

Querying Social Media to Track People's Movements and Discussions Via a Browser Interface

¹ Hanan A. Tahir, ² Hoger Mahmud

¹Deptment of Information Technology Management, Duhok Polytechnic University, Kurdistan Region of Iraq

²Department of Computer Science, University of Human Development, Sulaymaniyah, Kurdistan Region of Iraq

ABSTRACT

Social networking is micro-blogging site where users post short text-based messages. These messages often relate to a particular topic, event or product and are frequently coupled with the authors' opinion. This information provides an exciting opportunity for individuals and companies who might benefit from being able to collate and analyze the data to discover opinion on an item of interest. This paper reports on the outcome of a project that aims to produce a web page that queries the social Web (specifically Twitter.com and Foursquare.com.) through the available APIs to track people's movements and the topic of their discussions. Tweets will be the main data source and queries results will be ranked and sorted out by total frequency in a Web accessible MySQL database. The result shows that the system has been successful in achieving the set aims.

KEY WORDS: Social Media, Social Networking, Browser Interface, Data Analysis, Data Query

1. Introduction

It is a known fact that for millions of people, social networking has become a significant part of everyday life.

Social networking is facilitated by social media which is described as web-based and mobile-based applications that provide the medium where users create, access and exchange content (Kaplan & Haenlein, 2010).

Social media allows users to share their thoughts, ideas and experiences to people in their social network. Online social media websites, such as Twitter and Facebook, are growing year on year at an exponential rate. Humanity generates a huge amount of data through expressing feelings and thoughts on social media. The availability of this real and freely available data, which can be acquired with minimal efforts, has encouraged researchers to analyse it and harness its potential in areas such as marketing, decision making and predicting the future (Phillips, Dowling, Shaffer, Hodas, & Volkova, 2017) (Derczynski, Yang, & Jensen, 2013).

Analyzing social media data is carried out for many purposes, one of which is sentiment analysis, where users' comments posted on social media about a particular topic is processed and classed into positive, negative and neutral opinions. Based on the analysis results, further actions can be taken, for example change the course of a process or continue or not continue with the process (Manasa & Padma, 2019).

Opinions are important, this has always been the case, but modern technology has allowed people to share their opinions on just about any topic with anyone who have access to the World Wide Web. This database of information forms a unique and useful insight into the thoughts of millions of people, whether it is relating to a place, product, person, or organization. With all this information so readily available to almost everyone, it is no wonder that more and more people are now looking to the internet for a quick and easy way to make informed decisions.

There are many instances of social media in use, such as Facebook, Twitter and Instagram; according to (Manasa & Padma, 2019) Twitter is the most popular one for data collection purposes by researchers; due to the fact that it is a general-purpose medium which covers data from a wide spectrum of fields. Authors of (Ratkiewicz et al., 2011) claim that Twitter has grabbed researcher's attention due to its popular use and transparent data sharing policy. Twitters allow account holders to compose and post short messages (up to 140 characters) known as "tweets". Tweets rapidly spread and publicly available for everyone to see unless the user opt for making a private tweet. Since the launch of twitter in July 2006, the number of regular Twitter users has sky rocketed. As of the first quarter of 2019, Twitter averaged 330 million monthly active users. Figure 1 shows the number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019 (in millions). It is interesting to see how Twitter has evolved over the past years to get a feeling of the growth potential of this social media.

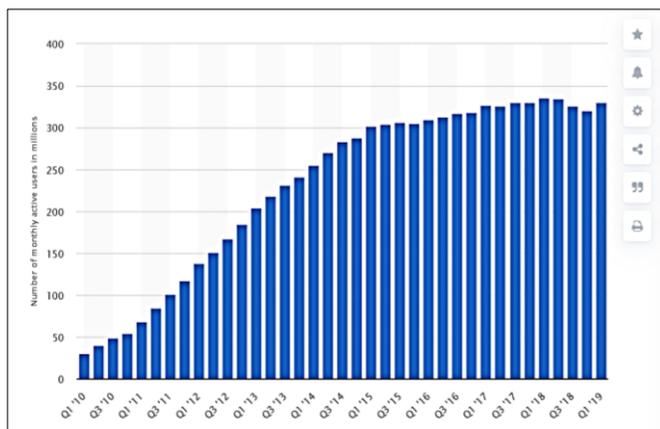


Fig. 1. Statistics of Twitter Usage from 2010 to 2019

This paper reports on the outcome of project designed and implemented to store and analyze information about users that have been tracked, the keywords they have used, the people they have contacted, and the venues they have visited based on their tweets. The

information is stored in a Web accessible MySQL database. This paper will explain in detail the Issues, Design Choices, Requirements and Limitations of the system. The paper explicitly discusses the details of the interaction between the database queries and the live queries to Twitter/Foursquare. In particular the ways database is used to save Twitter/Foursquare calls.

