

# ScavTweets

Pravina Bhatt<sup>#1</sup>, Laveen Vasnani<sup>#2</sup>, Hemal Samant<sup>#3</sup>, Yixian Hao<sup>#4</sup>, Rushi Patel<sup>#5</sup>

*Masters of Applied Computing, University of Windsor  
Windsor, ON, Canada*

{bhatt9<sup>#1</sup>, vasnani<sup>#2</sup>, samant<sup>#3</sup>, hao119<sup>#4</sup>, patell11t<sup>#5</sup>}@uwindsor.ca

**Abstract**—Twitter is a social system that permits people to post messages that can be read by everyone. We are using this well-known technology as a source to plot research based study according to user requirement. For implementation driven concept, a web application can be designed which will assist its users to search for particular keywords. On the basis of which, they can put up a statistical estimation for their research. This application will gather details from the database which collects information among top trending topics. Thus, user can articulate the product based reviews emerging in top trending tweets.

**Keywords**— Social media, Json, Twitter, ScavTweets, UML, Sentimental analysis, DFD, Interaction diagram

## I. INTRODUCTION

With the expansion of web technology and its evolution, there is a considerable volume of facts present in the web for internet users and a lot of data is generated too. Internet has become a podium for online education, trading philosophies and sharing views. Social networking sites like Twitter, Facebook and Google+ are speedily gaining acceptance as they allow individuals to share and express their interpretations about matters which can be debated with different communities, or post messages across the world. There has been lot of toil in the field of sentiment analysis and link prediction of twitter data. This survey has emphases mainly on investigation of twitter data which is supportive to analyse the data in the tweets where views are highly unstructured, heterogeneous and are either constructive or destructive, or unbiased in some cases. Analysis of twitter data based on reviews or discussions of various people may encounter the importance of that event for such time period. Mainly view sharing and virtual communication for all formal and informal purposes are increasing day by day on large scale which is demonstrated in figure 1. The main business goal is to mine opinion from social media data and design the framework that cluster similar text belong to the same topic. This can be helpful for industrialists and market people to get the idea about demand of their growth and services provided and may plan accordingly for future market race.

### A. Problem Definition

Twitter is a social system that permits people to post messages that can be read by everyone. Every day around million or more tweets are posted on Twitter. Through this, the keyword linked with a hashtag (#) can be searched throughout the tweets sent by people in the particular location and duration. Data is very significant for many researchers in statistics reclamation. A dataset of an article used on the internet is essential for the cataloguing and probing tasks.

These datasets should be dependable and a representation of the real world. Tweets provide information about social interactions, which is valuable for social research and hence we are going to define a problem wherein we will try and extract information on the basis of the keywords enter by the user for further assistance in Research-based decisions.

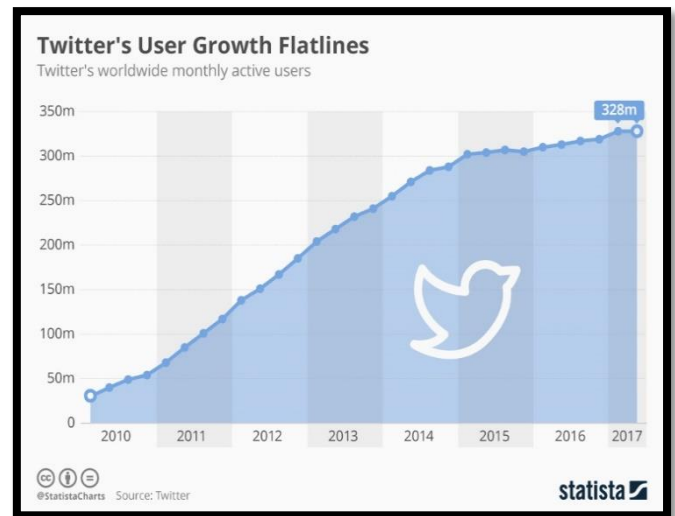


Fig 1: <https://www.statista.com/chart/10460/twitter-user-growth>

### B. System Analysis

Research is vital to collect facts and statistics about a company's products, clients, staffs and competitors. On the basis of these figures, companies are able to make healthier executive verdicts. The collected statistics are structured into reports and the management crew uses them to take action. A worthy research mechanism is crucial, regardless of the extent of the company and its client base. Research is imperative for remaining competitive in the marketplace. Without this application, businesses would collect data for research either from their customers directly (regarding their product) or begin an extensive search of the data from magazines and journals. With this application, they can readily gather the required research information first hand by entering the keyword, duration and location.

### C. Mapping Solution

In our project we can inspect and gather data from social media application of Twitter. Furthermore, we can analyse this data and plot it graphically. The main purpose is to study about the latest trending topics which people are tweeting

globally. With zillions of users and over 500 million tweets being sent each day, Twitter offers an influential prospect for people to reach a worldwide audience. No matter what type of promoter you are — a CEO reinforcing their private brand, a vendor endorsing fresh merchandise — you can use Twitter to form significant acquaintances with an appropriate and involved audience. These connections can lead to arrangements across a grid of trustworthy clientele for your industry.

