

**Middlesex  
University**

# VisualThoughts: Visualizing online social information

---

**CMT4141**

**Constantia Chairepeti**

**M00328813**

Supervised By: Dr. Kai Xu

***Abstract***

“Social network analysis” and “Information Visualization” are two upcoming and highly coming terms which can be used in many fields of science, to help scientists in gaining a better understanding of complex situations. In this paper, it is described how a visualization tool named VisualThoughts is designed and implemented but also in which means it can be used. VisualThoughts application enables the user to visualize Tweets related to a topic defined by the user and interact with the created visualized graphs.

## Table of Contents

Chapter 1: Introduction .....	5
Chapter 2: Background .....	7
Information Visualization.....	7
Social Network Analysis .....	8
Twitter.....	9
Technologies Used .....	10
Related Work .....	11
Chapter 3: Requirements Specification .....	13
VisualThoughts System Requirements .....	13
Use Case.....	14
Chapter 4: Design.....	15
System Architecture.....	15
VisualThoughts System .....	15
UML Class Diagram .....	16
Data Flow Diagram.....	16
Sequence flow diagram.....	18
Software/Hardware Decisions .....	18
Intended Functionality/User Interface Design.....	19
Chapter 5: Implementation & Testing .....	21
System Process Description .....	21
Retrieving Popular Search.....	21
Setting Popular Search.....	21
Generating a Graph.....	22
Display Tweets for a single month and day .....	22
Click on the Graph.....	23
System Testing .....	24
Testing Strategy .....	24
Testing Plan.....	24
System Test Cases .....	25
Testing Implementation.....	27
Chapter 6: Demonstration & Evaluation.....	28

System Walkthrough.....	28
Request a Graph.....	28
Display selected graph .....	29
Interacting with Graphs .....	30
Retrieving/Setting Popular Search .....	32
System Evaluation.....	34
Chapter 7: Critical Evaluation .....	37
Review of Aims.....	37
Revisions to Design and Implementation .....	37
Lessons Learnt.....	37
Chapter 8: Conclusion .....	38
Further Work/Research .....	38
Concluding Points .....	38
References .....	39
Appendices.....	42
Appendix A – Use Cases .....	42
Appendix B - Questionnaire .....	48

## Chapter 1: Introduction

---

The rapid growth of the Internet during the last few decades enables people to communicate with other people and share their opinions, knowledge or social networks. Thus, our experiences and as a result, our lives become digital, online social networks turn out to be too crowded, and enormous new data collections are becoming available. As a consequence, while hardware capabilities make analysis of large data sets feasible, more people get attracted and interested in social network analysis and visualization. Social network analysis and visualization focus on studying, measurement and mapping large non numerical sets of data gathered from social networks into complex graph-based structures.

Twitter is a combination of an online social networking service as well as a micro-blogging service growing extremely fast the last couple of years.<sup>[1][2][3]</sup> Thousands of people use Twitter on a daily basis to share opinions and posts among their friends. But have you ever wondered how many people use Twitter to share an opinion or make a statement about a critical subject and from where do these people post? VisualThoughts application might be the key to what you look for, a visualization tool written in Java programming language. The main aim of this project is to visualize Tweets in proportion of two factors: the geographic location and time.

VisualThoughts visualization tool gives the user the ability to visualize Tweets by creating its desired graph which may be either WorldMap graph, a representation of the geographic locations of the Twitter members using the OpenStreetMaps maps, or RiverTheme graph, a representation of the Tweets in a fraction of time, or both. The user, in order to create the visualization, will need to enter into the system the search keywords. Search keywords may consist of one or more words, and can be either typed in by the user, or selected from a 10 items list that represents the 10 most popular search keywords. VisualThoughts, in order to access and visualize all this Tweet information, is connected to a database where search keywords are sent to, so all the information regarding the relevant Tweets is collected.

VisualThoughts besides visualizing Tweets, it enables the user to interact with the graphs and the information that is displayed. The initial visualization that is created represents all the Tweets around the year that include the search keywords. However, the user is capable of selecting a single month and day of the year modifying in this way the graph to display only the Tweets for that single month/date. Moreover, both graphs are clickable so when the user clicks on any of them, all the Tweets corresponding to that region and their information are displayed. Last but not least, if the displayed graph is WorldMap, easy to use features including zoom and slider options are enabled.

As it has been mentioned before, VisualThoughts visualization tool is an application written in Java programming language using a combination of Java libraries. Java programming language is a high level, object oriented language that is designed to make programming simple and easy. Programming in Java language also allows maximum portability with minimum dependencies during implementation.<sup>[5][6][7]</sup> Java Libraries are a fundamental component of Java programming. They are enormous collections of code that provide a wide range of API, make programming even easier but also enable the prevention of re-creating commonly used algorithms. SwingX-WS, which enables programmers to interact with video,

images, maps etc.,<sup>[11]</sup> and Lucene Apache, a high-performance, which provides users with a text search engine<sup>[8]</sup>, are the two Java libraries used for the implementation of VisualThoughts application. Besides Java programming language, Processing programming environment is also used for the implementation of VisualThoughts. Processing is an open source development environment for programmers who want to create animated and interactive applications.<sup>[9]</sup>

People often think ‘too big’, when it comes to their contribution. However, an MSc project contribution is nothing else but only a tiny little step in the right direction.<sup>[4]</sup> Visualizing information gathered from a social network is not revolutionizing computing, but still using known resource to develop an application may have a small but sensible contribution to research. Information Visualization allows users to underlie, analyze and understand data quickly thus motivates human cognition.<sup>[12][21]</sup> Besides that, social network analysis and visualization can be used in many fields of science such as psychology, epidemiology, politics, anthropology etc and can help scientists as well as non-scientists users to gain a better understanding of complex situations by exploring and analyzing useful information.<sup>[13][20]</sup> Scientists, through social network analysis, have the chance to find solutions by observing users’ reaction on crucial issues, predict their interests, explore how easily they get influenced, understand how human communication support or prevent the spread of diseases etc. Research in a number of science areas has shown that social network analysis and visualization plays a significant role in solving key problems, determining up to what degree individuals are achieving their goals and how organizations run.<sup>[21]</sup>

When developing a system, there are three main principles to take into account including respect for people, beneficence and justice. In this project, there is nothing that can offend people either as individuals or community. Moreover, although the information gathered from the social networks deals with users’ personal data, the system is entirely authorized to.

In the Chapters following, a detailed analysis will be given for the VisualThoughts application. Some background research as well as reading analysis will be covered in the second chapter while, in the third one, the requirements of the system will be specified. In chapters number four, the design decisions and justifications are fully described. Moreover, an overview of the implementation and the testing of the application can be found in chapter five. Finally, in chapters six and seven, the system but also a critical evaluation of can be found.

Before closing up, there are some people that need to be mentioned, as they had offered a lot in developing the software and writing the report. These people are Dr Kai Xu, a brilliant man and even better professor, as well as my friends and family that truly supported my throughout the project.

## Chapter 2: Background

### INFORMATION VISUALIZATION

Information Visualization is the interdisciplinary study which focuses on how enormous collections of non numerical information are visually represented.<sup>[21][22]</sup> The study field of Information Visualization was born after some research made in HCI, psychology, computer science, business methods, graphics and visual design. However, Information Visualization, which has been used for studying and solving scientific problems, started to becoming well known during the first years of computer graphics. Nowadays Information Visualization can be found in digital libraries analysis, financial analysis, drug discovery, data mining, market studies, network of relation on the internet etc. with social network analysis and representation being one of the most active areas of research.

Carson Kai-Sang Leung et al,<sup>[23]</sup> proposed a visualizer known as SocialViz(Figure 1) which can function as a standalone tool or as a complement one form exploring social networks. The main aim of SocialViz is to provide users with frequency information on social relationships between multiple entities in the network. The visualizer is connected to a frequent pattern mining engine like Apriori Algorithm of FP-Growth Algorithm. SocialViz visualizes the mined patterns in a 2-dimensional space where X-axis represents the social entities and Y-axis their frequencies.

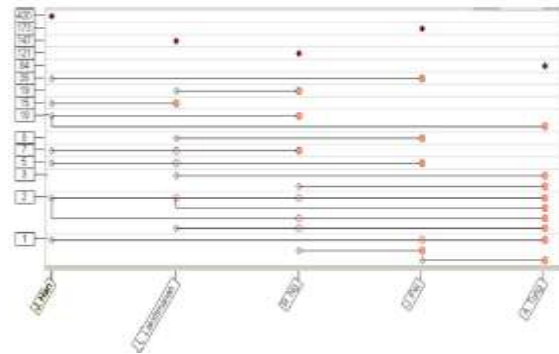


Fig. 1 – SocialViz visualizer

One significant feature of Information Visualization is that is not all about drawing a picture and then just watching it. Information Visualization is also about how people interact with the data included in the visualization and the entire model. Interacting with an information visualization application allows conclusions to be done more easily than it used to be. Moreover, allows users to underlie, analyze and understand data quickly thus motivates human cognitions and helps scientists as well as non-scientists users to gain insights about useful information gathered from complex systems.<sup>[13]</sup> However, Information Visualization research mostly focuses on using techniques for one isolated data type, while there are applications that have to deal with massive collections of unstructured and abstract data of different data types. Thus, a lot of innovative visualization techniques are used in some application domains.

The process of analyzing enormous collection of data and the visualizing information takes a lot of effort and consists of four different stages and are illustrated in Figure 2.<sup>[24]</sup> The four stages are:

- Collection and storage of data
- Process the collected data and transform them into information that we can understand
- Create the display hardware and the software algorithms that produce the visualization

- Create the human perceptual and cognitive system

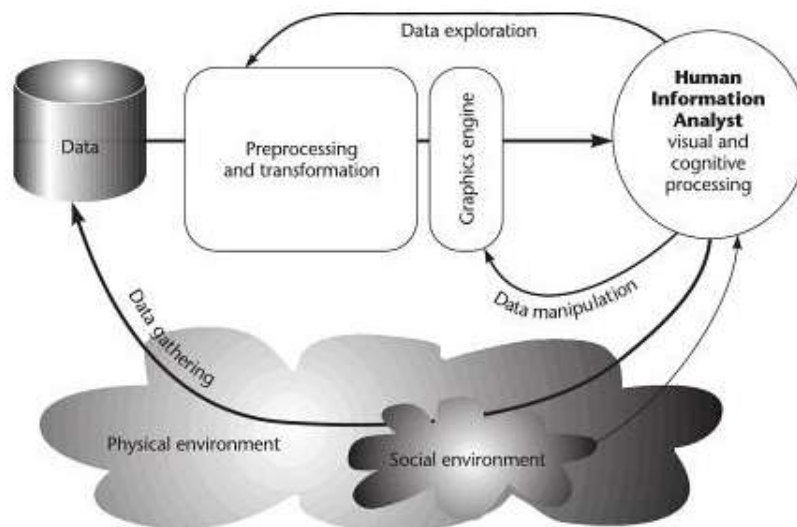


Fig. 2 – Information Visualization Process

The longest stage involves data gathering and storing as the more data you gather, the most fascinating the visualization will be. The next stage controls entire processing that takes place so that the collected data are transformed information before their visualization. The expert analyzes the gathered data in an effort to give data meanings that are understandable to the users. The third stage is all about representing information in an image – visualization using the appropriate hardware and software equipment. Finally, the expert combines the visualization with some interactive features creating in this way an understandable visualization to both experts and non-experts.

## SOCIAL NETWORK ANALYSIS

Social networking, as well as social network analysis, has attracted people's curiosity and interest in the last few decades. A social network is a social complex consisted of nodes, which might represent individuals or organizations, connected by at least one types of interrelationship. In simpler words, social networking has to do with grouping individuals into communities depending on their connection such as friendship, relationship, dislike etc.<sup>[18][19][20]</sup> Social networks exist in workplaces, universities, schools, but the most popular are online social networks.

Online social networking can be found in some websites like Facebook, MySpace, Twitter and Classmates. Being a member of an online social network gives you the opportunity of meeting new people creating in this way a relation between them. The set of these relations is inestimable as the "relation" word may have different meanings. As stated above, some meanings of the term relation may be kinship, acquaintance and evaluation of another person, physical connection, relationship of beliefs and opinions, or even just the existence of a website/profile link in a website.



Being a member of a social network means nothing until you start connecting with other members and creating relations; however, individuals and their attributes are not the subjects under examination. Conversely, the connections between these individuals and their structure are objects under observation in such a way that analysis of social processes as a result of the connections between individuals is permitted.

Social network analysis studies, measures and maps into complex graph-based structures, which consist of nodes and links, the social relationships and flows in terms of network theory.<sup>[20][25]</sup> The nodes in the network are the individual entities like members, groups, and organizations while the links in the network are the connections or flows between the nodes. Social network analysis studies two kinds of variables: Structural, which represent the different types of connection between the nodes, and composition, which represent the nodes' attributes.

Paul Mutton,<sup>[27]</sup> describes a way which implies that a group of IRC users in a channels is also a social network (Figure 3). For this aim, an IRC bot is used to control the channel and at the same time to perform a heuristic analysis of the events taking part. A mathematical approximation of the social network is created using the analyzed data, and a visualized graph of the inferred social network is produced. This visualization represents the structure of the network over time as well as the connectivity and strengths of relationships between users.



Fig3. A Shakespearean social network in IRC

Social network analysis has been used in many fields of science such as psychology, epidemiology, politics, anthropology etc. Scientists, using the social network analysis, have the chance to find solutions by observing users' reaction on critical issues, predict their interests, explore how easily they get influenced, understand how human communication support or prevent the spread of diseases etc. Research in a number of science areas has shown that social network analysis plays a significant role in solving crucial problems, determining up to what degree individuals are achieving their goals and how organizations run.

TWITTER

Twitter is a mixture of an online social networking service, messaging and a micro-blogging service that gives the user the ability to send text based posts less than 140 characters and images or access and read other users' posts called Tweets. Users can send or read Tweets via SMS, instant messaging or email to the Twitter website, or via an application like "Twit This".<sup>[1][2]</sup>

The whole idea of Twitter created during a brainstorming session held by some members of Odeo Podcasting Company. A chief executive officer, Dorsey, suggested the idea of using an SMS service in order to communicate within a small group. The service was firstly given the name “twtrr”, which is

inspired by Flickr –an image and video hosting online community- and the American 5 characters length SMS short codes. “10958” was initially considered by the developers as a short code. However, “40404” took its place as, it was much easier to be used and remembered than the previous one. Dorsey was the first who published the first Tweet, “just setting up my twttr”, on 21st of March, 2006. However, the first prototype was used only by the Odeo employees and was not introduced into publicity until 15th of July, 2006.

Twitter online service experienced rapid growth since its first year. In 2007, 400,000 tweets were posted per quarter with these growing up to 100 million tweets per quarter in 2008. In early 2010, more than 50 million tweets per day were posted with the company recording more than 70,000 registered applications. The focal point for this rapid growth was the “South by Southwest” festival in 2007 where two 60-inches plasma screens displaying Tweets were placed in the hallways of the festival. As a result to this, Twitter popularity was increased from 20,000 to 60,000 tweets per day.

