News Map Visualization

Interactive visualization and analysis of social media

Angelos Cleanthous

M00222257

CMT4141

Supervised By: Dr. Kai Xu

Abstract

The last few years social network sites like Facebook and twitter are becoming more commonly used and the data being transferred over the web are getting enormous. The availability of these data has given an interest for scientists, from different fields for creating various applications called visualization tools and understanding complex situations. This paper demonstrates the implementation of a visualization tool called news map visualization and by what means it can be useful for a user.

Table of Contents

Chapter 1 - Introduction

In the last few years the web has become larger and more people around the world are being able to use any of the available social network sites like Facebook, Twitter for sharing and giving their opinions on different subjects, more easily. While all the people are constantly using these technologies the data that are created on the web are getting enormous and giving an interest for the people of computing to create a set of tools for analyzing them and using them for different purposes. Different kinds of companies and scientists are using this technique to gain a better understanding for various situations. According to Case and King (2011) twitter has been used from the Center for Disease Control and Prevention in United States to provide updates regarding the H1N1 situation. Twitter has also been used in Businesses as a promotional tool. For example a marketing firm might want to know what are the latest trends from around the world and concentrating on upgrading their products and services that will bring more profit to their company Wikipedia (2012).

This project demonstrates the implementation of the "news graph visualization" that processes information of the tweets coming from a large database of Social Network Site (SNS) called Twitter. Processing involves the extraction of the metadata of the tweet for example what the user has posted, what is the location of the tweet, when the user has made the post, etc. The document makes use of the NER (Names Entity Annotation) which is the task of extracting annotated data from a document such as specific patterns of name locations. After processing and extracting the information from the tweets the system is able to place those results on a geo-graphical map by giving a better understanding of the query the user has requested from the twitter database.

The application introduced is to let people visualize tweets on a geographical map, generated from users and from the news agencies themselves. More importantly the system will have support of showing tweets from different countries like Spain, France, England, Greece and USA and giving the flexibility for the user to view news from he's/her's favorite news channel as well. Besides visualizing the news on the map the system will have the capability of distinguishing positive and negative news by analyzing the text return from twitter so that the reader will have more interaction with the system. Additionally a user will be able of interacting with the available data shown on the screen, like opening internal windows and seeing who made the post, at what time, what he has said inside that tweet, etc. Both of the tweets shown on the map are also laid out on the two left and right columns of the page so that he/she is able to view them all more clearly and interacting with them in various ways.

News map visualization is a web application written in JavaScript using twitter's API and Google Maps JavaScript API 3. It also uses Yahoo's placemaker service for identifying locations within different types of texts in case the user doesn't provide a location on their profile.

Twitter API provides three ways of allowing users to gain access to their core database, the REST API, the Search API and the Streaming API. Each one of them has different uses like the Search API returns relevant tweets that match a predefined query while the Streaming API gives real time continuous

stream of all Tweets, although it doesn't filter Tweets that are relevant. News graph visualization will make use of the search API and the REST API since it will have to make two kinds of requests. The search API will give as relevant tweets that match a specific query like "Euronews" and the REST API will return the most recent posts of the news company.

In the following chapters a background will be given to describe the context of this project while in the third chapter a specification of the requirements will be specified. On the fourth chapter analysis and design will come into place that will document the requirements of the "News Map Visualization" design using appropriate methods. In most cases the documenting of the system's requirements is done by using UML diagrams (e.g Class Diagrams) and including commentaries on how design decisions are made. On the following chapter implementation and testing will give details on how the design is actually implemented and how the testing of the system has been planned and conducted. After showing how the application has been implemented a demonstration will occur to show the system in action and evaluate if it fulfills the initial requirements. On the seventh chapter and before concluding with the report there will also be a critical evaluation to give further and possible improvements to the system.

Before continuing I would like to thank some very important people that had a significant role in this project. First to mention is my loving family who were very supportive and Dr. Kai Xu, a very helpful supervisor who has shown me every step for completing this project.

Chapter 2: Background and literature review

2.1 Information visualization

Information can be defined as a process of transforming the vast amount of abstract information or data into human readable form with the use of computer software. Abstract data include both numerical and non-numerical such as text and even geographic information. Visualization of data can give an ease of use by enabling users to retrieve and view the information they want and understand it in a short period of time.

In the society we are leaving news reporting plays an important role for the people since it keeps them informed with the latest news that are happening around the world. According to Ningjing Ma(2012) "*Facing a large number of news information, people want to know more about the relevancy of them and find the essential characteristics and rules behind them*". Because now there are plenty of websites and portals that offer the service of showing news to the public audience, most of them list the news line by line which makes it more unimaginative and less accessible for them. However news have various assets over their text such as time, location, text, images, links etc. so it takes them more time for the users to read. Therefore it is best if we find a way of representing the different news on a timeline or even better on graphical map such as Google maps. Because the news happen in a different location on the map news readers can easily find out where the events are happening along with all the rest information that an article holds inside.

2.2 Social Network Analysis

Social networks and even social network analysis has engaged with people's curiosity in the last few years. A social network is consisted with actors or nodes that each one of them has at least one connection with each other. In other words social networking has to with the grouping of individuals into specific groups like neighborhood or communities. However social networking is most commonly known over the internet where people get connected with each other by sharing their contact details, sharing information on different subjects like politics, cooking, religion, sports etc. The good thing about online networking is its diversity since anyone from around the world either that is from India or Romania could easily access a social website and develop an online friendship and learn many things about others cultures [1]. As mentioned from before a social network is a term used to study relationships between individuals or even entire social groups that became known from phycology, sociology and graphs theory, while Jacob Moreno was the first one of creating sociograms for studying interpersonal relationships during the 1930(Wikipedia 2012).

Some famous social networking sites are called Facebook, Twitter and meetup. Becoming a member in these sites gives you the opportunity of creating new relationships with people although relationship could only mean the existence of a profile link over a website or the evaluation of another person during

an online meet up. Nevertheless online social networking means nothing for someone until he/she starts connecting with other people and creates relationships with them.

Social network analysis is not about examining the physical connection between these people but their individual structure as objects and the nodes or links that connect each other over the network.

Based on Paul Matton (2004) findings *"a social network analysis concerns with the modeling and measuring of the relationships and flows between entities"* . Paul Matton describes an IRC bot shown on figure 1 that is able of connecting to an IRC network and identifying the nodes and the edges where each node represent a user and an edge is the relationship between the two nodes. After identifying those elements a graphical map is produced to visualize how the network changes over time and to identify the similarities between other channels.

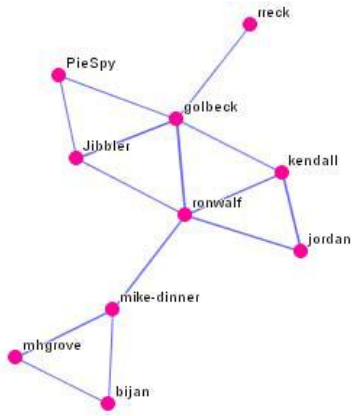


Figure 2: A simple social network

2.3 Twitter

Twitter is a free micro blogging service used from millions of people from around the world for engaging in different conversations, keeping informed on upcoming events, reporting news to families', member's, co-workers and friends, within the use of any device that has internet connection. Twitter was discovered from Dorsey during a brainstorming session held by a podcasting company called Odeo. The idea was that an individual could communicate within a group using an SMS service. After deciding the project code name which was inspired from Flickr and the five character length SMS code 40404, twitter was initially launched on 21st of March of 2006 with Dorsey's first message *"just setting up my twttr".* However, twitter was publicly released on the 15th of July 2006 since Odeo employees wanted to firstly experiment with twitter prototype and then reveal a full version.

It is estimated that twitter has been one of the most common used sites in the world with over 44.5 million users worldwide and increasing to 1382% year by year(Takeshi et.al (2010)). It is said that twitter "retweet activity" is the equivalent of email forwarding while users are posting messages originally posted from others.

Ever since twitter's inception it has been notable used in various environments like governments, policing, news reporting, in emergencies, in survey opinions, in space mission news etc. Case King (2011). There are many reasons why governments use twitter. One of those reasons is to facilitate the

public with up to date emergency information, posting key events, amending new content via hash tags and the competency to be used on wireless devices (F. Dianne 2010). Some other benefits that twitter could bring to the government departments is by providing to their public domain ways of subscribing to their latest updates via RSS SMS or even email and keeping an open relationship with their journalists and bloggers from around the world.

2.4 Accessing tweets from the Twitter

Twitter API provides three ways of allowing users to gain access to their core database, the REST API, the Search API and the Streaming API. Each one of them has different uses like the Search API returns relevant tweets that match a predefined query while the Streaming API gives real time continuous stream of all Tweets, although it doesn't filter Tweets that are relevant. News graph visualization will make use of the search API and the REST API since it will only need to make two kinds requests. The search API will give as relevant tweets that match a specific query like "Euronews" and the REST API will return the most recent posts of the news company. There is evidence that twitter API has a limit of number of requests for each white listed IP address of 10,000 requests per hour (Haewoon, lee & Housung, 2010). It is also advised to keep the request rate around 10,000 and maintaining a time dealy between each request for not having any duplication on the output of data.

Because users are not required to provide their location on their profile preferences some tweets will not be exclusively geo-tagged (geographic coordinates latitude and longitude) and some search API's responses will not have all the necessary information. Twitter API has clearly explained that it is purely optional for the users to provide their location while tweeting through twitter (Twitter Dev 2012).

Twitter API (Twitter dev 2012) provides with different formats their responses to requests depending on the application needs like XML, JSON and ATOM. I nowadays most of the programmers try to use JSON for interchanging information between their applications since it is very easy to implement and easy to use. JSON is a lightweight data interchange in text format which makes it more human readable and it is based on JavaScript's object literal notation.

2.5 Geocoding data

Geocoding data play a significant role on the development of this project since it provides the service of finding the geo coordinates like latitude and longitude of a given query and representing them on an animated graphical map. The most common service of geo-coding information is Google's API and Yahoo's place finder. Both APIs require authorization for using their services and both of them have similarities on their HTTP requests and their response formats are JSON and XML. Because their services are used in commercial basis they have given a limitation on how many requests a user can make for each. Google API limits their requests into 2,500 per day from each whitelisted IP address and Yahoo place finder limits their requests into 50,000 per day for every user application. Mano Marks (2010) guidelines of the policy clearly state the use of geocoding without plotting them on a map is prohibited. The only difference between the two APIs Google geocode gives better results that Yahoo's place finder.

2.6 Technologies used

Visualization tools can be built from any existing programming languages which support graphics like JavaScript, Action Script or PHP. However the question is which one them is the most appropriate to choose from and some important factors need to be taken into consideration, like functionality and performance. It is known that JavaScript will definitely perform faster than PHP since is a client side scripting language and it will not have to send and perform actions back and forth the server side.

2.7 Related work

Hangzai et.al (2006) have created a novel visualization framework for extracting keywords from text, regarding their interestingness, and help audiences find news stories of their own interest from a variety of news channels more easily. This kind of analysis helps businesses, politics and even the military for different purposes. On the one side of the story business analysts or even security experts can benefit from this information from cultural analysis, decision making and on the other hand simple people could just have fun only by reading or watching those interesting news. Although providing a summarization of large scale -videos are very important for the general audience it is also very expensive when trying to examine all those available news videos and reproduce a summary of them. Nevertheless the framework that they have developed is able of categorizing news stories into their corresponding importance and giving them weight of interest. Another advantage of their novel news visualization framework is being able to organize all those large scale news videos more efficiently and attract the user's attention. In figure 2 an example of videos news visualization is given to demonstrate the presentation of the most relevant news stories where users can have an overview while observing the news tendency of changing over time.

