DRIVING INNOVATION IN CRISIS MANAGEMENT USING SOCIAL MEDIA APPS

PROJECT REPORT

Submitted in partial fulfillment of the requirement for the degree of

Bachelor of Technology

In

Computer Science & Engineering

By

Abhay Garg (121231)

Harmeet Singh Dhot ((121234)

Under the Supervision of

Ms. Ruchi Verma

to

Jaypee University of Information and Technology

Waknaghat, Solan – 173234, Himachal Pradesh

Certificate

I hereby declare that the work presented in this report entitled **"Development of prototype in Crisis management Resilience"** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering/Information Technology** submitted in the department of Computer Science & Engineering and Information Technology**,** Jaypee University of Information Technology Waknaghat is an authentic record of my own work carried out over a period from January 2016 to June 2016 under the supervision of **Ms. Ruchi Verma** (**Assistant Professor (Grade-I) Computer Science & Engineering**).
The matter embodied in the report has not been submitted for the award of any other degree or diploma.

Abhay Garg
(121231)

Harmeet Singh Dhot
(121234)

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

Ms. Ruchi Verma
Assistant Professor (Grade-I)
Computer Science & Engineering

Dated:

ACKNOWLEDGEMENT

This project work would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

First and foremost I offer my sincerest gratitude to my supervisor Ms. Ruchi Verma who has supported me throughout my project work with his patience and knowledge whilst allowing me the room to work in my own way. I attribute the level of my Masters degree to her encouragement and effort and without him this thesis, too, would not have been completed or written. One simply could not wish for a better or friendlier supervisor.

The work was performed at Department of Computer Science and Engineering at the University, and I would like to thank all the people there for their hospitality and support.

Last, but not the least, my family and the one above all of us, the omnipresent God.

Date:

ABSTRACT

Social Media has been active in immense from last decade in assorted domains and obviously it is one of the catalyst for crisis management by the live media. In classical way, the social media is being used for broadcasting, sharing, like and distribution of content to the masses and personal groups. Still, this domain of social media and live tweets is under research. In this project work, a unique application and methodology of integrating Twitter Apps with the programming constructs is implemented so that the fetching, cleaning and prediction of the news and its impact on social community can be analyzed. In this project work, the apps of Twitter are developed using Java for fetch the live tweets so that cumulative prediction and detailed interpretation can be done. Twitter is used in this project as key social media as it is used widely for personal, corporate as well as social applications for assorted variety of devices and platforms. Twitter has mobile apps for iPhone, iPad, Android, Windows 10, Windows Phone, BlackBerry, Firefox OS, and Nokia Series. There are also specialized versions of the website available for the mobile devices, SMS and MMS service. For many years, Twitter has limited the use of third party applications accessing the service by implementing a 100,000 user limit per application. It is one of the excellent cases of social media mining or Twitter mining in specific. In this work, the OAuth is used with Twitter. OAuth is one of the effective and widely used concepts to wrap the code with any programming language. In this work the implementation of twitter mining is done using assorted technologies so that the overall sentiment score can be predicted and analyzed.

Keywords - Crisis Management using Social Media, Driving Innovation, Twitter API and Tweets Fetching, Java Based Twitter Integration

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 INTRODUCTION

Twitter describes itself as a "real-time information network," a network that primarily connects one person to many in short, 140-character messages. Using it requires only a free Twitter account and a computer with Internet access (or a phone with SMS capability).

While Twitter has been around since 2006, it really came to prominence when a user posted the first picture of a plane floating on the Hudson River with its passengers standing on the wings. Initially, Twitter was used to tell others what one was doing or thinking. Celebrities, notably Ashton Kutcher, helped it garner public attention for that kind of use. Politicians then used it to tweet their thoughts and make calls for action. Businesses soon followed with introductions of their products, coupons or special Twitter promotions, and integration with other social media messages. Jack–in-the-Box integrated it with a Super Bowl video and Facebook page where Jack was hit by a bus, his condition reported as tweets along with an insider story about a takeover. It has also been used for posting job announcements and for backchanneling during meetings. And today companies are constantly monitoring it for public and customer relations purposes. Clearly, Twitter is an evolving and growing communication medium for a variety of social messaging purposes.

The education community is also using it for delivery to technical sessions and tutorials. Some have used it to post content in small chunks, some post review or test questions, and more are beginning to use it to create conversations (or discussions). Surprising many, Twitter can work quite well as a conversational tool. It forces one to think clearly in order to express thoughts both concisely and clearly. Furthermore, those involved in the conversation are more willing to read others short messages, helping everyone stay focused.

As with any tool, knowing how to use it effectively makes it more valuable. Below are three types of information that should help you master the basics, know where to look for more good information, and use a few advanced techniques to become even more efficient.

Quick Start Guide: This section includes basic instructions for setting up a Twitter account along with some tips for reading (called *following*), posting, forwarding (called *retweeting*), and replying to messages.

Resources: In addition to a list of some basic tweeting terms, this section includes links to third-party apps along with an annotated list of links to videos, tutorials, and blogs.

Advanced Tips: Learning a few advanced tips can often give you a distinct advantage over others. Included here are a few good ones culled from research and users' comments. As you become experienced with Twitter yourself, you're likely to add a few more to your list of favorites.

1.2 SETTING UP A TWITTER ACCOUNT

1. Go to Twitter.com (http://www.twittter.com/) and sign up by entering your name, email address, and a password for the account you are creating. Twitter will display a confirmation page that includes a suggested username along with suggestions for other usernames that are available. You can enter another name if you like; however, for class purposes using your real name is a good idea.

2. Click Create my account to display the Twitter Welcome. You can begin to select those you'd like to follow by clicking on the + Follow button if you like.

3. Click on Next Step: Interests > to select others to follow based on topics that interest you.

4. Click on Next Step: Friends > to find others to follow.Twitter will search through your Yahoo, Gmail, Hotmail, AOL, or LinkedIn contacts to find others to follow. For this class account, click Skip Import > to display the Twitter start page.

5.You could immediately begin tweeting here, but it's also a convenient place to do a few more things. Each of the options is linked to a page that helps you complete that task. If you plan to use only your phone to tweet with text messages, you'll want to set it up next. However, most users prefer tweeting from their desktops, laptops, tablets, or smartphones. But everyone should setup a personalized profile with a picture and short bio. This appears in your public profile, in search results, and more. The picture can easily be added by clicking on the Browse button and pointing to the photo file you want to use. If you describe yourself in your bio as a student at your school name majoring in your major, it'll help others with similar interests, including recruiters, find you. When you are finished, click Save. .

6. By now you may have noticed more tabs at the top of the screen.In addition to the Profile tab, you'll see one for Account. Clicking on it opens a screen with some settings you can adjust for language, time zone, location, and more. If you want to keep your tweets private or if your instructor recommends that you do, check the Tweet Privacy box to Protect my tweets.
. When you are finished, click Save.

1.3 FOLLOWING (READING) AND FINDING POSTS

A good way to start with Twitter is first to read others public posts as well as find what others are posting on topics that interest you.

There are several ways to start reading posts. You can use the gray search box that is embedded at the top of the Twitter start page to find a word, multiple words, or an exact phrase on a topic that interests you. You can also click on one of the trending topics shown in the right sidebar, which you can also change to another location by clicking on the word change. Twitter will display the most recent tweets on the left side, and when you click on a tweet, it'll expand it in the right sidebar.

Here you see labeled in green callouts some typical components of a tweet (a Twitter message). These include the user's picture that was uploaded at sign in (and can be changed at any time in Settings under the Profile tab), the user's full name, and the actual message, including in this case a complete URL to a video. If there is a picture, video, map, or other link in the message, you'll see an icon for it included in the upper right corner of the tweet display.

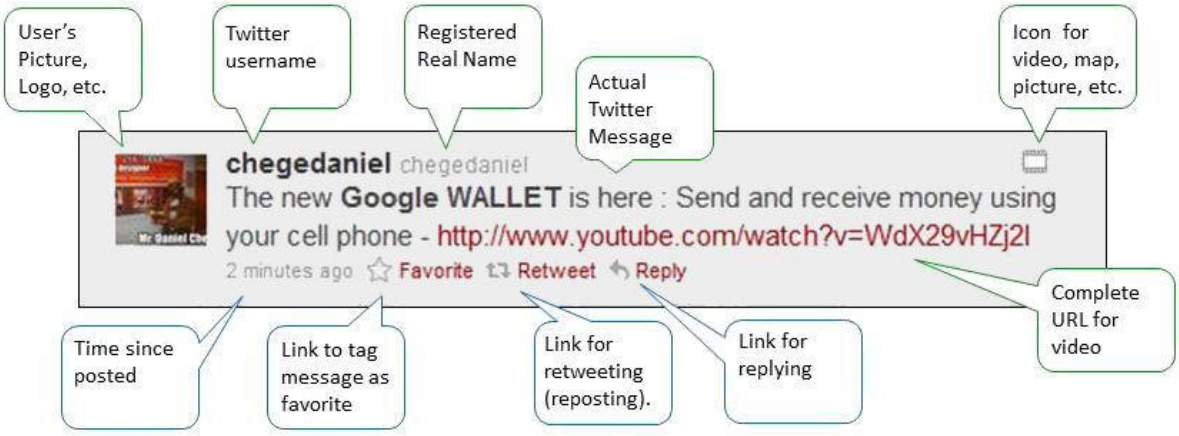


Figure 1.1 –Twitter Tweet Components

Twitter gives you some additional information and options shown here with blue callouts. It tells you how old the post is and gives you the option to save it as a favorite, retweet (repost or forward) it, or reply. If you click on a post, Twitter will expand it in the right sidebar. Below you can see how this post with its video link looks when it is expanded.

Another way to find posts is through the use of a hashtag (#) and another term or abbreviation entered in the search box. In the illustration below, you see the search results from using two hashtags together, #MLB and #Strasburg. The characters used in hashtags are part of the 140 allowed in the message. They can be typed anywhere in the message. Hashtags help you find tweets on the same topic; your instructor will likely assign hashtags when creating a discussion topic. Be sure you spell them exactly right or you won't find the discussion. If one had searched for #MLB #Strassberg, no tweets would have been found or only ones where another person spelled it *Strassberg*, too.

After you are comfortable with finding posts and have read enough of them in your areas of interest to understand conventions used there, you are probably ready to begin posting. Just remember that what you post may become part of the public record. It could even go viral quickly.

1.4 FETCHING LIVE DATA FROM SOCIAL MEDIA

In the same way, the twitter live feeds can be fetched using Python APIs. Using twitter developer account, the new app can be created and then the Python Script is mapped with the Twitter App