## II. RELATED WORKS

On the basis of several researches we came up with many related works in this field of twitter data analysis.

- Many of them were dealing with the type of content that were fetched like sentimental analysis.[2]
- Others focused on retrieving timeliness of events for particular reading like population survey or some other use case studies. (Refer instance [3]).
- For estimating unemployment crisis [1].
- Data analysis using twitter API can also be used for categorical data analysis[5]
- Twitter data can also be used in medical field for retrieval of crucial comparisons.

## III. OVERVIEW

As described in above sections we have clear view of how twitter data can be used for various research fields. To ease those researchers job we are appealing to implement a web application which will help them to scrutinize their searches, keep those records and use it for future references. This web application is an initiative for appending the result oriented user interface in which user can keep record of their searched data too along with its search. So in this paper we will elaborate our idea along with practically implemented demonstration of web application.

- Firstly, in section IV we will discuss about the motive behind building this application with the use of various types of goals we are planning to achieve under the roof of ScavTweets.
- Secondly, in section V the actual basic requirements will be listed out which will work as user guide to set up this application.
- Previous requirements will be followed by enhanced technical implementation portion which includes diagrammatical and implemented view of our application.
- Finally we will end up with specifying future scopes and conclude the purpose of this application development at the end of this paper.

## IV. PROJECT GOALS

The main motive of designing this web application is to build a bridge between people reviews and potential users way of converting such reviews into useful data. This is a new product which can be dealt with an innovative perspective to be of assistance to Business statisticians. It our initiative to view research based market from the perception of marketers

as shown in figure 2. It is not limited for specific kind of users, thus we expand into range of various vital potential customers like organizations, researchers, companies and other more similar users.

### A. Functional Goals

The project will gather information from social media of Twitter by using provided API. Record the activeness of different events, for example how long had it been popular before the event really started and how long did it last. Then we will plot the recorded time and to see the changes with time or topics. By analysing the data in many ways we can define the effectiveness of the events.

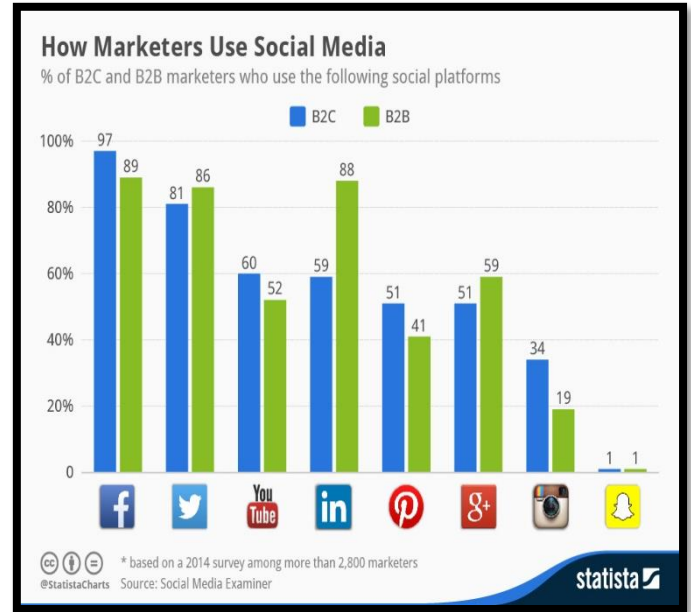


Fig 2: <https://www.statista.com/chart/2289/how-marketers-use-social-media>

### B. Strategic Goals

Main strategy depends on handling data, which should be reliable, optimal and efficient. Twitter data should not just be bounded within a limitations of follows and followers, but such valuable data sharing must be delivered as a crucial form of data for future analysis and researches. This data will be emerging in top trending topics and that will illustrate actual market value for particular product, company or similar kind of facts.

### C. Business Goals

Analysis of twitter data based on reviews or discussions of various people may encounter the importance of that event for such time period. The main business goal is to mine opinion from social media data and design the framework that cluster similar text belong to the same topic. This can be helpful for industrialists and market people to get the idea about demand of their growth and services provided and may plan accordingly for future market race.

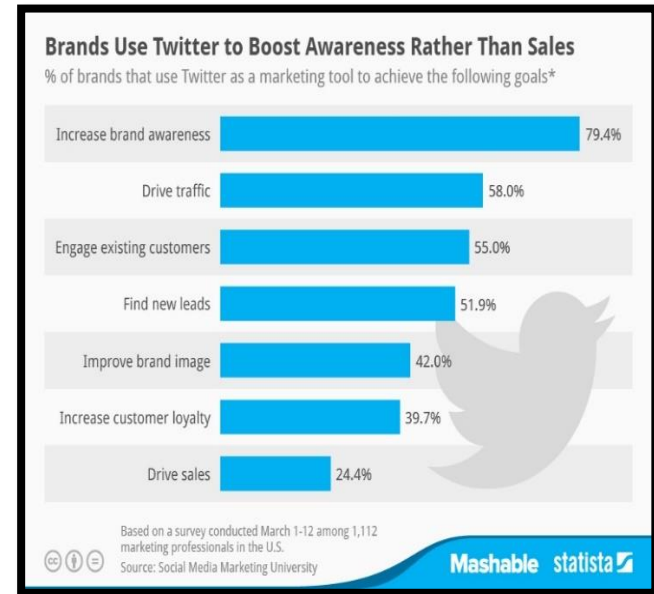
### D. Technological Goals

Fetching data into system database from twitter database

using Java based API namely Twitter 4J. User will enter into the keywords which is assumed to be existing in top trending discussion topics. This keyword will be searched from ScavTweets database and the searched result will be plotted for statistical visualisation of data. This statistical reclamation can also be used for future graphical demonstrating visualization of a result or in a tabular form. Plot the activeness with respect to time and define the relationship. Judge the behaviour of event either it is or will be successful or not.