Figure 2: An example of video news visualization.

Chapter 3: Requirements Specification

In this chapter all the necessary requirements are gathered together for visualizing how the system will work and what features it will hold. At the same time a use case will be created to help in our future testing plans of the system.

Ideally News map Visualization will give the user the ability to view which news events are most tweeted over the internet and how many of those users are tweeting for the same news in relation to the channels RSS feeds. Users must be able to select and view news from their preferred or origin country either that is from England, Spain, Italy or Greece etc and afterwards select from the available news channels. After selecting their country and their channel they will need to choose the amount of tweets that they want to be shown on the map and if they want they can also choose from which date they want visualize tweets but that is optional. Until the user submits the form and requests the tweets from the Twitter database, the system must check if the user has given all the necessary input for the application to run otherwise an error message will pop up indicating the mistake he has made. There are two types of tweets being requested from twitter's API. One type of them is the one posted from the users and the other one is the news tweets which are posted from the news channels as RSS feeds. Consequently when the user clicks on any of the news on the page with the button "show user tweets" he/she will instantly get all the tweets which are posted from user's that talked about the same event.

All those tweets which are returned from twitter and plotted on the map they will need to be shown on a separate div or column so that the user will have a better understanding of the available information. However the information that is represented on the columns will have the name of the user, the date of the tweet posted on twitter, the text and the URL where the user will be able to open and view on the same page.

Another very good feature of the application that will hopefully possess with by the end of this project is its ability to distinguish between positive and negative tweets returned from the twitter so that users will have a better interaction with the system and knowing which the bad and good news are. More or less this will limit their cognitive work load memory interacting with the application and trying to read all the available information.

If it's possible to produce a sentiment analysis on the available text of the tweets it will be also good to actually take all those bad news and put them over an animated news ticker showing them as breaking news.

Finally when the system first loads it will provide the user with a brief explanation on how to use and interact with the application.

Aims

1. Visualize user and news tweets on a geo-graphical map

2. Show both type of tweets on the same page
3. Embed user interaction between the tweets of the page and the map.
4. Produce a sentiment analysis on the text to distinguish positive and negative news.
5. Show which users post for the same news in relation to the channels postings on twitter

Objectives

1. Attain a Google and a Yahoo API key for being able to make the HTTP GET requests.
2. Find which twitter APIs will be necessary for the requests.
3. Retrieve and save user's physical location.
4. Geocode user's physical address.
5. Use other available services to extract locations from text in case users don't provide location on their profile page.
6. Construct the two columns that will show the two types of tweets.
7. Show all user tweets when clicking on specific news event.
8. Show different icons (markers) on map for very bad news like explosions, crashes and accidents (Part of text analysis).

3.1 – Requirements

3.1.1 - Project Scope

"News map visualization" attempts to provide users with an easy way of viewing large amount of user and news tweets of their own choice over the same page. For enhancing the visual side of the application the system will have adapted a text analysis approach for searching specific keywords and phrases inside the returned texts from twitter. Hence this will give users the ability of recognizing the bad news from the good ones. Because in times the system will have returned a big number of tweets to the screen it will have to be as easy as possible to use and to have light colors of layout for easier human reading of the data.

3.1.2- Software Requirements

1. Dreamweaver CS5(HTML, JavaScript, PHP IDE)
2. XAMPP server
3. JavaScript library (Jquery)
4. Microsoft Word 2010

3.1.3 Functional Requirements

1. Request news tweets from twitter from a specific news channel and save results to the system.
    a. Save the metadata of each tweet
        i. news_name
        ii. news_date
        iii. news_text
        iv. news_url
        v. news_profile_img
2. Request user tweets mentioning the name of the news channel

    a. Save the metadata of each tweet
        i. User_name
        ii. User_date
        iii. User_text
        iv. User_url
        v. User_profile_img
3. Request user tweets with the same text of each news tweet from the first request.
    a. Save the metadata of each tweet
        i. User_news_name
        ii. User_news_date
        iii. User_news_text
        iv. User_news_url
        v. User_news_profile_img
4. Find the geo-location for each user inside the user_tweets array
    a. Save result back to array
5. Use Yahoo placefinder to find places and locations from text for each user who doesn't provide location as a preference on his profile.
    a. Save result back to array
6. Use Yahoo placefinder to find locations from all the texts of news_array
    a. Save result back to the array
7. Search specific keywords of all texts to distinguish between the good and bad news.
8. Show different markers on the map for news which have very negative phrases inside their text.
9. Show all the bad news as "breaking news" category over an animated jquery news ticker.
10. Show the rest of the results on separate columns and render them on a map.

3.1.4 Non - Functional Requirements

1. Register with Google API and obtain an API key for loading maps on a page.
2. Register with yahoo API and obtain an application ID
3. No more than 2.0GHz processor and 2GB of R.A.M. is needed for the system to run since all the functions will be running on client side.

To sum the documentation of the system's requirements will play a significant role on the development of this project since it will be a necessity to know whether the end product fulfills them at the end.

3.2 Use case

According to Grady Booch (2005) use case diagrams are used for the presentation of the system requirements to the project's stakeholders for understanding the interaction between the actors (users) and the system.

Below there is the list of use cases which are briefly discussed afterwards.

1. Request tweets without selecting any of the dropdown selections.
2. Request to display tweets by selecting all the required dropdown selections.
3. Request to display tweets by selecting all the required dropdown selections along with the option of date.
4. Click on a user's name.
5. Clicks on a channel's name.
6. Click on a "show tweets" button.
7. Click on a URL link over the text of the user's/news tweet.
8. Click on the "help" icon.

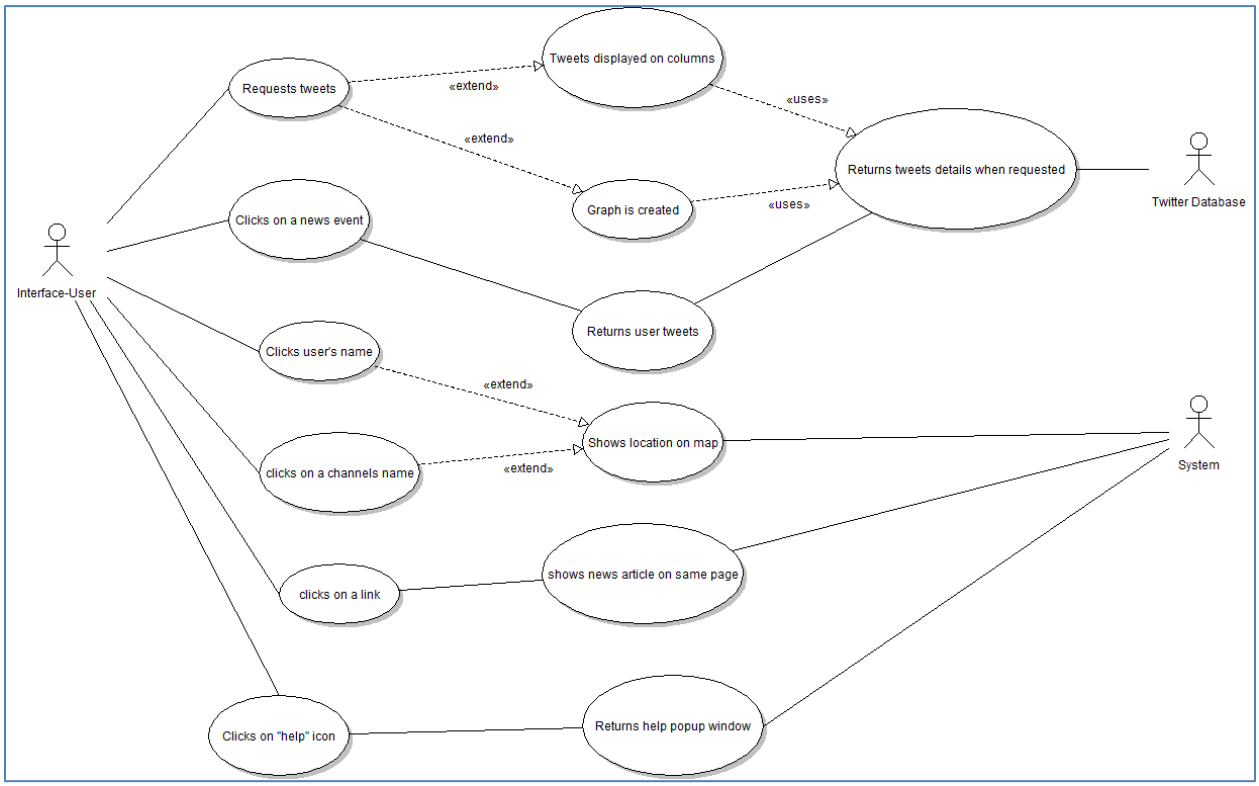


Fig.3- Use Case

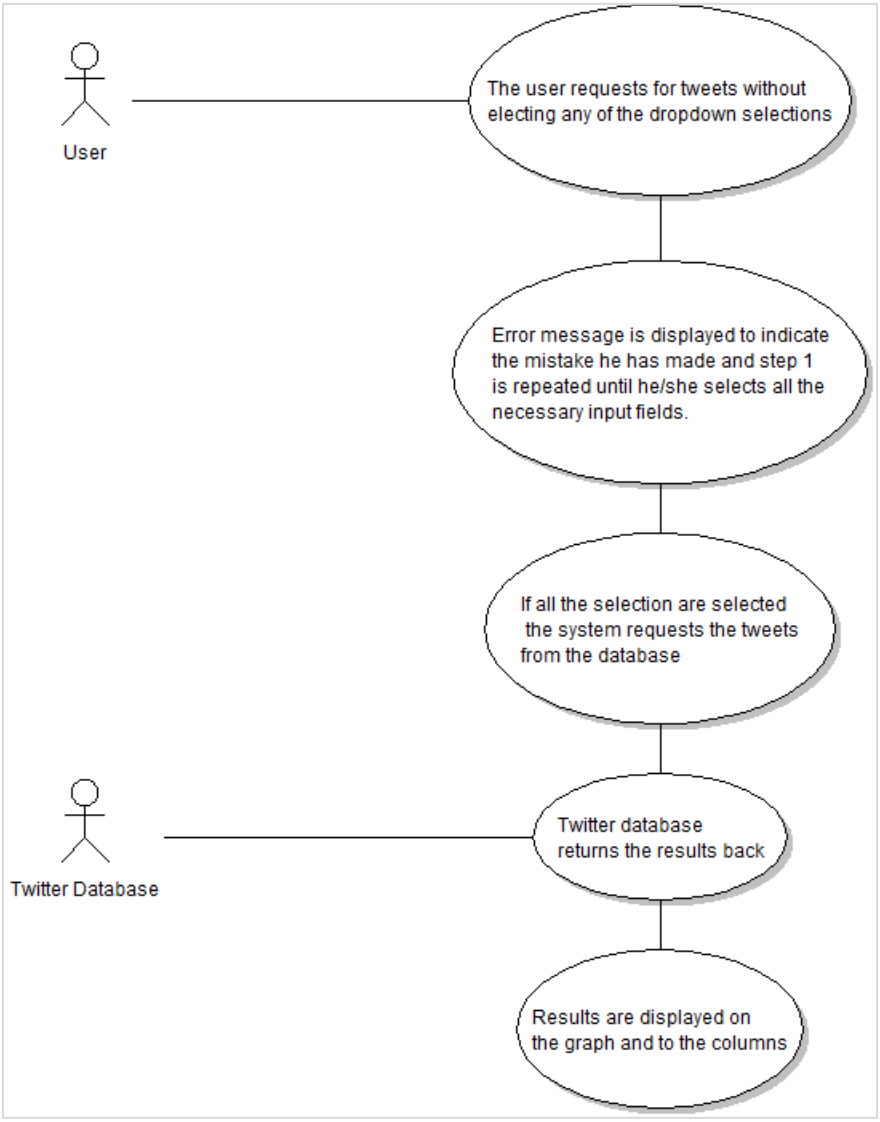
Use case 1- Request tweets without selecting any of the dropdown selections.