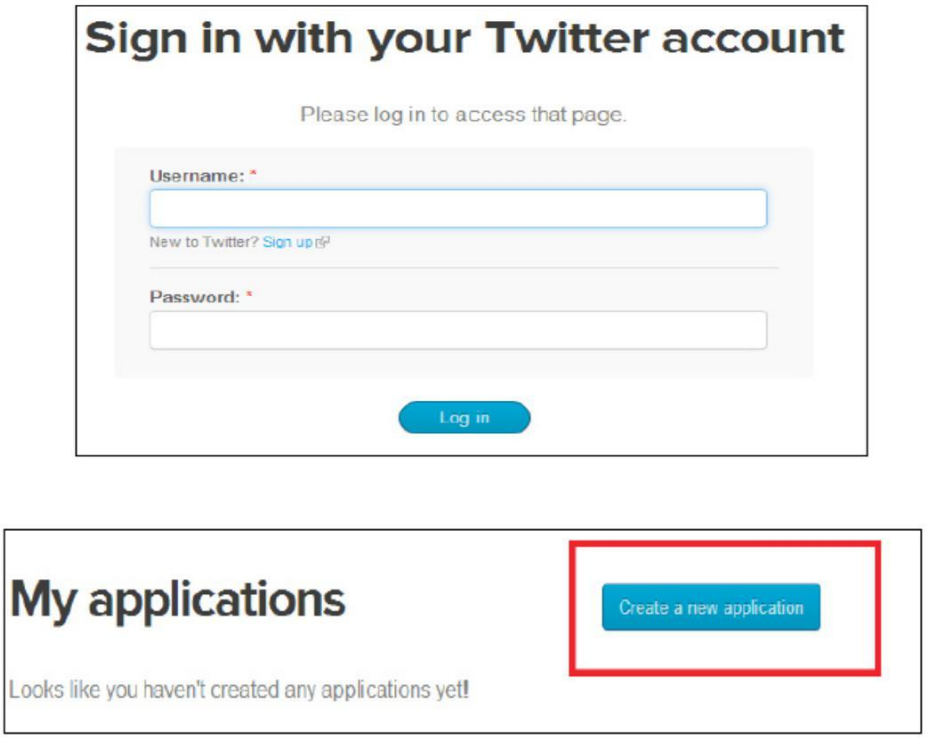


Figure 1.2 – Creating New App in Twitter

Figure 1.3 – Generation of Authentication Tokens from Twitter

Figure 1.4 – Fetching Live Tweets from Twitter in JSON Format

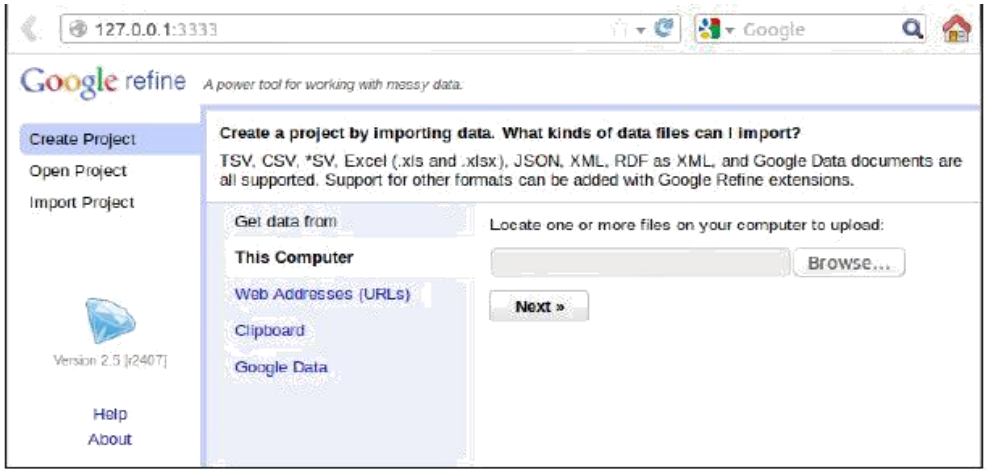


Figure 1.5 – Parsing of JSON using Google Refine

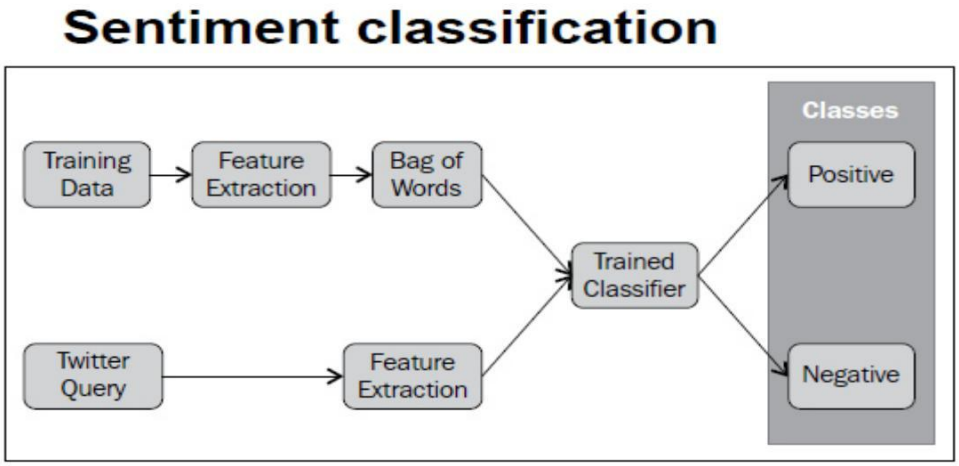


Figure 1.6 – Generation of Sentiments in Different Classes

1.5 WHY THIS TOPIC

As per the international statistical reports from Statista, there are around 1 million new user registrations on Whatsapp. Besides this, 700 million active users present on Whatsapp. Around 30 Billion messages are sent and 34 billion messages are received everyday. If we analyze the statistics of Twitter, 350 Million Tweets daily and more than 500 Million Accounts. There is the huge and rapid growth in the unstructured data every moment. The production and

generation of data is predicted to be 44 times in 2020 as compared to the data in 2009. All these figures and statistical data are amazing and growing in exponential pattern. Such data is unstructured in nature which means the data of different and heterogeneous formats. This concept is classically known as Big Data. The deep investigation of intelligence and meaningful patterns from Big Data is known as Big Data Analytics. A number of researchers and scientists are working in this domain of Big Data using assorted technologies and tools. There are number of approaches by which the live data can be obtained for research and development. One of these approaches is getting data from Open Data Portals. The open data portals provide authentic data sets for research and development in multiple domains. The data sets can be downloaded from these portals in multiple formats including XML, CSV, JSON and many others. In this research work, the sentiment data analysis and prediction shall be done on the big data fetched using Python Scripts. Using prediction tools and algorithms, the effective and accurate results shall be measured.

REAL WORLD APPLICATIONS

- Banking and Finance
- Marketing and Business Policies
- Consumer Behavior
- Market Basket Analysis
- Military and Defense
- Log Files Analysis
- Forensic Investigation
- Bio-Statistics and Bio-Informatics
- Web Usage Mining
- Medical Data Mining
- Expert Systems
- Knowledge Discovery

1.6 POSTING, RETWEETING, AND REPLYING TO TWEETS COLLECTIVELY AND INDIVIDUALLY

While most is done publicly, you can send private messages to individuals and followers. However, those messages could be retweeted, so it's best to think of both your primary reader and perhaps even the public as secondary readers whenever you tweet. The newest version of Twitter also lets you delete messages and undo retweets. Although they may stay in Search for a while, the folks at Twitter say they will clear eventually.

As you know by now, your tweets are limited to 140 characters including spaces, hashtags, and retweeting characters. So writing clearly and concisely is critical. But overdoing the shortcuts at the expense of clarity is a mistake since there are ways to include a link to a longer message if more words are absolutely necessary. However, many users will only read the first part, guessing or skipping the linked material. At the moment Twitter doesn't spellcheck or have predictive technology, so accuracy and completeness are the writer's responsibility, too.

While Twitter's prompt, What's happening?, accurately reflects its form as a real-time communication medium, you can enter whatever you like in the box – a thought, a question, a recommendation. You can also add your location, which might be especially useful to your followers if you were tweeting about the great street pitas at a new Greek restaurant. As you enter the text, Twitter will let you know how many characters are left. Click Tweet when you are ready to post your message. It will be public unless you've set your account up to allow only approved people to follow you. If that's the case, it will only be available to your followers, but any one of them could retweet it.

In the latest version of Twitter, when a message has been retweeted Twitter inserts a retweet icon and the word *by* along with the original tweeters name at the top of the tweet. Until now it's been conventional in third-party apps as well as for long-time tweeters to insert *RT @username* to credit the source. Not only did it become problematic when users retweeted already retweeted messages, but in doing so it also become a part of the character count, limiting the message length.

1.7 ADVANCED TECHNIQUES

Once you are familiar with the basics of tweeting, you might find you'd like easier or better ways to find information, shorten long URLs, tag and view favorite posts, post pictures, or even cross-post to Facebook, LinkedIn, and other social media sites as well as manage tweets and multiple accounts. All of these are possible; some can be done with a use of some Twitter features and others with easy-to-use third-party apps. Here are some brief descriptions of some of these advanced techniques you can try.

Use Advanced Search	https://twitter.com/#!/search-advanced	Go to this Twitter page to refine your tweet search. In addition to fine-tuning the way it searches for words, you can search to, from, and mentions along with locations and a few other options.
Shorten URLs	http://is.gd http://tinyurl.com http://bit.ly	Make more characters available for the message in your tweets by shortening long URLs with one of these easy-to-use apps. Type or paste in the long URL, and the app will immediately generate a new, short one to use.
Tag and View Favorites	http://www.twitter.com	When you are logged in and viewing tweets, you can tag a tweet as a favorite by clicking the star, which will fill with color ( ). To view your tagged tweets, go to your Profile page and select the Favorites tab. All the tweets you've tagged as

		a favorite will be shown there with the most recent one on top.
Post Pictures	http://www.twitpic.com	Post a picture or video with this app. It's easy to use and available on smartphones, making it one of the most widely used photo sharing apps.
Cross-post	http://support.twitter.com/articles/3 1113-how-to-use-twitter-with-facebook	Cross-post to save time and present a consistent message to various readers. Third-party apps are probably the easiest way to set up this feature, but you can do it yourself by following the instructions in Twitter's support. However, at one time, Facebook removed this capability, causing many members to complain. As a result, Facebook reinstituted the feature, but it needs to be enabled in Facebook now.

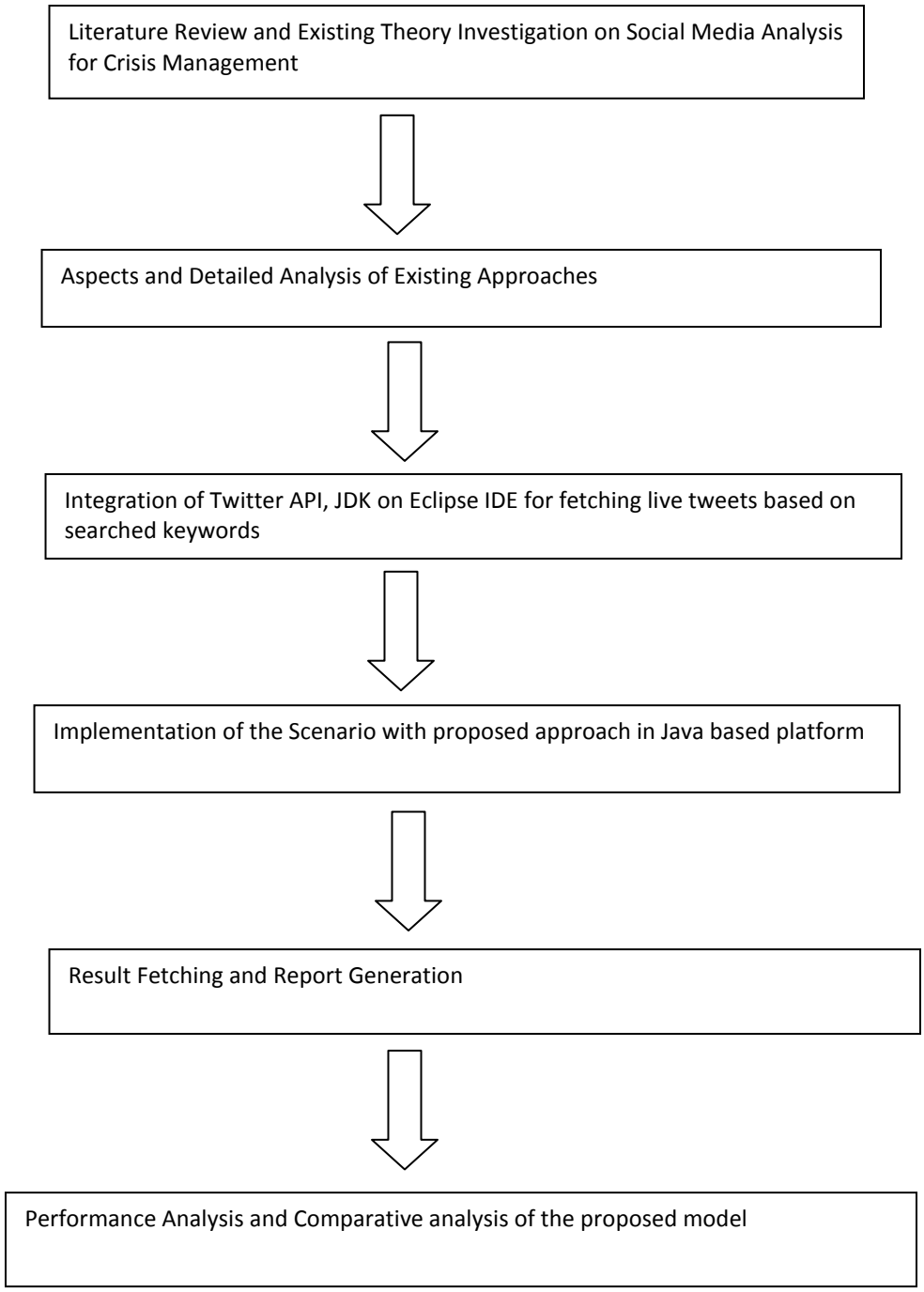
Use Third-party Clients	http://www.hootsuite.com http://www.twhirl.org http://www.tweetdeck.com http://www.socialoomph.com/	Install one of these full-featured clients, which are apps that work with Twitter and integrate many of the advanced features here. Additionally, they include other features for scheduling, cross-posting, managing multiple social media accounts, and more. Some have free and professional versions, so looking them all over to find the one that best meets your needs is important. Once you've selected it, becoming thoroughly familiar with its features will enhance your twittering experience.

1. 8 PROBLEM STATEMENT

- The social media is traditionally used for personal groups, talks and distribution of the thoughts, emotions and related aspects.
- Currently, the news and media agencies are using social media for generation of breaking and crisis related news after fetching the tweets from global locations.
- In this project work, an effective approach is proposed which is being used by the media and government agencies to find out the most interesting patterns from live tweets
- In addition, these live tweets can be used for forensic information including finding out any particular person, event, news or related patterns which are hidden in traditional way.
- Twitter and other social media apps can be used and implemented for detailed investigation and prediction on any event or news.

1.9 OBJECTIVES

1. To perform the deep investigation of live twitter feeds and its impact on the social or related news predictions.
2. To perform a comparative analysis of various APIs used by assorted social media applications
3. To propose and implement an effective algorithm and work that is Java Based Application in association with Twitter APIs for extraction of the live tweets based on the keywords.
4. The tweets will be searched based on hashtag or string with a database connectivity using JDBC to MySQL Database Engine.
5. The keywords density and its effect on overall system shall be analyzed. The parameters of analysis shall be
   - Keywords Density
   - Performance and Efficiency
   - Cost Factor
6. Creation of Java Based GUI for fetching the tweets
   a. Integration of Twitter API Twitter4J
   b. Extraction of Live Tweets for Twitter Portal based on multiple dimensions
      i. Keyword
      ii. Location
7. Implementation of Data Mining Algorithms
   a. Association Rule Mining
   b. Clustering



METHODOLOGY

ORGANIZATION

Chapter 1 highlights and underlines the assorted aspects of social media mining. In this chapter, the introduction to various approaches and technologies for sentiment score are covered. The key focus on Social Media Mining and Prediction is done so that the proposed work can be defended.