As mentioned before, users can send their posts via the Twitter website, SMS if it is applicable in your country of residence, or via external applications like mobile phones. Tweets posted are visible to anyone, even to visitors that are not Twitter members, unless users select to deliver their post just between their followers. Twitter as a free online social network that gives everyone the right to join it. In the last few years, it was noticed that newly registered members create fake profiles and pretend to be a celebrity. To prevent this, Twitter launched a verification feature in June 2008 that allows celebrities to verify the originality of their accounts. In the following years, this verification feature was also been used by businesses or public figures who may not tweet but wish to preserve the entire control and rights over the account that represents their name.

Twitter besides being an online social network service can be also used as a marketing tool in some cases. Politicians, businesses, and celebrities are using Twitter to keep their fans updated on the latest news either by hiring an outside writer either in their own voices. In this way, all these public figures and businesses give people the ability to approach them as well as to speak to them as comfortably as like walking next to them.

## TECHNOLOGIES USED

Visualization tools can be built in any existing programming language that supports graphics like Java, Lingo, and ActionScript etc. However, which is the most appropriate one for visualizing information? Factors like functionality, performance and quality can be taken into account before selecting a programming language. By the first point of view, Java programming language seems to be the right one for developing a visualization tool like VisualThoughts as it predominates over the other ones.

First of all, it is particularly essential for a graphics application to execute fast. Programs written in Java programming language do execute much faster than other programs written in any other programming languages. Besides this, Java programming language is designed to allow maximum portability with minimum dependencies during implementation as possible as well as being simple to be used by many programmers.<sup>[5][6][7]</sup> This can be achieved by the use of Java libraries; enormous collections of code that

enables the ease of use for generality and flexibility, as well as the prevention of re-creating commonly used algorithms.<sup>[8]</sup>

However, Java programming language is not that strong as being combined with the Processing programming language; a programming environment created in Java for developing graphics application.<sup>[9]</sup> Processing enables developers to build shorter and easier to read programs although their functionality remains the same as if the program would be written in any other language. Moreover, any application written in Processing are translated into Java application while compilation; thus they usually perform faster than applications written in scripting languages.<sup>[10]</sup>

## RELATED WORK

During the last few years, as our lives and experiences become more digital, enormous new data sets are becoming available, and while hardware capabilities makes the analysis of large data collections feasible, the need to visualize these large data collections is growing. The existence of the Internet allows people to share their social networks or to communicate with other people for all over the world using email or instant message.

Social Network Fragments (SNF) and PostHistory are two visualization tools that give users the right to access high-level data of someone's email habits.<sup>[28]</sup> Both of these visualizations represent emails' data in two dimensions, people and time. However, SNF and PostHistory differ in some ways. SNF focuses on grouping people that come out from a person's email exchanges. On the other hand, PostHistory represents all the dyadic email relationships by highlighting stimulating changes in the interaction between the person and its contacts over time.

Social network analysis is also a growing area during the last few years as even more people than before prefer to control their social lives electronically.

NodeTrix is a network visualization based on the nodelink diagram where matrices are used to represent communities (Figure 5).<sup>[29]</sup> NodeTrix representation consists of two networks: the raw underlying network which acts as the opening input and the aggregated network which is the result of a raw underlying network. The nodes of the aggregated network can communicate with a unique underlying node or to a community of underlying nodes. On the other hand, underlying nodes cannot be shared or connected by aggregated nodes. Users of NodeTrix have the option to add or merge aggregated nodes.

Network representations, in the last few years, consist of dense graphs because of the enormous data collections that are gathered. For this reason, NodeTrix aggregates dense diagrams and displays them as

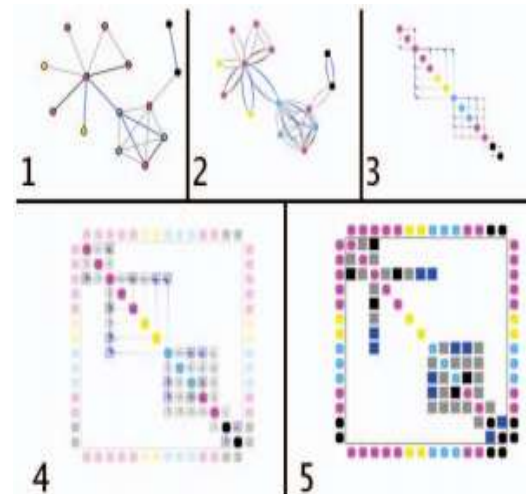


Fig. 5 – NodeTrix matrix

matrices whereas a single connection between aggregated nodes and underlying nodes are displayed as a single node.

Vizster Visualization tool (Figure 6) provides a representation of Friendster social network that enables users to explore their expressed network in a playful manner.<sup>[30]</sup> Vizster acts as a network by itself as it represents an individual and all the relationships with its direct friends by using node-link

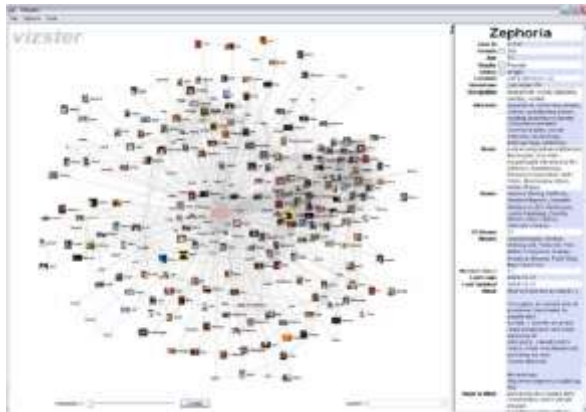


Fig. 6 – Vizster Visualization tool

representations to present members of the system (nodes) and their friendship links (links). Members of the social network are presented by their name and profile picture. Users can select a node, so all the personal information of the selected individual is shown at the right hand side. Moreover, the display is expanded, and all the friends of the selected individual are visible to the selected user. In this way, Vizster engages a fun online space with the ability to discover new people and connections.

## *Chapter 3: Requirements Specification*

---

In this Chapter, all the necessary requirements for the development of VisualThoughts system will be specified and described in as much detail as possible. Moreover, Use Case diagrams are going to be designed forming the basis of future testing plans.

### **VISUALTHOUGHTS SYSTEM REQUIREMENTS**

During the last decade, online social network and especially Twitter face a rapid growth with an enormous amount of information being available for analysis and visualization. VisualThoughts project is an attempt to build a visualization tool which main aim is to visualize Tweets in proportion of two factors: the geographic location and time.

VisualThoughts application must be able to allow user creating graphs that visualize information gathered from Twitter social network. With the intention of creating Tweets visualization, the user must input to the system search keywords that will represent the range of Tweets being visualized. Inputting the data to the system can be done in two different ways; either by typing in the keywords through a text field or by selecting keywords from the “Popular Search” representing the 10 most popular keywords used for visualizing Tweets. As soon as the user requests the graphs to be created, the system must check the case that no data were input. In this case, the system must display an error message to the user informing him to input the necessary data.

Moreover, before creating the graphs, the user should be able to select between two types of graph, WorldMap or RiverTheme, which one desires to be displayed. WorldMap graph must be able to represent Tweets’ information in proportion of the geographic location at where the Tweet was posted. Each Tweet displayed on the graph, must be represented by a random word included in the Tweet’s text. On the other hand, RiverTheme graph must be able to represent Tweets’ information in proportion of time and quantity.

What's more, the user should also be able to interact with the displayed graph/graphs. In more details, the system should enable the user to modify the range of Tweets visualized by selecting a particular a date. The system after gathering the selected date should modify the graphs, so only the Tweets posted during that date are visualized. In addition to this, the user should be able to click on any graph in order to gain more insight about the Tweets visualized like the entire Tweet’s text, the Username and Profile Picture of the Tweet’s owner as well as the date and time when the Tweet was posted.

To sum, the requirements specification of a project plays a vital role in its development. In a point of fact, the success of building a visualization tools depends on whether the developing project meets 100% the requirements being specified.

## USE CASE

A use case diagram is a behavioral diagram that describes the interaction between a user and a system.<sup>[31][35]</sup> The main purpose of a UCD is to show the functions are performed by the system and for which actor. The term actor represents the user or even another system that interacts with the modeled system. A UCD is a view of how a system performs some actions in order to complete a task and can be used for revealing requirements and planning the project.

In figure 10 displayed below, a Use Case Diagram representing the entire VisualThoughts system can be found.

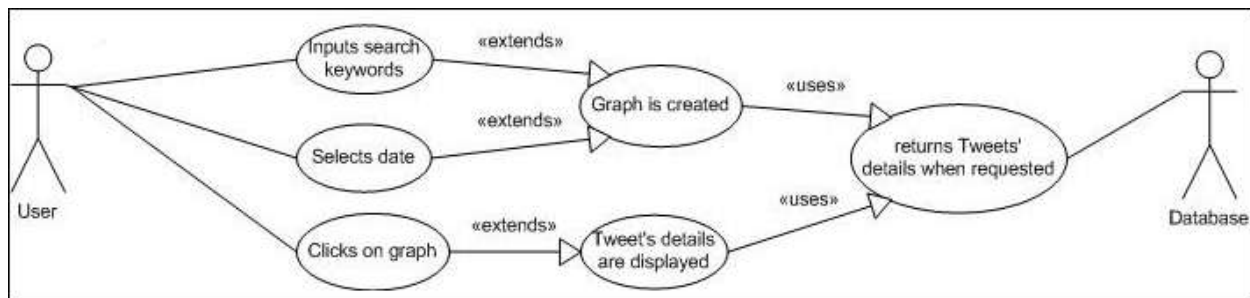


Fig. 10 –Use Case Diagram of VisualThoughts System

Besides the system's Use Case diagram, the following use cases have been also drawn up for the new system forming the basis of testing plans as the software is going to be tested according to their saying. The Use Case diagrams stated below are described in Appendix A section.

- (1) Request for a graph with no fields completed.
- (2) Request for a graph with the fields filled.
- (3) Try to click on the graph in a section that no Tweets are represented.
- (4) Try to click on the graph in a section that Tweets are represented.
- (5) Request for a graph that represents on Tweets posted in a single month.
- (6) Request for a graph that represents on Tweets posted in a single month and day.

## Chapter 4: Design

---

In the following pages, a possible design of the VisualThoughts visualization tool is discussed. The System Architecture and the layout of the User Interface will be designed as well as some software and hardware decisions will be taken. Besides that, data and component modeling will be examined using Unified Modeling Language (UML) techniques like Data Flow Diagram (DFD), Class Diagram, Use Cases etc.

Modeling is an essential part when developing large software projects as it deals with designing the software before coding. Modeling a large software project helps people, who are responsible for success of the project, to ensure themselves that the user needs and requirements are met but also that project's functionality is correct and complete.

Unified Modeling Language (UML) is a graphical language that helps software engineers in specifying, visualizing, and documenting a software system, including the system's structure and behavior.<sup>[35][36]</sup>

Data Flow Diagram, Sequence Flow Diagram and Class Diagram are some of the diagrams that are going to be used for the design of VisualThoughts visualization tool.

### SYSTEM ARCHITECTURE

VisualThoughts system is a visualization tool that enables the user to visualize Tweets by creating its desired graph which may be either WorldMap graph, a representation of the geographic locations of the Twitter members, or RiverTheme graph, a representation of the Tweets in proportion of time, or both. The user inserts into the system some search keywords in order to create the visualization.

VisualThoughts, in order to access and visualize all this Tweet information, is connected to a database where search keywords are sent to so all the information regarding the relevant Tweets are collected. VisualThoughts besides visualizing Tweets, it gives the user the ability to interact with the graphs and the information that is displayed.

The basic structure of the system consists of an application and a database that communicates with the application. In point of fact, the application does not just communicate with the database; however, the entire system depends on it as it keeps all the information of the Tweets. When a user runs the application and requests the graphs to be displayed, the system connects with the database and gathers all the information needed for visualizing the Tweets' information into graphs.

### VISUALTHOUGHTS SYSTEM

VisualThoughts system is managed by the user through the User Interface. However, data parsed by the user through the User Interface, are processed by the back-end of the application into operations. These operations are to be passed around the system, and each one performs a specific function. Functions in

some cases gather information from the database and display them to the User Interface, while in other cases store the parse data into the database depending on the function's needs.

## UML CLASS DIAGRAM

A class diagram is a static, structured diagram that describes the contents of a system, their types as well as their relationships<sup>[32][35]</sup> Class diagrams uses design elements like packages, classes and objects to describe the structure and the contents of the system. Class diagrams are used to identify both implementation and conceptual details of the system; thus, it helps developers in building other UML diagrams such as sequence diagrams.

The UML Class diagram for VisualThoughts application, which includes all the classes and objects of the system, can be found in figure 7.

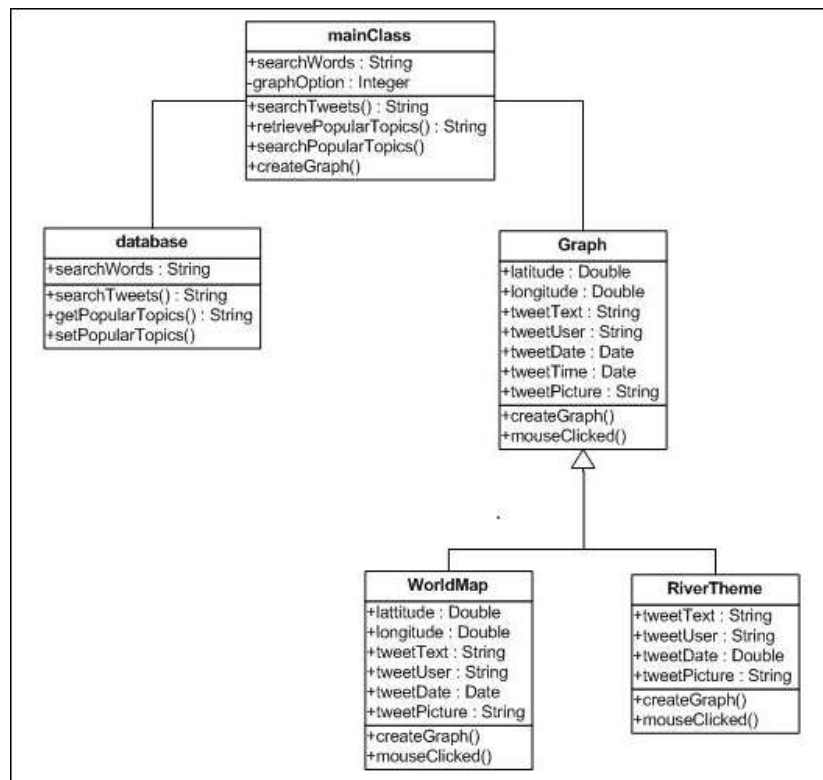


Fig. 7 – Class Diagram of VisualThoughts System

## DATA FLOW DIAGRAM

A data flow diagram (DFD) is a visual representation of how the data "flow" through a software system and its process attitudes.<sup>[33][35]</sup> DFD are used to create an overview of the system which can later be expanded; what types of data are input and output, how they are processed and where are they stored.