The rest of this paper is organized as follows. In section 2 we present a literature review relating to the topic, in section 3 we describe our research approach, in section 4 we list and discuss the requirements identified for the Twitter mining system that is implemented. In section 5 we discuss design and implementation, in section 6 we present the results of mining Twitter data using the implemented system and discuss findings. Finally we conclude in section 7 and provide future work opportunities.

2. LITERATURE REVIEW

Researchers have been using data obtained from social media widely; in this section we provide a brief review of a number of publications which have used data from social media for different purposes.

The authors in (Rashidi, Abbasi, Maghrebi, Hasan, & Waller, 2017) used social media data such as hash-tags and check-in, to study the use of social media for transportation purposes with the aim to provide guidance to the transport industry to improve services. Spatial social media data such as check-in has also been used to predict movements (next check-in location) by (Chang & Sun, 2011) and life-style location-based activity choices by (Hasan & Ukkusuri, 2015). Data from Social media and in particular Facebook is used by governmental bodies such as the EU to track and make sense of migrations in the block (Spyratos et al., 2018). Monitoring sentiments of social media users is growing for commercial purposes which is used as a tool to attract consumers and improve products (Maynard, Bontcheva,

& Rout, 2012). Authors of (Saumya, Singh, & Kumar, 2016) used social media data to predict stock movement in Indian market.

As mentioned, Twitter is one of the most widely used social media used by researchers. The authors in (McClellan, Ali, Mutter, Kroutil, & Landwehr, 2017) have used data obtained from Twitter to develop a model that measures the level of interest in topics related to mental health. The authors in (Ratkiewicz et al., 2011) have used Twitter data to detect political abuses from individuals who work in politics and have used the result to gauge political support for a particular candidate in 2010 US election. Others have analyzed acquired data from Twitter for purposes such as consumer opinion on a particular brand (Jansen, Zhang, Sobel, & Chowdury, 2009) and people's reaction to a given headline news (Kwak, Lee, Park, & Moon, 2010). Twitter data has been used to predict real-world outcomes such as box-office movies revenue by researchers in (Asur & Huberman, 2010) using a simple model that counts the rate of tweets about a particular movie. Authors in (Zhu, 2017) analyzed data obtained from Twitter to study the impact of social media on public discourse in Hong Kong and drawing conclusions based on social movement theories. They have concluded that social media can amplify shared grievances to an extent that can influence the breakout of public discourse in countries with similar grievances.

Researchers have utilized various tools and approaches to analyze social media data, a survey of tools and techniques used for different purposes such as mental health diagnostics to marketing and movement tracking is presented in (Batrinca & Treleaven, 2015). The authors of (Manasa & Padma, 2019) have studied the sentiment analysis tools and processes and have proposed a number of steps to guide the analysis process. They also highlight the use of machine learning as an important

tool for sentiment analysis. The authors in (Ratkiewicz et al., 2011) have used a similar method to what we have used in this paper in which they have implemented a system which they call "Truthy" to detect specific words and memes in Twitter data streams and generate statistics for a given word. The approach we have applied differs from the approaches used in the literature. In our approach social media mining process (Twitter in particular) is made easy through the use of a simple interactive web page that is multi-purpose. The mining result is organized in the form of tables which can be read and understood easily.

3. METHODOLOGY

Generally speaking, there are four main approaches in sentiment analysis which are machine learning, lexicon-based, statistical and rule-based (Manasa & Padma, 2019). In this paper we use rule-based approach to analyze the opinion of Twitter users regarding a particular topic. This is where users' comments are analysed based on a number of specific words and present the result for each word. The main task of the project reported in this paper is to query different kinds of information from the social media (Twitter and Swarm) by Foursquare. Their Application Program Interface (API) is used to collect the information that is needed and create a website that will present the data collected. The availability of Twitter APIs through which researchers can easily obtain data is one of the reasons that Twitter has become the preferred social media data source as claimed by (Batrinca & Treleaven, 2015).

3.1 Extracting Twitter Data

Twitter provides two main application programming interfaces (APIs) for accessing a wide range of data. Representational state transfer (REST) APIs are available for tasks such as searching for recent tweets about specific topics or retrieving a specific tweet by its unique ID. Streaming APIs focus on retrieving data as soon as

possible after it is created by a user.

Although a substantial number of REST API methods are made available by Twitter for the use of application developers, the search/tweets method or Search API is perhaps the most useful for retrieving significant quantities of tweets. As with all of Twitter’s APIs, requests to the Search API must be authenticated using the authorization framework (“AOuth Authorisation Framework,” n.d.), and each registered application is limited within 15 minutes intervals to 450 requests, with an additional 180 requests per authorized user of the application. According to Twitter these limits protect the performance. The Twitter REST API allows access to past data from Twitter i.e. tweets that have already occurred. You can request specific data by searching with some criteria. This could be keywords, hash-tags, usernames, locations etc.