The detailed literature review from the research paper, books, journals and conferences are done in **Chapter 2**. In this chapter, the extracts from assorted research papers on Social Media Mining and Prediction are taken and depicted.

Chapter 3 covers the system development which is the key aspect of this work. In this chapter, the proposed model, algorithm and related parameters are emphasized.

The simulation of implementation results with the relative performance analysis is shown in **Chapter 4 and 5**. In this chapter, the simulation results and screenshots are revealed to depict and defend the work

Chapter 6 ends with the detailed conclusion and scope of the future work which guides the upcoming students and research scholars to enhance the current work with higher efficiency and effectiveness.

CHAPTER 2
LITERATURE SURVEY

For completion, justification and solving the problem definition, a number of research papers, magazines, journals and online links are investigated in details.

In this chapter, the details of research papers and journals are specified from where we have analyzed the content and formulated the problem.

A number of research scholars and scientists has written a number of research papers and found excellent results. This section underlines all those research papers and their extracts

Bollen (2010) - Behavioral economics tells us that emotions can profoundly affect individual behavior and decision-making. Does this also apply to societies at large, i.e. can societies experience mood states that affect their collective decision making? By extension is the public mood correlated or even predictive of economic indicators? Here we investigate whether measurements of collective mood states derived from large-scale Twitter feeds are correlated to the value of the Dow Jones Industrial Average (DJIA) over time. This work analyze the text content of daily Twitter feeds by two mood tracking tools, namely OpinionFinder that measures positive vs. negative mood and Google-Profile of Mood States (GPOMS) that measures mood in terms of 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy). This paper cross-validate the resulting mood time series by comparing their ability to detect the public's response to the presidential election and Thanksgiving day in 2008. A Granger causality analysis and a Self-Organizing Fuzzy Neural Network are then used to investigate the hypothesis that public mood states, as measured by the OpinionFinder and GPOMS mood time series, are predictive of changes in DJIA closing values. The results in this paper indicate that the accuracy of DJIA predictions can be significantly improved by the inclusion of specific public mood dimensions but not others. The authors find an accuracy of 87.6% in predicting the daily up and down changes in the closing values of the DJIA and a reduction of the Mean Average Percentage Error by more than 6%.

Bollen (2009) - Microblogging is a form of online communication by which users broadcast brief text updates, also known as tweets, to the public or a selected circle of contacts. A variegated mosaic of microblogging uses has emerged since the launch of Twitter in 2006: daily chatter, conversation, information sharing, and news commentary, among others. Regardless of their content and intended use, tweets often convey pertinent information about their author's mood status. As such, tweets can be regarded as temporally-authentic microscopic instantiations of public mood state. In this article, we perform a sentiment analysis of all public tweets broadcasted by Twitter users between August 1 and December 20, 2008. For every day in the timeline, we extract six dimensions of mood (tension, depression, anger, vigor, fatigue, confusion) using an extended version of the POMS, a well-established psychometric instrument. The authors in this paper compare the results to the values recorded by stock market and crude oil price indices and major events in media and popular culture, such as the U.S. Presidential Election of November 4, 2008 and Thanksgiving Day. This work finds that events in the social, political, cultural and economic sphere do have a significant, immediate on the various dimensions of public mood. The authors speculate that large scale analyses of mood can provide a solid platform to model collective emotive trends in terms of their predictive value with regards to existing social as well as economic indicators.

Asur (2010)– Sentiment Analysis is important part for social networking and content sharing. And yet, the content that is generated from these websites remains largely untapped. In this paper, the authors demonstrate how social media content can be used to predict real-world outcomes. In particular, this work uses the chatter from Twitter.com to forecast box-office revenues for movies. This paper shows that a simple model built from the rate at which tweets are created about particular topics can outperform market-based predictors. This work further demonstrates how sentiments extracted from Twitter can be further utilized to improve the forecasting power of social media.

Tan (2011)– The authors show that information about social relationships can be used to improve user-level sentiment analysis. The main motivation behind the approach is that users that are somehow “connected” may be more likely to hold similar opinions; therefore, relationship

information can complement what we can extract about a user's viewpoints from their utterances. Employing Twitter as a source for our experimental data, and working within a semi-supervised framework, we propose models that are induced either from the Twitter follower/followee network or from the network in Twitter formed by users referring to each other using "@" mentions. The proposed transductive learning results reveal that incorporating social-network information can indeed lead to statistically significant sentiment classification improvements over the performance of an approach based on Support Vector Machines having access only to textual features.

Saif (2012) - Sentiment analysis over Twitter offer organisations a fast and effective way to monitor the publics' feelings towards their brand, business, directors, etc. A wide range of features and methods for training sentiment classifiers for Twitter datasets have been researched in recent years with varying results. In this paper, we introduce a novel approach of adding semantics as additional features into the training set for sentiment analysis. For each extracted entity (e.g. iPhone) from tweets, we add its semantic concept (e.g. "Apple product") as an additional feature, and measure the correlation of the representative concept with negative/positive sentiment. The authors apply this approach to predict sentiment for three different Twitter datasets. The results show an average increase of F harmonic accuracy score for identifying both negative and positive sentiment of around 6.5% and 4.8% over the baselines of unigrams and part-of-speech features respectively. We also compare against an approach based on sentiment-bearing topic analysis, and find that semantic features produce better Recall and F score when classifying negative sentiment, and better Precision with lower Recall and F score in positive sentiment classification.

Davidov (2010) - Automated identification of diverse sentiment types can be beneficial for many NLP systems such as review summarization and public media analysis. In some of these systems there is an option of assigning a sentiment value to a single sentence or a very short text. In this paper the authors proposes a supervised sentiment classification framework which is based on data from Twitter, a popular microblogging service. By utilizing 50 Twitter tags and 15 smileys as sentiment labels, this framework avoids the need for labor intensive manual annotation, allowing identification and classification of diverse sentiment types of short texts. The authors

evaluate the contribution of different feature types for sentiment classification and show that the proposed framework successfully identifies sentiment types of untagged sentences. The quality of the sentiment identification was also confirmed by human judges. This paper also explores dependencies and overlap between different sentiment types represented by smileys and Twitter hashtags.

Saif (2012) - Twitter has brought much attention recently as a hot research topic in the domain of sentiment analysis. Training sentiment classifiers from tweets data often faces the data sparsity problem partly due to the large variety of short and irregular forms introduced to tweets because of the 140-character limit. In this work the authors proposes using two different sets of features to alleviate the data sparseness problem. One is the semantic feature set where this work extracts semantically hidden concepts from tweets and then incorporate them into classifier training through interpolation. Another is the sentiment-topic feature set where we extract latent topics and the associated topic sentiment from tweets, then augment the original feature space with these sentiment-topics. Experimental results on the Stanford Twitter Sentiment Dataset show that both feature sets outperform the baseline model using unigrams only. Moreover, using semantic features rivals the previously reported best result. Using sentiment topic features achieves 86.3% sentiment classification accuracy, which outperforms existing approaches.

Bifet (2009) - Micro-blogs are a challenging new source of information for data mining techniques. Twitter is a micro-blogging service built to discover what is happening at any moment in time, anywhere in the world. Twitter messages are short, and generated constantly, and well suited for knowledge discovery using data stream mining. The authors briefly discuss the challenges that Twitter data streams pose, focusing on classification problems, and then consider these streams for opinion mining and sentiment analysis. To deal with streaming unbalanced classes, this work propose a sliding window Kappa statistic for evaluation in time-changing data streams. Using this statistic this work performs a study on Twitter data using learning algorithms for data streams.

Social networking, blogging and online forums have turned the web into a vast repository of comments on many topics, generating a potential source of information for social science

research (Thelwall, Wouters, & Fry, 2008). The availability of large scale electronic social data from the web and elsewhere is already transforming social research (Savage & Burrows, 2007). The social web is also being commercially exploited for goals such as automatically extracting customer opinions about products or brands. An application could build a large database of web sources use information retrieval techniques to identify potentially relevant texts, then extract information about target products or brands, such as which aspects are disliked.From a social sciences perspective, similar methods could potentially give insights into public opinion about a wide range of topics and are unobtrusive, avoiding human subjects research issues

The sheer size of the social web has also made possible a new type of informal literature-based discovery .The ability to *automatically* detect events of interest, perhaps within pre-defined broad topics, by scanning large quantities of web data. For instance, one project used time series analyses of (mainly) blogs to identify emerging public fears about science and businesses can use similar techniques to quickly discover customer concerns. Emerging important events are typically signalled by sharp increases in the frequency of relevant terms. These bursts of interest are important to study because of their role in detecting new events as well as for the importance of the events discovered. One key unknown is the role of sentiment in the emergence of important events because of the increasing recognition of the importance of emotion in awareness, recall and judgement of information as well as motivation associated with information behaviour .

The research field of sentiment analysis, also known as opinion mining, has developed many algorithms to identify whether an online text is subjective or objective, and whether any opinion expressed is positive or negative (Pang & Lee, 2008). Such methods have been applied on a large scale to study sentiment-related issues. One widely-publicised study focused on the average level of sentiment expressed in blogs (as well as lyrics and US presidential speeches) in order to

identify overall trends in levels of happiness as well as age and geographic differences in the expression of happiness. A similar approach used Facebook status updates to judge changes in mood over the year and to assess "the overall emotional health of the nation" and another project assessed six dimensions of emotion in Twitter, showing that these typically reflect significant offline events. Nevertheless, despite some research into the role of sentiment in online communication, there are no investigations into the role that sentiment plays in important online events. To partially fill this gap, this study assesses whether Twitter-based surges of interest in an event are associated with increases in expressed strength of feeling. Within the media, a well-established notion is that emotion is important in engaging attention, as expressed for violence by the common saying, "if it bleeds, it leads" and through evidence that audiences emotionally engage with the news. It seems logical, therefore, to hypothesise that events triggering large reactions in Twitter would be associated with increases in the strength of expressed sentiment, but there is no evidence yet for this hypothesis.

According to Alexa, and based upon its panel of toolbar users, Twitter had become the world's ninth most popular web site by October 2010 despite only beginning in July 2006. The rapid growth of the site may be partly due to celebrities tweeting regular updates about their daily lives Also according to Alexa, amongst Internet users people aged 25-44 were slightly over-represented in Twitter and those aged 55+ were much less likely to use it than average; women were also slightly over-represented. Thus, despite the mobile phone connection Twitter is not a teen site, at least in the US: "Teens ages 12-17 do not use Twitter in large numbers, though high school-aged girls show the greatest enthusiasm"

2.1 INFORMATION DISSEMINATION AND SOCIALISING WITH TWITTER

Twitter can be described as a microblog or social network site. It is for microblogging because the central activity is posting short status update messages (tweets), via the web or a mobile phone. Twitter is also a social network site because members have a profile page with some personal information and can connect to other members by "following" them, thus gaining easy

access to their content. It seems to be used to share information and to describe minor daily activities although it can also be used for information dissemination, for example by government organisations. About 80% of Twitter users update followers on what they are currently doing, whilst the remainder have an informational focus.There are clear differences between users in terms of connection patterns: although most seem to be symmetrical in terms of having similar numbers of followers to number of users followed, some are heavily skewed, suggesting a broadcasting or primarily information gathering/evangelical function.Twitter displays a low reciprocity in messages between users, unlike other social networks, suggesting that its primary function is not as a social network, but perhaps to spread news (including personal news) or other information instead.

An unusual feature of Twitter is retweeting: forwarding a tweet by posting it again. The purpose of this is often to disseminate information to the poster's followers, perhaps in modified form and this reposting seems to be extremely rapid. The reposting of the same (or similar) information works because members tend to follow different sets of people, although retweeting also serves other purposes such as helping followers to find older posts. The potential for information to flow rapidly through Twitter can also be seen from the fact that the average path length between a pair of users seems to be just over 4. Moreover, if retweeted, a tweet can expect to reach an average of 1000 users.Nevertheless, some aspects of information dissemination are not apparent from basic statistics about members. For instance the most followed members are not always the most influential, but topic focus within a Twitter account helps to generate genuine influence Moreover, an important event can be expected to trigger more informational tweeting, which suggests that it would be possible to detect important events through the automatic analysis of Twitter. In support of this, Twitter commentaries have been shown to sometimes quite closely reflect offline events, such as political deliberations.

Another communicational feature of Twitter is the hashtag: a metatag beginning with # that is designed to help others find a post, often by marking the Tweet topic or its intended audience

This feature seems to have been invented by Twitter users, in early 2008 The use of hashtags emphasises the importance of widely communicating information in Twitter. In contrast, the @ symbol is used to address a post to another registered Twitter user, allowing Twitter to be used quite effectively for conversations and collaboration (Honeycutt & Herring, 2009). Moreover, about 31% of Tweets seem to be directed at a specific user using this feature emphasising the social element of Twitter rather than the information broadcasting function associated with hashtags.