Actors: User, Twitter Database

Normal Flow:

1. The user requests for tweets without selecting any of the dropdown selections.
2. Error message is displayed to indicate the mistake he has made and step 1 is repeated until he/she selects all the necessary input fields.
3. If all the selection are selected the system requests the tweets from the database.
4. Twitter database returns the results back.

5. Results are displayed on the graph and to the columns

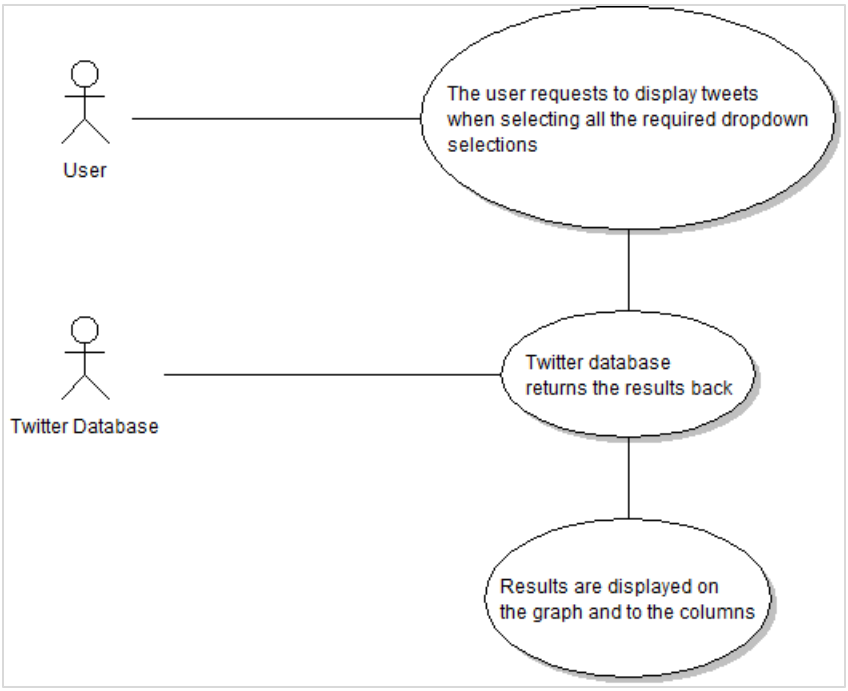


USE CASE 2- Request to display tweets by selecting all the required dropdown selections.

Actors: User, Twitter Database

Normal Flow:

1. User requests to display tweets when selecting all the required dropdown selections
2. Twitter database returns the results back.
3. Results are displayed on the graph and to the columns

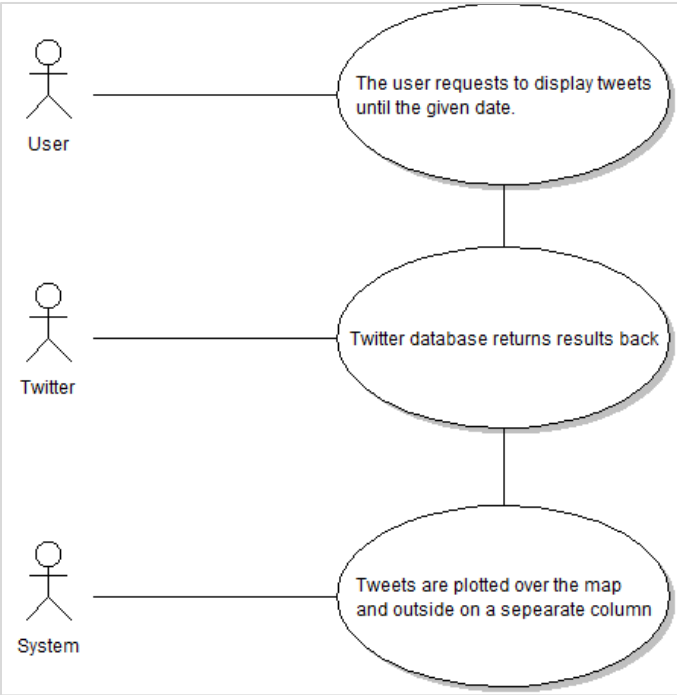


USE CASE 3 - Request to display tweets by selecting all the required dropdown selections along with the option of date.

Actors: User, Interface

Normal Flow:

1. User requests to display tweets when selecting all the required dropdown selections and the date option.
2. Twitter returns results onwards the date given.
3. User tweets are plotted over the map and outside on a separate column.

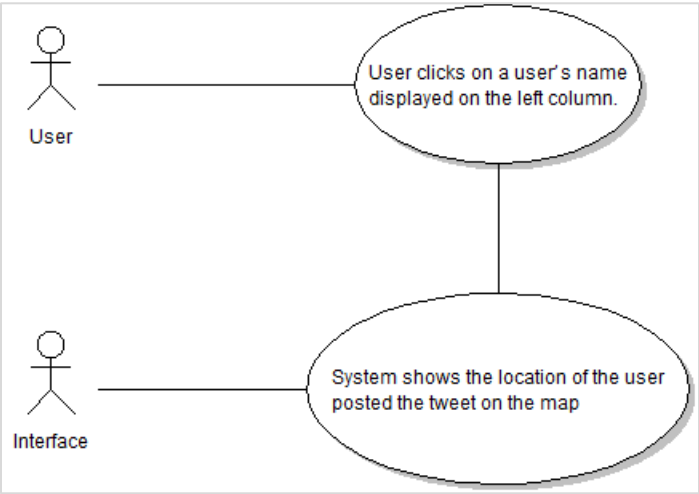


USE CASE 4 - Click on a user's name.

Actors: User, Interface

Normal Flow:

4. User clicks on a user's name displayed on the left column.
5. System shows the location of the user posted the tweet on the map.

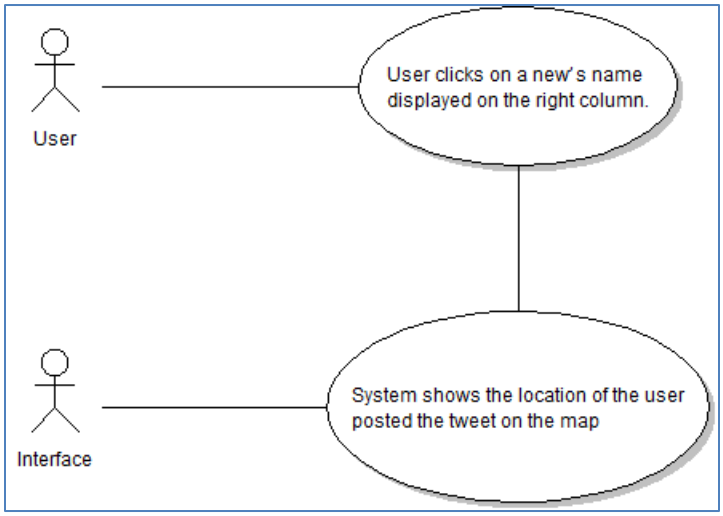


USE CASE 5-User clicks on a channel's name

Actors: User, Interface

Normal Flow:

1. User clicks on a news's name displayed on the right column.
2. System shows where the news event has occurred on the map.

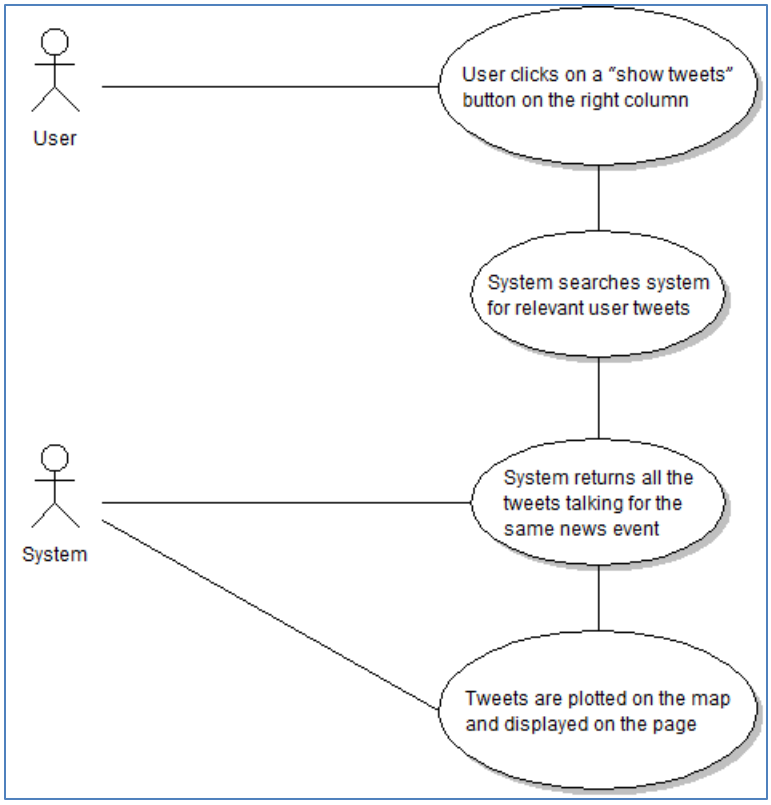


USE CASE 6- Click on "show tweets" button.

Actors: User, Interface

Normal Flow:

1. User clicks on a "show tweets" button on the right column.
2. System searches itself to find relevant user tweets.
3. System returns all the tweets talking for the same news event.
4. Tweets are plotted on the map and displayed on the page as well.