The Twitter streaming API differs to the REST API in that it gives you access to tweets happening in near real-time. Using the same criteria as before keywords, hashtags, usernames etc. Tweets that match the query will be returned as they happen. However, once again the data received is limited. The percentage of total tweets that you are able to extract using this API depends on the current traffic, as well as the nature of the criteria you requested. One important limitation to consider if using the streaming API is that it requires keeping a consistent HTTP connection open. The following diagram taken from the Twitter developer site, illustrates how the code for maintaining the Streaming connection is typically run in a process separate from the process which handles HTTP requests to avoid connection loss.

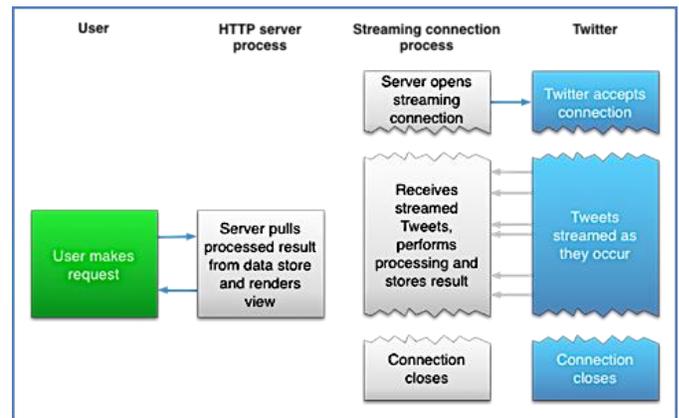


Fig. 2. Twitter Streaming API application setup

3.2 SYSTEM REQUIREMENTS

The project aim was to design, implement and evaluate a system that query the social web, specifically Twitter.com and Foursquare.com. Before the system design and implementation begins, it is a normal practice in system development to define a number of requirements to be fulfilled by the system. Below, we provide the main functions that had to be accommodated in the system. The functions were defined with the main task of the system in mind.

3.3 Tracking Discussion

This task requires tracking public discussion based on the user input. The user is expected to input hash-tags or keywords, the output will be a list of messages containing the input hash-tag. In addition, the user may limit the query to specific geographic region which means that the user can put the latitude/longitude coordinate to filter the message. The result of the query must be able to show up to 10 users who have retweeted. Table 1 shows a query example for this task and its result.

TABLE 1
Query example for tracking discussion

	London, Traffic, Monday
Output:	- Date: 31.1.2018: from: @fabcirca: “never seen such traffic in London on a Monday morning - Date: 31.1.2018: from: @noonreally: “Where is everybody? You would expect London to be gridlocked at this time on a Monday

3.4 Queries about specific users

a) **What specific users discuss about?** This task enables the user to find out the most frequent keywords used by up to 10 users (the ids/screen names of the users must be specified in the input) in the last Z days (must be specified by the user). Table 2 shows a query example and output would be as the following:

TABLE 2
Specific user query example

Input	Screen_names: fabcira, tbl, matodel, sipaca					
	How many keywords: 2					
	How many days: 5					
Output	fabcira	tbl	matodel	sipaca	total	
	London	10	1	15	15	41
	Music	8	7	20	3	38

As shown is the table, the words London and Music are the two most frequent words used by fabcira, tbl, matodel, sipaca in the last 5 days. The ranks are sorted by total frequency.

b) **What points of interest?** This task enables the user to search for points of interest (Foursquare venues) a specific user has visited in the last X days. The user must be able to input the user id plus the number of days so the result will only return places up to that day. The result must include a description about the place, photo, and the place name. If the number of days inserted was zero, we use a streaming API to capture the information. Table 3 is a query example for this task and its result would be as the following:

TABLE 3
Point of interest query example

	User: fabcira
Input:	Number of days: 3
	- Loch Fyne Restaurant
Output:	- St Peter & St Paul School

c) **Who is visiting venues in a specific geographic area?** In this final task, the user is required to input the location name and the number of days to read all the tweets around the desired area. The result here will be a list of tweets of this location. This task also has streaming ability. This means that new tweets would appear as soon as it is published. Generally, in this project, the user will have the ability to click the username and it will be directed to the twitter profile with all the general information. Tweets, up to 100, will also be shown in the profile. Table 4 is a query example for this task and the result would be as the following:

TABLE 4
Visiting venues query example

	Location: Sheffield
Input:	How many days: 5
	a list of user
	Information about users - when displayed -
Output:	include at least name, location where they live in (if available), a profile picture (if available) and a description (if provided).