Twitter use as an information behaviour: The affective dimension

Whilst the above subsection describes Twitter and usage patterns, responding to an external event by posting a tweet is information behaviour, and therefore has an affective component, in the sense of judgements or intentions, irrespective of whether the information used is subjective Of particular interest here is whether individuals encode sentiment into their messages: a topic that appears to have attracted little research.

A useful theoretical construct for understanding how people may react to events is the concept of *affordances*, as also used by Nahl..Instead of focusing on the ostensible purpose or function of something, it also makes sense to consider what uses can be made of it to suit the goals of the person concerned. For a use to occur, however, its potential must be first perceived. In the context of Twitter, this suggests that an event reported in the media may be perceived by some Twitter users as affording an opportunity to satisfy unrelated goals, such as to create humour, show analytical skill or declare a moral perspective. Hence, whilst an emotional event might seem likely to elicit intuitive reactions, such as declarations of pleasure or disgust, this is not inevitable. This analysis aligns with the uses and gratifications approach from media studies which posits that people do not passively consume the media but actively select and exploit it for their own goals. Borrowing an idea from computer systems interface design, it seems that non-obvious affordances need a culture of use to support them and so if there is a culture of using information in non-obvious ways for Twitter posts then this culture can be passed on by its

originators to other users and could become the main explanation for the continuation of the practice.

2.2 SENTIMENT ANALYSIS OF ONLINE TEXT

Sentiment analysis is useful for research into online communication because it gives researchers the ability to automatically measure emotion in online texts. The research field of sentiment analysis has developed algorithms to automatically detect sentiment in text Whilst some identify the objects discussed and the polarity (positive, negative or neutral) of sentiment expressed about them other algorithms assign an overall polarity to a text, such as a movie review Three common sentiment analysis approaches are full-text machine learning, lexicon-based methods and linguistic analysis. For standard machine learning a set of texts annotated for polarity by human coders are used to train an algorithm to detect features that associate with positive, negative and neutral categories. The text features used are typically sets of all words, word pairs and word triples found in the texts. The trained algorithm can then look for the same features in new texts in order to predict their polarity The lexicon approach starts with lists of words that are pre-coded for polarity and sometimes also for strength and uses their occurrence within texts to predict their polarity A linguistic analysis, in contrast, exploits the grammatical structure of text to predict its polarity, often in conjunction with a lexicon. For instance, linguistic algorithms may attempt to identify context, negations, superlatives and idioms as part of the polarity prediction process In practice, algorithms often employ multiple methods together with various refinements, such as pre-filtering the features searched for, and methods to cope with changes in data over time

A few algorithms detect sentiment strength in addition to sentiment, including some for informal online text.These work on the basis that humans can differentiate between mild and strong emotions in text. For instance, hate may be regarded as a stronger negative emotion than

dislike. Sentiment strength algorithms attempt to assign a numerical value to texts to indicate the strength of any sentiment detected.

In addition to academic research, sentiment analysis is now a standard part of online business intelligence software, such as Market Sentinel's Skyttle and sysomos's Map. The direct line provided by Twitter between customer opinions and businesses has potentially valuable implications for marketing as a competitive intelligence source.There are also now web sites offering free sentiment analysis for various online data sources, including tweetfeel and Twitter Sentiment.

CHAPTER 3

SYSTEM DEVELOPMENT

PROPOSED MODEL

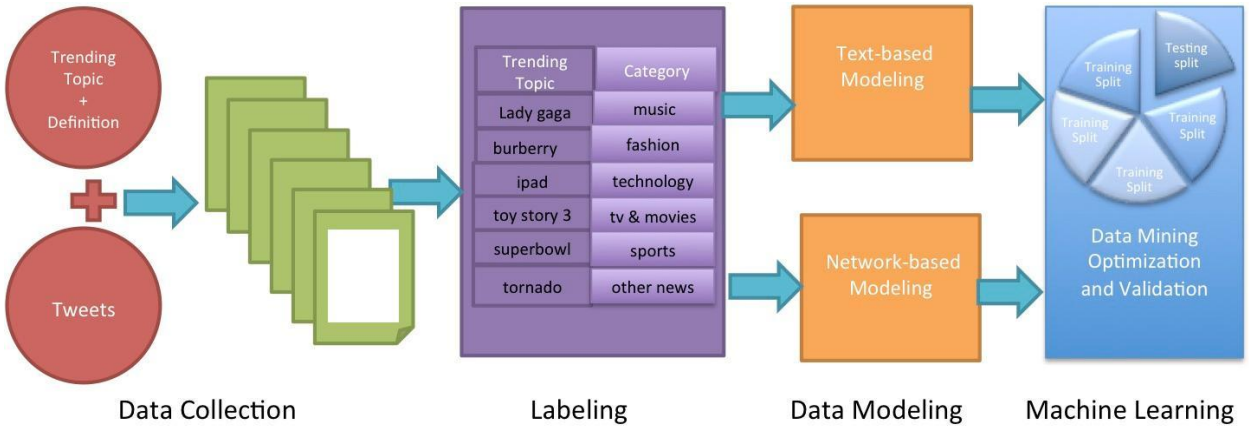


Figure 2.1 – Machine Learning and Data Modeling Process

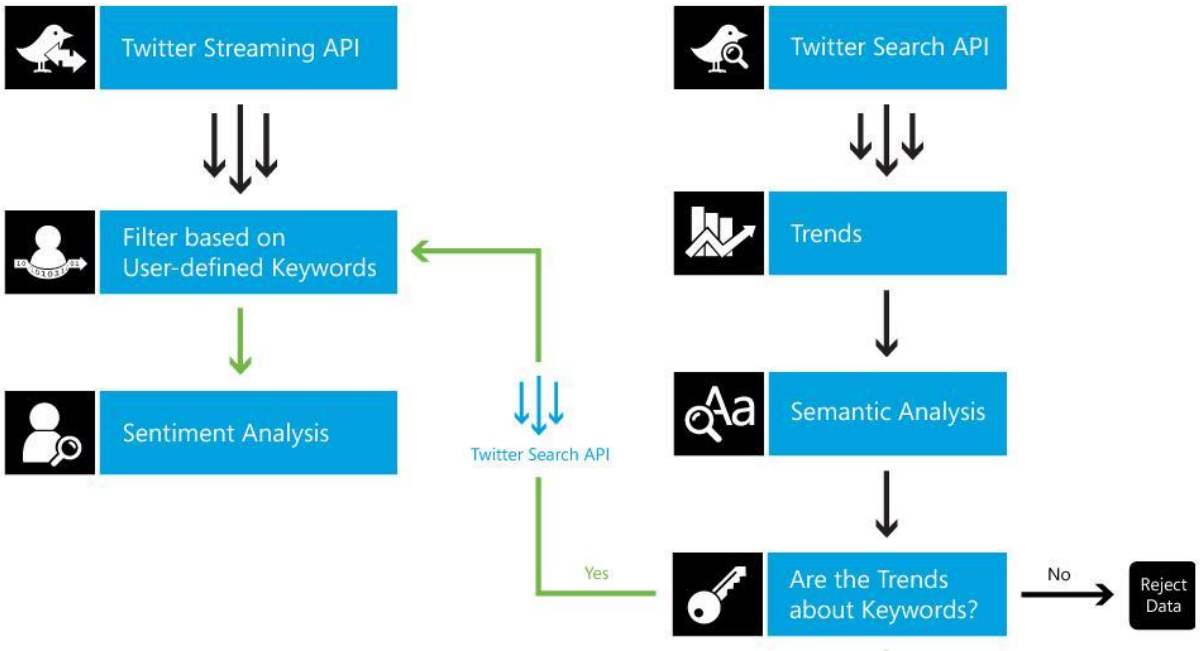


Figure 2.2 – Sentiment Analysis

3.1 HARDWARE REQUIREMENTS

The minimum requirements needed to perform operations are

- Intel Pentium Processor at 2 GHz or Higher
- RAM 256MB or more
- Hard disk capacity 10GB or more

3.2 SOFTWARE REQUIREMENTS

The software required to perform the implementation are

- Windows or Linux Operating System (Ubuntu, Fedora)
- Advance Java
- Twitter APIs
- AJAX
- MySQL Database Engine
- Notepad++
- Twitter4J
- Eclipse IDE
- GEPHI
- JSON (JavaScript Object Notation)
- WEKA - Data Mining and Machine Learning Tool

CHAPTER 4

PERFORMANCE ANALYSIS

Many social networks and apps have their own interface that programmers can work with. These interfaces are called APIs (short for Application Programming Interface).

Following is the results of Twitter-Java integration

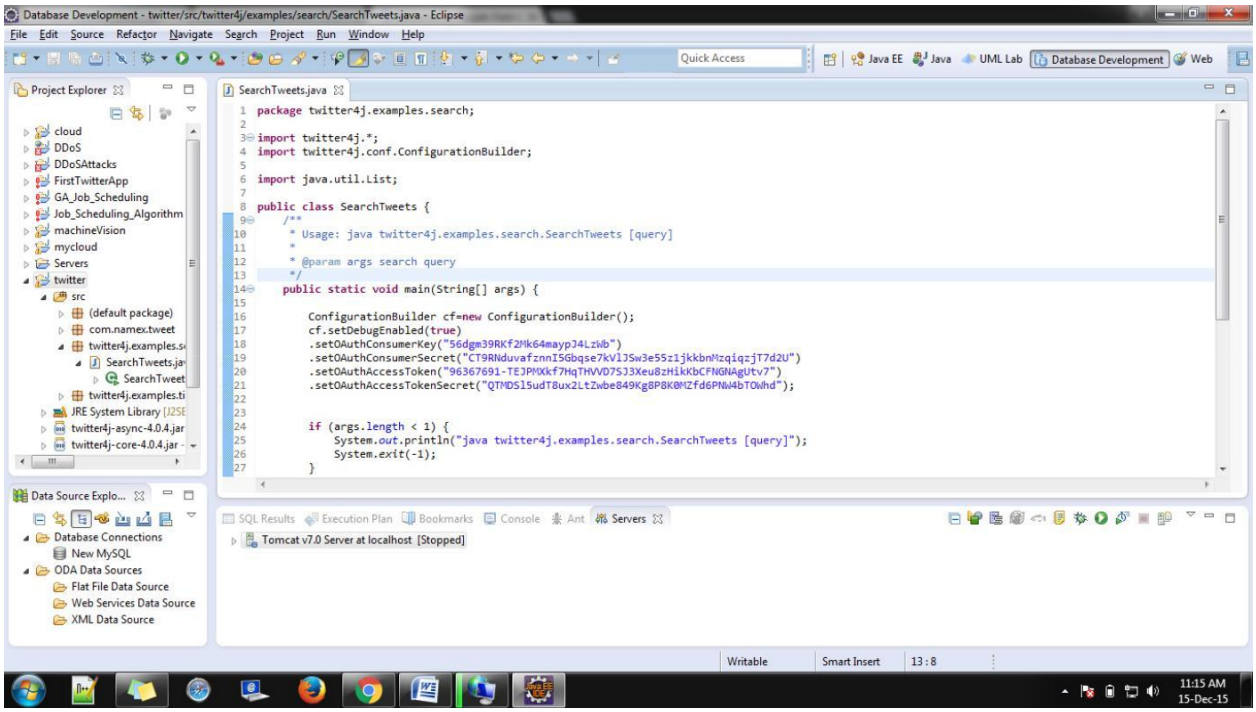


Figure 4.1 – Machine Learning and Data Modeling Process

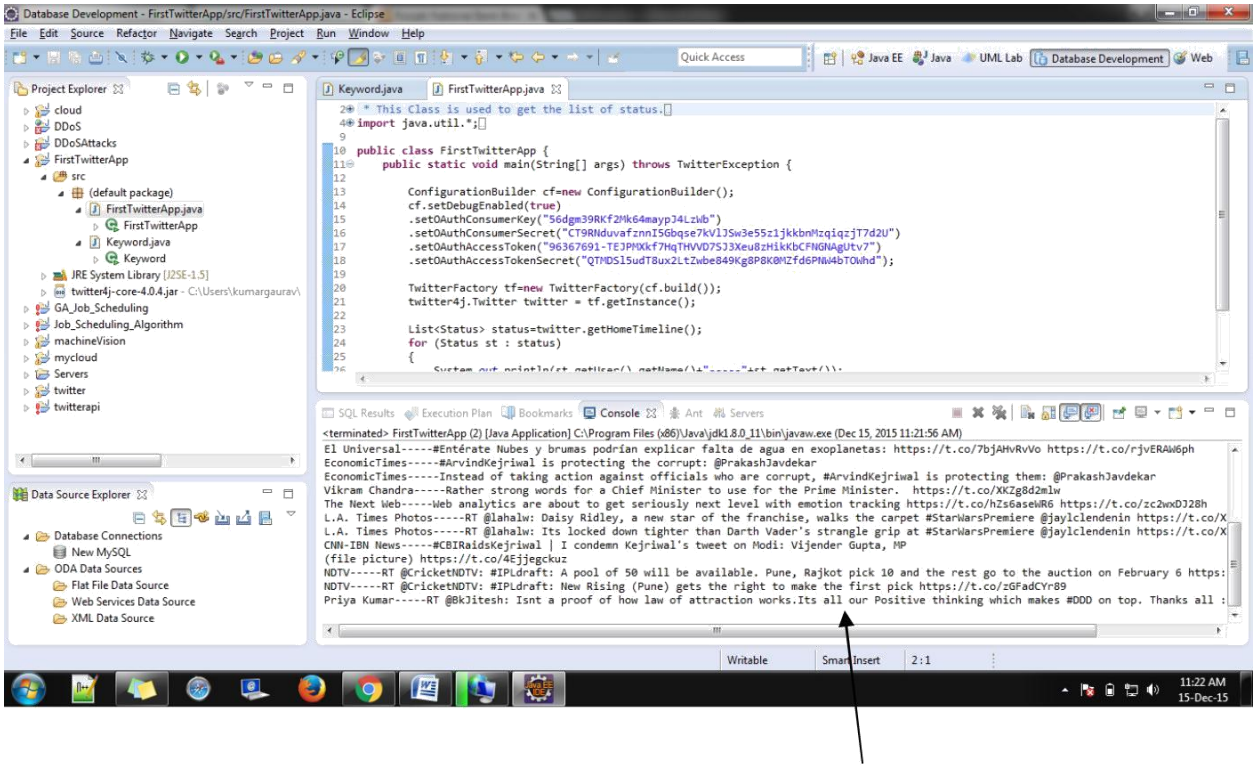


Figure 4.2 - Fetching of Tweets in Java Console of Eclipse IDE

4.1 CONFIGURATION OF TWITTER4J

Twitter4J is an unofficial Java library for the Twitter API.