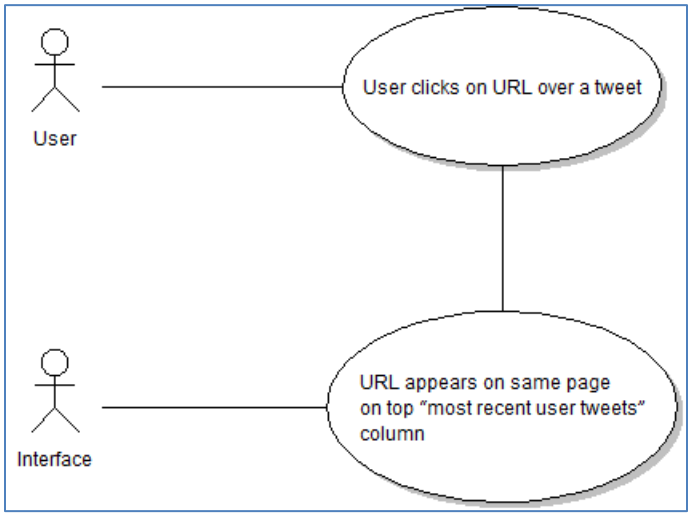


USE CASE 7- Click on a URL link over the text

Actors: User, Interface

Normal Flow:

1. User clicks on URL over a tweet
2. URL appears on same page on top "most recent user tweets" column.

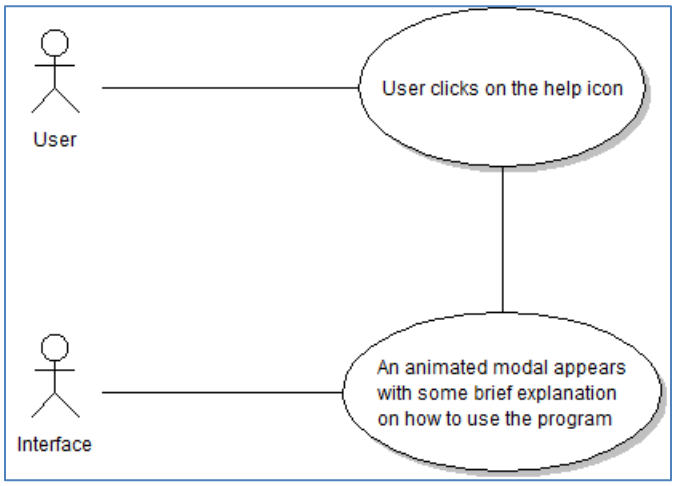


USE CASE 8- Click on the help icon

Actors: User, Interface

Normal Flow:

1. User clicks on the help icon.
2. An animated modal appears with some brief explanation on how to use the program.



Chapter 4- Analysis and Design

On the following pages possible designs will be provided to show how the system will work by describing its functionality and the features it will hold.

4.1-System Architecture

News map visualization is a visualization tool that engages the user to request tweets from his origin country and view them on a geographical map. In order for visualization map to work needs to have access into twitter's database for collecting and showing all the relevant and necessary information to the user. Aside from visualizing tweets on the map the user will be able to interact with the available information placed on the interface. Map visualization map is controlled from the user interface although the data analyzed from the interface are processed from the back end operations of the application.

4.2 – UML Class Diagram

Off all the UML diagrams in object oriented systems modeling, Class diagram is one of the most common. A class diagram can be used to show a static view of the system but also for visualizing, documenting and specifying structural models of executable programs.(Grady Booch 2005) Nevertheless a class diagram consists with a set of classes, interfaces and their relationship between those classes.

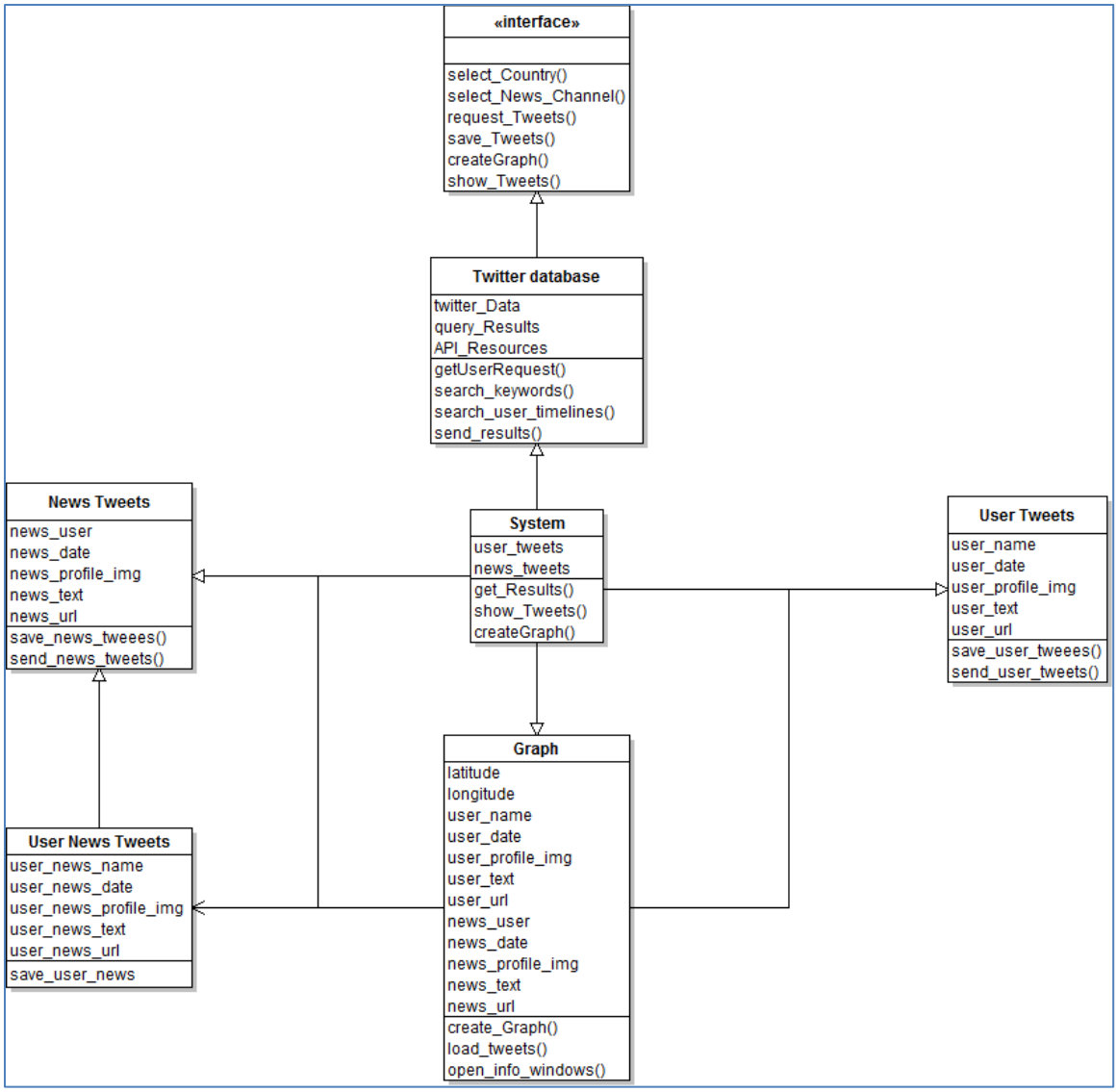


Fig.4 -Class Diagram

4.3 – Sequence Diagram

According to Grady Booch (2005) sequence diagrams are used to model the interaction between the different object instances of the system by showing the messages that are exchanged between them. Sequence diagrams can be very helpful for software developers and for the business staff as well because it shows how actions and messages of a system interact with each other.

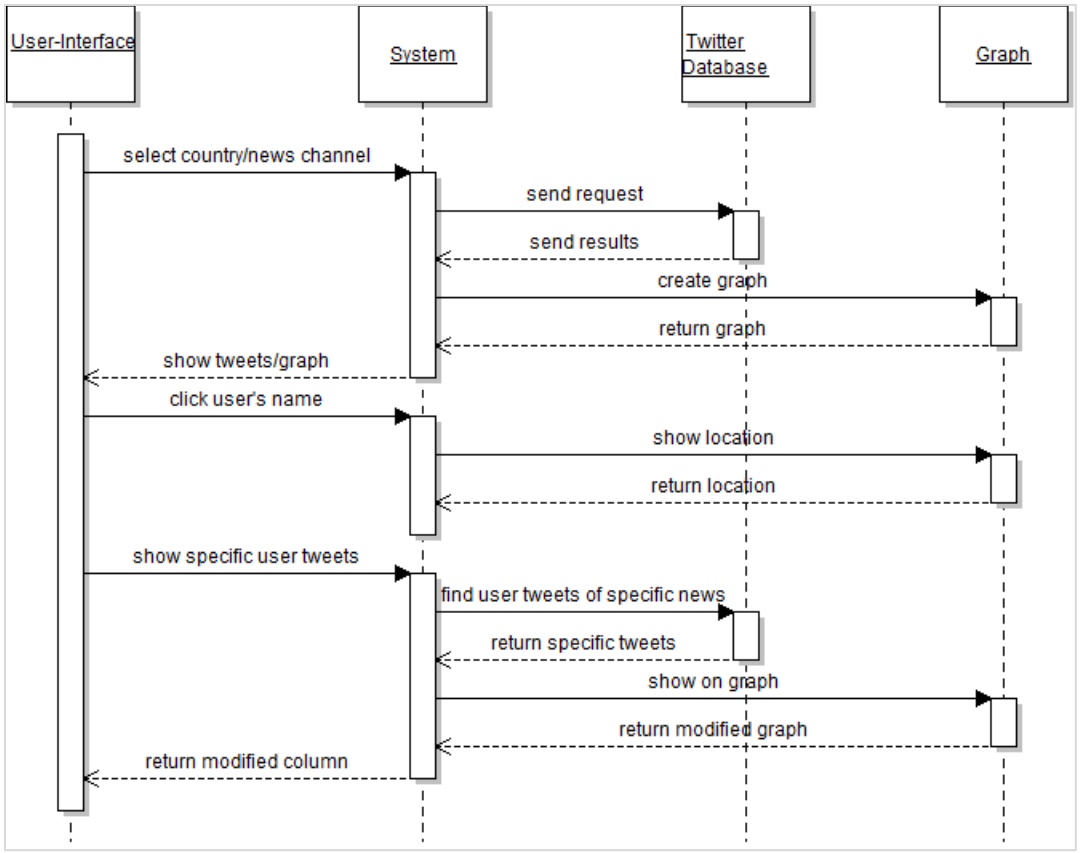


Fig.5 Sequence Diagram

The following figure shows how the user interface of the system looks like and the functionalities which are provided to the end user. All the functionalities of the system will be described on the following paragraphs.

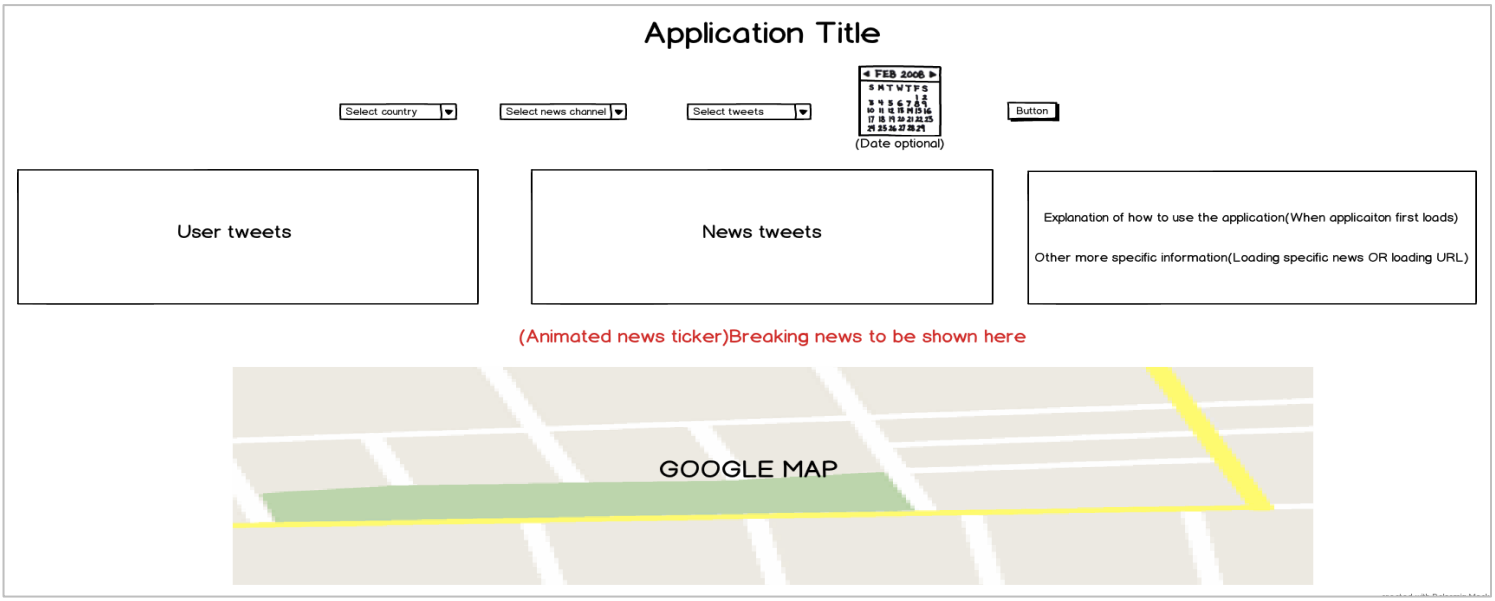


Fig.6 User Interface

Figure 6 shows the layout of the system and the functionalities which are provided to the end user. However users are capable of requesting and viewing a small or a big number of news and user tweets on a geographical map.