4. IMPLEMENTATION AND RESULTS

The program is controlled through a browser-based interface. The interface enables the user to perform the queries. The system has been implemented as a HTML/JavaScript set of files. The interface is served by a node.js server and all the interaction with the social web also have been done through a node.js server. Web Pages mainly have been designed in Jade. Jade is a templating engine that compiles to HTML. Although with jade we can type faster and it improves readability

which means maintainability and productivity is better. We have used Node JS and Express for the communication between the client and the server.

The structure of the system is divided into three parts. The first part is JavaScript file that deals with the API (library folder). The second part is JavaScript file that deals with the information and route of our website (routes folder). The third part Jade file that presents the user interface (views folder). The way this work is that our views communicate with routes and routes help to retrieve information from the API used. In order to implement and maintain the quality of solution; JavaScript, JSON, Rest API and node.js have been used.

Figure 3 shows the [architecture of the system](#). The following languages are used in developing the system:

- JavaScript to check on the client that the input is properly provided. Ajax has not been used in order to leave the client active while the server retrieves the answer.
- JSON-based exchange of information between client and server.
- Node.JS
- Jade, Jade is a brief language that extends from HTML language.
- REST API to retrieve data from twitter, foursquare.
- Express.js, a web development framework for Node.js
- Ajax to leave the client active while the server retrieves the answer.



Fig. 3. Client server architecture interaction building on the node.js, Json and Ajax.

4.1 Twitter API Features

This section will illustrate the twitter API features that have been used for each type of query. In the first query (task 1), “GET search/tweets” is used which allows searching for keywords and then returning corresponding matching tweets. It obtains a location of relevant tweets matching a specified query. In getting the retweeters, we used the “GET statuses/retweets/id” Twitter API which takes tweet status id as parameter and then returns user Ids for users who retweeted that particular tweet. Figure 4 shows the code used for twitter API search/tweets

```

36 // use twitter api search/tweets
37 client.get("search/tweets", q, function (err, data) {
38     users = [];
39     // handle err
40     var err_msg = null;
41     if (err) {
42         err_msg = err.message;
43     } else {
44         for (var i in data.statuses) {
45             var tweet = data.statuses[i];
46             var user = tweet.user;
47             users.push({
48                 id_str: user.id_str,
49                 name: user.name,
50                 screen_name: user.screen_name,
51                 location: user.location,
52                 photo: user.profile_image_url,
53                 text: user.description
54             });
55             // add each tweet to array
56             tweets.push(tweet);
57         }
58     }
59 }

```

Fig. 4. Twitter API search code

Figure 5 is the home page interface. The public discussions are tracked by getting the keywords or hash-tags as input and then the text and date of the tweets are displayed. It is a pictorial evidence for the success implementation of the requirement defined in section 4.1

(Tracking Discussions).

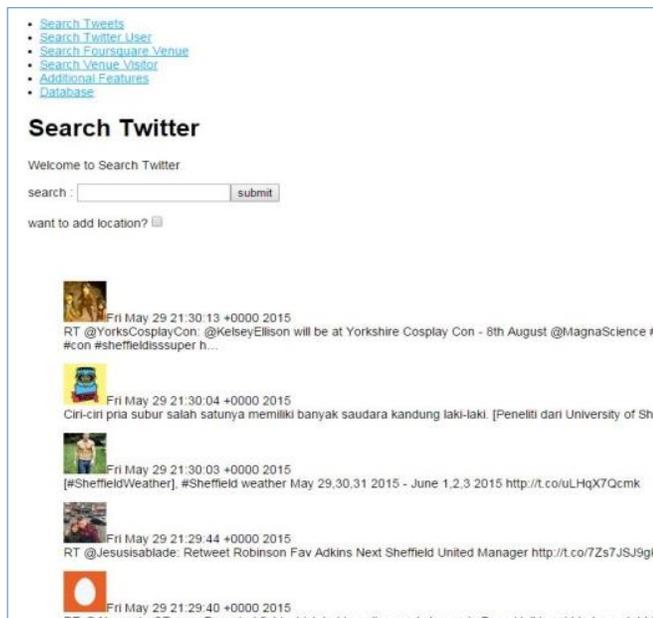


Fig. 5. Home Page / Tracking Discussions

For requirement (4.2a), In getting queries about specific users also “GET search/tweet” end point of REST API of twitter is used to search tweets from required users and since last x days. Tweets for all the users mentioned are obtained from API. Then keywords used in each tweet are extracted by splitting the string on every whitespace. An inverted index is formed with main index as keywords. Frequency of a keyword used by a particular user is also stored in index. With help of this index the most frequent x (given by user) keywords are found. The frequency of each of these x keywords for specific user and the total frequency are returned.