With Twitter4J, you can easily integrate your Java application with the Twitter service. Twitter4J is an unofficial library.

Twitter4J is having following features -

- 100% Pure Java - works on any Java Platform version 5 or later
- Android platform and Google App Engine ready
- Zero dependency : No additional jars required
- Built-in OAuth support
- Out-of-the-box gzip support
- 100% Twitter API 1.1 compatible

4.2 GEPHI

A network is made of two components : a **list of the actors** composing the network, and a **list of the relations** (the interactions between actors). As part of a mathematical object, actors will then be called *vertices* (***nodes***, in Gephi), and relations will be denoted as *tiles* (***edges***, in Gephi).

Figure 4.3 - Gephi Integration for Social Network Analysis

Here left, a **very simple directed social graph, with both lists explicited**. Two attributes are attached to the nodes : a **label** (his or her "name") and a numeric attribute (here, a distinction between boys and girls). In the edge list, "Source" and "Target" entries refer to the nodes' identifiers (Id).

In our example, the attribute determines the color of the nodes. The size of a node depends on the value of its "degree centrality" (its number of connexions). The **centrality measures** are essential metrics to analyze the position of an actor in a network. They come in many variations, as shown at right (A = *Degree centrality*, number of connexions ; B = *Closeness centrality*, closeness to the entire network ; C = *Betweenness centrality*, bridges nodes ; D = *Eigenvector centrality*, connexion to well-connected nodes).

Step 1: Import the Data Into Gephi

1. Download the file "little-review-1918-09.csv" onto your computer (http://www.modernist-magazines.org/content/downloads). Be sure to place it where you'll remember to find it (i.e. on the Desktop or in a subfolder of Documents).
2. Start Gephi and go to File > Open...
3. Browse to little-review-1918-09.csv and open it.
4. A dialogue window will ask if you want certain parameters, so make sure Graph Type is

set to Directed, and that Auto-scale, Create missing nodes, and Add full graph are all checked. These options should already be set as the default.

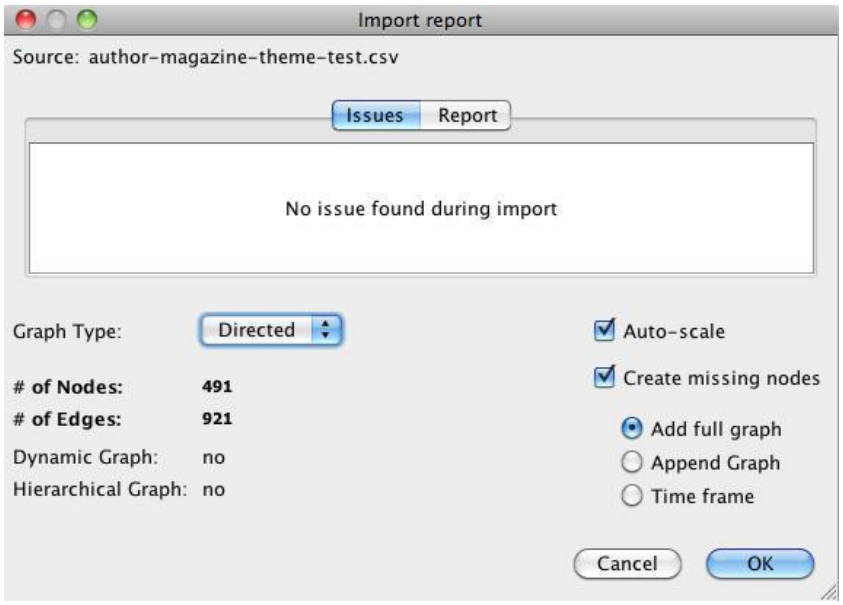


Figure 4.4 – Data Import in Gephi

5. Once the dialogue appears as in this picture, click OK.

Step 2: Manipulating the Graph and Adding Labels

- Once the data are imported, you'll see a square-ish rat's nest of black dots and lines, some of which are darker than others. Mouse over them to see how their immediate neighbors become highlighted within the graph. These dots represent different nodes within the data (such as author names and magazine titles), and show you the connecting links to other data.
- You'll notice that Gephi doesn't immediately tell you what these dots are, so we need to make their labels visible. Click on the little rectangular arrow at the bottom right of the Graph area to open some options.
- Once the options pane opens across the bottom, click on the Labels tab and then on the

checkbox next to Node.

- After a moment, the Graph area will fill with large text, which you should make smaller by clicking on the Font button (under the Node checkbox) and selecting a smaller font size like 8pt.
- Now that the text is smaller, try zooming in to get a closer look, or pulling nodes around, and get a sense of what's in the network model.
- To try a different layout, go to the Layout panel in the left sidebar, select one of the options (such as Yifan Hu) and then click on the Run button.

Step 3: Adding Interpretive Features with Node Size and Color

Gephi can change the visual aspects of its nodes and edges to represent the levels of connectedness among the nodes. For example, the nodes that have more connections, such as the magazine *BLAST*, will appear larger or in different colors to provide a more immediate conception of where the dominant nodes in the network lie. These features are based on statistical properties that Gephi will compute.

1. In the right sidebar, click on the Statistics tab. If it is not there, add it by clicking on the Window menu at the top of the screen and then selecting Statistics.
2. Next to Average degree, click on the Run button. You will see some output about the degree densities of the graph.
3. Now, in the left sidebar, click on the Ranking tab, and then on the Nodes sub-tab.
4. To the right of the Nodes subtab, the Color palette should already be selected.
5. Select Degree from the dropdown menu and then click on the Apply button. The nodes should now appear on a scale from red to black (or whatever colors were pre-selected on your version of Gephi).
6. Click on the Weight button (looks like a red diamond) to the right of the color palette.
7. Select Degree again, leave the default size parameters in place, and click on the Apply button.
8. Some of the nodes should now appear larger.

Step 4: Creating Ego Networks

The large-scale graph can be very interesting. However, we might wish to drill down and look at smaller networks in order to enhance our thinking about a particular author, magazine, or theme. Filtering for an ego network (a network centered on a single node) is one way to do this.

- In the right sidebar, click on the Filters tab.
- Expand the Topology folder.
- Drag Ego Network down to the Queries area and then click on it.
- In the Node ID field at the bottom, type in a term from our spreadsheet that you wish to examine, such as BLAST.
- Click on OK and then on Filter to start the Ego Network.
- You will then see a smaller network for those nodes most immediately connected to BLAST.

You can then adjust the number of degrees out from BLAST shown in the network by selecting 1, 2, 3, or Max from the Depth menu under the Node ID

CHAPTER 5

IMPLEMENTATION RESULTS

- Download Twitter4J from http://twitter4j.org/archive/twitter4j-4.0.4.zip

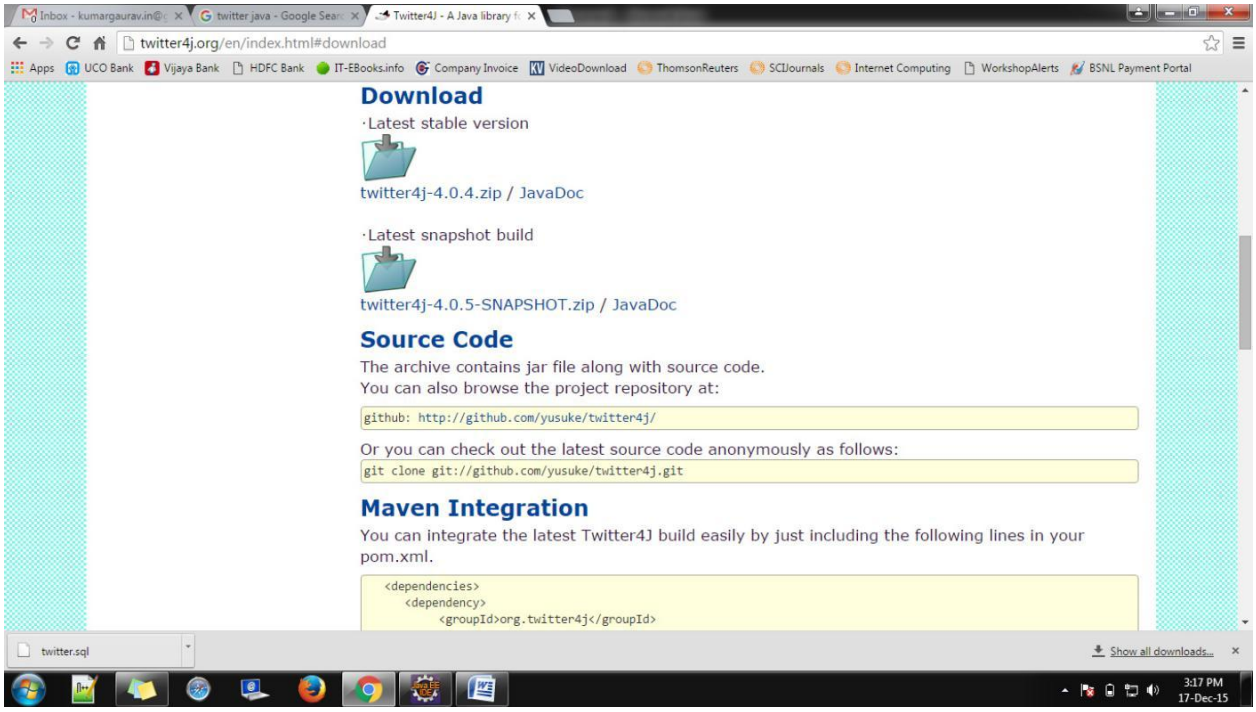


Figure 5.1 – Twitter4J for Twitter Mining

- Uncompress and copy to any location
- In Eclipse, Create a New Project
- Right Click the Project Name, Select Build Path -> Configure Build Path