#### E. Quality Goals

Considering strategic goal as a pillar, need to focus on the quality and reliability if data. Defining new way of vital study related to facts emerging according to diversified population. Statistical optimality of knowledge driven with respect to follower's perception. We also need to focus on negative tweets instead of ignoring them as it can help researches to overcome the drawback related to their product, company or more as shown in figure 3. Draw plot in an acceptable time and in a brief form. Leading a bridge to make proper judgements based on driven data and fall into potential users trust factor.



.Fig 3: <https://www.statista.com/chart/2058/reasons-to-use-twitter-as-a-marketing-tool/>

#### V. REQUIREMENTS

Our application is mainly focusing on all the virtual devices that supports any web services. We are implementing this system as both web and mobile application, for which we are using latest frameworks that makes our system portable for both personal computers and mobile based application. So as a specific hardware requirement this application ScavTweets can be able to run on any hardware that supports web services. Further detailed requirement specifications for this application is listed out in table number 1. This table covers all basic required specifications for this application usage.

TABLE I  
SYSTEM HARDWARE SPECIFICATION

Font Size	Hard ware Requirements	
	Component	Specifications
1	Processor	Dual core
2	RAM	2 GB
3	Network	56 Kbps or faster connection between client computers and server For connections between computers in your server farm, 1 gigabit per second (Gbps) connection

Coming to software requirements, our application requires supportive operating system which can accord our application. Due to latest framework bootstrap our application will be compatible for mobile devices and personal computers both the only requirement is hardware should be supportive. Moreover, it includes software support for html files and mere network connectivity that connects to system database which will extract data from twitter database.

#### VI. TECHNICAL IMPLEMENTATION

Here comes actual implementation part of our application that includes flow of system, front end and back end used for designing this application and finally the output view for system users.

##### A. UML demonstrating flow of system

UML is a standard language of demonstrating the design model for any application. It includes various types of modelling diagrams like data flow diagram, interaction diagram, class diagram and ER diagram. It is majorly categorised into behavioural model and structural model. Thus, the above stated diagrams are part of structural model on which we are focusing on this segment.

##### 1. DFD (Data Flow Diagram):

For every software the relationship between its entities and its actual transition flow can be demonstrated using interaction diagram and data flow diagram for a system. Thus DFD that is Data flow diagram for level 0 given in figure 4.

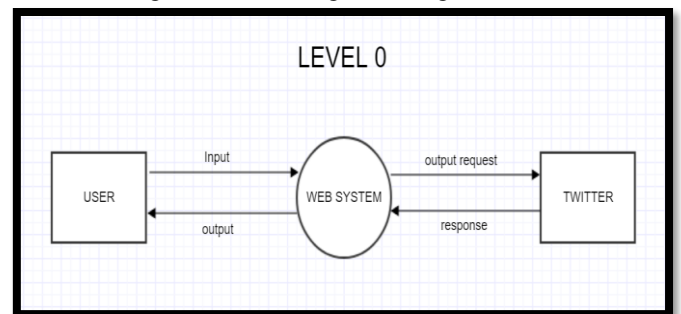


Fig 4: Data Flow Diagram level 0

Above given level 0 DFD is just an overview of the flow of system whereas, the deep connection between the interface and entities is demonstrated in level1 view of DFD. The upcoming every levels of any data flow diagram will be the extension of entities and their responsibilities. Just for above level 0 it is shown that connection between client and twitter server is handled by the use of web system, moreover web system acts as an interface in this scenario. Furthermore in level 1 this sole responsibilities and its functionalities will be extended to some extent, that includes its internal behaviour and processing tasks. As given below in below figure 5 we can visualise that further web server is responsible for connecting to login page and also fetches the data from system database.