The page interface in in figure 6 lets users to enter multiple screen names (up to 10) and number of keywords to search as well as number of days. The resulting output is shown in figure 7. It shows frequency of word not only common to all users, but also for each user. The result in figure 7 indicates a successful implementation of the requirement.



Fig. 6. Web interface for multiple keyword search

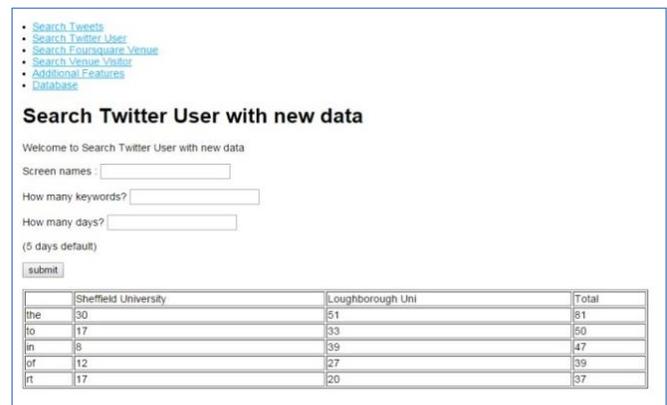


Fig. 7. Results in getting queries about specific users

For the requirement (4.2b), it is for foursquare which gets the specific venues visited/checked in by user over certain number of days. Combine Twitter API and Foursquare API are used. Stream API is used to catch real-time information. When the number of days is equal to zero, we used “STREAM statuses/tweet” Twitter API search twitter for certain user and catch the tweets only from foursquare. Then the “checkins/resolve” foursquare API is used to get venues which user has visited or checked in. At last, “venues/VENUE_ID” foursquare API is used to get photo of those venues. But in the situation in which the number of days is more than zero, “GET search/tweet” Twitter API is used for certain user and catch the tweets only from foursquare. Then the “checkins/resolve” foursquare API is used to get venues which user has visited or checked in. At last, “venues/VENUE_ID” foursquare API is used to get photo of those venues. Figure 8 is the code used to

configure API request and using get method.

```

15 // configure the request
16 var options = {
17   url: 'https://api.foursquare.com/v2/',
18   method: "GET",
19   headers: headers,
20   qs: {
21     oauth_token: config.access_token,
22     v: 20150529
23   }
24 }
25
26 // define get method for all get from foursquare api
27 foursquare.get = function(path, params, callback) {
28   options.url = 'https://api.foursquare.com/v2/' + path + "?";
29   if (params.shortId) {
30     options.qs['shortId'] = params.shortId;
31   }
32   for (var i in params) {
33     options.url += i + "=" + params[i] + "&";
34   }
35   console.log(options.url);
36   request(options, function (error, response, body){
37     callback(error, response, body, params);
38   });
39 };

```

Fig. 8. The code used to configure API request.

The page interface shown in Figure 9 allows user to enter user id to search the venues a specific user visited in the specified number of days. The resulting output is shown in figure 10 which shows all the venues visited by that user within the days requested (name, picture, address, and description of the venue).

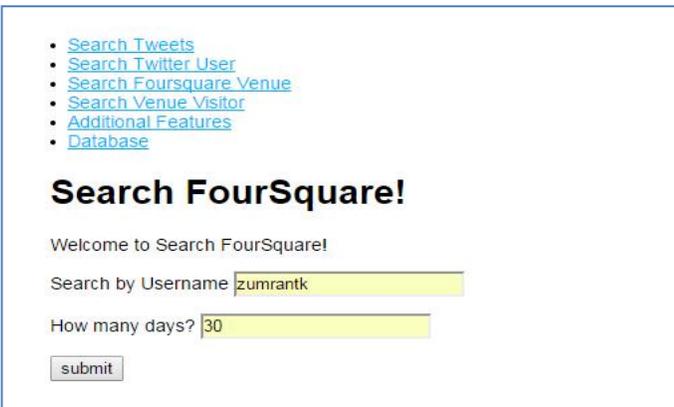


Fig. 9. Searching Foursquare Venue web interface

Search FourSquare!

Welcome to Search FourSquare!

Search by Username

How many days?

	name	categories	address
	Deutsch-Wagram	City	2232, Österreich
	Rathaus Wien	City Hall/Monument / Landmark/Bldg	Rathausplatz 1, 1010 Wien, Österreich
	Kärntner Straße	Street	Kärntner Str., 1010 Wien, Österreich
	Café de l'Europe	Café/Italian Restaurant	Graben 31, 1010 Wien, Österreich
	Stephansdom St. Stephen's Cathedral	Church/Historic Site	Stephansplatz 3, 1010 Wien, Österreich
	Lugner City	Mall	Gablenzgasse 5-13, 1150 Wien, Österreich

Fig. 10. searching Foursquare Venue/ results

In requirement (4.2c), the user is able to search users who visited a particular location within specific number of days. Figure 11 is a screen shot of the page interface. When number of days are 0, get ('geo/search') Twitter API is used to get coordinates (area) of this location. Then, stream ('statuses/filter') Twitter API is used to get tweets which are sent within the specified area. According to the tweets, get user information. When number of days is more than 0, get ('search/tweets') Twitter API is used to get coordinates (area) of this location with a particular query. This query contains the feature of a tweet which is from foursquare. The resulting output is a list of users who visited the location together with their details shown in Figure 12.