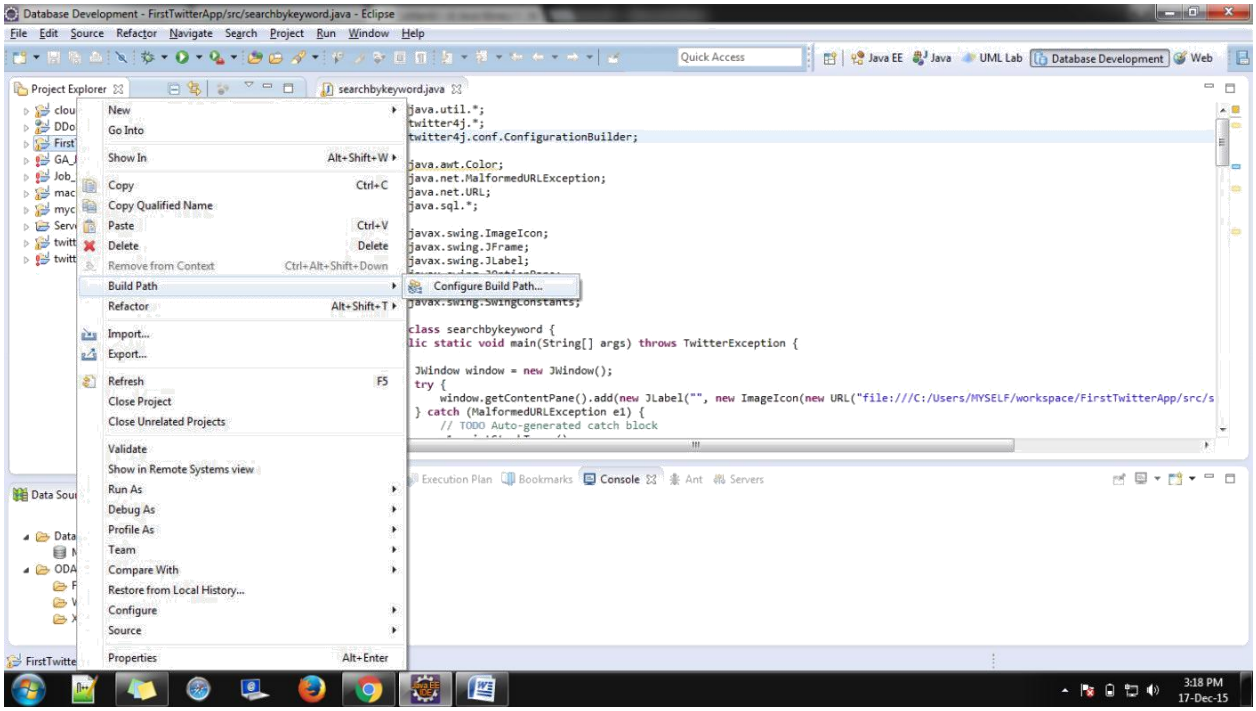


Figure 5.2 – Eclipse IDE Setup and Configuration with Twitter4J

GoTo Libraries, click Add External JAR and give the path of all JAR files in lib

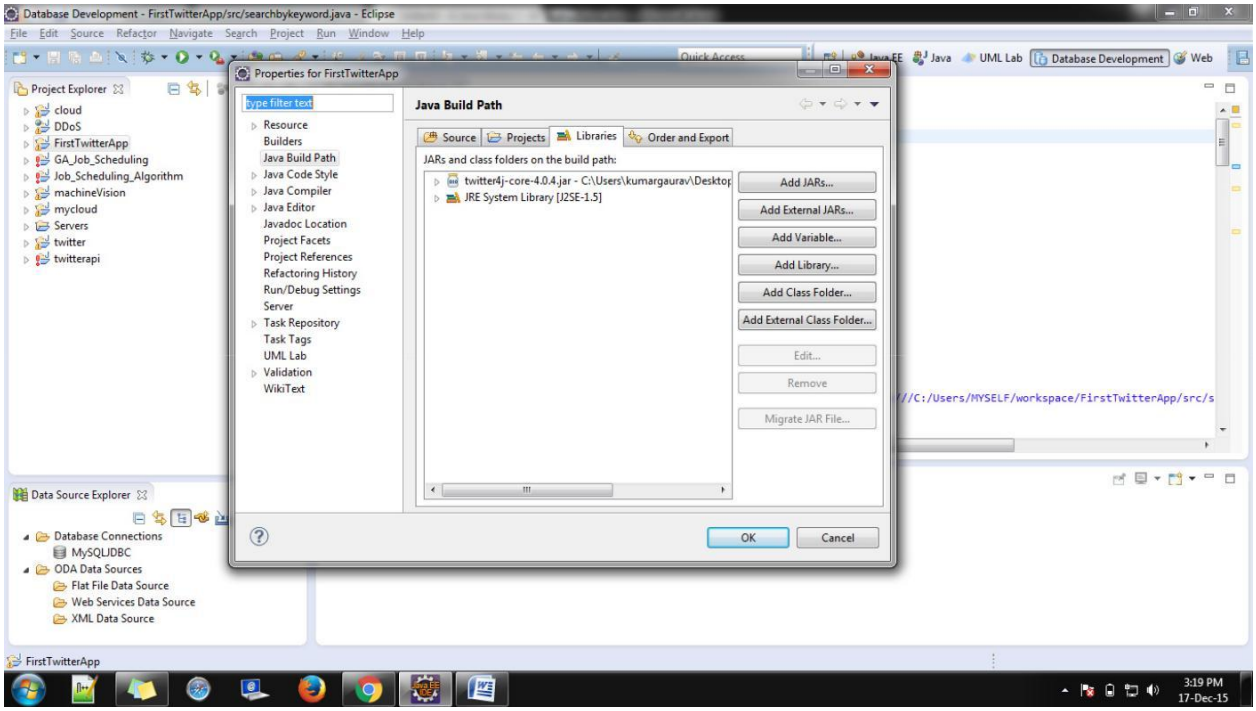


Figure 5.3 – Build Path Settings for JAR files

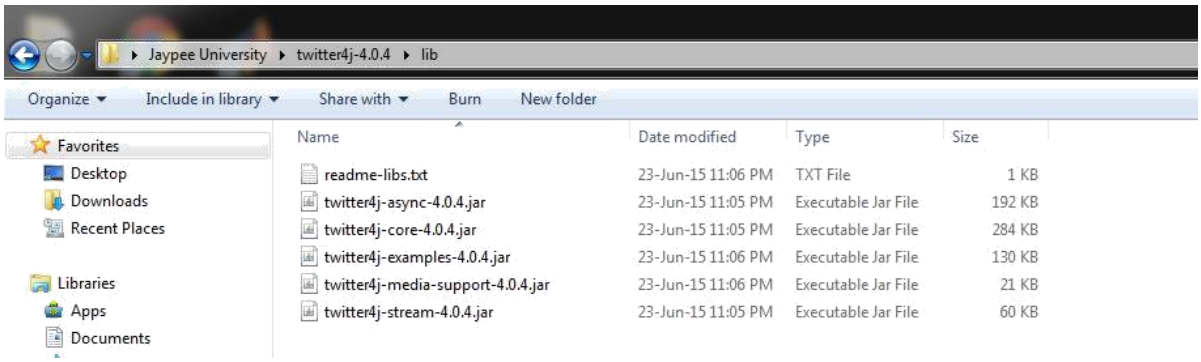


Figure 5.4 – Libraries in Twitter4J

Download and Install WAMPSERVER 2.1 from -> http://sourceforge.net/projects/wampserver/files/WampServer%202/WampServer%202.2/

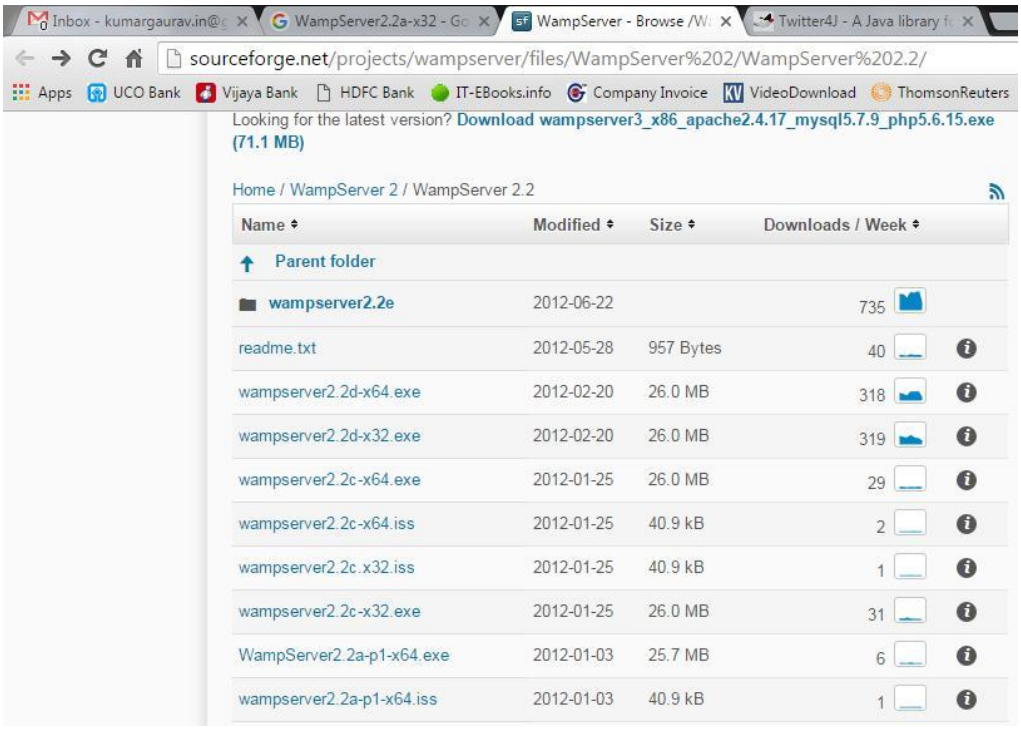


Figure 5.5 – WAMP Server for MySQL Database

Install Tomcat 7 and follow the instructions Code inside eclipse

1. Windows Menu => open perspective=>other=>Database Development
2. Inside Database Development, on left side you will find **Data source explorer**
3. Under this select Database connections, right click and select **New**.
4. Select **MySQL** as Connection Profile Types, click next.
5. In next screen, under general tab write 'twitter' in DataBase
6. And url will be: jdbc:mysql://localhost:3306/twitter

JDBC Twitter Integration Results

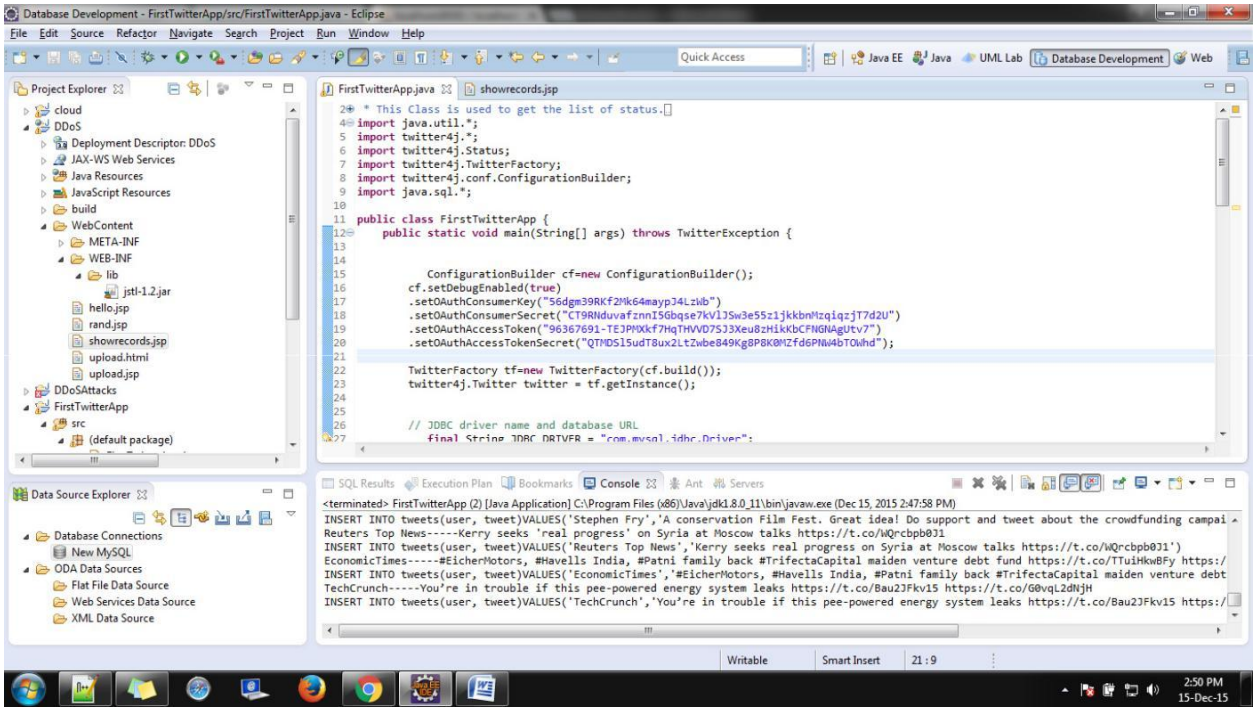


Figure 5.6 – Setup of Live Tweets

SQL result

Host: localhost
Database: twitter
Generation Time: Dec 15, 2015 at 09:19 AM
Generated by: phpMyAdmin 3.4.5 / MySQL 5.5.16-log
SQL query: SELECT * FROM `tweets` LIMIT 0, 30 ;
Rows: 22

id	user	Tweet
1	ABS-CBN News Channel	NOW on ANCs Top Story: PAGASA presser on #NonaPH. https://t.co/X9vkm8nAfc
2	Huffington Post	24 guacalicious gifts for avocado lovers ?? https://t.co/8nrtEERZxq https://t.co/7M1pYXfy2I
3	Bloomberg Business	Rio Tinto "well positioned" for a tough 2016, CEO Sam Walsh says https://t.co/Yj09XpoAnE https://t.co/yOUB8jxeQW
4	NDTV	PM Modi presides over commanders conference aboard INS Vikramaditya https://t.co/EopFl3zyVX https://t.co/2IE2itxNTh
5	ABS-CBN News Channel	PAGASA: #NonaPH expected to be a low pressure area by Friday.
6	Huffington Post	24 guacalicious gifts for avocado lovers ?? https://t.co/8nrtEERZxq https://t.co/7M1pYXfy2I
7	Sarah ?????? ?	@zbleumoon I dont understand how someone can be so cruel to an animal!!! They have no heart!!! #BoycottFoieGras
8	EconomicTimes	.@Paytm ropes in former #Amazon India executive Vikas Purohit to lead payments business https://t.co/uZRVktZGBq https://t.co/iXpT5HLxHC
9	ABS-CBN News Channel	PAGASA: #NonaPH has slowed down.
10	Hootsuite	So many new features came out in 2015. Check out the ones you missed: https://t.co/19oL0SrQ34 https://t.co/NiDEgXmiY7
11	ABS-CBN News Channel	NOW on ANCs Top Story: PAGASA presser on #NonaPH. https://t.co/X9vkm8nAfc
12	EconomicTimes	#Congress avoids #GST talks during luncheon, says formal engagement once Mallikarjun Kharge returns https://t.co/0NX9U5Vfje
13	Wall Street Journal	Watch: Star Wars finally awakens in Hollywood https://t.co/lkTbFpTuxT https://t.co/iUVnKWYPOr
14	The Associated Press	Audiences greet new and old faces at "The Force Awakens" premiere, @derrikjlang writes https://t.co/48EVLVfHtN https://t.co/CnmekEDPCo
15	ABS-CBN News Channel	The storm toppled trees and cut electricity to at least seven provinces. https://t.co/1KPy1lnPWX
16	CNN-IBN News	Odisha Congress MLA suspended for watching porn inside Assembly, party red-faced https://t.co/1uQtx8GstQ https://t.co/wbm48WhRv8
17	Sarah ?????? ?	@zbleumoon Thanks so much #friends for signing petition & RT to raise world awareness. ??