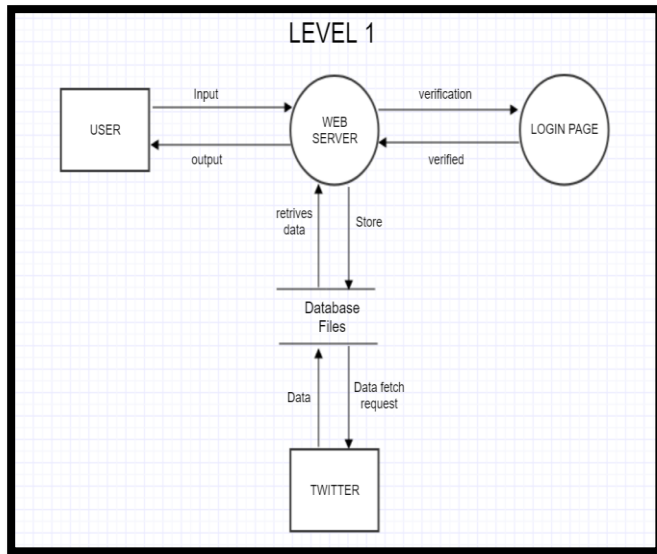


Fig 5: Data Flow Diagram level 1

## 2. Interaction diagram(Sequence diagram):

The following part includes interaction diagram, specifically it is sequence diagram. In this diagram vertical line indicates objects whereas horizontal lines shows connection between those objects and transition of activities. Main purpose of this interaction diagram is to visualise users about how their command to application is being executed, with the usage of which entities. Interaction diagram for our application is shown in figure number 6, where objects are mainly modules of web application those are users, web server, ScavTweet database and Twitter database. The transition of user command, their communication to knowledge stored in database will be retrieved with the help of web servers.

## 3. Class diagram:

According to above specified UML diagrams we can estimate the behavioral part of our system, which includes the interactive flow of tasks between various modules of application. Now, we will understand about the structural part of our application named as class diagram. Class diagram is a part of structural unified modeling language, which shows the relationships of classes created within the application specifying its attributes and methods. We have implemented

our whole system fully threaded that can help to execute parallel tasks for variant users. Under the module of user we create main thread that allows parallel execution of 10 individual searches. Each search corresponds to individual separate thread that is 10 threads for each search. This separate thread lead to 10 separate queries to generate, in which we allow various tweets. Each generated queries leads to 100 tweets for each query, thus overall for 10 queries we can retrieve 1000 tweets. The main focus on threading development of our application is due to our target of allowing 16 parallel users to be allowed to implement their search per 30 minutes time span. This parallel processing of 16 users will lead to execution of 16 parallel threads. Each thread will constitute 10 threads into it which will further leads to parallel 1000 tweets for each queries generated under those threads, thus overall system runs on 160 queries leading to 16000 tweets to be fetched from twitter server at a time.

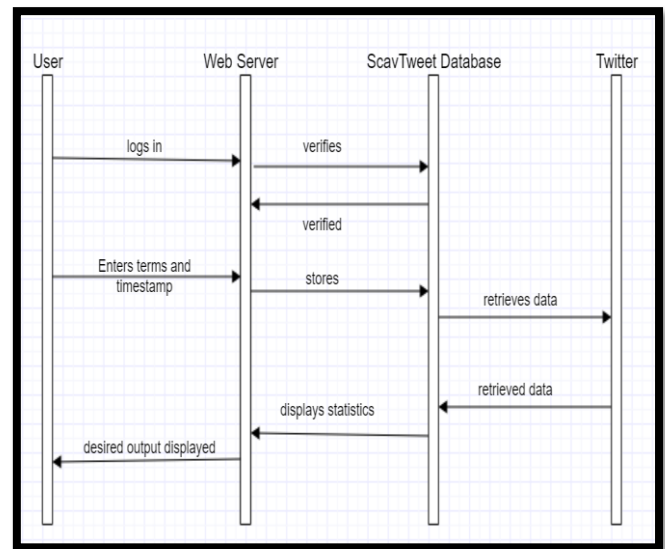


Fig 6: Interaction diagram

As shown in figure 7 we can estimate that dispatch.java class is an identical class which is not actually an interface but it behaves to be an interface for searchtask.java class and count.java class. Dispatch class will append the process of accepting the keyword that will be processed under search task and delivers the result related data to user. Previously we discussed about our system will be threaded for handling multiple tweets parallel at a time, this threading process will be handled within tweet.java class. This input keyword will be compared with their emerging in top trends topics of twitter database and will result their occurrence in the search using comparision.java class.

## B. Technical demonstration of ScavTweets:

This section of paper will cover the implementation part of our application. It will include listing of technologies used for execution using the demo of application webpages. User will able to view the actual GUI which will lead them to input their search keyword. Below listed are the technologies used for enactment, which covers front end and back end section too.