Chapter 5 - Implementation and Testing

Unfortunately not all requirements specified from the beginning have been fulfilled by the end of this project. The reason behind this had to do with my father's health issue, where I had to go back in Cyprus and see him and also some other programming problems as well. The functionalities that haven't been implemented during the development concerned with the analysis of text for checking certain words like explosions, disaster, bomb, etc and re-customizing the visual perspective of the interface so that users could distinguish between the very bad news from the good ones. This would also give us the competence to show different icons for the markers on the Google Map and maybe show all the negative news over an animated jquery news ticker between the three columns and the Google Map.

5.1 - Requesting tweets

Once the application is executed the main screen is being displayed to the user. The user must select one selection of each dropdown list to start using the program. On the first dropdown he selects the country, on the other one the news channel and on the last one how many tweets he wants to get. Once the application accepts all the inputs it starts sending the requests to the twitter API. Twitter responds back with the results which are instantly saved into arrays. One array is populated with tweets generated from users mentioning the query that was submitted during the submission e.g. "Euronews", and the other one includes the tweets posted from the news channel, as RSS feeds, received from their public profile. There was also another set of tweets that needed to be imported into an array called user_news_news. This array included all the tweets from users, who have posted for the same news's event of the second array, once a user has clicked on a news tweet shown on the screen.

I must say that I have encountered many difficulties while attempting to save the tweets into the last array but eventually I managed to get it working. Nevertheless on the effort of trying to request tweets with the same text of the second array called "news_array", the browser's console was giving me an error message on each runtime. After a lot research from online discussion groups and blogs people have explained me that I was trying to access data outside from an asynchronous request but instead I had to wait for the request to finish. However people suggested me of using "jQuery's Deferred object" which makes handling of this kind requests much easier.

According to JQuery API(2012) jQuery's Deferred object is a *"is a chainable utility object that can register multiple callbacks into callback queues, invoke callback queues, and relay the success or failure state of any synchronous or asynchronous function*." In other words multiple simultaneously requests are invoked within the same function and output their results when each of them are resolved or failed to execute until the end.

```
function news_tweets(query, user_id, count) {
    var news_array = [],
        user_tweets = [];
    return $.getJSON("https://api.twitter.com/1/statuses/user_timeline.json", {
        include_entities: "true",
        include_rts: "false",
        user_i: user_id,
        count: count
    }).then(function (data) {
        return $.when.apply(null, $.map(data, function (item) {
            news_array.push({
                news_user: item.user.name,
                news_date: item.created_at,
                news_profile_img: item.user.profile_image_url,
                news_text: item.text,
                news_url: item.entities.urls.length ? item.entities.urls[0].url : ''
            });
            return $.getJSON("http://search.twitter.com/search.json", {
                q: item.text,
                rpp: 10,
                include_entities: "true",
                result_type: "mixed"
            }).done(function (data) {
                $.each(data.results, function (i, item) {
                    user_tweets.push({
                        user: item.from_user,
                        user_id: item.from_user_id,
                        date: item.created_at,
                        user_profile_img:item.profile_img_url
                       url:item.entities.urls.length ?
                    item.entities.urls[0].url : '',
                        text: item.text
                    });
                });
            });
        }));
    }).then(function() {
return [news_array, user_tweets,query,count];
    })
}
```

Fig.7 Using Jquery Deferred object for chaining callbacks.

Thankfully their suggestion has given me a great relief by solving my problem and continuing with the rest of the project.

5.2 - Geocoding limits

Google map have limitations when it comes to geocoding addresses. Besides the 2,500 geolocation requests per day and per user usage limitation, Google also limits requests per second. While trying to geocode more than 10 requests in a loop the browser was giving me an error saying "OVER_QUERY_LIMIT". After researching the problem over the web I have found that the reason behind this was that between each request there has to be 1 second apart as seen on the following figure.

```
function geocode_user_tweets(user, date, profile_img, text, location,query) {
  var geocoder = new google.maps.Geocoder();
  geocoder.geocode({
    address: location
  }, function (response, status) {
    if (status == google.maps.GeocoderStatus.OK) {
      var x = response[0].geometry.location.lat(),y = response[0].geometry.location.lng();
      var myLatLng = new google.maps.LatLng(x, y);
      var marker = new google.maps.Marker({icon: profile_img,title: user,map: map,position: myLatLng});
      var infowindow = new google.maps.InfoWindow({content: templates[0].replace('user',
      user).replace('text', text).replace('date', date)});
      var $tweet = $(templates[1].replace('%user', user).replace(/%profile_img/g,
      profile_img).replace('%text', text));
      $('#user-banner').css("visibility", "visible");
       $('#user-banner').text('Users mentioned @'+query+'');
      $('#user-tweets').css({'overflow-y': 'scroll','overflow-x': 'hidden'}).append($tweet);

      function openInfoWindow() {
        infowindow.open(map, marker);
      }
      google.maps.event.addListener(marker, 'click', openInfoWindow);
      $tweet.find(".user").on('click', openInfoWindow);
      bounds.extend(myLatLng);
    } else if (status === google.maps.GeocoderStatus.OVER_QUERY_LIMIT) {
      setTimeout(function () {
        geocode_user_tweets(user, date, profile_img, text, location);
      }, 500);
```

Fig.8 Avoiding geocoding limits

5.3 - Geographic Information Retrieval

Sometimes users from twitter weren't providing locations over their profile, so an alternative way for plotting the rest of the tweets on the map needed to be found. At the beginning of the project Yahoo Placemaker has been used where its main role was for identifying and extracting location names out of unstructured and structured textual content, like RSS feeds, news articles or status updates (Yahoo Dev). Although after several times trying to implement the service on the system, it was found that it couldn't handle simultaneous requests so an alternative service was used called placefinder.

According to Yahoo Dev (2012) Placefinder helps developers to make their programs "location aware" by converting different place names or street addresses into geographic coordinates like latitude and longitude so that they can be used to center a map or place a marker on the map.

Placefinder web service provides developers with two types of requests. First request is to find the coordinates of one's street address and the other is to find coordinates of place names which is been used for this application. To perform this kind of operation jquery getJSON request has been used to geocode and place all the rest of the tweets that users didn't have location on the profile. The following figure shows the implementation of placefinder which is also preceded from the last code shown in the last figure.

```
} else {
$.getJSON("http://where.yahooapis.com/geocode?q=" + text +
"&flags=J&appid=Y7pwNojV34HFg6fmPML_2_YDetrgip_ZFLNaq3cetV6waFtW3O1eF2wJOcU.FMkbHO5iU9R7DxFZ", function (data)
{
  var latitude = data.ResultSet.Results[0].latitude;
  var longitude = data.ResultSet.Results[0].longitude;
  var myLatLng = new google.maps.LatLng(latitude, longitude);
  var marker = new google.maps.Marker({  icon: profile_img, title: user, map: map, position: myLatLng  });
  var infowindow = new google.maps.InfoWindow({
              content: templates[2].replace('user', user)  .replace('news_title', news_title) .replace('date', date)
  });
  var $tweet = $(templates[3].replace('%user', user).replace(/%profile_img/g, profile_img).replace('%news_title', news_title));
  $('#user-banner') .css("visibility", "visible");
  $('#news-banner') .css("visibility", "visible");
  $('#user-news-tweets') .css({'overflow-y': 'scroll', 'overflow-x': 'hidden' }) .append($tweet);

  function openInfoWindow() {
              infowindow.open(map, marker);
  }
  google.maps.event.addListener(marker, 'click', openInfoWindow);
  $tweet.find(".user").on('click', openInfoWindow);

  bounds.extend(myLatLng);
});
}
```

Fig.9 Using alternative services for geocoding user's and news locations

5.4 - Showing tweets

There are two ways of showing the tweets on the main screen of the application and that is within the two columns called the user tweets and the news tweets as well as on the geographic map. On the left column one can identify that there are two types of tweets with different color. The ones geocoded with Google code have blue color and the others returned with a lighter color are from yahoo's placefinder service, where users are not providing location over their profile.

Fig.10 Showing tweets

5.5 - Clicking on a user's name

Once you click on a user's name of the tweet the system shows the location on the Google map, indicating where the user had made the post or where the news event has happened. As I said from before there are two types of tweets returned from the system. Consequently the corresponding marker of the tweet opens up over the map. The infoWindow that opens shows the same details which are shown on the left column.