id	user	Tweet
18	Hindustan Times	RT @htshowbiz: Shah Rukh Khan ignored farmers' plight, so boycott Dilwale: MNS https://t.co/JQJTRYYsJg https://t.co/zmrn3kOde5
19	Stephen Fry	A conservation Film Fest. Great idea! Do support and tweet about the crowdfunding campaign for the Wildeye Festival: https://t.co/b6j4RytneR
20	Reuters Top News	Kerry seeks real progress on Syria at Moscow talks https://t.co/WQrcbpb0J1
21	EconomicTimes	#EicherMotors, #Havells India, #Patni family back #TrifectaCapital maiden venture debt fund https://t.co/TTuiHkwBFy https://t.co/wBKNrYtTiH
22	TechCrunch	You're in trouble if this pee-powered energy system leaks https://t.co/Bau2JFkv15 https://t.co/G0vqL2dNjH

tweets

Column	Type
id	int(11)
user	varchar(255)
tweet	text

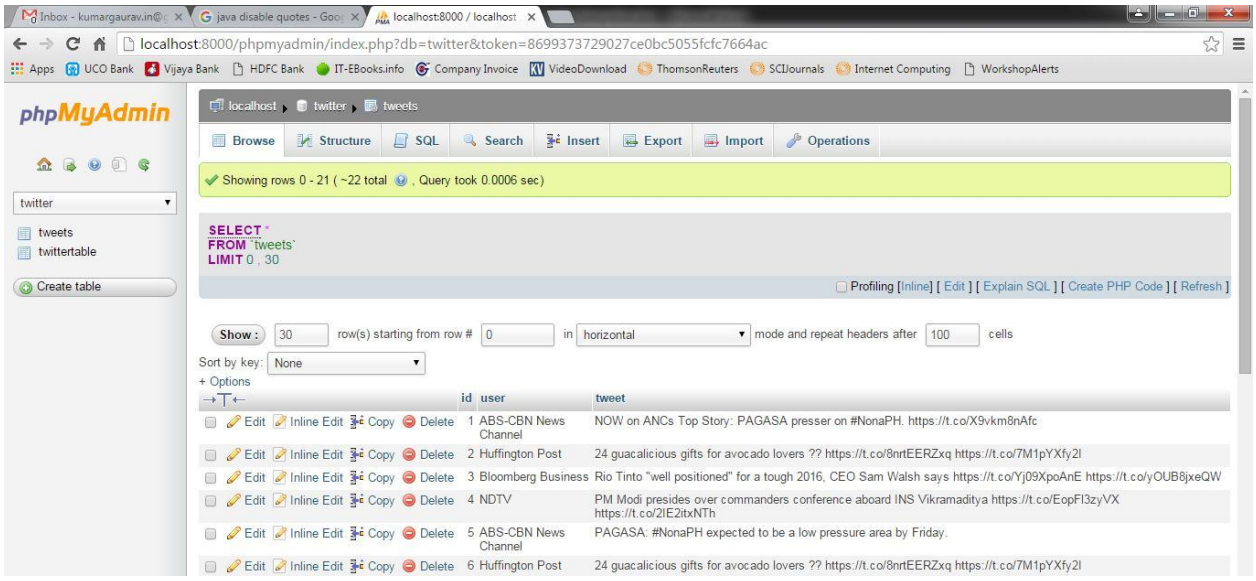


Figure 5.7 – MySQL database with the fetched tweets

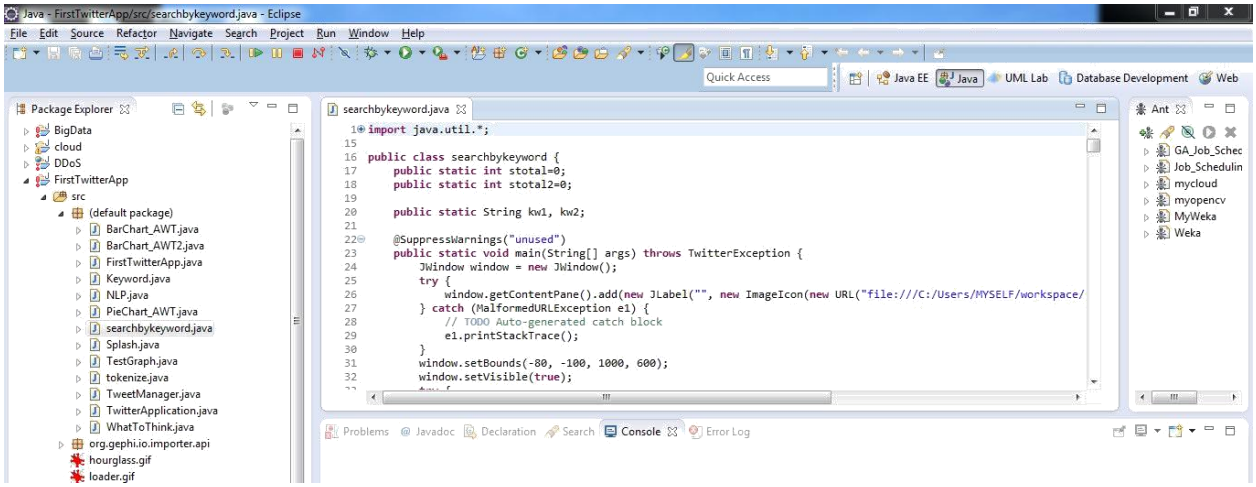


Figure 5.8 – S earching by Keywords from Twitter and Java

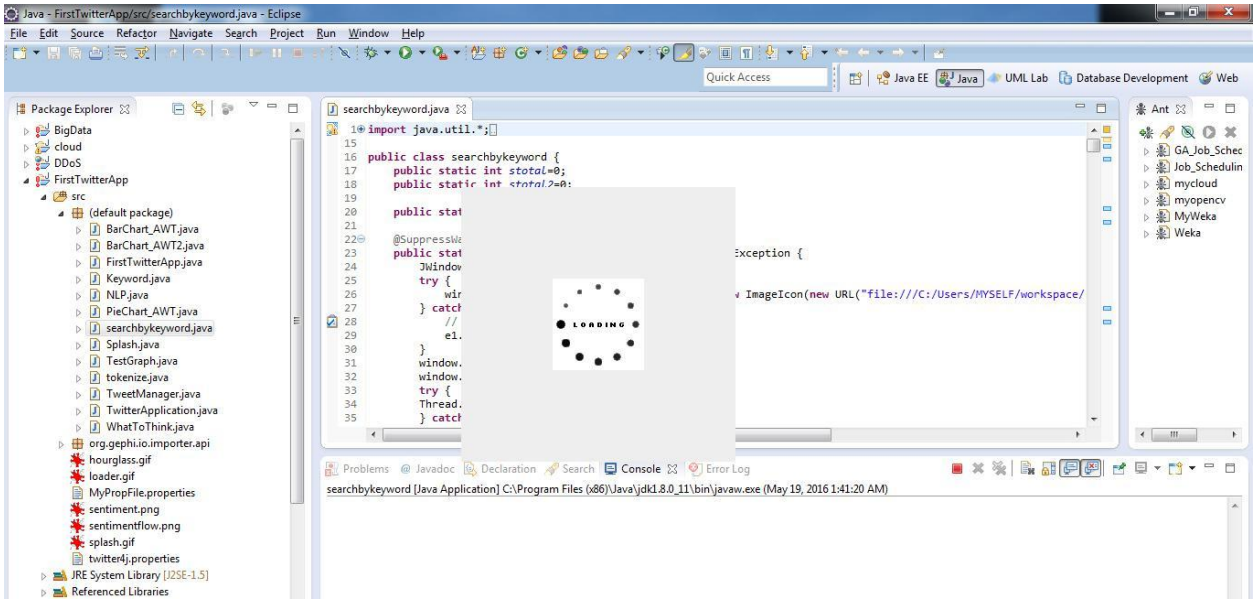


Figure 5.9 – Splash Loader in Java

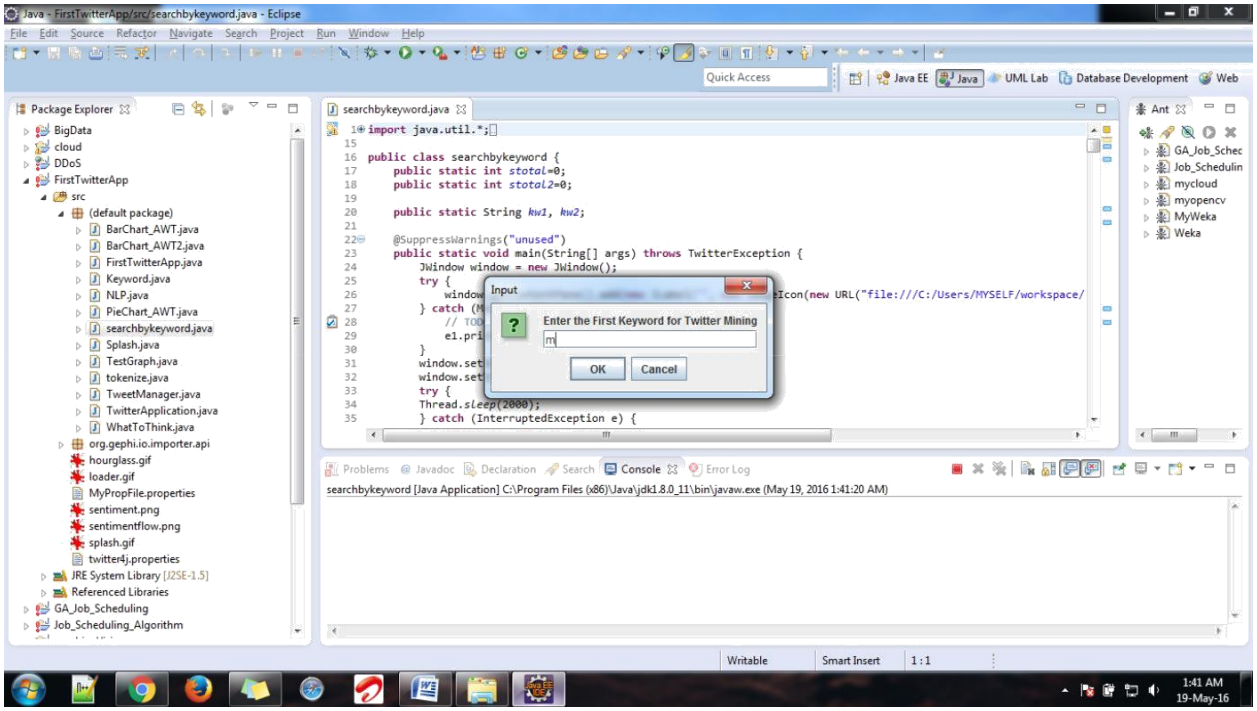


Figure 5.9 – Splash Loader in Java

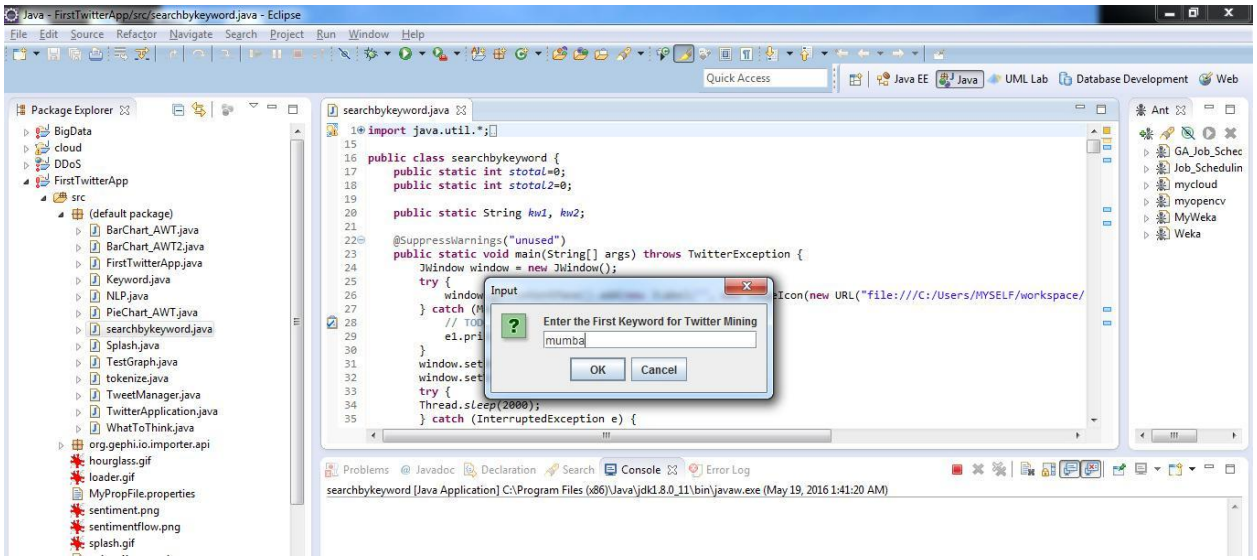


Figure 5.10 – Searching Panel

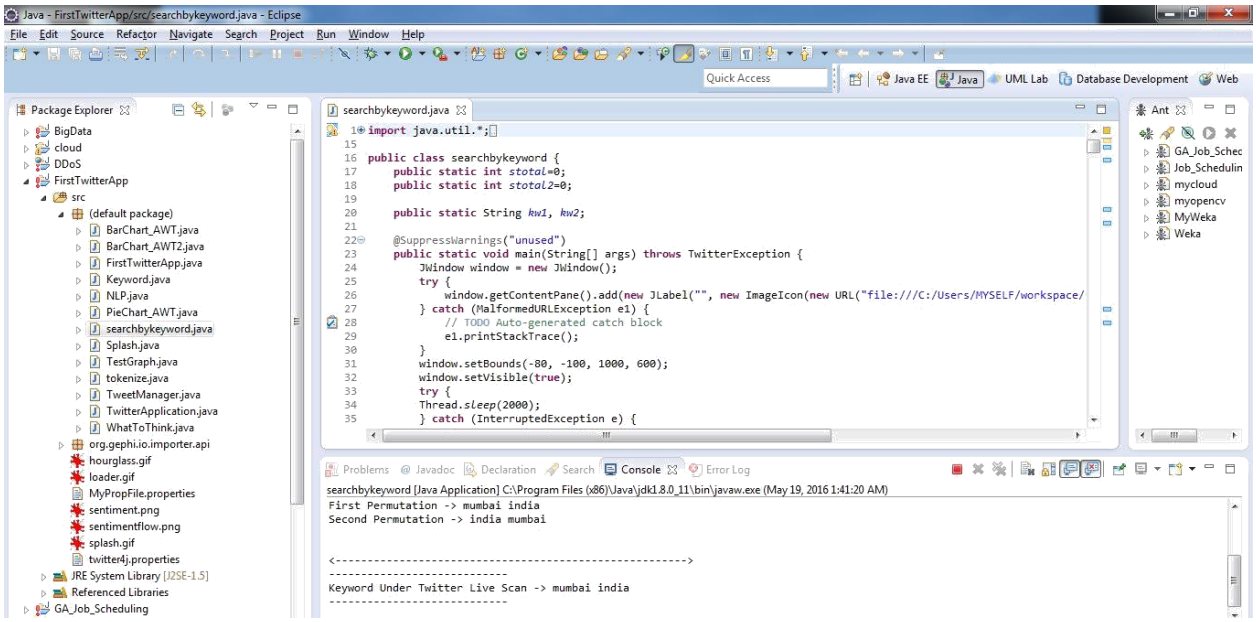


Figure 5.11 – Generation of Permutation of Search Strings

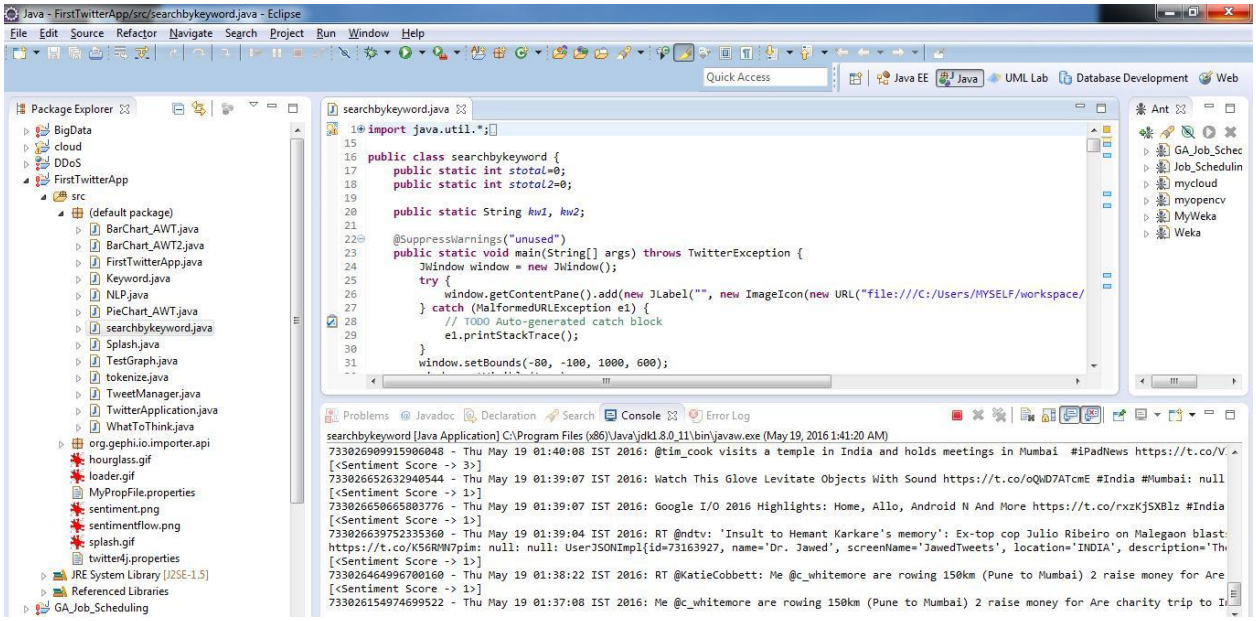


Figure 5.11 – Live Tweets from Twitter Platform

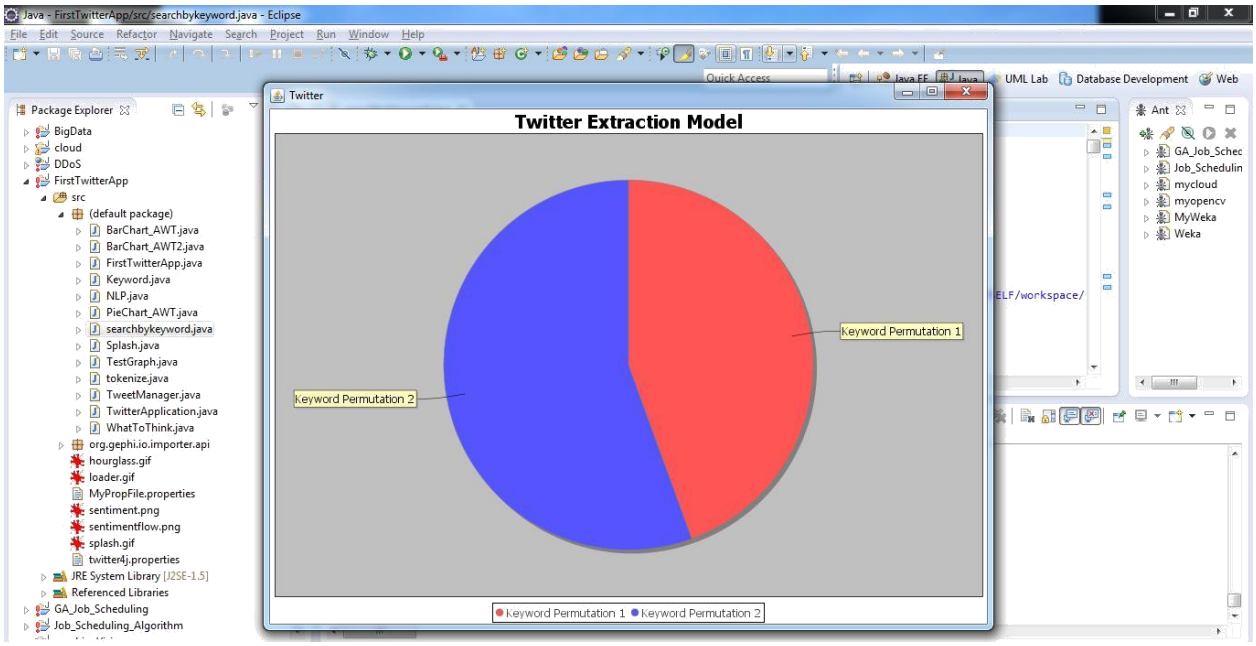


Figure 5.1 2 – Pie Graph of the Results on Sentiment

CHAPTER 6
CONCLUSION

Fetching the live social media or related dimension sentiment analysis is under research from a long time for detailed analysis and prediction of the events with respect to the social cause. There is huge scope of research and development using Java scripts and specialized APIs for assorted applications including cyber security, data mining, Internet of Things, cloud simulation, grid implementation and many others. Java is one of the effective programming languages that can process and handle any type of data stream.

REFERENCES / BIBLIOGRAPHY

[1] Allan, J., Papka, R., & Lavrenko, V. (1998). On-line new event detection and tracking In Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval (pp. 37-45). New York, NY: ACM Press.

[2] Archak, N., Ghose, A., & Ipeirotis, P. G. (2007). Show me the money!: Deriving the pricing power of product features by mining consumer reviews. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 56-65). New York, NY: ACM Press.

[3] Balog, K., Mishne, G., & Rijke, M. d. (2006). Why are they excited? Identifying and explaining spikes in blog mood levels. 11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL 2006), Retrieved July 8, 2010 from: http://staff.science.uva.nl/~mdr/Publications/Files/eacl2006-moodsignals.pdf.

[4] Bansal, N., & Koudas, N. (2007). BlogScope: a system for online analysis of high volume text streams. In Proceedings of the 33rd international conference on Very large data bases (pp. 1410-1413). New York, NY: ACM Press.

[5] Bassett, E. H., & O'Riordan, K. (2002). Ethics of Internet research: Contesting the human subjects research model Ethics and Information Technology, 4(3), 233-247.