The movement of data in the system can be separated into the information input to the system by the user and used for creating the visualization, as well as the information exchanged between the User Interface, the back end and the database.

The flow of data in the VisualThoughts (figure 8) system is mostly cyclic. When a user starts using the application, he inputs data in the system like word keywords for creating the Visualization. The system then processes this data and translates them into a database query for gathering information from the database.

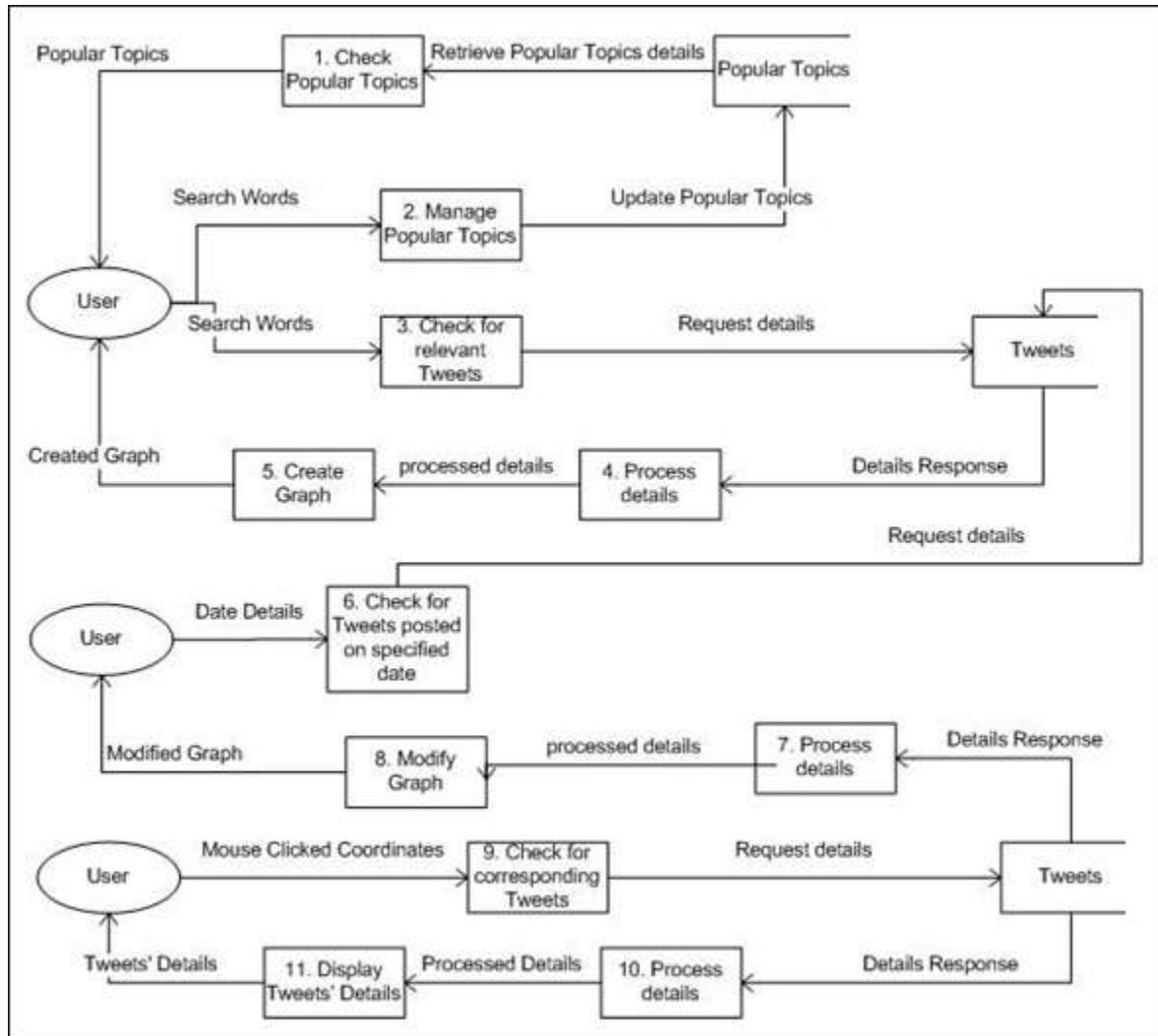


Fig. 8 – Data Flow Diagram of VisualThoughts system

Once the visualization is created, the user can interact with the graphs created by passing some data in the system such as mouse-clicked data, selection data, slider-changed data etc. Depending on the user's action, the system collects the data inserted by the user, processes and transforms them into information that can be displayed to the User Interface. For example, the user can modify the created graph to represent Tweets for a single month or date by using a slider. The system gathers the data passed from the user through the slider and uses them to get the information needed for the graphs' modification.

## SEQUENCE FLOW DIAGRAM

Sequence flow diagram is a visual representation or modeling of how actions, messages and events interact with each other in a sequential order that these interactions take place. <sup>[34][35]</sup> Sequence flow diagrams can be used to help developers of software applications, as well as an organization's business staff by showing how different objects of the business interact.

All the events, actions and messages which are occurred while the VisualThoughts application is running as well as the interaction between them are represented in figure 9. The sequence of the interactions is represented by the time in the vertical direction, while header elements are represented in the horizontal direction at the top of the diagram.

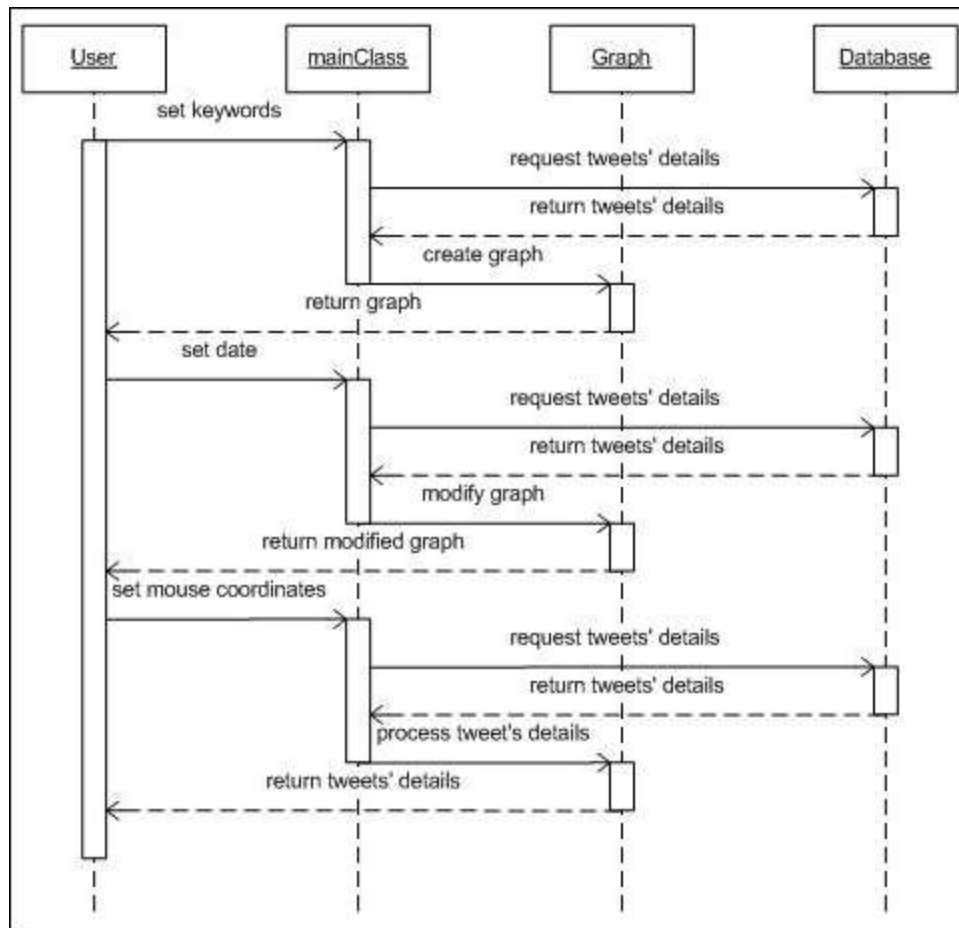


Fig. 9 – Sequence Diagram of VisualThoughts System

## SOFTWARE/HARDWARE DECISIONS

As mentioned in a previous chapter, visualization tools like can be built in any programming language. VisualThoughts application is going to be built using a combination of programming languages and their components. Java programming language along with some Java libraries is the main programming language used during the development of VisualThoughts application. Besides Java programming language, Processing programming environment is also used to support the creation of the graphics.

**Java programming language** is a high level, general purpose and object oriented language that is designed to be simple to use by many programmers. The Java language is also designed to allow maximum portability with minimum dependencies during implementation as possible. Java was first known as a programming tool for creating applets for the World Wide Web, which are mini applications that run inside a web page. Safety features and ease of programming give the ability to programmers to produce quickly working applications while maximum portability gives users the right to run and share Java applications anywhere within a Java runtime environment.<sup>[5][6]</sup>

During programming courses, in an attempt to minimize complexity programmers are taught a subset of the programming language and its libraries. As a result, programmers learn a new way of thinking and gain new skills. **Java Libraries** are a fundamental component of Java programming.<sup>[7]</sup> They are enormous collections of code that provide a wide range of API such as graphics, I/O, data structures etc. The use of Java libraries enables the ease of learning, the ease of use for generality and flexibility as well as the prevention of re-creating commonly used algorithms.

SwingX-WS is a Java Library example designed having in mind graphical features like video, image, maps etc.<sup>[11]</sup> SwingX-WS enables programmers to interact with web services and embed into a Java application the services. JXMapView is a SwingX-WS component used for accessing and displaying satellite data from OpenStreetMaps, as well as interacting with the data with some easy to use features including zoom and slider options.

**Processing** is an open source programming environment written in Java Programming Language for creating animations, images, visualizations, and interactions.<sup>[9]</sup> Processing was firstly developed to be a software sketchbook and help users to learn the fundamentals features of programming applications within a visual context. However, popularity of Processing grown remarkably quickly, evolving it into a tool used into finished works even by professionals. Besides professionals, there thousands of artists, researchers, students and designers who use Processing environment for prototyping and producing finished installations.

As mention before, the Processing programming environment, is created in Java, so applications written in Processing are translated into Java while compilation, and then executed as Java applications. Thus, applications written in Processing, usually execute faster than applications written in other languages like Flash which is extremely necessary for graphics applications. However, programming using Processing environment is much simplified rather than Java as users are not required to understand advanced concepts of Java programming like objects, classes as they are already integrated into Processing. Moreover, developing programs using Processing, results in making programs shorter and easier to read, although their functionality remains the same as if the program would be written in Java.

## INTENDED FUNCTIONALITY/USER INTERFACE DESIGN

In the next few paragraphs, all the functions been provided to users through the User Interface (Figure 11) will be discussed.

The diagram illustrates the VisualThoughts User Interface, which is organized into three main sections:

- Control Panel:** Located in the top right, it contains an "Input Search" text field, an "Or" label, and a "Select Popular Topic" section with a list of radio buttons for "topic 1", "topic 2", "topic 3", and "...". Below this is a "Get Graph" button. At the bottom of the panel are two sliders labeled "Select Month:" and "Select Day:".
- Graphs:** On the left side, there are two large rectangular areas labeled "WorldMap Graph" (top) and "RiverTheme Graph" (bottom).
- Tweet Text List:** A large rectangular area at the bottom left, labeled "Tweet Text List", intended for displaying tweet details.

Fig. 11 – VisualThoughts User Interface

Through the main screen of the VisualThoughts system, which is displayed in Figure 12, the user is capable of requesting the visualization of Tweets relevant with a specific topic. The user is the one who selects the topic of the visualization by inserting to the system search keywords, which consists of text-based keywords either typed in or selected from the “Popular Topics” list by the user. In order to get the visualization, at least one of the items in the list must be selected, or text-based keywords must be typed in otherwise, an error message will be generated and displayed. Moreover, the user besides of creating a general visualization, he can select to visualize information gathered from a single date by dragging the sliders provided to the desired month and day. Finally, the user can interact with the visualization by clicking on some areas of the graphs and get more details about the Tweets displayed such as the entire Tweet text, the username and the profile picture of the Tweet’s owner as well as the date and time when the Tweet was posted.

## *Chapter 5: Implementation & Testing*

---

The implementation of the VisualThoughts visualization tool mostly follows the design plan even though there are some small differences to the original design plans. Implementing the VisualThoughts application was not completed by the time we planned it as some issues came across such as house moving, health issues, programming problems or just bad luck.

The basic structure of the system consists of a software application written in Java and a database that communicates with the application. In fact, the application does not just communicate with the database. However, as discussed in the design chapter, the entire system depends on a database that keeps all the information of the Tweets collected. When a user runs the application and requests the graphs to be displayed, the system creates a query and connects with the database. The database then finds all the needed information and returns them for visualizing the Tweets' information into graphs.

During developing software, a lot of programming errors may come up preventing in this way the successful development of the application. However, testing your code once a single module is created can lead to the best results possible. In the next few sections, the system's process and testing is going to be discussed.

### SYSTEM PROCESS DESCRIPTION

#### **Retrieving Popular Search**

Once the application is being executed, the system displays the main screen of the application to the user. The user can input data to the system either by typing in the search keyword or by selecting the search keywords from the "Popular Search" list. "Popular Search" list consists of the 10 most popular keywords used for generating the visualization graph. However, where does the "Popular Search" come from? When the user runs the application, the system creates a database query, gets connected to the database and gathers all the information listed in the popularSearch table in a descending order based on their publicity and stores them in an array. Finally, after the information is gathered the system displays on the screen the 10 first entries of the array.

#### **Setting Popular Search**

Once the user inputs the search keywords and requests to create the graphs, the system firstly runs a function for updating the database table that holds information about "Popular Search" list. The system gathers the data inserted by the user and for each word contained in the search keywords creates a database query. The database query searches through the entire database if the word already exists. If the word exists, the system increases its popularity by 1 point otherwise the system creates a "Create" database query to insert the word in the table with its popularity value set to 1.

### Generating a Graph

Generating a graph is not an easy thing to do. Firstly, the user has to input the search keywords in the system either by typing the keywords in a text field or by selecting multiple keywords from the “Popular Search” list. After inputting the search keywords and before pressing the “Get the Graph” button the user must select which graph is going to be generated; WorldMap graph, RiverTheme graph or both of the WorldMap and RiverTheme graphs.