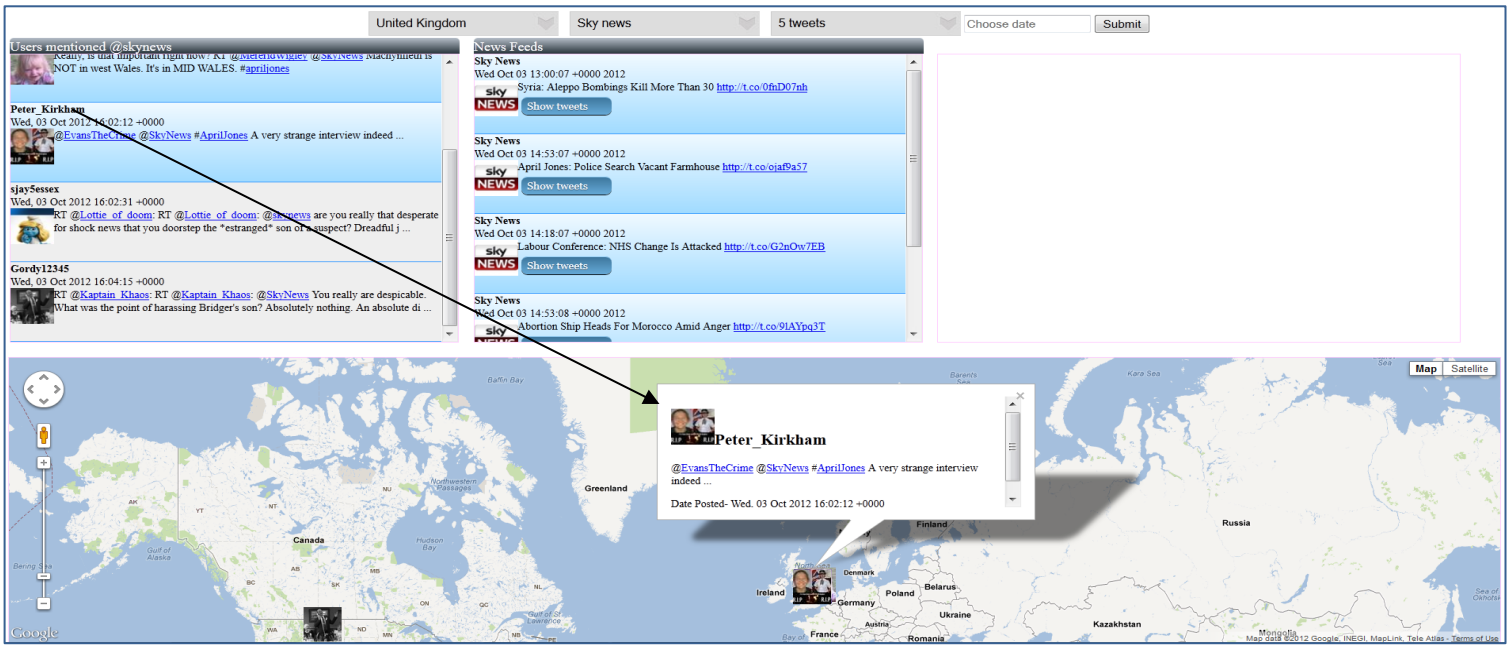


Fig.11 Revealing the tweets

5.6 - Opening a URL

The system offers a way of viewing the articles of the news events which are inside the text of the tweet, on the same page. This makes it easier for the viewer to read the news of each tweet without opening new windows on his/her browser. Once opening the URL he can see it on the right hand side and when he/she finishes reading he may close it by clicking the green cursor on the top corner.

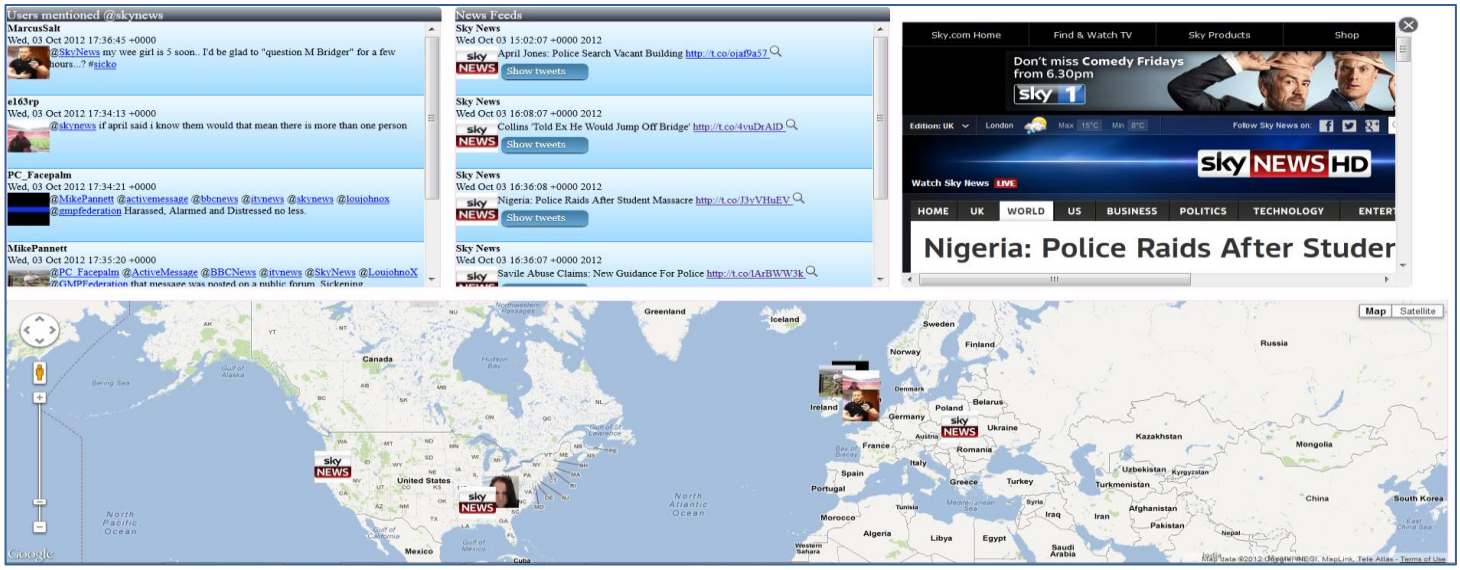


Fig.12 Reading the news

5.7 - Showing related user tweets

The system offers another functionality to the end user by showing all the tweets which are talking for an individual news tweet. This is achieved when the user clicks on any of the buttons named “show tweets”. Once the tweets are shown on the right of the page, there markers are also populated on the map as well.

Fig.13 Showing users who talk about specific news events

5.8 - How to use the system

The application provides the user with a brief step by step tutorial on how to use the system by clicking on the “help” icon on the upper right hand side of the screen.

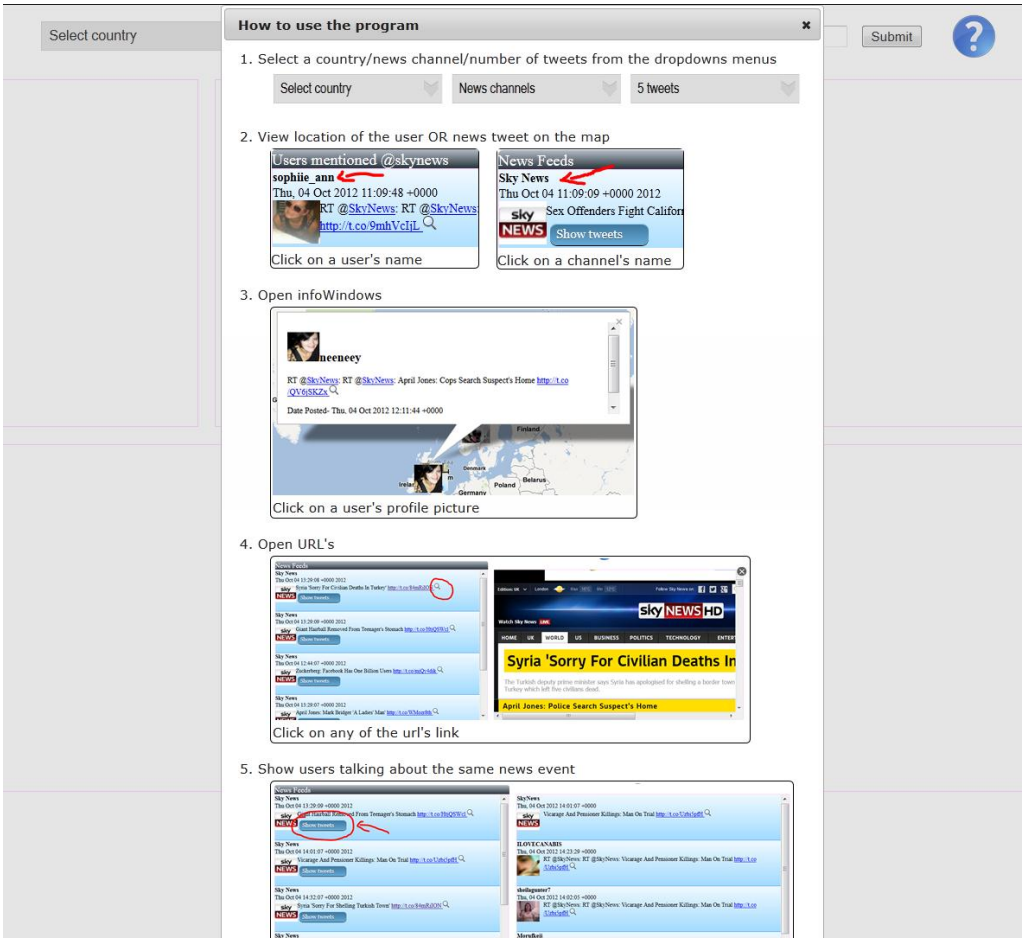


Fig.13

Figure 13 shows the help window loaded on the screen. This is very important for guiding the user on how to use the program and decreasing the time to learn it.

5.9 - Testing

In almost every stage of the software development of a project, testing is always required. When a block of code is created the developer must check whether those functions produce the expected results. Testing may also check whether classes of the system exhibit the proper behavior and everything works as initially planned. Lastly when the entire program is developed a "system testing" is produced for checking if all specified requirements are fulfilled.

The testing of News map visualization is accomplished by using the principles and guidelines of "Test Driven Development". TDD encourages developers to create small iterated unit tests on every new feature developed. Unit testing means when you have a new task to do (i.e., some bit of functionality to implement) you need to write the code that will test whether that functionality will work as required before you implement the functionality itself(David Astels 2003). Writing code (failing unit tests) requires only enough passing the test and putting a limit on the code you will write afterwards.

There are three levels which are recommended from TDD and used for "news map visualization" system. Their names are Module testing, Integration testing and System Testing. News map visualization was tested depending on these three levels of testing. Therefore each time a new functionality was introduced a new test was generated, for checking, and later when classes where built they were also checked whether they were communicating in the right way and producing the expected results.

Finally when all the classes were built, system testing was taking place for ensuring that all of them were executed, as they should, and if any errors were occurring further improvements would have taken place. Below there are the system test cases listed.

5.9.1 - System test cases

Test number: 1

Description: Request tweets without selecting any of the dropdown selections.

Properties:

Input data: None

Expected result: Alert box appears showing the user the error he/she has made.

Actual result: As expected

Test number: 2

Description: Request to display tweets by selecting all the required dropdown selections.

Properties:

Input data: Country selection, News channel selection and the number of tweets

Expected result: The results are returned from twitter and displayed on both the columns and the map

Actual result: User tweets are displayed normally on the left column but news column only outputs one news feed

Corrective actions: Changed Yahoo's placemaker service with placefinder since placemaker couldn't execute multiple simultaneous requests and only the last tweet was returned from the data result set.

Test number: 3

Description: Request to display tweets by selecting all the required dropdown selections along with the option of date.

Properties:

Input data: Country selection, News channel selection and the number of tweets.

Expected result: Return tweets until the given date.

Actual result: No results returned

Corrective actions: The solution of this function has been postponed for later supervision due to some other, more important programming problems.

Test number: 4

Description: Click on a user's name.

Properties:

Input data: String(user's name).

Expected result: Show location of the user on the map.

Actual result: The marker's InfoWindow does not open.

Corrective actions: Save map and marker variables outside the function and put the html strings into set of templates for not declaring them in every loop.

Test number: 5

Description: Click on a news channel name.

Properties:

Input data: String(channel's name).

Expected result: Show location of the event on the map.

Actual result: The marker's InfoWindow does not open.

Corrective actions: Save map and marker variables outside the function and put the html strings into set of templates for not declaring them in every loop.

Test number: 6

Description: Click on “show tweets” button on a news feed.

Properties:

Input data: String(News text).

Expected result: Show all recent posts from users for the same news feed.

Actual result: As expected.

Test number: 7

Description: Click on a URL link over the text of the user’s/news tweet.

Properties:

Input data: String(URL link).

Expected result: Open page on top “most recent user tweets”.

Actual result: The page opens on a different positon on the page.

Corrective actions: Change “position” attribute in CSS to absolute.