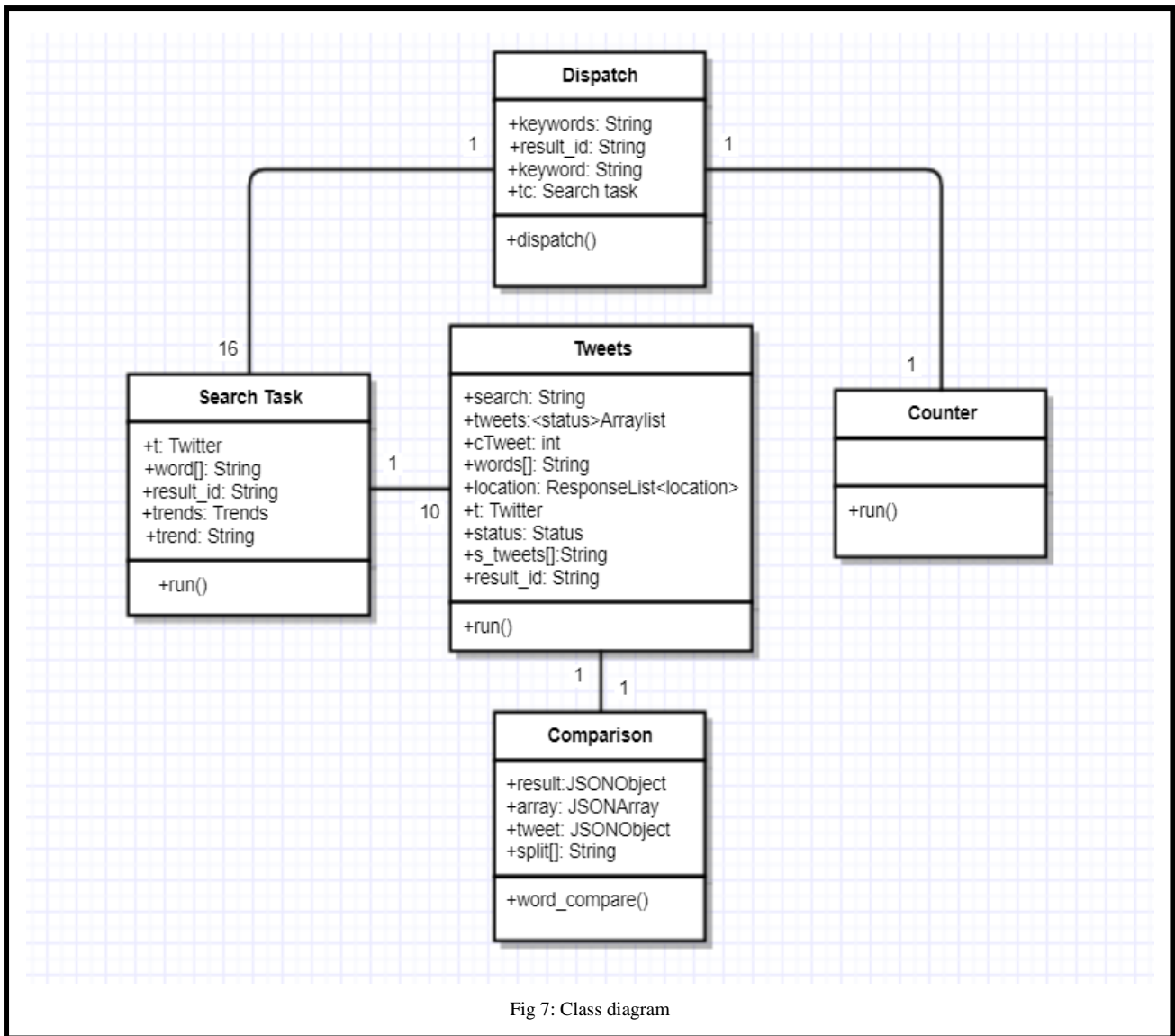


Fig 7: Class diagram

## 1. Front-end:

- **HTML:** It is abbreviated as Hypertext mark-up language, is considered as traditional mark-up language for designing web pages for web applications.
- **CSS:** Abbreviated as Cascading Style Sheet, is a style sheet language used for describing the presentation of a document written in a mark-up language.
- **JavaScript:** It is a high-level, dynamic, prototype-based, multi-paradigm, and interpreted programming language used to make web pages interactive. Validations for user input is implemented using JavaScript. In case of any web application validating user data is the most crucial task to carry on the reliable data delivery and efficiency in the flow of application.
- **Bootstrap:** This frame work is used to support the

the device compatibility of a web site. Just for an instance if a web application is designed for personal computer then it will also be supported in the view of mobile devices. This framework is used in our web application to fulfil user requirements in case of both mobile devices and personal computers.

## 2. Back-end:

- **MySQL:** This is well known standardized query language used to handle database for any web application. MySQL will help to handle, modify or discard the content of database.
- **Twitter4J:** This is java language based API (Application Program Interface) used to retrieve tweets from twitter database on demand. We mainly selected java language based API due to its

compatibility, easy usage , simple way of compiling of compiling and debugging and the most efficient it is user friendly programming language constituting vast object oriented programming concepts.

- **Java programming language:** We have used this Java programming language in our web application development. We mainly selected this programming language just because of its thread support. In our system we will be having multiple users requesting for their queries on same time, and number of tweets will be also in large number. Thus, we need to use such high level programming language which will support us implementing thread and will also be distributed. Moreover Java is platform independent language which supports “write once run anywhere” methodology, this quality will support to implement our web application on any virtual web supporting devices.

### C. Technical demonstration of ScavTweets:

Here is the actual implemented view of our application, it includes the clear path for transition of user request. Initialization of web application use will start with login page where user will input there log in data to create identity as shown in figure 8. After user log in it will redirect to homepage as shown in figure 9 which will introduce the application to its users, constitutes module of functionalities that is search and result. If user wants to start the search will select search tab from home page and then command will be conveyed to initiation page where initiation of search will be done as shown in figure 10. Here, user will enter into the term to be searched from twitter database as per its presence in top trending topics. After entering into the search keyword or terms user will be leaded to result page as shown in figure 11. This result page will show the resultant table showing the number of terms searched by till along with starting and ending time of the search.

