# INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATION & MANAGEMENT

IJRCM


A Monthly Double-Blind Peer Reviewed (Refereed/Juried) Open Access International e-Journal - Included in the International Serial Directories

*Indexed & Listed at:*

Ulrich's Periodicals Directory ©, ProQuest, U.S.A., EBSCO Publishing, U.S.A., Cabell's Directories of Publishing Opportunities, U.S.A.,

Open J-Gage, India [link of the same is duly available at Inflibnet of University Grants Commission (U.G.C.)],

Index Copernicus Publishers Panel, Poland with IC Value of 5.09 & number of libraries all around the world.

Circulated all over the world & Google has verified that scholars of more than 3480 Cities in 174 countries/territories are visiting our journal on regular basis.

Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA

http://ijrcm.org.in/

# CONTENTS

# CALL FOR MANUSCRIPTS

We invite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the areas of Computer Science & Applications; Commerce; Business; Finance; Marketing; Human Resource Management; General Management; Banking; Economics; Tourism Administration & Management; Education; Law; Library & Information Science; Defence & Strategic Studies; Electronic Science; Corporate Governance; Industrial Relations; and emerging paradigms in allied subjects like Accounting; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Rural Economics; Co-operation; Demography: Development Planning; Development Studies; Applied Economics; Development Economics; Business Economics; Monetary Policy; Public Policy Economics; Real Estate; Regional Economics; Political Science; Continuing Education; Labour Welfare; Philosophy; Psychology; Sociology; Tax Accounting; Advertising & Promotion Management; Management Information Systems (MIS); Business Law; Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labour Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; International Relations; Human Rights & Duties; Public Administration; Population Studies; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism & Hospitality; Transportation Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic; Web Design and emerging paradigms in allied subjects.

Anybody can submit the **soft copy** of unpublished novel; original; empirical and high quality **research work/manuscript** ***anytime*** in ***M.S. Word format*** after preparing the same as per our **GUIDELINES FOR SUBMISSION**; at our email address i.e. infoijrcm@gmail.com or online by clicking the link **online submission** as given on our website (***FOR ONLINE SUBMISSION, CLICK HERE***).

# GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION:**

**DATED: _____**

***THE EDITOR***
IJRCM

Subject: **SUBMISSION OF MANUSCRIPT IN THE AREA OF .**

**(e.g. Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)**

**DEAR SIR/MADAM**

Please find my submission of manuscript entitled '_____' for possible publication in your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published elsewhere in any language fully or partly, nor is it under review for publication elsewhere.

I affirm that all the author (s) have seen and agreed to the submitted version of the manuscript and their inclusion of name (s) as co-author (s).

Also, if my/our manuscript is accepted, I/We agree to comply with the formalities as given on the website of the journal & you are free to publish our contribution in any of your journals.

**NAME OF CORRESPONDING AUTHOR**:
Designation:
Affiliation with full address, contact numbers & Pin Code:
Residential address with Pin Code:
Mobile Number (s):
Landline Number (s):
E-mail Address:
Alternate E-mail Address:

**NOTES**:

a) The whole manuscript is required to be in ***ONE MS WORD FILE*** only (pdf. version is liable to be rejected without any consideration), which will start from the covering letter, inside the manuscript.
b) The sender is required to mentionthe following in the **SUBJECT COLUMN** of the mail:
**New Manuscript for Review in the area of** (Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/ Engineering/Mathematics/other, please specify)
c) There is no need to give any text in the body of mail, except the cases where the author wishes to give any specific message w.r.t. to the manuscript.
d) The total size of the file containing the manuscript is required to be below **500 KB**.
e) Abstract alone will not be considered for review, and the author is required to submit the complete manuscript in the first instance.
f) The journal gives acknowledgement w.r.t. the receipt of every email and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of manuscript, within two days of submission, the corresponding author is required to demand for the same by sending separate mail to the journal.

2. **MANUSCRIPT TITLE:** The title of the paper should be in a 12 point Calibri Font. It should be bold typed, centered and fully capitalised.

3. **AUTHOR NAME (S) & AFFILIATIONS:** The author (s) **full name**, **designation**, **affiliation** (s), **address**, **mobile/landline numbers**, and **email/alternate email address** should be in italic & 11-point Calibri Font. It must be centered underneath the title.

4. **ABSTRACT:** Abstract should be in fully italicized text, not exceeding 250 words. The abstract must be informative and explain the background, aims, methods, results & conclusion in a single para. Abbreviations must be mentioned in full.

5. **KEYWORDS:** Abstract must be followed by a list of keywords, subject to the maximum of five. These should be arranged in alphabetic order separated by commas and full stops at the end.

6. **MANUSCRIPT:** Manuscript must be in ***BRITISH ENGLISH*** prepared on a standard A4 size ***PORTRAIT SETTING PAPER***. It must be prepared on a single space and single column with 1" margin set for top, bottom, left and right. It should be typed in 8 point Calibri Font with page numbers at the bottom and centre of every page. It should be free from grammatical, spelling and punctuation errors and must be thoroughly edited.