Test number: 8

Description: Click on the “help” icon.

Properties:

Input data: none

Expected result: Open popup window to show explanation on how to use the system”.

Actual result: As expected.

5.10 - Testing Implementation

Various problems have occurred during the development of this project but fortunately, almost all of them have been resolved by the end of development. Some of them were easy to untangle but some other needed some extra research. Some of the biggest problems that took me more time to solve had

to do with the results returned from Yahoo's placemaker service and trying to show them on both columns and the map.

Afterwards another problem came along and it had to do with opening the markers of tweets, outside the map. Since this problem was part of the most important functionalities of the program, In the meantime the problem has been resolved and gradually I could continue with the rest of the implementation.

Nevertheless those were the problems which halted the implementation of the remaining program, and not making it more interactive for the user. The goal trying to achieve had to do with distinguishing positive and negative news between them and customizing differently the way tweets were represented on the page.

Chapter 6- System Evaluation

Evaluation's main purpose is for answering questions whether the system is easy for the user to use, whether it has fulfilled the requirements specified at the beginning of the project and their results give suggestions for further improvements for the system.

To evaluate "news map visualization" think aloud protocol has been used. In this type of evaluation technique the user performs a number of tasks and is asked to think aloud and explain what he is doing at each stage, and why. The evaluator records the user's actions (sometimes using tape recordings or video or computer logging or via user notes) though for this evaluation no audio or video recordings have been used but only written notes. Think-aloud has the advantage of simplicity while it requires little expertise to perform, and can provide a useful insight into the problems with an interface. However, there are a number of problems too. The information is necessarily subjective, and can be selective depending on the tasks chosen. Being observed and having to describe what you are doing can also affect the way in which you do something.

6 users have been asked to use the system for testing and after finishing them they had to complete two different questionnaires. First questionnaire consisted with six generic questions to learn the background of the user's and the second one was focusing on the usability testing of the system. During the testing users were also provided with a task list on which they had to complete and later on when all data have been gathered, they would be analysed to find out how much the system is responding to the user needs and if any problems have occurred during the testing. The questionnaires with all the questions can be found in the appendix.

The questions of the first questionnaire deal with understanding the user's background with some generic questions like "What is your age?", "Which social networks sites are you using?", "Have you heard of any visualization tool?" Below there are the main questions of the first questionnaire.

1. How old are you?
2. Have you heard of the terms social networking and online networks?
3. Which social networks sites are you using?
4. Have you heard of any visualization tools?
5. Can you name one visualization tool you have used?
6. What is the main purpose of using Social networks?

As I mentioned before when users are performing the testing of the system they are also provided with a task list which they have to perform step by step the tasks and answer accordingly to the question provided.

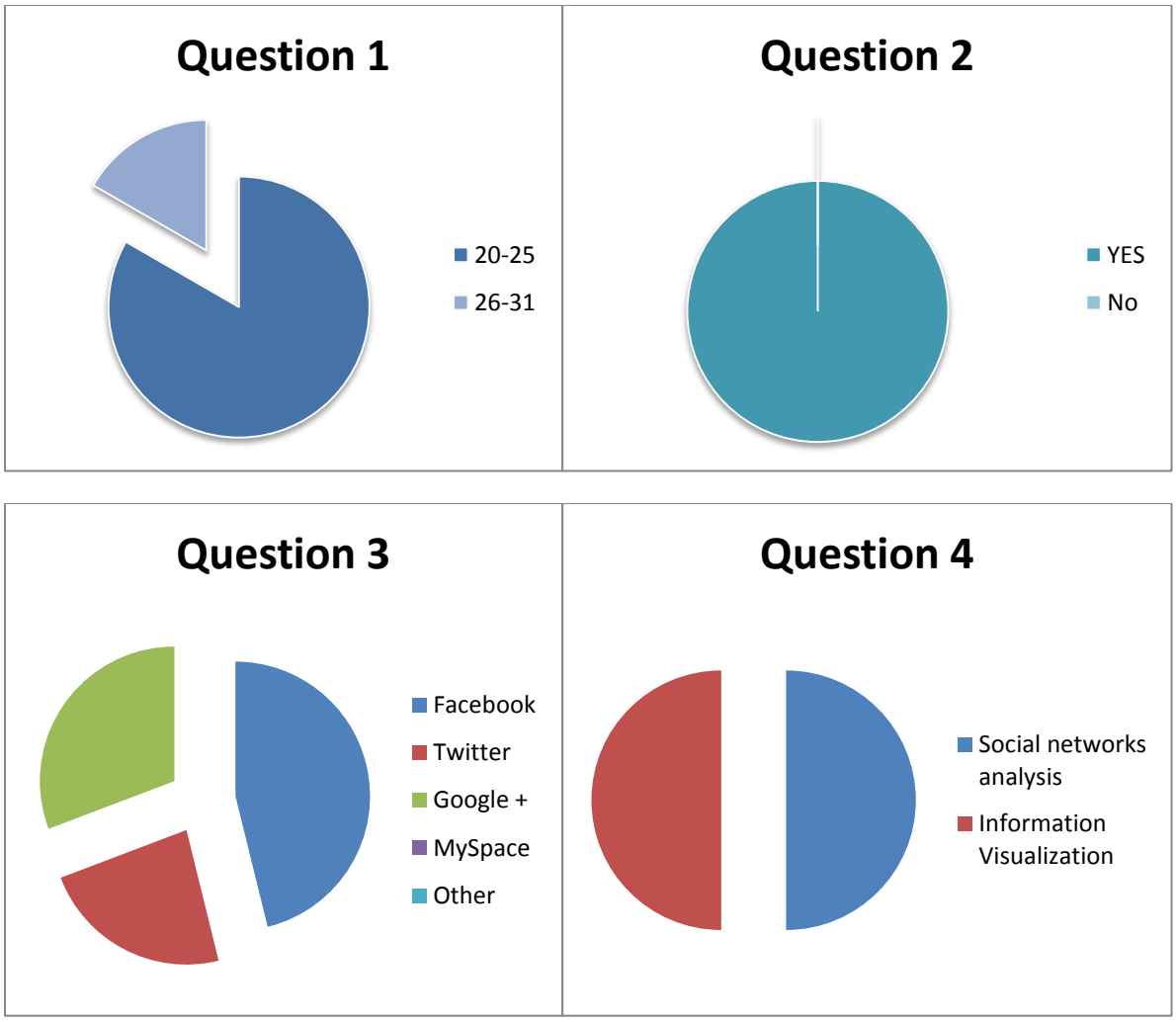
6.1 - Task list

1) Select a country.
2) Select a news channel.
3) Select the number of tweets you want to view.
4) Press submit to start the visualization.
5) Can you identify which are the user tweets and which are the news tweets on the page?
    a. Yes b. No
6) Press the name of any user from the left column on the page to view their location on the map.
7) Can you identify the location of the user on the map?
    a. Yes b. No
8) Click on the profile image of the user on the map to open an infowindow.
9) Can you clearly read what's inside the window you opened?
    a. Yes b. No
10) Click on a new tweet's name on the right column of the page to view where the event has occurred on the map.
    Locations from news tweets shown on the map are extracted from the text of the tweet
11) Does the location shown on the map correspond to a place mentioned in the text of the tweet?
    a. Yes b. No
12) Can you identify a link from the text of a tweet, shown from either the left or the right panel?
    a. Yes b. No
13) Click on any URL (link) to open a page.
14) Do you see link opening on the page?
    a. Yes b. No
15) From the news panel on the right, click on the "show tweets" button to view similar posts from users

Testing is done! Fill out the last questionnaire.

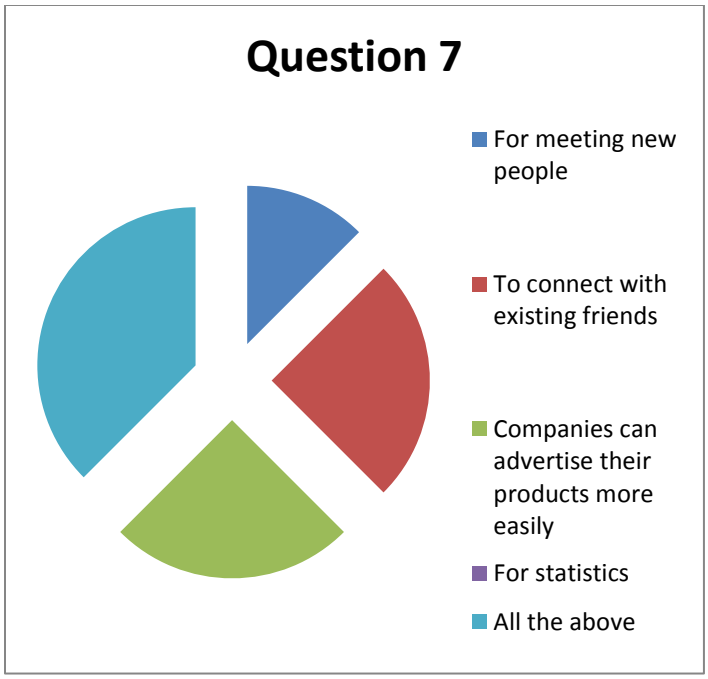
6.2 - Evaluation results

According to the answers of the questionnaires most of the users been tested where between 20 and 25 years old and most them where using Facebook and Google+ social networking sites. Although most of them who answered yes to if they are aware of the terms social networking, only two of them knew the term Information visualization and social network analysis.



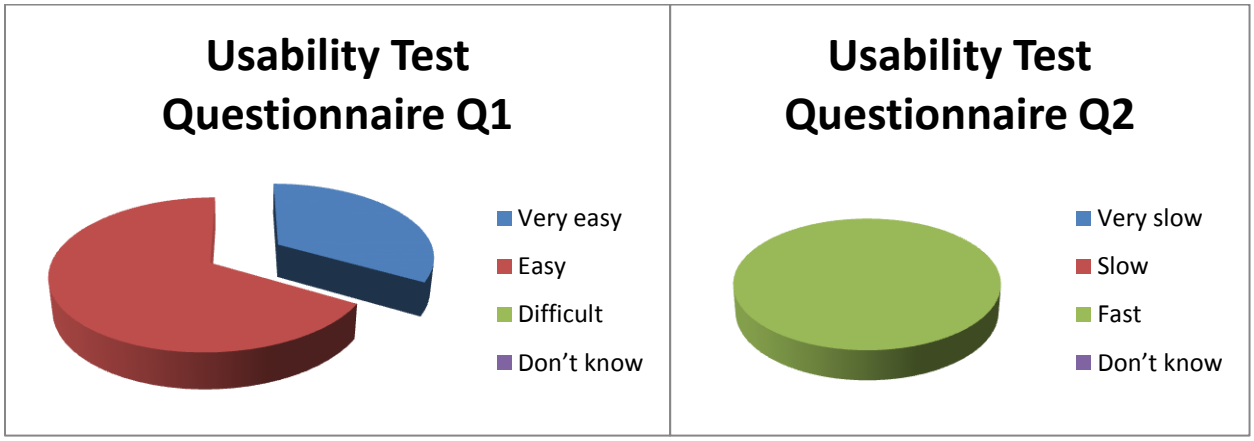
Additionally from the two users who answered yes for knowing the terms social network analysis and information visualization no one heard of any visualization tool from before.

On the last question almost everyone has given the answer b (To connect with existing people) but some other people have vote for more uses as well.