When the user presses the button to generate a graph, the system firstly checks if no data are input. In this case, it generates a dialog error message and presents it to the user. If all the required fields are filled, the system using the randomGenerator function generates different colors of blue to represent each search keyword and then creates a database query for gathering Tweets’ information that contains the search keywords. The gathered information, which consists of the Tweet geo-location in terms of latitude and longitude, the date and time when the Tweet was posted, the Tweet text as well as the Tweet’s user and user’s profile picture, are locally stored in a 2-dimension array. After the system collects the necessary details, it gets ready for creating the graph/graphs. The graphs to be displayed are selected by the value parsed from the user through the “Select Graph” combo box.

The RiverTheme graph is created using the date when each collected Tweet was posted. The system uses a function for separating the Tweets’ dates into a 2-dimension array where each column represents each different search keyword. After the Tweets’ dates are stored in different 2-dimension arrays for each search keyword, the function splits the dates into the 12 months of the year. For each month, the function calculates the frequency of Tweets and stores the frequencies in another 2 dimension array where each column represents each month of the year. This is where Processing programming language is going to be used to represent in a flow graph the Tweets’ frequencies for each month using Vertex curve. For each search keyword, a new sequence of Vertex curves is created on top of the previous sequence, each one painted with its representative pre-selected color.

The WorldMap graph is created using the geo-location and text details of the tweet. All the provided geo-location details are translated into a set of map waypoints and displayed to the OpenStreetMaps map provided by the SwingX-WS java library. However, besides displaying the waypoints, the system modifies how the waypoints look like using a setRenderer function. For each waypoint created, the system gets the analogous Tweet text and uses a random word from the text to be displayed in the waypoint icon. In addition to the modification of the displayed text in each waypoint, the system using the pre-selected colors paints each waypoint with the color that represents the contained search keyword.

### Display Tweets for a single month and day

The user of VisualThoughts application has the ability to select a particular range of Tweets to be visualized by selecting a specific month and day. When the user changes the value firstly of “Select Month” jSlider and then of “Select Day” jSlider, the system is informed about this change in order to take the necessary measures.

In the case that the user changes only the value of “Select Month” jSlider, the system must visualize the Tweets posted during the selected month. The system gets the changed value of the jSlider and using the analogous function tries to gather all the Tweets posted during that month. The only way that this can be achieved is by extracting the month digits for each Tweet’s date and then compare them with the value parsed through the jSlider. As soon as the necessary Tweets’ details are gathered, the system follows the same procedure with creating the graph in order to get the modified graphs displayed. Of course, the user is capable of changing the graph, so it visualizes the Tweets throughout the year by setting the “Select Month” jSlider value to 0(zero).

On the other hand, when the user changes both value of “Select Month” and “Select Day” jSlider, the system must visualize the Tweets posted during the selected month and day. The system gets the changed valued of the “Select Month” jSlider and follows the procedure described above to collect all the Tweets posted during the selected month. Once the Tweets of the selected month are collected, the system gets the changed value of the “Select Day” jSlider. The system uses the jSlider value to call a function for searching through the Tweets of the selected month, in order to collect the Tweets posted during the selected day. This can be achieved by extracting the digits representing the day for each Tweet’s date and then compare them with the value parsed through the “Select Day” jSlider. As soon as the necessary Tweets’ details are gathered, the system follows the same procedure with creating the graph in order to get the modified graphs displayed. Still, the user is capable of changing the graph, so it visualizes the Tweets throughout the selected Month by setting the “Select Day” jSlider value to 0(zero).

### Click on the Graph

When a user clicks on a generated graph, the system displays in a jList the related Tweets’ text. As soon as the user clicks on the graph, the system gathers the mouse coordinates in proportion of the screen’s size.

If the mouse clicked event happens in the WorldMap graph, the system compares the coordinates of the mouse with the coordinates of each waypoint displayed in the graph. If no waypoints match with the mouse coordinates, then no Tweets’ text is displayed in the jList. Otherwise, the system gathers the Tweets’ details related with the matching waypoints, stores them in a 2-dimension array and displays the Tweets’ text in the jList.

On the other hand, if the mouse clicked event occurs in the region of RiverTheme graph, the system checks if the color of the clicked pixel exists in the array of the colors used to represent the different search keywords. If no color, matches with the pixel’s color then no Tweets’ text is displayed in the jList. Otherwise, the system uses the x-coordinates of the mouse to determine in which month, day or hour the user refers to, stores the related Tweets’ details in a 2-dimension array and displays them in the jList.

Displaying the Tweets’ text in a jList gives the user the ability to select a specific Tweet’s text, in order to get displayed more detailed information about the Tweet. When a Tweet’s text is selected, the system

uses the value of the selected index to search through the associated array and get the Tweet's details. Once the details are collected, the system prints them in labels contained in the "Tweet Info" panel.

## SYSTEM TESTING

### Testing Strategy

A reliable software system requires testing almost at any stage of development, so it can be developed successfully. When a particular function or block of code is created, it should be tested so the best results are guaranteed. Same as with individual blocks of code and functions, when a number of classes are completed, the communication between these classes should be evaluated and tested to ensure that everything works based on the plan. Last but not least, when the entire system is developed, it should be comprehensively tested. In this way, developers are confident that the system meets all the specified requirements but also that errors and bugs are prevented.

The testing of VisualThoughts system is mainly based on Test Driven Development Methodology (TDD). TDD is an evolutionary approach that uses automated unit tests to encourage the development of systems.<sup>[37][38]</sup> Using TDD, the developer firstly writes some failing tests that correspond to some desired functions or improvements. Once the tests are written, the developer builds the necessary code that will be able to pass the test. The developer continues to write new tests and their implementation code until the entire system is completed with no failing tests exist.

### Testing Plan

As mentioned before, testing is one of the most important components of developing a software system as it ensures that everything was successfully built. The three levels recommended and used for testing the VisualThoughts application are Module testing, Integration Testing, System Testing

At the stage of Module testing, each time a single module is created in the system, like a function or a block of code, has to be tested to confirm that each command is successfully executed and provides the correct results. The tests performed should cover a wide range of trial data like real life data as well as some incorrect data which might cause errors to the program.

When a number of modules that communicate between them are completed, an Integration Testing should take part. During Integration Testing, completed modules, such as multiple functions or sets of code included in a class, are combined and tested to make sure that they successfully communicate as a group. Integration testing takes as input the modules that have been tested during Module testing, groups them and applies various tests and finally returns as output a combined system for System Testing.

System Testing is the final level of testing and is focused on testing a combined, and complete system, in order to make sure that the system meets the entire specified requirements. At this point of testing, all the test cases that came up from the Use Case diagrams are going to be tested. System Testing takes as input all the combined modules tested during Integration Testing, and its purpose is to find any disagreements between the integrated modules.



VisualThoughts application was tested according to these three levels. At the beginning, each time a new functionality was built a test was run to check that the expected results were returned. At a later level when all the classes were built, testing was run to ensure that these classes communicate in the right way and that the results returned are the expected ones. Finally, the System Testing occurred once all the necessary classes were built. During the System Testing, all the Test Cases stated in the section below were executed to guarantee that everything execute as expected otherwise the necessary improvements had to take part.

### System Test Cases

**Test Number: 1**

**Description:** Request a Graph with no data input.

**Properties:**

**Input Sting:** None

**Expected Result:** An error message is displayed to the user asking him to input the necessary data, so a graph could be created.

**Actual Result:** As expected

**Test Number: 2**

**Description:** Request both graphs, WorldMap and RiverTheme, to be created and displayed with all the necessary data input.

**Properties:**

**Input Sting:** Search Keywords, Graph Type

**Expected Result:** The selected type of Graph is created and displayed to the user

**Actual Result:** As expected

**Test Number: 3**

**Description:** Request WorldMap graph to be created and displayed with all the necessary data input.

**Properties:**

**Input Sting:** Search Keywords, Graph Type

**Expected Result:** WorldMap graph is created and displayed to the user

**Actual Result:** Both graphs, WorldMap and RiverTheme, are created and displayed to the user

**Corrective Actions:** myInit() function that was called when the "Request a Graph" button was pressed, was setting the Graph Type value to the initial, which displays both graphs. The value of Graph Type is parsed to a new variable before is set to the initial one, in order to prevent the problem stated above.

**Test Number: 4**

**Description:** Request RiverTheme graph to be created and displayed with all the necessary data input.

**Properties:**

**Input Sting:** Search Keywords, Graph Type

**Expected Result:** RiverTheme graph is created and displayed to the user

**Actual Result:** Both graphs, WorldMap and RiverTheme, are created and displayed to the user

**Corrective Actions:** myInit() function that was called when the “Request a Graph” button was pressed, was setting the Graph Type value to the initial, which displays both graphs. The value of Graph Type is parsed to a new variable before is set to the initial one, in order to prevent the problem stated above.

**Test Number: 5**

**Description:** User clicks on the graph in a section that no Tweets are represented.

**Properties:**

**Input Sting:** Mouse Coordinates

**Expected Result:** No Tweets are displayed in the jList

**Actual Result:** As expected

**Test Number: 6**

**Description:** User clicks repeatedly on the graph in a section that Tweets are represented.

**Properties:**

**Input Sting:** Mouse Coordinates

**Expected Result:** Each time, the selected Tweets, are displayed in the jList

**Actual Result:** The first time the user clicks on the graph, the selected Tweets, are displayed in the jList. However, the second time the user attempts to click on the graph, the jList does not remove the previously selected Tweets and displays all the selected Tweets from both two attempts. To correct this, each time the user clicks on a graph, the jList clears all the items displayed in before new items are added.

**Test Number: 7**

**Description:** User selects a Tweet from the jList to display more details.

**Properties:**

**Input Sting:** Selected Tweet’s Index

**Expected Result:** Displays in “Tweet Info” panel all the available Tweet’s details

**Actual Result:** As expected

**Test Number: 8**

**Description:** User selected a particular month using the “Select Month” jSlider to visualize Tweets posted during the selected month

**Properties:**

**Input Sting:** jSlider Value

**Expected Result:** Graph is modified and represents all the Tweets posted during that month

**Actual Result:** if the jSlider value is less than 10, then the Graph does not select the correct Tweets. For example if the jSlider value is “2” then Tweets posted during both 2<sup>nd</sup> and 12<sup>th</sup> months are visualized. To prevent this, if the jSlider is less than 10 like 2, then the system searches for Tweets posted during the “02” month instead of “2” month.

**Test Number: 9**

**Description:** User selected a particular day using the “Select Day” jSlider to visualize Tweets posted during the selected day, after he has already selected a month using the “Select Month jSlider

**Properties:**

**Input Sting:** jSlider Value

**Expected Result:** Graph is modified and represents all the Tweets posted during that month and day

**Actual Result:** if the jSlider value is less than 10, then the Graph does not select the correct Tweets. For example if the jSlider value is “2” then Tweets posted during 7<sup>th</sup>, 17<sup>th</sup> and 27<sup>nd</sup> days are visualized. To prevent this, if the jSlider is less than 10 like 7, then the system searches for Tweets posted during the “07” day instead of “7” day.

**Test Number:** 10

**Description:** Items in “Popular Search” are less than 10 and the user inputs search keywords to the system. One of those keywords does not exist in the Popular Search list.

**Properties:**

**Input Sting:** Search Keywords

**Expected Result:** System checks if each of the keywords exists in the table. If the keyword exists, then the system increases its popularity. Otherwise, inserts the new keyword in the table with its popularity equals to 1. The new keyword is displayed in the “Popular Search” list.

**Actual Result:** As expected

## Testing Implementation

During software implementation, some problems and errors rise. Some of them, where easy to be solved whereas some other needed some more research in order to be solved.

The first major problem, which came up, had to do with importing the JXMapKit wrapper of JXMapView component into the project. Once the JXMapKit was imported, the interactive map was not displayed on the screen. The reason was that the initial value of the Provider is set to SwingLabsBlueMarble and had to be changed to OpenStreetMaps.

Since this problem was fixed, the development flow was running by plan until Processing and the development of RiverTheme graph came along. In the beginning, problems arise with setting up a Processing project in Java Netbeans whereas during implementing RiverTheme graph in Processing a problem came up during the attempt of drawing evenly consecutive curves, which represent the RiverTheme graph.

Nevertheless, except for the two main problems stated above and some regular programming errors, no other problem was big enough to affect the flow of the system’s development.

## Chapter 6: Demonstration & Evaluation

In the next few pages, the system's walkthrough will be discussed in order to gain a clear idea of what the VisualThoughts visualization tool looks like and does without executing the code. Besides the system's walkthrough, the entire system will be evaluated to find out if it accomplishes all the specified requirements.

### SYSTEM WALKTHROUGH

In the next few pages, a walkthrough of the system will show what the system is like when is used

#### Request a Graph

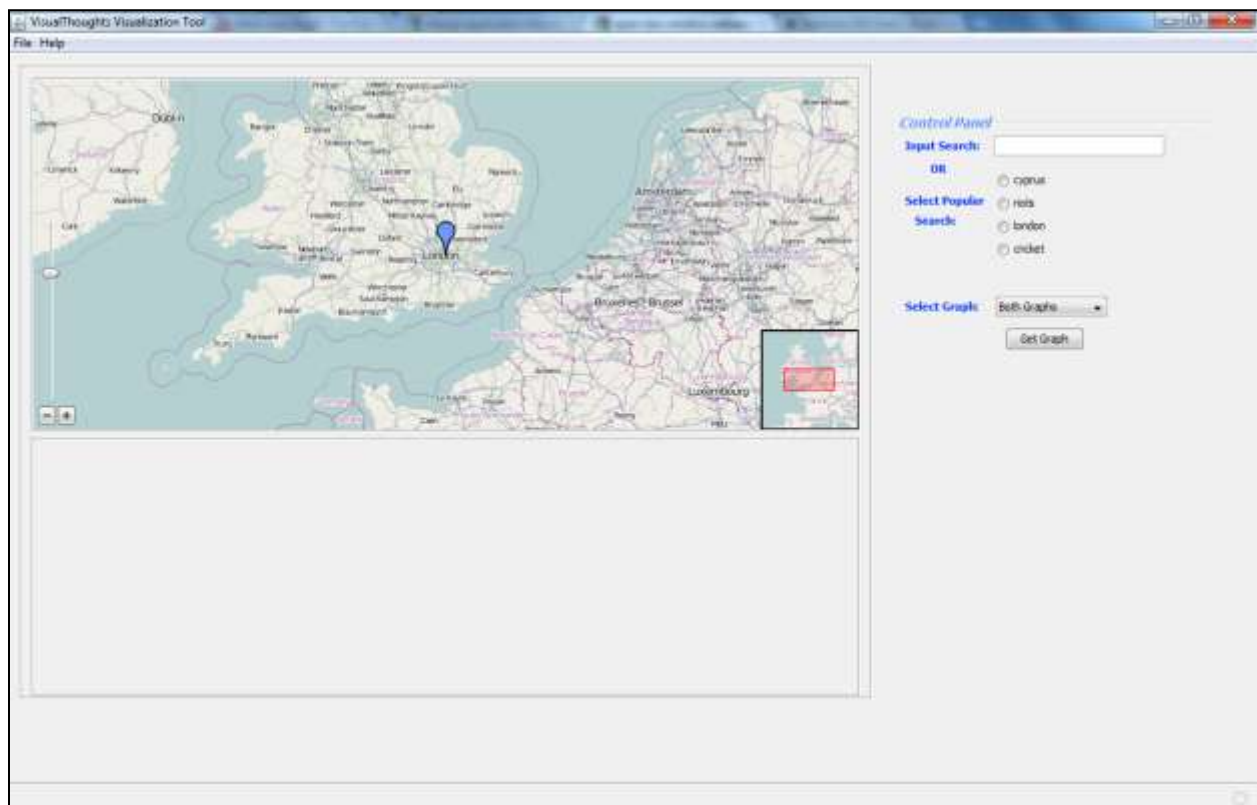


Fig. 12 – The main screen of the VisualThoughts application

The screenshot above (figure 12) shows the main screen of the VisualThoughts application. Through the main screen, the user can insert data in order to create the graph. As illustrated, there are two possible ways of inputting the search keywords; either by typing the keywords in a text field or by selecting multiple keywords from the “Popular Search” list. Just below the “Popular Search” list, there is a “Select Graph” combo-box that enables the user to select which graph will be created; WorldMap graph, RiverTheme graph, or both of the WorldMap and RiverTheme graphs. The “Select Graph” combo box

value is set by default to “Both Graphs” so in case that the user does not select anyother option both of the graphs are going to be displayed.