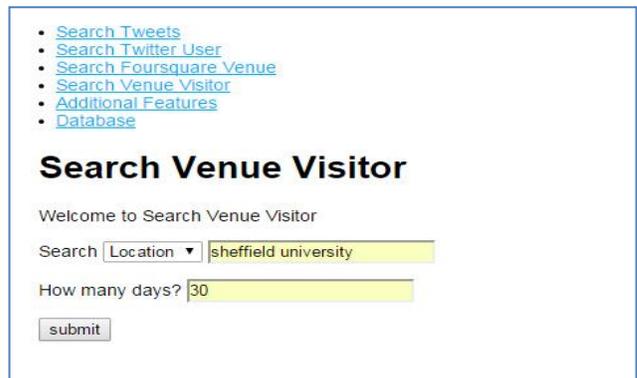


Fig. 11. Searching Venue Visitor web interface

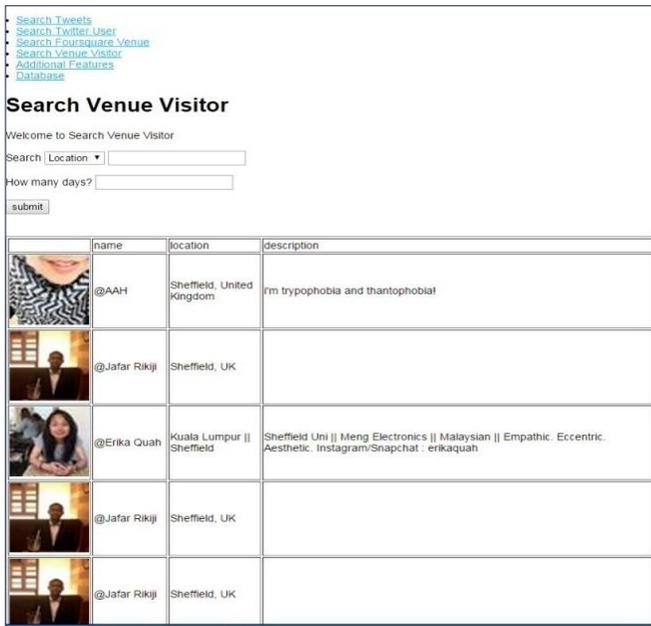


Fig. 12. results / who is visiting venues in a specific geographic area?

5. EVALUATION AND DISCUSSION

In this work we have used a simple table like displays to present and visualise the results of a given search results from which users can easily read and obtain their needed information. The authors of (Derczynski et al., 2013) state that analyzing spatial data which is related to tracking movements is a challenging task as the content of a message on social media may contain qualitative spatial data such as the name of a restaurant or quantitative spatial data such as the GPS coordinates of the restaurant. According to (Sadilek, Kautz, & Silenzio, 2012) less than half of mobile tweets contain GPS coordinates, which means in order to track movements the rest of 75% of tweets have to be analysed to produce accurate or near accurate results using techniques such as author profiles (Hecht, Hong, Suh, & Chi, 2011). With the simple web interfaces developed in this project we have addressed this challenge through which users can obtain spatial data of a queried user.

The website gives user a lot of experience in querying information from the social web. The System uses

different APIs. The result delivered in a user interface of a sophisticated website that is user friendly. The system succeeds in fulfilling all the requirements defined. In terms of data integrity, the data used in this study is taken from a source of sound integrity, and the data's accuracy can be relied upon.

In order to obtain the perception of users about the developed system a survey using the questionnaire method was carried out. The questionnaire contained 12 questions; the first 4 questions of the survey were designed to identify the system testing audience. The rest of the questions were designed to test the system performance from the audience's pint of view. The testing was conducted with 10 adult volunteers (the mode age range was 26-35, 7 males and 3 females), all of them having experienced social network sites. Figure 13 shows the results of questions 1, 2, 3, 4, 5 presented in the form of pie charts.

