neutral

Please explain why:

Question: 4. How useful is classifying languages of real-time tweets?

Answer: 3. Neutral

Please explain why:

Question: 5. How useful is visualising popular hashtags of real-time tweets?

Answer: 2. Useful

Please explain why:

Question: 6. How useful is visualising top trends in a selected country?

Answer: 2. Useful

Please explain why:

Question: 7. How useful is visualising tweets related to a selected trend?

Answer: 3. Neutral

Please explain why:

Question: 8. How useful is visualising tag cloud based on tweets?

Answer: 3. Neutral

Please explain why:

Question: 9. How useful is viewing a tweet Card by clicking on a tweet?

Answer: 2. Useful

Please explain why:

Question: 10. How useful is visualising selected tweets from a particular area on the map?

Answer: 2. Useful

Please explain why:

Question: 11. How useful is highlighting real-time tweets based on your own keywords?

Answer: 3. Neutral

Please explain why:

Question: 12. Please rate the application overall?

Answer: 2. Useful

Please explain why:

Question: 13. How often will you be willing to use this application?

Answer: 3. Neutral

Question: 14. For which purposes will you use this application? Please give reasons.

Answer: for finding data

Participant 8:
Question: 1. Please rate your experience in learning to use the application.

Answer: 2. Easy to Learn

Please explain why:

Answer: The buttons of this tool are easy to understand, and the layout of this tool is very simple. we do not feel confused when using this tool. besides, the functions are easy for use to learn

Question: 2. Please rate visualising real-time tweets on the map.

Answer: 2. Useful

Please explain why:

Answer: very interesting, we can read the up-to-date tweets in real time. and its easy for us to do that

Question: 3. How useful is visualising classification of platforms of real-time tweets on the map?

Answer: 2. Useful

Please explain why:

Answer: maybe not useful for the normal users, but it's helpful for those people who want to analyse the platforms. it's interesting that we can learn how real-time tweets are generated by different platforms.

Question: 4. How useful is classifying languages of real-time tweets?

Answer: 1. Very Useful

Please explain why:

Answer: we can find tweets that are in a specific language. through the classification, it's faster to find a tweet that we are interested in.

Question: 5. How useful is visualising popular hashtags of real-time tweets?

Answer: 3. Neutral

Please explain why:

Answer: too many things on the screen, may feel confusing what to select

Question: 6. How useful is visualising top trends in a selected country?

Answer: 1. Very Useful

Please explain why:

Answer: through this function, we can see those tweets in a specific country that we are interested in. also, we can see the hot topics and images, very useful for most people, and quite interesting

Question: 7. How useful is visualising tweets related to a selected trend?

Answer: 1. Very Useful

Please explain why:

Answer: through visualising tweets related to a selected trend, it's faster for us to find interesting tweets. and the visualisation helps understand that better.

Question: 8. How useful is visualising tag cloud based on tweets?

Answer: 1. Very Useful

Please explain why:

Answer: it's easy for us to see what tags are the most popular. compared to reading other types of tags, this one is more convenient and interesting

Question: 9. How useful is viewing a tweet Card by clicking on a tweet?

Answer: 1. Very Useful

Please explain why:

Answer: we can get access to that tweet easily. it saves time.

Question: 10. How useful is visualising selected tweets from a particular area on the map?

Answer: 2. Useful

Please explain why:

Answer: similar to other functions, this function meets our specific requirement that read the tweets from a particular area, the visualisation helps us to better understand that.

Question: 11. How useful is highlighting real-time tweets based on your own keywords?

Answer: 2. Useful

Please explain why:

Answer: it narrows the area,

Question: 12. Please rate the application overall?

Answer: 1. Very Useful

Please explain why:

Answer: with this tool, we can reads the tweets in a smarter way. those functions are fabulous!

Question: 13. How often will you be willing to use this application?

Answer: 1. Always

Question: 14. For which purposes will you use this application? Please give reasons.

Answer: to find particular tweets, or just for fun! its useful and interesting, we need to explore the functions more

10.4 Appendix 4: Screenshot of website created for facilitating usability study