Once the user types in or selects the data, which represent the search keywords, the “Get Graph” button, must be pressed in order to get the graphs created. Nevertheless, how does the system react if the user tries press the button without inputting any data? In such a case that the user does not type in any data, an error dialog message is displayed on the screen, and the user cannot get any graph displayed until he inserts the needed data. The screenshot (figure 13) below shows how the error message is displayed.

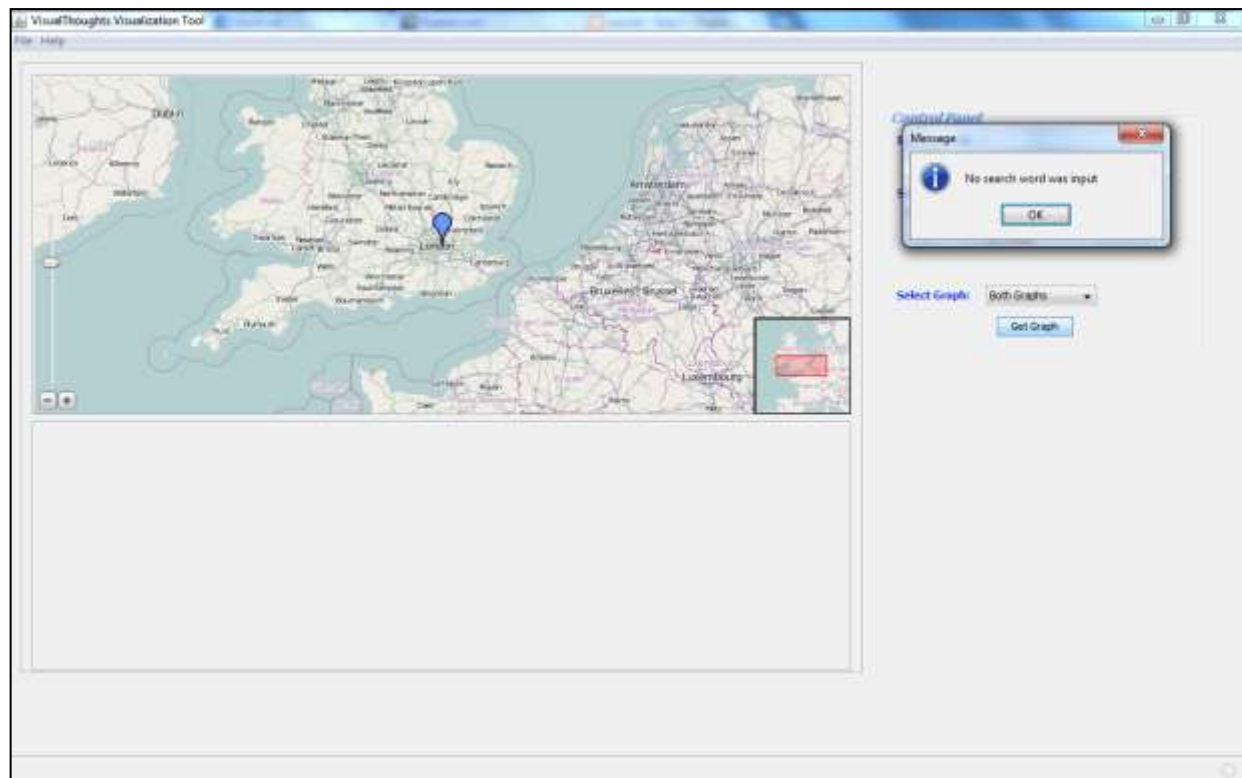


Fig. 13 – If a user does not type in any data, an error message is displayed to the screen.

### Display selected graph

Using VisualThoughts visualization tool, there are two different ways to visualize Tweets; thus, there are three different options to select from for displaying the graphs on the screen. The first option is to display both of the WorldMap and RiverTheme graphs (figure 14), while the other two options is to display either WorldMap graph (figure 14a) by itself or RiverTheme (figure14b).

In figure 14 listed below, the screens displayed to the user, once the graphs are generated, are demonstrated. It is obvious that new features and options have been added in contrast with the main screen. All these features give the user the ability to interact with the graphs and gain insight about Tweets' information.

The two jSlider buttons that have been added enables the user to select a specific month for start (the Day slider is disabled while the Month slider value is set to zero(0)) to visualize all the Tweets posted during the selected month. After selecting a specific month, the Day slider is enabled allowing the user to select a specific day as well to visualize all the Tweets posted during the selected month and day. Besides the jSlider buttons, a jList list and a “Tweet Info” panel have been added to the screen. The JList displays all the Tweets gathered from the selected section once the user clicks on one of the graphs displayed. In addition, when a user selected a Tweet displayed in the jList, all the details of the selected Tweet are displayed in the “Tweet Info” panel.

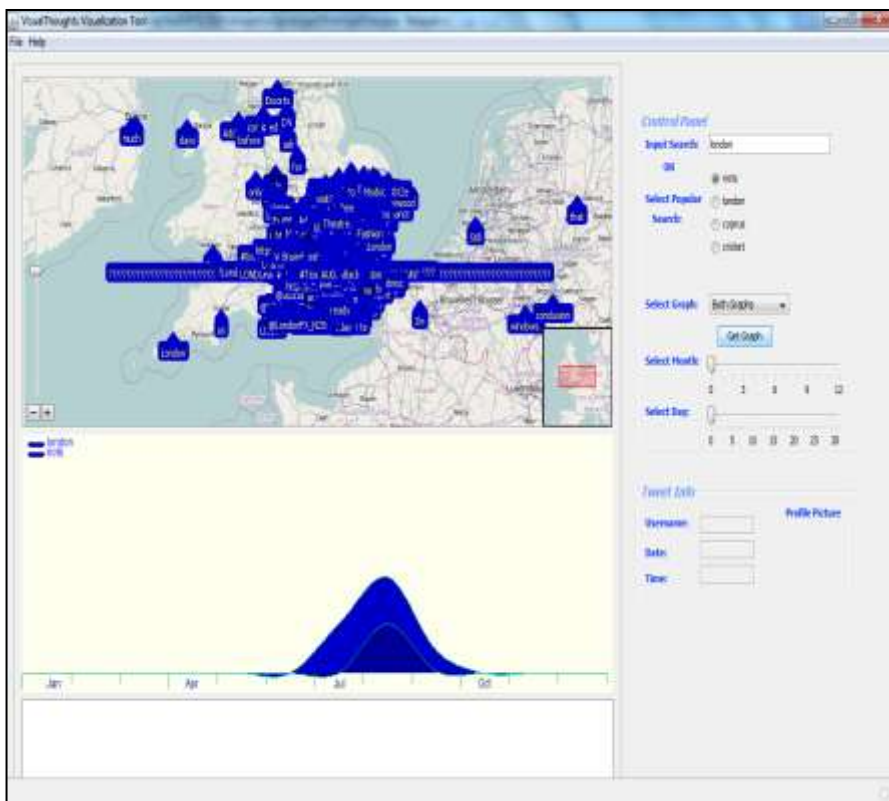


Fig. 14 – Screenshot of how screen looks like when graphs are generated

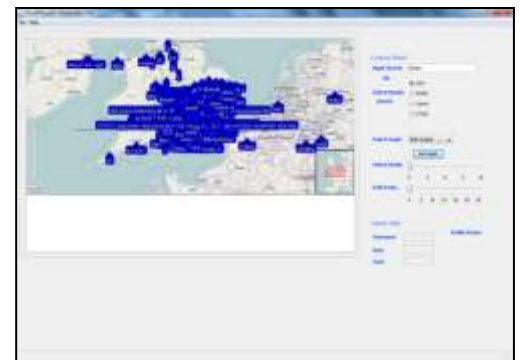


Fig. 14a – Screenshot of WorldMap

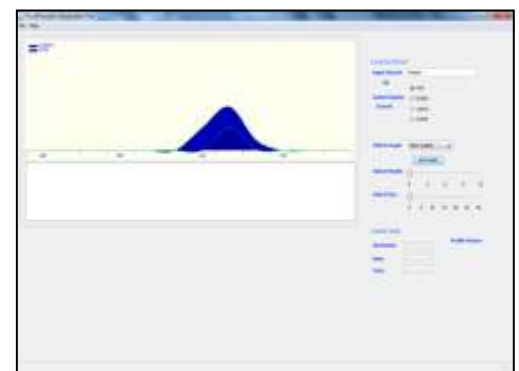


Fig. 14b – Screenshot of RiverTheme graph

### Interacting with Graphs

VisualThoughts visualization tool besides generating graphs representing Tweets, it enables the user to interact with the information visualized and gain more insights about visualized Tweets. For example, the user can request to get Tweets posted during a particular month and day visualized using the jSlider buttons. Moreover, the user is capable of interacting with the graphs by clicking on them, requesting in this way to get all the details analogous to the Tweets represented in the clicked section.



In figure 15, the user requests to get all the Tweets posted during 15<sup>th</sup> of July (15<sup>th</sup> day of 7<sup>th</sup> month). The user can achieve this by moving the jSlider buttons to the appropriate values, which in our case are 7 for the Month jSlider and 15 for the Day jSlider. In addition to this, if the user changes the value of Day jSlider back to 0(zero), the graph represents only the Tweets posted during the selected month (figure 15a). The graph represents all the Tweets relevant to the search keywords (figure 15b) if the user changes both Month and Day jSlider to 0(zero).

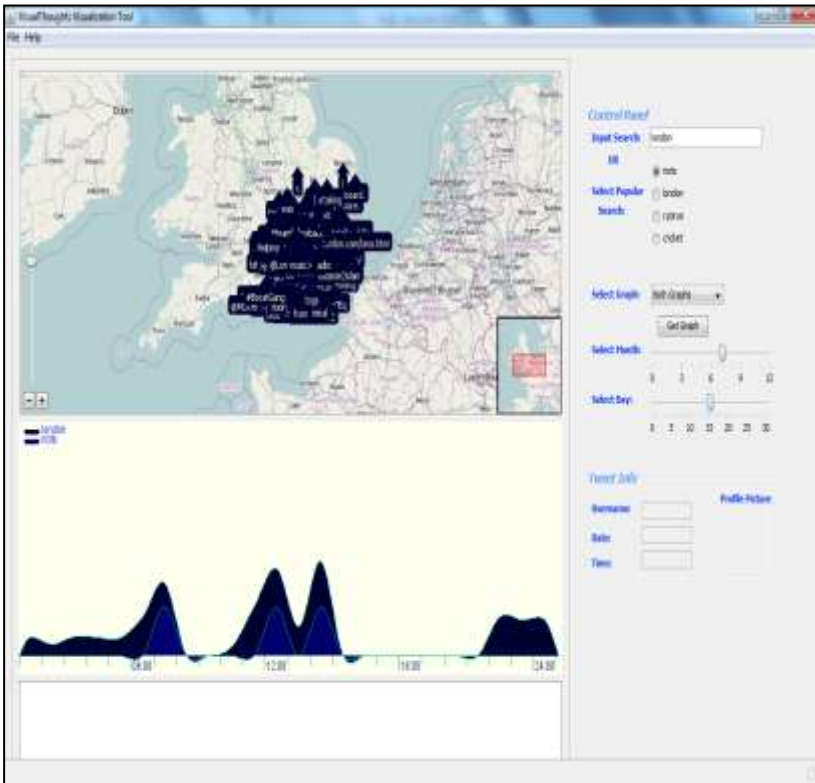


Fig. 15 – User interacts with the jSlider buttons to visualize Tweets posted during particular month and day



Fig. 15a – User set Day jSlider value to 0



Fig. 15b – User set Month jSlider value to 0

The user of VisualThoughts application can also interact with the graphs by clicking on them to get extra details regarding the Tweets being visualized in the clicked region. In figure 16, the user clicks on a waypoint included in the WorldMap graph. All the Tweets posted in the geo-location represented by the waypoint are listed in the jList displayed below the graphs. The user can also select a particular Tweet from the jList in order to get the Tweet's details displayed in the "Tweet Info" panel (figure 16a), such as the owner of the tweet and his profile picture as well as the date and time the tweet was posted.