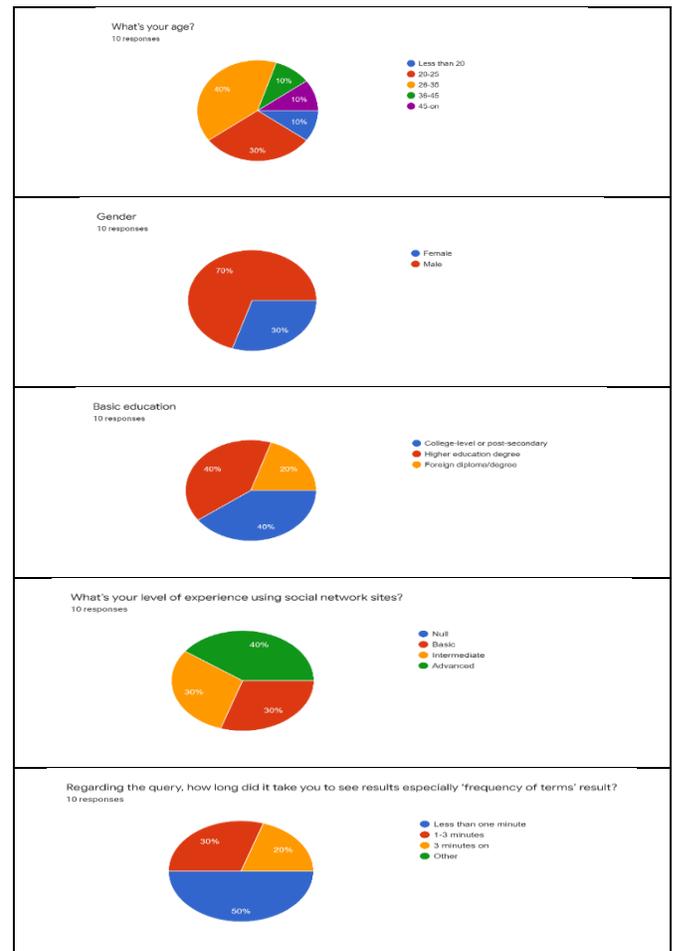


Fig. 13 Statistics for question 1, 2, 3, 4, 5 of the survey

The result shows that the majority (60%) own just one social media account and 20% owns between 2-3 social media accounts, and the other 20% more than 3 social media accounts. Table 5 shows questions 6 to 12 of the survey.

TABLE 5
System performance relating questions

Questions	Yes	No
6 Do you think the application interface is easy to use?	8	0
7 Are you able to limit the query to a specific geographic region via latitude/longitude?	8	0
8 When querying about specific users, in the last Z days (must be specified by the user), the user is able to see list of the most frequent keywords used by up to 10 users	7	1
9 When searching for venue visitors in a specific area, the result is list of tweets of this location. When user clicks the username, it will be directed to the twitter profile with all the general information	8	0
10 Accessibility online via the World Wide Web, The site is fully accessible online within all web browsers	6	2
11 Compatibility on Mobile and Tablet, the system can be run via mobile version and tablet	5	3
12 Low cost, it does not require any license or purchase	8	0

All the survey participants shared the opinion that the system succeeded in fulfilling the general requirements

according. The overall results for each question is shown in figure 14.

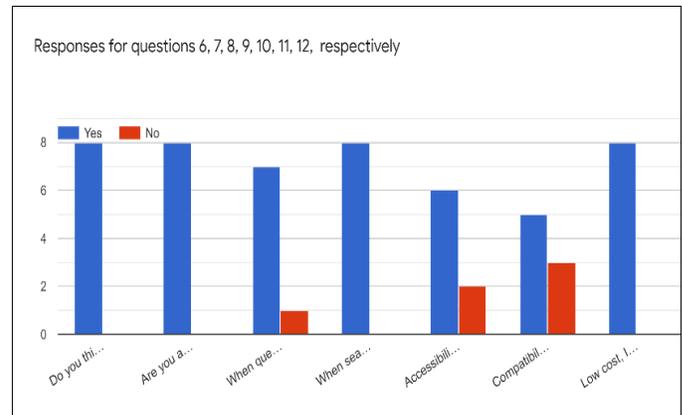


Fig. 14 Statistics for questions 6-12 of the survey

Finally, it is important to mention some of the limitations of the system. One of the limitations was the rate limitation we faced when using Twitter API and Foursquare. The rate limitation in Twitter API version 1.1 is considered on a per-user basis. If a method allows for 15 requests per rate limit window, then it allows the user to make 15 requests per window per leveraged access token. As for Foursquare, it defaults to 5,000 requests/day, if the users need more, they need to contact them which is again a limitation imposed on our system externally.