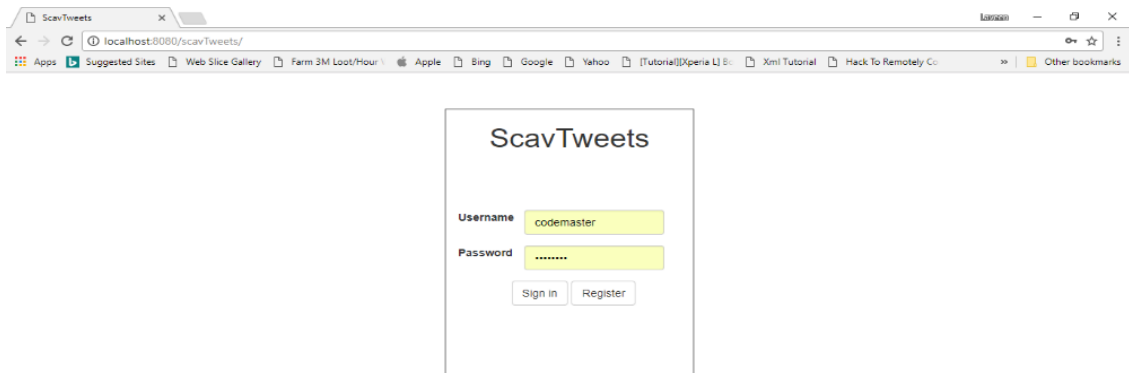


Figure 8: Login Page

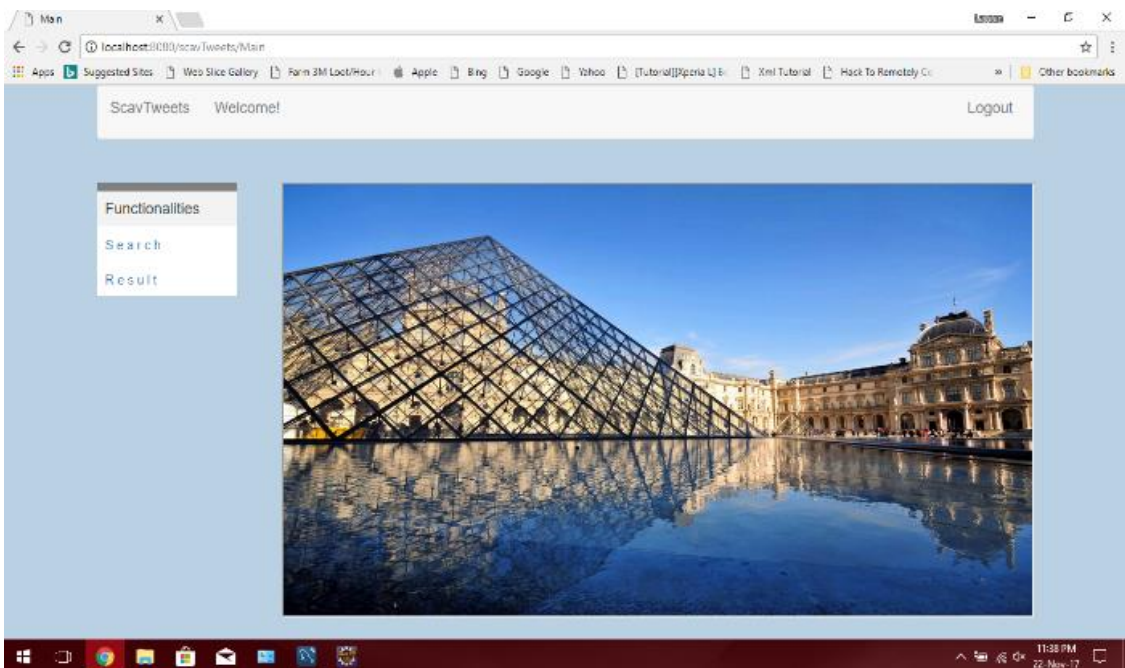


Figure 9: Home Page

**Gnerate a new search**

name: Search Demo

term: Nashville

location: Toronto

Start\_time: 2017-11-23 00:06:46

End\_time: 2017-11-23 00:06:49

close add

Figure10: Initiate Search

ScavTweets Welcome! Logout

Functionalities

- Search
- Result

Name	Term	Start_time	End_time	Detail
Search Demo	Nashville	2017-11-23 00:06:46	2017-11-23 00:06:49	<a href="#">i</a>

Figure11: Search display

This page will also include search bar that is used to help user for easy search for their keywords which we entered previously. This searches will and there corresponding searched tweets will be analysed in twitter database and the data will be stored in application database which can be used for future references. In case of twitter database normally we cannot get access the previous tweets, thus this data accumulation will help us to use those data for future analysis

and researches. After entering into the search keywords those search can results can be visualised after specified time span into result table as given in figure 12. This page will have a resultant table of search and its content too. If user intend to see the content of query the need to select content section of the table which will pop up tweet content emerging in top trending topics at top of the screen as visualised in figure number 13. We can see the conformed content of searched term in this pop up box.