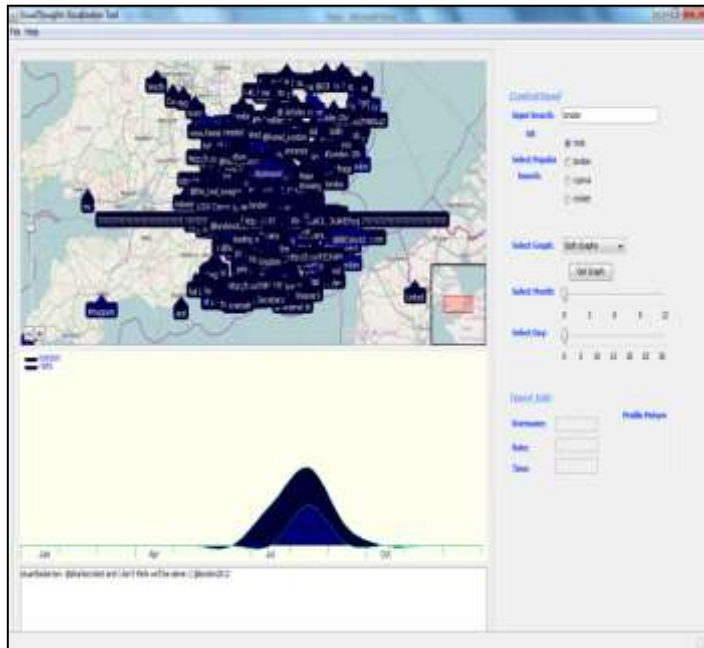


Fig. 16 – User Clicks on the WorldMap graph and gets the analogous Tweets displayed in the jList

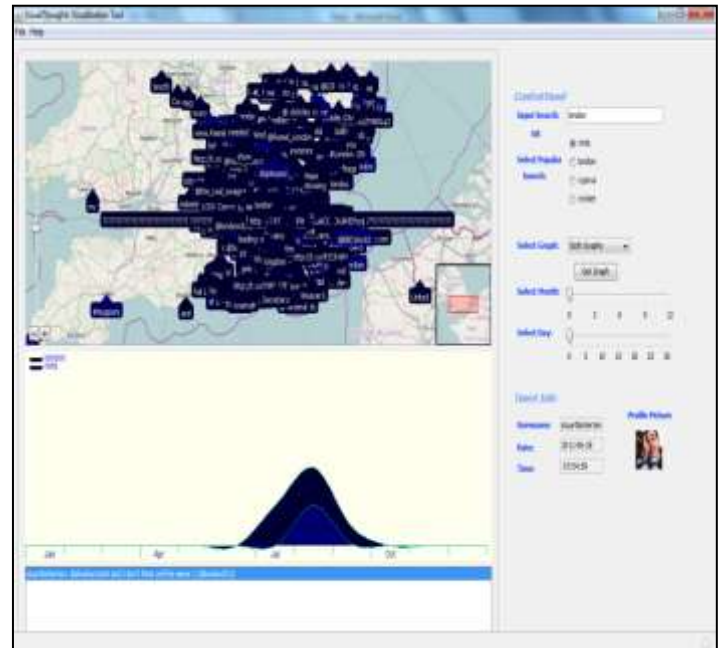


Fig. 16a – User selects a particular Tweet from jList to get its details displayed in “Tweet Info” panel

## Retrieving/Setting Popular Search

When a user runs the VisualThoughts application to generate a visualization graph, there are two

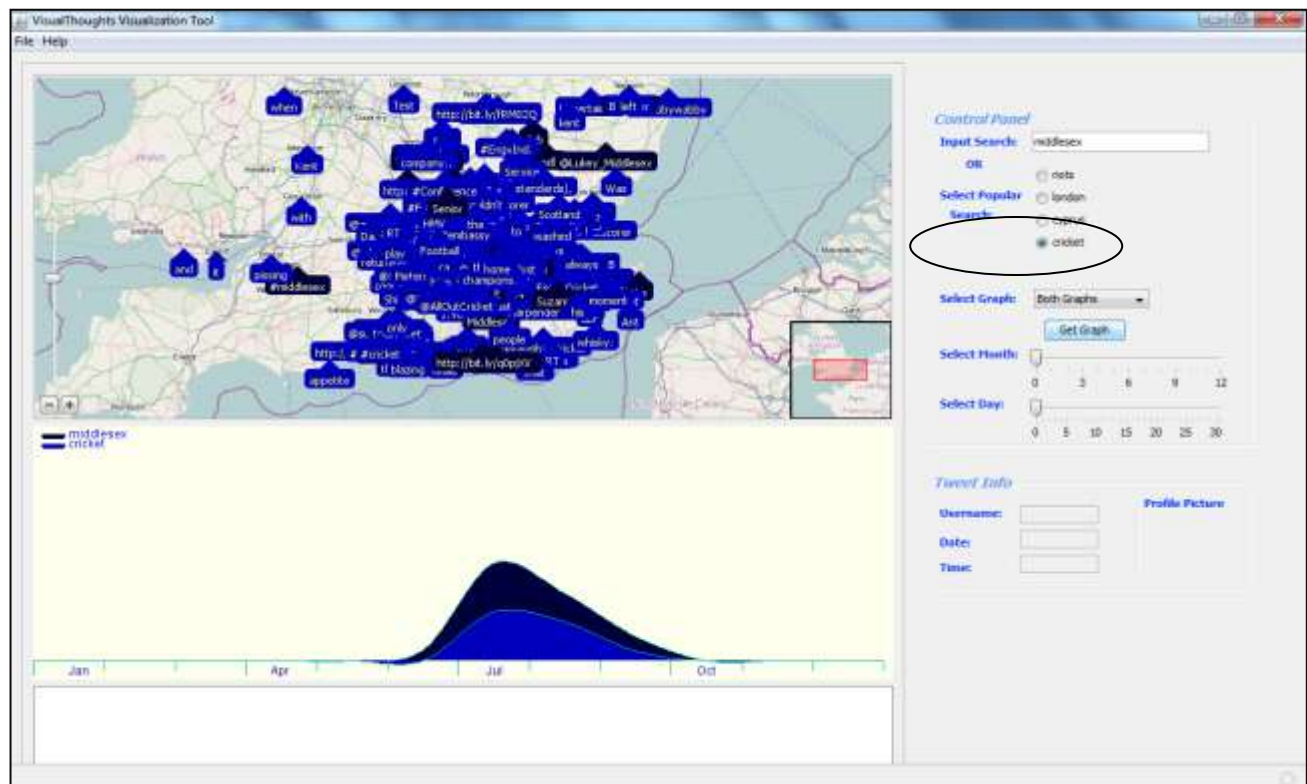


Fig. 17 – The user requests a graph to be generated by typing in a keyword that does not exist in “Popular Search” list



possible ways of inputting the search keywords; either by typing the keywords in a text field or by selecting multiple keywords from the “Popular Search” list. “Popular Search” list represents the 10 most popular keywords used as search keywords in VisualThoughts application. Once the application is executed, the system retrieves from the database the 10 most popular keywords and displays them on the “Popular Search” list.

In figure 17, a user tries to generate a graph by typing in a keyword. The system stores the keyword in the database if it does not already exist in the popular searches table; otherwise, it just increases its popularity by one point. In figure 17a when another user executes the application after the previous user, the new search keyword is displayed in the “Popular Search” list as it only consists of 5 popular searches.

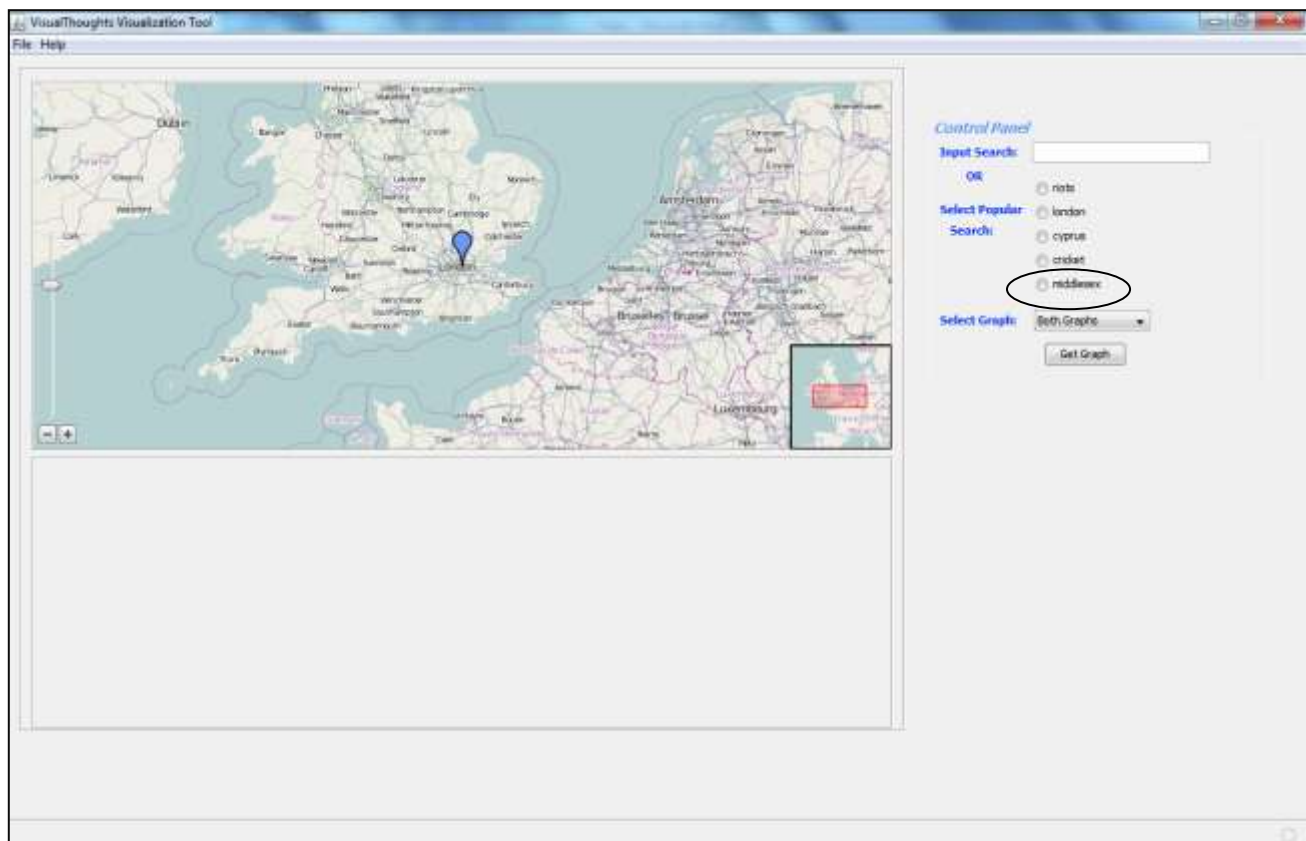


Fig. 17a – When a new user executes the application the new keyword is included in the “Popular Search” list

## SYSTEM EVALUATION

VisualThoughts application is a visualization tool with the main aim of visualizing specific topics of Tweets in proportion of their geo-location and time. Nevertheless, does the VisualThoughts application meet all the specified requirements? Is the application easy to use and what about its performance? The answers of these questions may define whether the development of the project was successful or not. In order to find out the answers to the above questions, 30 people have been asked to use VisualThoughts application and answer a related questionnaire. The questionnaire with all the questions included can be found in Appendix B.

The first six questions of the questionnaire deal with some social statistics like the age of the users, if they are familiar with “Social Network”, “Social Network Analysis” and “Information Visualization” terms, if they are members in any online social network etc. In more details, the first six questions are listed below.

**Question 1:** How old are you?

**Question 2:** Are you familiar with the terms “Social Network” and “Online Social Network”?

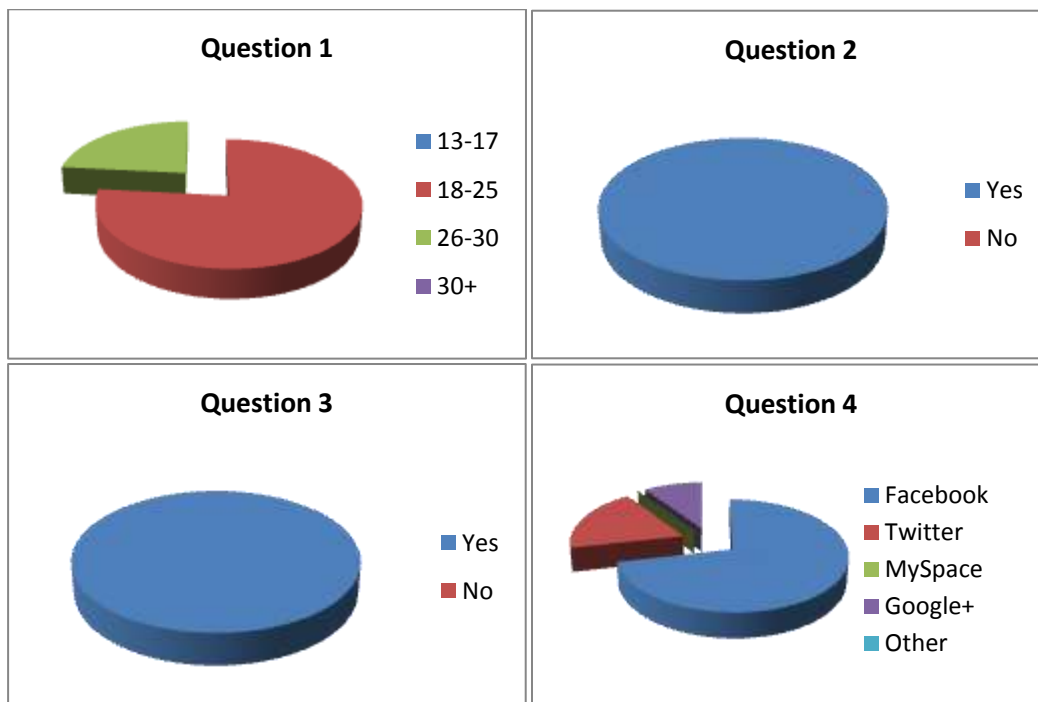
**Question 3:** Have you ever been a member of an Online Social Network?

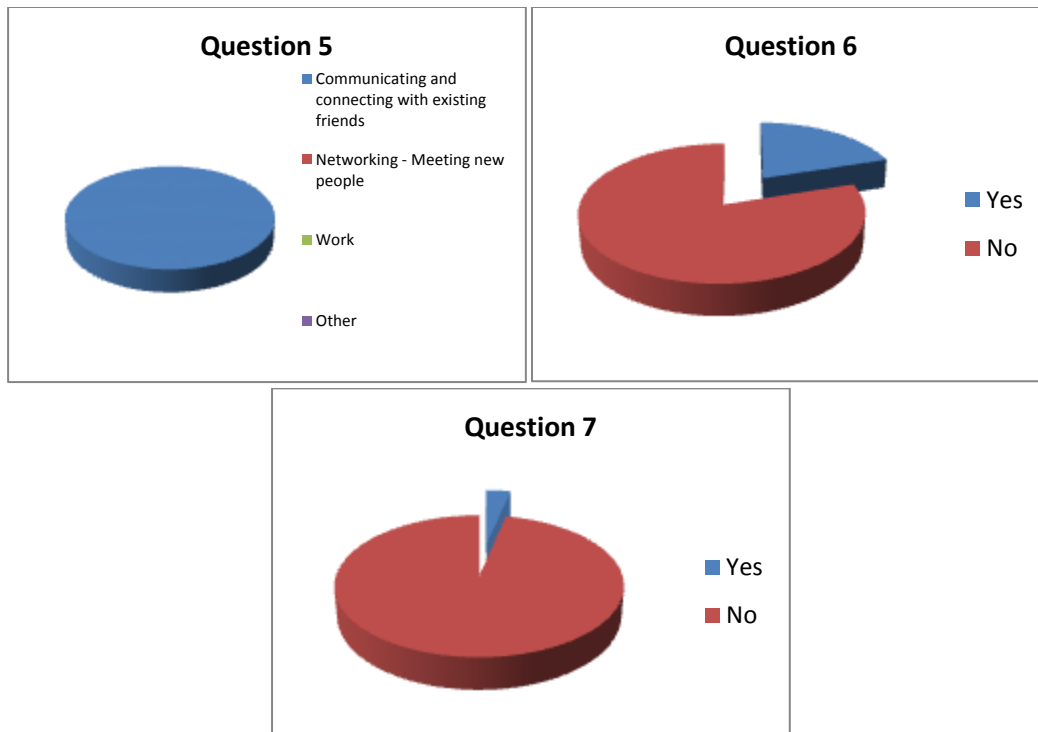
**Question 4:** If yes, mark the Online Social Networks which you are registered to.