[6] Bifet, A., & Frank, E. (2010). Sentiment knowledge discovery in Twitter streaming data. In Proc 13th International Conference on Discovery Science (pp. 1-15). Berlin: Springer.

[7] Blumler, J. G., & Katz, E. (1974). The uses of mass communications: Current perspectives on gratifications research. Beverly Hills, CA: Sage.

[8] Bollen, J., Pepe, A., & Mao, H. (2009). Modeling public mood and emotion: Twitter sentiment and socioeconomic phenomena. arXiv.org, arXiv:0911.1583v0911 [cs.CY] 0919 Nov 2009.

[9] boyd, d., Golder, S., & Lotan, G. (2009). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. Proceedings of HICSS-43, Retrieved November 12, 2009 from: http://www.danah.org/papers/TweetTweetRetweet.pdf.

[10] Bruza, P., & Weeber, M. (2008). Literature-based discovery. Berlin: Springer.

[11] Case, D. O. (2002). Looking for information: A survey of research on information seeking, needs, and behavior. San Diego, CA: Academic Press.

[12] Cataldi, M., Caro, L. D., & Schifanella, C. (2010). Emerging topic detection on Twitter based on temporal and social terms evaluation In Proceedings of the Tenth International Workshop on Multimedia Data Mining table of contents (pp. A4). New York, NY: ACM Press.

[13] Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. P. (2010). Measuring user influence in Twitter: The million follower fallacy. In Proceedings of the International AAAI Conference on Weblogs and Social Media (pp. Retrieved August 9, 2010 from: http://an.kaist.ac.kr/~mycha/docs/icwsm2010_cha.pdf).

[14] Choudhury, M. D., Sundaram, H., John, A., & Seligmann, D. D. (2008). Can blog communication dynamics be correlated with stock market activity? In Proceedings of the Nineteenth ACM conference on Hypertext and Hypermedia (pp. 55-60). New York, NY: ACM Press.

[15] Culotta, A. (2010). Detecting influenza outbreaks by analyzing Twitter messages. arxiv.org, Retrieved September 1, 2010 from: http://arxiv.org/PS_cache/arxiv/pdf/1007/1007.4748v2011.pdf.

[16] Diakopoulos, N. A., & Shamma, D. A. (2010). Characterizing debate performance via aggregated twitter sentiment. In Proceedings of the 28th international conference on Human factors in computing systems (pp. 1195-1198). New York, NY: ACM Press.

[17] Dodds, P. S., & Danforth, C. M. (2010). Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. Journal of Happiness Studies, 11(4), 441-456.

[18] Efron, M. (2010). Hashtag retrieval in a microblogging environment. In Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval (pp. 787-788). New York, NY: ACM Press.

[19] Enyon, R., Schroeder, R., & Fry, J. (2009). New techniques in online research: Challenges for research ethics. 21st Century Society, 4(2), 187-199.

[20] Fox, E. (2008). Emotion science. Basingstoke: Palgrave Macmillan.

[21] Gamon, M., Aue, A., Corston-Oliver, S., & Ringger, E. (2005). Pulse: Mining customer opinions from free text. Lecture Notes in Computer Science, 3646, 121-132.

[22] Gaver, W. W. (1991). Technology affordances. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 79-84). New York: ACM.