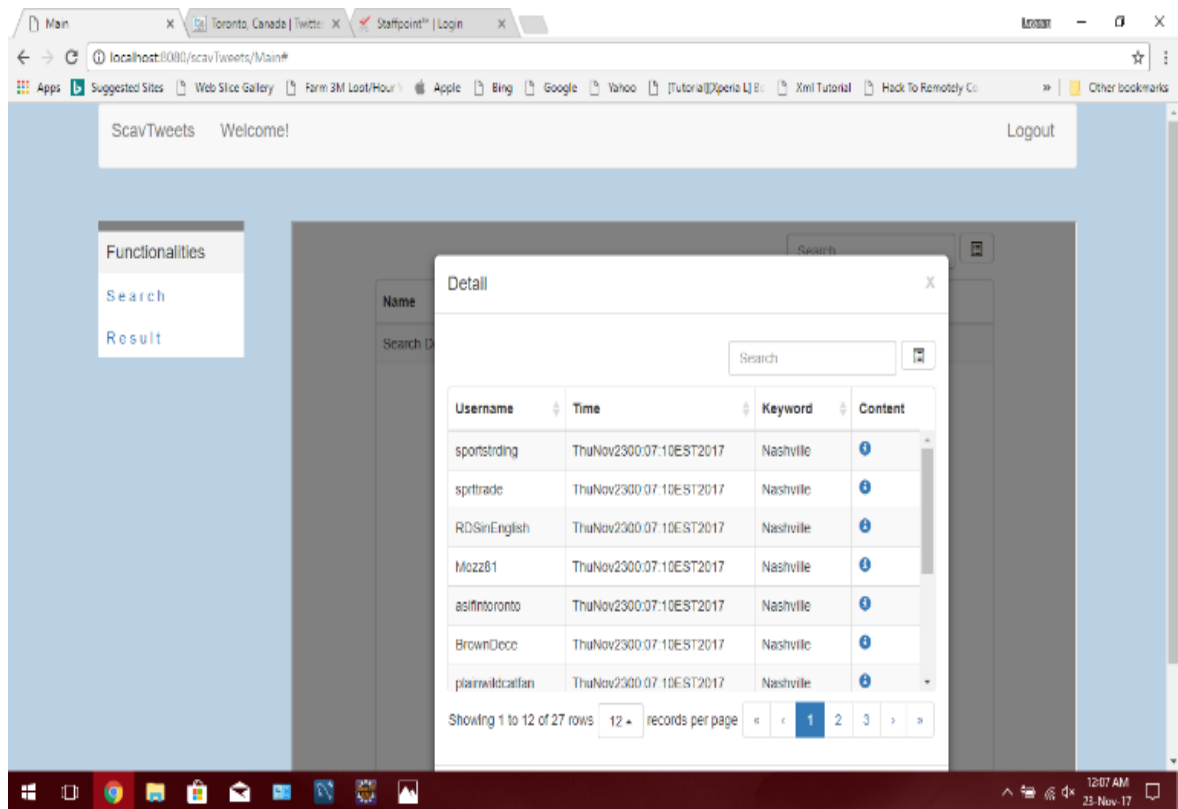


Figure12: Result Table

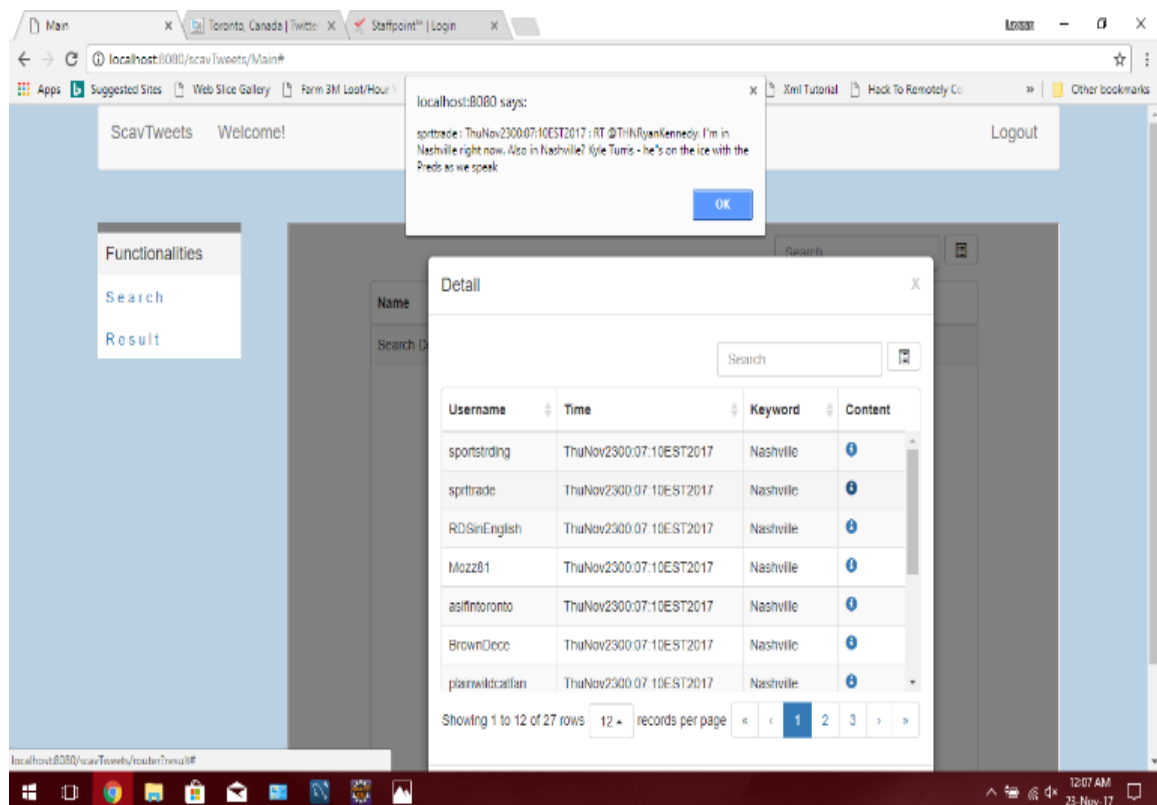


Figure 13: Tweet for selected result



Above mentioned figures 12 and 13 are result module that shows when user want to see result of their search after specified time can select this module. When user requests for result table it will be displayed and on selection of content column of the result table, particular content can be seen. As a next step this result will be plotted in a form of graph on which occurrence of specific term with respect to its time span will be plotted. In future this can be extended to check the liveliness of particular event or a product under the discussion of people tweeting around the world on the basis of time stapes. The graph will also carry outs the process of comparison among location based survey, as right now our application is focusing only for North American Countries which will be expanded for all worldwide Nations.

## VII. FUTURE SCOPE

Twitter data analysis can be used in various fields for data retrieval on the basis of specific purpose. As in case of company people they can use this data processing methodology to collect reviews of their product and plan accordingly for market strategy. This survey can be implemented for launching versions of their products or to introduce some new inventions to reach achieve customer's satisfactions. Speaking about some other applications twitter data analysis can also be used for sentimental analysis. Many multinational companies usually implement sentimental analysis for employers' survey on the basis of resume study. As similar to sentimental analysis this wide spread range of utilization of twitter can also be used for business consumer communication. There is massive potential in partnerships between Twitter and firms eager to tap into the peaks of information from which to tactically interconnect to their respective consumer base.