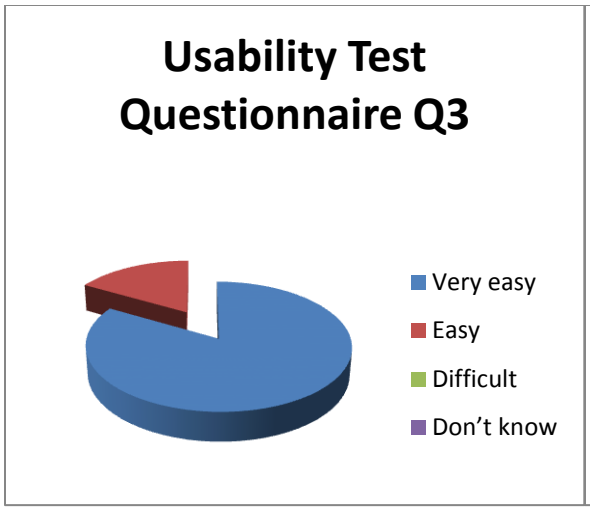


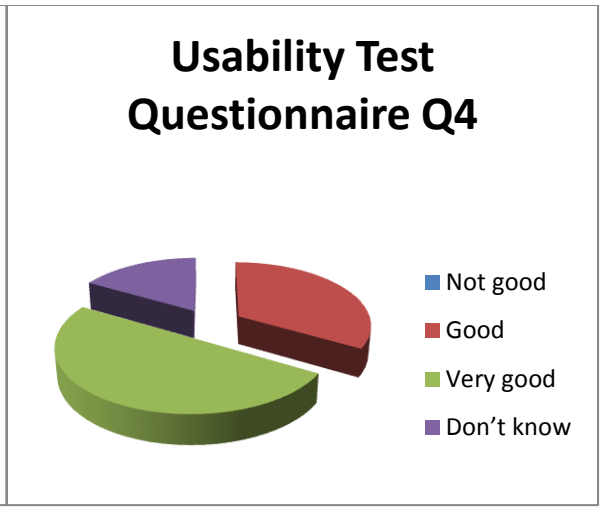
6.3 - Evaluation results from the Usability Test

From the first question of the usability's questionnaire ("Was news map visualization easy to use?") many of the users have answered easy and the rest very easy. Additionally on the second question everyone said that the application was fast.

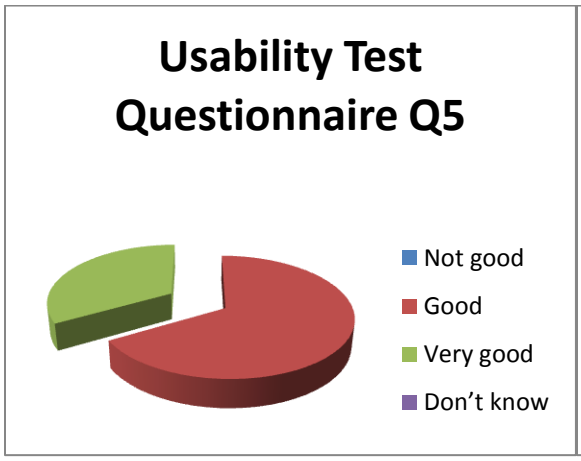


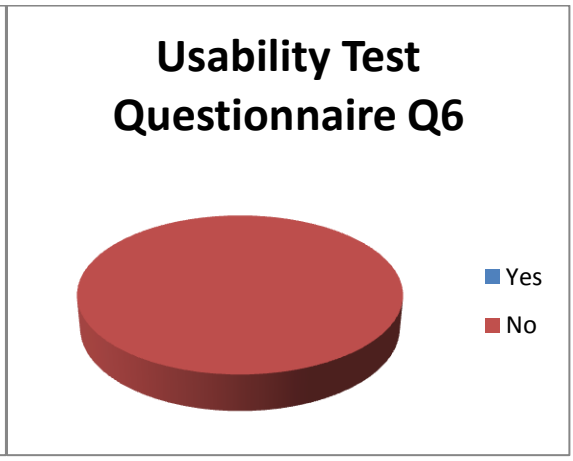
The next question of the questionnaire was asking whether finding and reading tweets was easy or difficult. From this question almost everyone has answered very easy and only one answered easy. On the fifth question many users have found "very good" the idea of customizing the interface of the system and differentiating the bad news from the good ones so that the application becomes more friendly and interactive.





Progressing to the next question of whether the colours and the graphics of the system's interface are good or not almost everyone answered they are "good". According to the answers given on the last question before users started filling their comments, no one has clicked on the "help" icon on the page, to get instructions to use the system hence this makes my application easy to use.





After the part of questions, users had the chance to state their problems, they have faced while using the system. Because most of the users I have conducted the testing with, where from Cyprus, most of them were selecting to view news from Greece but the application was systematically failing to show the correct location on the map, due to Yahoo's Placefinder inability of supporting the Greek language. Another problem that has been observed and mentioned in the questionnaire from users was the inadequate clue of indicating which link they had to open when they were asked in the questionnaire. One mentioned that "All links needed to be activated and opened on the same page" while trying to open a URL link from an "infoWindow" inside the map. Only the links from the two panels were working except from the map. Another person claimed for not specifying exactly which name he had to click on the tweet, while he was clicking on the different words with the symbol '@'.

Overall the system was easy enough to be used and very fast to perform the tasks. Everyone also found the idea of changing the interface of the system by showing negative news in different way, to be good.

In consequence this would enhance the interaction between the user and the system-interface. There were also some problems reported which had to do with not identifying the exact locations over the map since Yahoo placefinder service wasn't accurate enough with extracting the right places from the texts.

6.4 - Further improvements

Based on the evaluation results "news map visualization" needs to be replaced with a different geoparsing service while not all of the news tweets were plotted correctly on the map and channels from Greece were not working properly because of the service lack of language support. Further improvement to the system is the point which has been mentioned from before and the one requirement which hasn't been implemented on the application because of time constraint and some other problems. The point which I am referring to is the augmentation of the system for changing the interface and showing differently the negative news from the positive. This innovative change will definitely change the way users are using the application.

Another improvement that could be implemented over the application's features is to show the number of users which are talking for specific news, on a graphical chart. This will also give a boost to the system's interactivity by visualizing the same data to the users with a different way.

Chapter 7 - Conclusion

Finishing this project has given me a great opportunity for learning how to program with JavaScript language. Also be aware of the different services for finding and extracting locations from texts and parsing them to other mapping services and plotting them on a geographical map.

Though News map visualization might not be an ordinary site for others to use in their daily routine, it has been a big challenge for me to accomplish whereas at the start of the project I have no idea of the system's difficulty.

References

Case King,2011 Journal of Marketing Development & Competitiveness *Twitter Usage in the Fortune 50: A Marketing Opportunity?* [e-journal] 5 (3) Available through: EBSCOhost Midddlesex University http://web.ebscohost.com [Accessed 1/10/12]

David Astels,2003 *Test Driven Development: A Practical Guide* Prentice Hall PTR

Eden, Brad, 2005 Library Technology Reports *Information visualization*- [e-journal] 41(1) Available through: EBSCOhost Midddlesex University http://web.ebscohost.com [Accessed 1/10/12]

Grady Booch,2005 *The Unified Modeling Language User Guide* 2nd ed London: Addison-Wesley

Haewoon, lee & Housung, 2010 What is Twitter, a social network or a news media?.In: WWW '10 Proceedings of the 19th international conference on World wide web ACM New York, NY, USA ©2010

Hangzai Luo,2006. Exploring Large-Scale Video News via Interactive Visualization.In: *Visual Analytics Science And Technology, 2006 IEEE Symposium On.* Dept. of Comput. Sci., North Carolina Univ., Charlotte, NC. 12-14 April 2010

JQuery API(2012) *Documentation*[online] Available at: http://docs.jquery.com/Main_Page

F. Dianne,2010. Twitter in Government: Building Relationships One Tweet at a Time.In: *Information Technology: New Generations (ITNG), 2010 Seventh International Conference*

Mano Marks. (2010). *Geocoding Strategies.* [online] Available at: http://code.google.com/apis/maps/articles/geocodestrat.html#intro.

Ningjing Ma, 2012. Mapping News: The Service of Complementing Information Visualization based on The Interface of Google Map. *Software Engineering and Service Science (ICSESS), IEEE 3rd International Conference on* Beijing (2012)

Paul matton, 2004. Inferring and visualizing social networks on Internet relay chat.In: Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on. Comput. Lab., Kent Univ., Canterbury, UK July 2004

Takeshi Sakaki,2010. Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors.In: *WWW '10 Proceedings of the 19th international conference on World wide web,* ACM New York, NY, USA ©2010

Twitter Dev,2012. *Documentation* [online] Available at: https://dev.twitter.com/docs

Wikipedia, 2012 *Twitter usage*. [online] Available at: http://en.wikipedia.org/wiki/Twitter_usage

Wikipedia, *2012 Social network analysis.* [online] Available at: http://en.wikipedia.org/wiki/Social_network_analysis

Yahoo Dev(2012) *Documentation* [online] Available at: http://developer.yahoo.com/

Appendices

Appendix A

Questionnaire 1

Dear Participant,

Thank you for coming. The application you are about to use deal's with visualizing user and news tweets returned from twitters large database and plotted on a geographical map. The ideal reason behind this testing is to check the system's efficiency for showing locations of users posting their tweets and also from where news events have occurred, over the map.

The information given by you will be used for study purpose only. So please fill the Personal Information form below.

Name: .....

E-Mail: .....

Program Name: .....

Year of Study: .....

Subjects of study/ Job function: .....

Educational Level: .....

Computer Literacy: .....

Microsoft Windows Experience: .....

Number of systems/ Applications used: .....

Number of electronic support tools (email, Internet, Pc etc) used: .....

Questions

7. How old are you?

   15-19 20-25 26-31 32+

8. Have you heard of the terms social networking and online networks?

a. Yes b. No

9. Which social networks sites you are using?
   a. Facebook b. Twitter c. MySpace d. Google+ e. other

10. Have you heard of the terms Social Network Analysis" and "Information Visualization"?
    a. Yes b. No

11. Have you heard of any visualization tools?

    ……..

12. Can you name one visualization tool that you have used?

    ……..

13. What is the main purpose of using Social networks?
    a. For meeting new people.
    b. To connect with existing friends.
    c. Companies can advertise their products more easily.
    d. For statistics.
    e. All the above.

Questionnaire 2

USERS USABILITY TEST

Dear Participant,

I would like to take your valuable feedback after you have used the application “news map visualization” and for evaluating whether the system is functioning properly.

Please if you could answer all the following questions.

1) Is “news map visualization” easy to navigate?

   Very Easy Easy Difficult Don’t know

2) What do you think about the performance of the application?

   Very slow Slow Fast Don’t know

3) Finding and reading the tweets

   Very Easy Easy Difficult Don’t know

4) Do you think the idea of showing good and bad news on the site is

   Not good Good Very Good Don’t know

5) The colors and graphics of the site are

   Not good Good Very good Don’t know

6) Have you used the help for instructions how to use the application?

   a. Yes b. No

7) What problems you faced while using this website?

   .....

   .....

   .....

.....

.....

.....

8) How this application can be improved (Suggestions)?

.....

.....

.....

.....

.....

.....

9) In Your opinion which kind of information is missing?

.....

.....

.....

10) Any other comments

.....

.....

.....

Thank you.

Sincerely,
Angelos Cleanthous

Appendix B

Turnitin Receipt