There are also limitations in the ability of storing information in the database due to the large amount of resources that we were obtaining from the server. Without a third-party query builder, creating a database using manual methods will be a challenge since the manual solution requires repetitive query. Through migration, our database may have the ability to be implemented in another platform such as SQLite, Oracle and so on

6. CONCLUSION AND FUTURE WORKS

Social media sites produce a large volume of textual information which contains user's opinions and experiences. This form of data is extremely valuable to businesses and organizations. The aim of the project

reported in this paper was to query different kinds of information from the social web (specifically Twitter.com and Foursquare.com.) and create a website that presents the data collected.

The main contribution of this paper is presenting a simple but effective method to collect, analyze and display social media queries regarding tracking and movements and sentiments of social media users (Twitter in particular). The method offers a practical advantage as the techniques and technologies used require a low learning curve. This is particularly useful to business owners and advertising companies who would like to benefit from the vast amount of data generated by social media to improve their businesses.

We suggest the following three key areas that we intend to research in the future:

- Mixed queries: What users have been discussing about in the last $X > 1$ days in a radius of Y km around a specific geographic point: this will require querying geotagged messages (e.g. tweets) and extract the contained keywords; the most frequent Z keywords will have to be displayed.
- Queries about venues: What points of interests (i.e. venues) are in a specific geographic area (input: name of location or coordinates)?
- What other venues are around a specified venue (input: id of venue). Pictures of venues could be added, taken from Google or from Flickr (almost certainly the best online photo management and sharing application in the world).

REFERENCES

1. Drews, F.A., Pasupathi, M., & Strayer, D.L. (2008). Passenger and cell phone conversations in simulated driving. *Journal of Experimental Psychology: Applied*, 14, 392-400. doi: 10.1037/a0013119
2. OAuth Authorisation Framework. (n.d.). Retrieved 15-5, 2019, from <https://oauth.net/>
3. Asur, S. & Huberman, B. A. (2010). Predicting the future with social media. In *2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology* (Vol. 1, pp. 492-499).
4. Batrinca, B. & Treleven, P. C. (2015). Social media analytics: a survey of techniques, tools and platforms. *AI & Society*, 30(1), 89-116.
5. Chang, J. & Sun, E. (2011). Location3: How users share and respond to location-based data on social. In *Fifth International AAAI conference on weblogs and social media*.
6. Derczynski, L. R., Yang, B. & Jensen, C. S. (2013). Towards context-aware search and analysis on social media data. In *Proceedings of the 16th international conference on extending database technology* (pp. 137-142).
7. Hasan, S. & Ukkusuri, S. V. (2015). Location contexts of user check-ins to model urban geo life-style patterns. *PLoS One*, 10(5).
8. Hecht, B., Hong, L., Suh, B. & Chi, E. H. (2011). Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 237-246).
9. Jansen, B. J., Zhang, M., Sobel, K. & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60(11), 2169-2188.
10. Kaplan, A. M. & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59-68.
11. Kwak, H., Lee, C., Park, H. & Moon, S. (2010). What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web* (pp. 591-600).
12. Manasa, K. & Padma, M. (2019). A Study on Sentiment Analysis on Social Media Data. In *Emerging Research in Electronics, Computer Science and Technology* (pp. 661-667). Springer.
13. Maynard, D., Bontcheva, K. & Rout, D. (2012). Challenges in developing opinion mining tools for social media. *Proceedings of The@NLP Can U Tag \# Usergeneratedcontent*, 15-22.
14. McClellan, C., Ali, M. M., Mutter, R., Kroutil, L. & Landwehr, J. (2017). Using social media to monitor mental health discussions- evidence from Twitter. *Journal of the American Medical Informatics Association*, 24(3), 496-502.
15. Phillips, L., Dowling, C., Shaffer, K., Hodas, N. & Volkova, S. (2017). Using social media to predict the future: a systematic literature review. *arXiv Preprint arXiv:1706.06134*.
16. Rashidi, T. H., Abbasi, A., Maghrebi, M., Hasan, S. & Waller, T. S. (2017). Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part C: Emerging Technologies*, 75, 197-211.
17. Ratkiewicz, J., Conover, M. D., Meiss, M., Gonçalves, B., Flammini, A. & Menczer, F. M. (2011). Detecting and tracking political abuse in social media. In *Fifth international AAAI conference on weblogs and social media*.

18. Sadilek, A., Kautz, H. & Silenzio, V. (2012). Modeling spread of disease from social interactions. In *Sixth International AAAI Conference on Weblogs and Social Media*.
19. Saumya, S., Singh, J. P. & Kumar, P. (2016). Predicting stock movements using social network. In *Conference on e-Business, e-Services and e-Society* (pp. 567–572).
20. Spyrtatos, S., Vespe, M., Natale, F., Weber, I., Zaghenni, E. & Rango, M. (2018). Migration data using social media: a European perspective.
21. Zhu, Q. (2017). Citizen-driven international networks and globalization of social movements on Twitter. *Social Science Computer Review*, 35(1), 68–83.