**Question 5:** What is the main purpose for using Online Social Network?

**Question 6:** Have you ever been familiar with the terms “Social Network Analysis” and “Information Visualization” before using VisualThoughts?

**Question 7:** If yes, have you ever used a Visualization tool before?





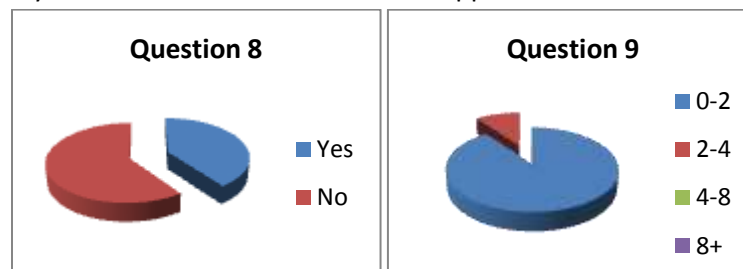
Graphs representing the answers gathered from questions 1-7

According to the answers, most of the users who used the VisualThoughts application are aged between 19 years old and 25 years old. Moreover, all of the users are members in Facebook for communicating and connecting with friends, whereas some of them are also members in Twitter and Google+. However, only 6 of the users were familiar with the “Social Network Analysis” and “Information Visualization” terms and just 1 user have used a Visualization tool before.

The next three questions deal with some application statistics like how many users would have used the application, if any of their friends would use it etc. These questions are listed below.

**Question 8:** Would you have used the VisualThoughts Application by yourself?

**Question 9:** How many of your friends would have used this application?



Graphs representing the answers gathered from questions 8-9

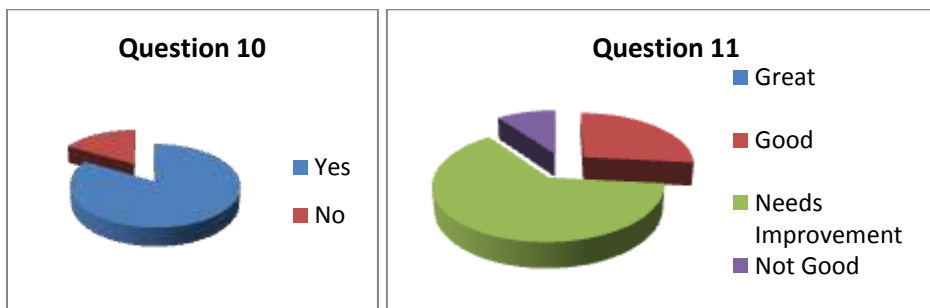
Twelve of users, who answered the questionnaire, stated that they would have used the VisualThoughts application with no one forcing them. However in question 9, as stated in the graph above most of the users answered that no more than 2 of their friends would have used VisualThoughts Application.

The last two questions have to do completely with the performance of the application, the usability of the application, as well as with some suggestions that may help in a future improvement of the application. These two questions are displayed analytically below.

**Question 10:** Is VisualThoughts Application easy to use?

**Question 11:** What is your opinion according to the performance of the application?

**Question 12:** Do you recommend any improvements about the application? If yes, please mention them.



Graphs representing the answers gathered from questions 10-11

An encouraging statistic gathered from the answers given to question 10, is the fact that 25 users believe that VisualThoughts Application is easy to use. Nevertheless, more than half of the users answered that the performance of the application can be described as "Needs Improvement". In addition to this, there are a few suggestions for improvements that I should take into account. Some of these suggestions include improving the application's decoration, adding a feature that will enable the user to share the graphs screenshot via Facebook or Twitter, as well as adding a button that the user may use it to send email to other friends and suggest them to use VisualThoughts application.

In view of the fact that the answers of the questionnaire have been analyzed and discussed, VisualThoughts application can easily be evaluated. Based on the testing occurred when the system was completed, VisualThoughts application meets all the specified requirements. Moreover, taking into account the answers given to the questionnaire, it can be said that VisualThoughts is also an easy to use application. However, some important upgrading must be done like improving the performance of the system and adding some more features to it.

## Chapter 7: Critical Evaluation

---

### REVIEW OF AIMS

The word successful can be given as an adjective to a project that reflects on the original aims.

VisualThoughts aim was to develop a visualization tool that would help users to visualize Tweets of a specific topic. Mercifully, all the specified requirements and aims of this project have been successfully developed in VisualThoughts application.

VisualThoughts application gives the user the ability to request a visualization graph by inputting to the system the desired topic to search for relevant Tweets. The user has the choice to select between two types of graphs, WorldMap and RiverTheme, which one he desires to be created. Moreover, VisualThoughts application enables the user to interact with the created graph and merge the range of Tweets to be visualized by selecting a particular month and day of the year. In addition to this, the user of VisualThoughts application can also interact with the created graphs by clicking on any section of them that represents Tweets and gain some insights about Tweets' details.

### REVISIONS TO DESIGN AND IMPLEMENTATION

Although the fact that the overall system was successfully implemented, if I had the chance to start the project all over the beginning I would be more focused on obtaining a robust system performance.

Spending much time on the design of the system had as a consequence to left out of time during implementation. Running out of time while coding resulted in focusing more on developing the specified requirements and less on how to achieve the best system performance.

### LESSONS LEARNT

Developing a project like VisualThoughts visualization tool has given me the opportunity to experience how a sizeable software system is produced. Moreover, during the developing of this project, I have learnt a lot of lessons. First of all, it taught me never to underestimate the difficulty that a looking simple task may hide. Besides that, I got familiar with new technologies like Processing programming environment and some Java libraries, such as SwingX-WS and Lucene Apache that really made my work easier. Additionally, I found out how much important is performance to the reliability of the system. Last but not least, I appreciated the fact that many programmers work exceptionally hard to successfully complete a project and learnt to think more closely about how users view systems.

## *Chapter 8: Conclusion*

---

### **FURTHER WORK/RESEARCH**

Although the overall system contains some strong enough characteristics, there are some other features that can be added so the VisualThoughts application is improved. For example, during the system's evaluation most of the users stated that the system's performance should be improved. Indeed, searching through the database for gathering the relevant Tweets takes a couple of minutes and it is something that should be improved in the future.

However, besides improving the performance of the system, some more development can be done and VisualThoughts application can be a more entertaining application. For example, some more features can be installed like allowing the user take a snapshot of the created graphs and share it among his friends via Online Social Networks. Moreover, VisualThoughts application can be modified to enable the user to insert the search keywords representing the Tweets' topic in some non-English languages such as Greek, Spanish, and Italian etc. In addition to this, VisualThoughts application may also allow the user to visualize all the Tweets posted by a particular Twitter member by inserting his username to the system.

Finally, VisualThoughts application could be more User Friendly, if it had been appropriately decorated. When a user visits an application with a fascinating decoration and graphic designs, he is more attracted in using it. For that reason, the application should be more elegant and attractive in the future.

### **CONCLUDING POINTS**

In this report, VisualThoughts application was presented and discussed. VisualThoughts application is a visualization tool enabling the user to visualize Tweets relevant to a specific topic and gain some insight information. Even though, there are some improvements that must take place in the future, VisualThoughts application has been a success as it fulfills all the specified aims and requirements.

While VisualThoughts application is not the kind of tool that normal people use on every day basis, the development of such a tool was a real challenge and an exceedingly fascinating topic of research. However, users of VisualThoughts are able to underlie, analyze and understand the visualized information quickly. Besides that, VisualThoughts application can be used by scientists in many fields of science in order to gain a better understanding of complex situations I believe that the project has been a success as I have demonstrated all the features I planned to.

## References

---

1. Jason Pontin, "From Many Tweets, One Loud Voice on the Internet," The New York Times, April 22, 2007
2. Twitter (2011) In Wikipedia, Retrieved October 10<sup>th</sup> 2011 from <http://en.wikipedia.org/wiki/Twitter>
3. Contact us (2011) In Twitter.com, Retrieved October 2011 from <http://twitter.com/about/contact>
4. Elke Dunker(2008), MSc Project Proposal Template, Middlesex University 2008
5. Arnold B. K., Gosling J., Holmes D., Arnold B. K., Gosling J., & Holmes D. (2005), THE Java™ Programming Language (Fourth Edi., pp. 29-30), Addison Wesley Professional
6. Gosling J.(2000), The Java™ Language Specification (Female Edi., pp. 1-6), Addison Wesley Professional
7. Winter V. L., & Mametjanov A. (2007), Generative programming techniques for Java library migration. Proceedings of the 6th international conference on Generative programming and component engineering - GPCE '07, 185, New York, New York, USA: ACM Press, doi: 10.1145/1289971.1290001.
8. Overview (2011) in lucene.apache.org, Retrieved September 14<sup>th</sup> 2011, <http://lucene.apache.org/>
9. Cover (2011) in processing.org, Retrieved October 2011, <http://processing.org/>
10. Java Comparison (2010) in wiki.processing.org, Retrieved October 2011, [http://wiki.processing.org/w/Java\\_Comparison](http://wiki.processing.org/w/Java_Comparison)
11. Bernabe S., & Plaza A. (2010), A new system to perform unsupervised and supervised classification of satellite images from Google Maps, Processing, 7810, 781010-781010-10, doi: 10.1117/12.863243.
12. Bresciani S., & Eppler M. J. (2009), The benefits of synchronous collaborative information visualization: evidence from an experimental evaluation, IEEE transactions on visualization and computer graphics, 15(6), 1073-80, doi: 10.1109/TVCG.2009.188.
13. Martino, F., & Spoto, A. (2006). Social Network Analysis : A brief theoretical review and further perspectives in the study of Information Technology. Social Networks, 4(1), 53 - 86.
14. Wasserman S. & Galaskiewicz J. (1994), Advances in social network analysis: research in the social and behavioral sciences(Focus Edi.), SAGE
15. Chen M., Ebert D., Hagen H., S R., & Van R. (2009). Visualization Viewpoints, Response.
16. Galaskiewicz J., & Wasserman S. (1993), Social Network Analysis: Concepts, Methodology, and Directions for the 1990s, Sociological Methods & Research, 22(1), 3-22, doi: 10.1177/0049124193022001001.
17. Wasserman S. and K. Faust (1994), Social Network Analysis, Cambridge: Cambridge University Press.
18. What is Social Networking? (2011) in whatissocialnetworking.com, Retrieved October 2011, <http://www.whatissocialnetworking.com/>

19. Social Networking Service (2011) In Wikipedia, Retrieved October 2011 from [http://en.wikipedia.org/wiki/Social\\_networking\\_service](http://en.wikipedia.org/wiki/Social_networking_service)
20. Social Network (2011) In Wikipedia, Retrieved October 2011 from [http://en.wikipedia.org/wiki/Social\\_network](http://en.wikipedia.org/wiki/Social_network)
21. Information Visualization(2011) in Wikipedia, Retrieved October 2011 from [http://en.wikipedia.org/wiki/Information\\_visualization](http://en.wikipedia.org/wiki/Information_visualization)
22. Card S. K., Mackinlay J. D. & Shneiderman B. (1999), Readings in Information Visualization: Using Vision to Think(pp. 1-35), Morgan Kaufmann
23. Leung C. K.-S. & Carmichael C. L. (2010), Exploring Social Networks: A Frequent Pattern Visualization Approach, *2010 IEEE Second International Conference on Social Computing*, 419-424, Ieee, doi: 10.1109/SocialCom.2010.67.
24. Ware C.(2004), Information Visualization: Perception for Design (Second Edi., pp. 4-5), Morgan Kaufmann
25. Social Network Analysis (2011) in orgnet.com, Retrieved October 2011 from <http://www.orgnet.com/sna.html>
26. Heer, J., & Boyd, D. (2005). Vizster: visualizing online social networks. IEEE Symposium on Information Visualization, 2005. INFOVIS 2005., 32-39. Ieee. doi: 10.1109/INFVIS.2005.1532126.
27. Mutton P. (1988) Inferring and visualizing social networks on internet relay chat, Proceedings. Eighth International Conference on Information Visualisation, 2004. IV 2004., 35-43, Ieee, doi: 10.1109/IV.2004.1320122.
28. Viegas F. B., Boyd D., Nguyen D. H., Potter J. & Donath J. (2004), Digital artifacts for remembering and storytelling: posthistory and social network fragments, 37th Annual Hawaii International Conference on System Sciences, 2004, Proceedings of the, 00(C), 10 pp. Ieee, doi: 10.1109/HICSS.2004.1265287.
29. Henry N., Fekete J.-D. & McGuffin M. J. (2007), NodeTrix: a hybrid visualization of social networks, IEEE transactions on visualization and computer graphics, 13(6), 1302-9, doi: 10.1109/TVCG.2007.70582.
30. Heer J. & Boyd D. (2005), Vizster: visualizing online social networks, IEEE Symposium on Information Visualization, 2005, INFOVIS 2005, 32-39, Ieee, doi: 10.1109/INFVIS.2005.1532126.
31. Use Case Diagram (2011) in Wikipedia, Retrieved October 2011 from [http://en.wikipedia.org/wiki/Use\\_case\\_diagram](http://en.wikipedia.org/wiki/Use_case_diagram)
32. Class Diagram (2011) in Wikipedia, Retrieved October 2011 from [http://en.wikipedia.org/wiki/Class\\_diagram](http://en.wikipedia.org/wiki/Class_diagram)
33. Data Flow Diagram (2011) in Wikipedia, Retrieved October 2011 from [http://en.wikipedia.org/wiki/Data\\_flow\\_diagram](http://en.wikipedia.org/wiki/Data_flow_diagram)
34. Sequence Diagram (2011) in Wikipedia, Retrieved October 2011 from [http://en.wikipedia.org/wiki/Sequence\\_diagram](http://en.wikipedia.org/wiki/Sequence_diagram)
35. Fowler M. (2004), UML distilled: A brief guide to the standard object modeling language, Addison Wesley Professional
36. UML (2011) in Wikipedia, Retrieved October 2011 from [http://en.wikipedia.org/wiki/Unified\\_Modeling\\_Language](http://en.wikipedia.org/wiki/Unified_Modeling_Language)



37. Test Driven Development (2010) in c2.com, Retrieved October 2011 from <http://c2.com/cgi/wiki?TestDrivenDevelopment>
38. March P. N. (2006), Introduction to Test Driven Development Methodology, (March), 1-8.