## VIII. CONCLUSION

Since its inception, Internet has been the promoter for innovative ways to communicate, gather information, and seek entertainment. Concurrently, Twitter has emerged as one of the top social media tools worldwide. We proposed and described an approach to cluster users in Twitter based on content. From our results and analysis, we conclude that the social connections are no doubt a dominating factor in the community detection process. The work can further be extended to achieve the idea of 'Sentiment Analysis' which has been gaining a lot of popularity in research. Thus through this application we will allow our users to search for specific terms among top trending topics which will further extended to statistical plotting of those data. This statistical plot may also be used for realizing liveliness of trending events to plane future predictions in market strategy, research sectors or for some similar purposes.

## IX. REFERENCES

1. Nirmala, C. R., Roopa, G. M., & Kumar, K. N. (2015, October). Twitter data analysis for unemployment crisis. In *Applied and Theoretical Computing and Communication Technology (iCATccT)*, 2015 International Conference on (pp. 420-423). IEEE.
2. Kharde, V., & Sonawane, P. (2016). Sentiment analysis of twitter data: A survey of techniques. *arXiv preprint arXiv:1601.06971*.
3. Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics online*, 1, 47.
4. Anber, H., Salah, A., & El-Aziz, A. A. (2016). A Literature Review on Twitter Data Analysis. *International Journal of Computer and Electrical Engineering*, 8(3), 241.
5. Agresti, A., & Kateri, M. (2011). Categorical data analysis. In *International encyclopedia of statistical science* (pp. 206-208). Springer Berlin Heidelberg.
6. Ravel, B., & Newville, M. A. T. H. E. N. A. (2005). ATHENA, ARTEMIS, HEPHAESTUS: data analysis for X-ray absorption spectroscopy using IFEFFIT. *Journal of synchrotron radiation*, 12(4), 537-541.
7. Xia, X., & Xie, Z. (2001). DAMBE: software package for data analysis in molecular biology and evolution. *Journal of heredity*, 92(4), 371-373.
8. Achrekar, H., Gandhe, A., Lazarus, R., Yu, S. H., & Liu, B. (2011, April). Predicting flu trends using twitter data. In *Computer Communications Workshops (INFOCOM WKSHPS)*, 2011 IEEE Conference on (pp. 702-707). IEEE.
9. Kumar, S., Morstatter, F., & Liu, H. (2014). *Twitter data analytics* (pp. 1041-4347). New York: Springer.
1. Nirmala, C. R., Roopa, G. M., & Kumar, K. N. (2015, October). Twitter data analysis for unemployment crisis. In *Applied and Theoretical Computing and*