## Appendices

---

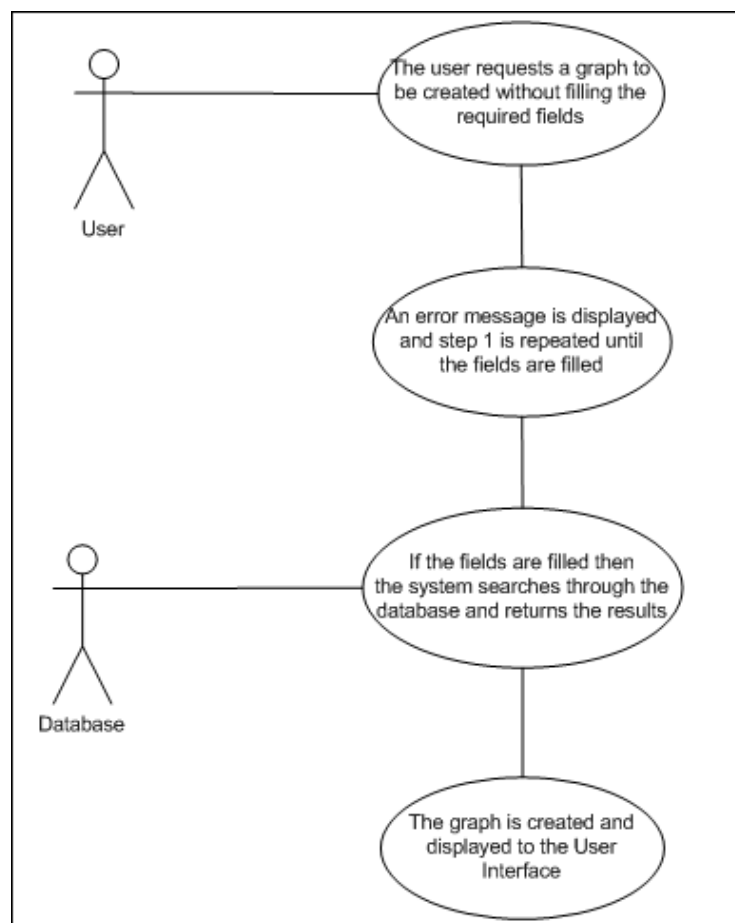
### APPENDIX A – USE CASES

#### Use Case 1: Request for a graph to be created with no fields completed.

Actors: User, Database

Normal Flow:

1. The user requests a graph to be created without filling the required fields.
2. An error message is displayed, and step 1 is repeated until the fields are filled.
3. If the fields are filled then, the system searches through the database and returns the results
4. The graph is created and displayed to the User Interface.

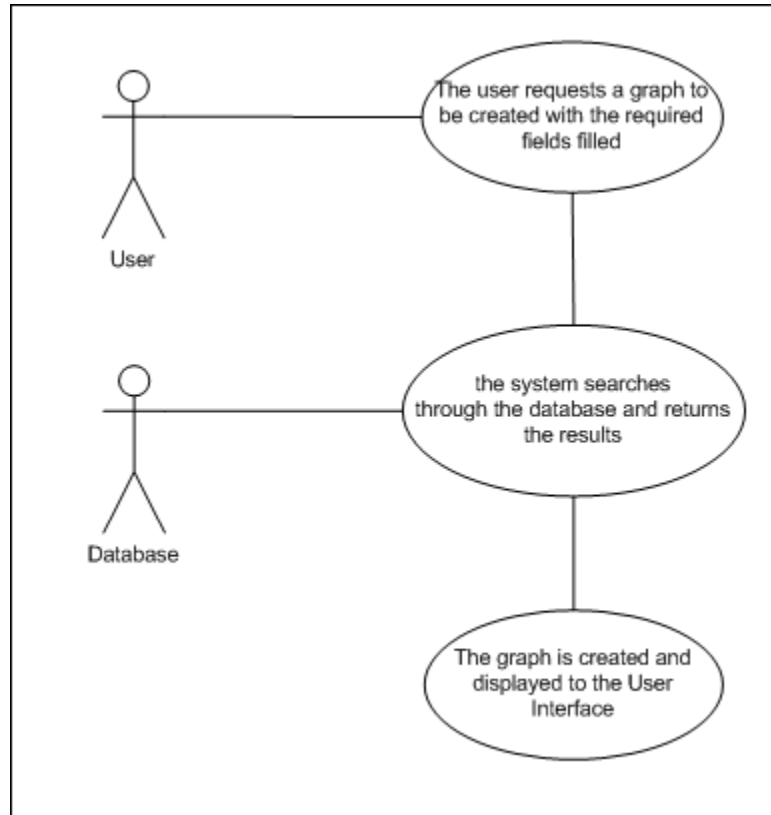


**Use Case 2: Request for a graph with the fields filled.**

Actors: User, Database

Normal Flow:

1. User requests a graph to be created with all the required fields filled.
2. The system searches through the database and returns the results
3. The graph is created and displayed to the User Interface.

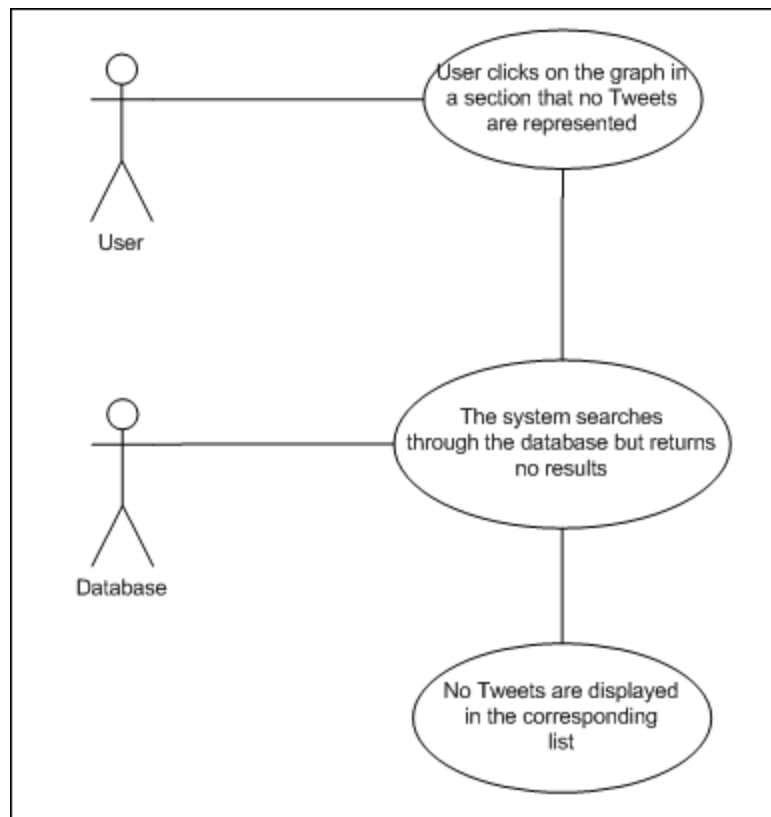


**Use Case 3: Try to click on the graph in a section that no Tweets are represented.**

Actors: User, Database

Normal Flow:

1. User clicks on the graph in a section that no Tweets are represented
2. The system searches through the database but returns no results
3. No Tweets are displayed in the corresponding list

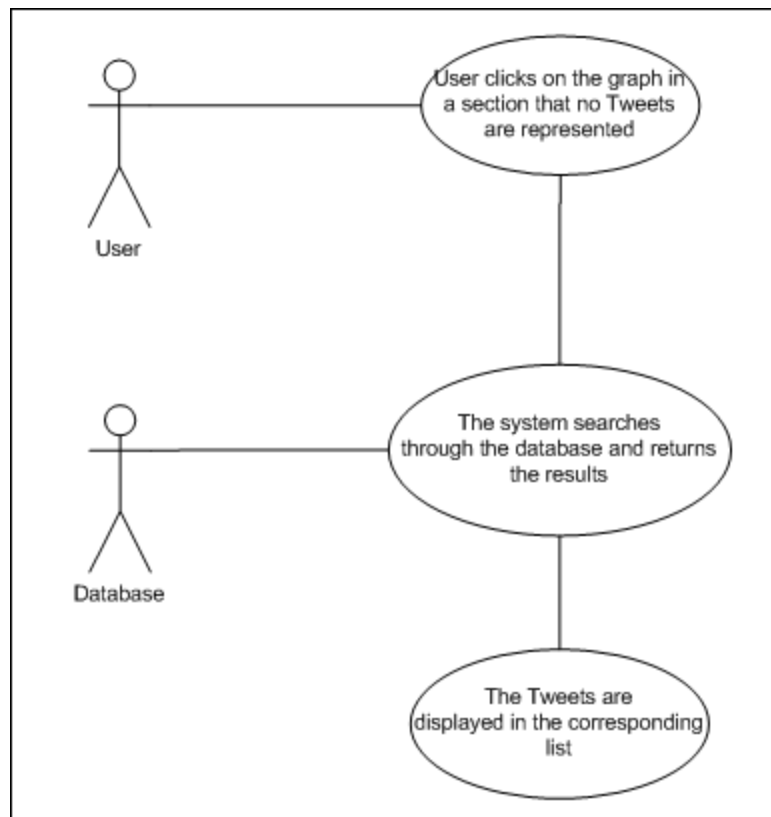


**Use Case 4: Try to click on the graph in a section that Tweets are represented.**

Actors: User, Database

Normal Flow:

1. User clicks on the graph in a section that no Tweets are represented
2. The system searches through the database and returns the results
3. The returned Tweets are displayed in the corresponding list

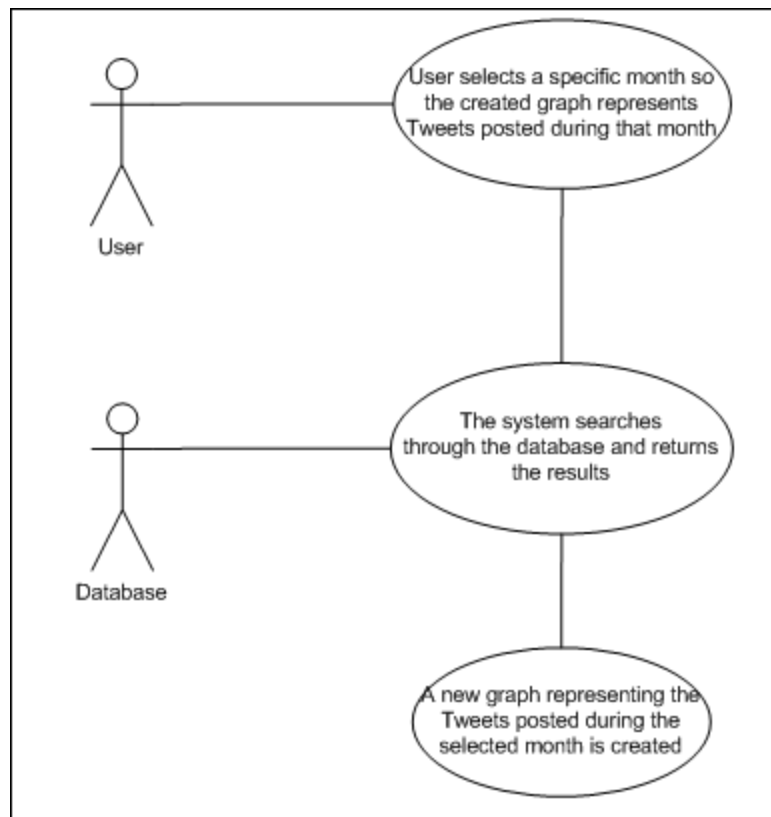


**Use Case 5: Request for a graph that represents on Tweets posted in a single month.**

Actors: User, Database

Normal Flow:

1. User selects a single month, so the created graph represents Tweets posted during that month
2. The system searches through the database and returns the results
3. A new graph representing the Tweets posted during the selected month is created

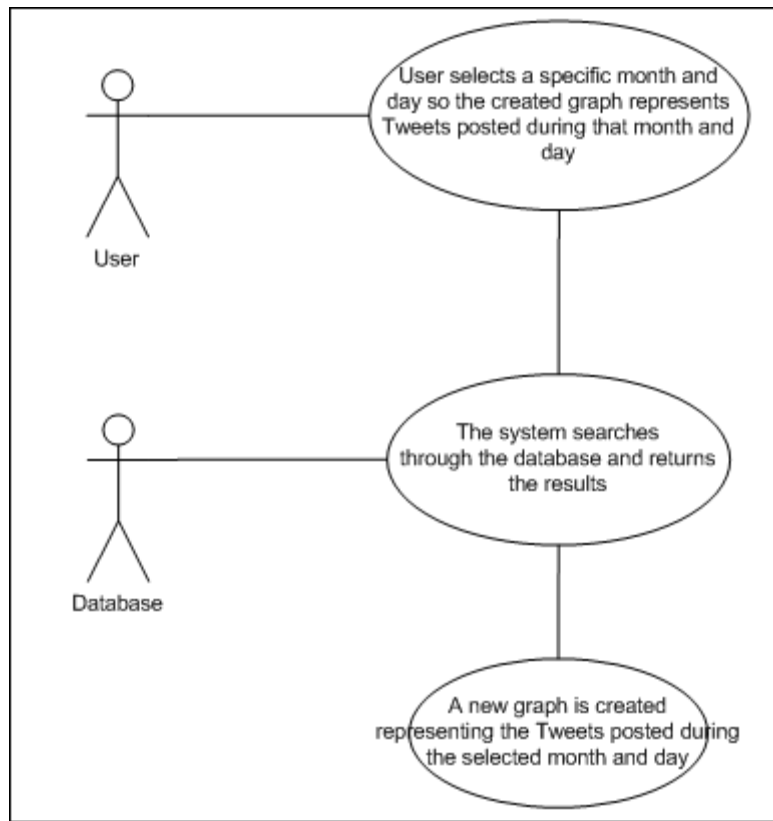


**Use Case 6: Request for a graph that represents on Tweets posted in a single month and day.**

Actors: User, Database

Normal Flow:

1. User selects a single month and day, so the created graph represents Tweets posted during that month and day
2. The system searches through the database and returns the results
3. A new graph is created representing the Tweets posted during the selected month and day



## APPENDIX B - QUESTIONNAIRE

1. How old are you?  
13-18                  19-25                  26-30                  30+
2. Are you familiar with the terms "Social Network" and "Online Social Network"?  
YES                  NO
3. Have you ever been a member of an Online Social Network?  
YES                  NO
4. If yes, mark the Online Social Networks which you are registered to.  
Facebook          Twitter                  MySpace                  Google+                  Other
5. What is the main purpose for using Online Social Network?  
a) Communicating and connecting with existing friends  
b) Networking – Meeting new people  
c) Work (business, school, college etc)  
d) Other
6. Have you ever been familiar with the terms "Social Network Analysis" and "Information Visualization" before using VisualThoughts?  
YES                  NO
7. If yes, have you ever used a Visualization tool before?  
YES                  NO
8. Would you have used the VisualThoughts Application by yourself?  
YES                  NO
9. How many of your friends would have used this application?  
0-2                  2-4                  4-8                  8+
10. Is VisualThoughts Application easy to use?  
YES                  NO
11. What is your opinion according to the performance of the application?  
Great                  Good                  Needs Improvement                  Not Good                  Awful
12. Do you recommend any improvements about the application? If yes, please mention them.  
\_\_\_\_\_  
\_\_\_\_\_