7. **HEADINGS:** All the headings should be in a 10 point Calibri Font. These must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.

8. **SUB-HEADINGS:** All the sub-headings should be in a 8 point Calibri Font. These must be bold-faced, aligned left and fully capitalised.

9. **MAIN TEXT:** The main text should follow the following sequence:

INTRODUCTION

REVIEW OF LITERATURE

NEED/IMPORTANCE OF THE STUDY

STATEMENT OF THE PROBLEM

OBJECTIVES

HYPOTHESES

RESEARCH METHODOLOGY

RESULTS & DISCUSSION

FINDINGS

RECOMMENDATIONS/SUGGESTIONS

CONCLUSIONS

SCOPE FOR FURTHER RESEARCH

ACKNOWLEDGMENTS

REFERENCES

APPENDIX/ANNEXURE

It should be in a 8 point Calibri Font, single spaced and justified. The manuscript should preferably not exceed ***5000 WORDS***.

10. **FIGURES &TABLES:** These should be simple, crystal clear, centered, separately numbered &self explained, and **titles must be above the table/figure**. **Sources of data should be mentioned below the table/figure**. It should be ensured that the tables/figures are referred to from the main text.

11. **EQUATIONS:**These should be consecutively numbered in parentheses, horizontally centered with equation number placed at the right.

12. **REFERENCES:** The list of all references should be alphabetically arranged. The author (s) should mention only the actually utilised references in the preparation of manuscript and they are supposed to follow **Harvard Style of Referencing**. The author (s) are supposed to follow the references as per the following:

- All works cited in the text (including sources for tables and figures) should be listed alphabetically.
- Use (**ed.**) for one editor, and (**ed.s)** for multiple editors.
- When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order.
- Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
- The title of books and journals should be in italics. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
- For titles in a language other than English, provide an English translation in parentheses.
- The location of endnotes within the text should be indicated by superscript numbers.

**PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:**

**BOOKS**

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

**CONTRIBUTIONS TO BOOKS**

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

**JOURNAL AND OTHER ARTICLES**

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

**CONFERENCE PAPERS**

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–22 June.

**UNPUBLISHED DISSERTATIONS AND THESES**

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, Kurukshetra University, Kurukshetra.

**ONLINE RESOURCES**

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

**WEBSITES**

- Garg, Bhavet (2011): Towards a New Natural Gas Policy, Political Weekly, Viewed on January 01, 2012 http://epw.in/user/viewabstract.jsp

# CRAWLING TWITTER DATA

*A. PAPPU RAJAN*
*RESEARCH SCHOLAR*
*DEPARTMENT OF COMPUTER SCIENCE & RESEARCH CENTER*
*ST.XAVIER'S COLLEGE (AUTONOMOUS)*
*PALAYAMKOTTAI*

*S. P. VICTOR*
*DIRECTOR*
*DEPARTMENT OF COMPUTER SCIENCE & RESEARCH CENTER*
*ST.XAVIER'S COLLEGE (AUTONOMOUS)*
*PALAYAMKOTTAI*

**ABSTRACT**

*The aim of this paper is to present an outline for how extracting data from twitter. The explosion of Web 2.0 and due to the large volume of data from web resources such as discussion forum, review sites, blogs, digital forum and social media. A substantial amount of research has been directed towards mining these texts and concludes on the overall meaning of the users and to assign a grade to the result under discussion. Mine this available huge data to make it proper use and presentable , giving right solution to a particular problem is a big real challenge .After getting the result form twitter , the data can be analyzed and predict future trend or taking efficient decision on the particular problem. In this paper deals with an introductory idea about the Twitter, TWITTER API, Streaming API, Open Authorization (OAuth) Workflow, Register OAuth credentials to twitter R session.*

**KEYWORDS**
Twitter Data, Extracting data from Twitter, Twitter API.

## INTRODUCTION

In an era of technology with increasing number of smart gadgets and decreasing cost of internet service has led people to interact in virtual worlds, universally known as social networking sites. Sites like Facebook, Twitter, LinkedIn, Hi5, Myspace, etc., has made use of internet as a communication point which surpassed the traditional communication platform of real world. The ease of sharing information in various fields like education, social issues, job related issues and entertainment had made it still more popular. Seventy five per cent of internet users in India are in the age group of 15-34. Adding to it the young population all over the world see it as a newest online trend to express all their feelings, emotions, humiliations, sadness, and joy in the form of posts, tweets, status, videos, pictures etc., without considering the aftermath of their actions.

The objective of this paper is to throw lime light on how can we extract or crawl data from Twitter. This paper also concludes with the set of guidelines for social networking users to crawl data from Twitter, the crawled data can be used for further research on opinion mining or sentiment analysis

Twitter was created in March 2006 and was officially launched in July 2006. The growth of Twitter has been phenomenal, currently having reached over 200 million users all over the world and handling over 200 million tweets per day. Users sign up for an account on Twitter, and once they have an account they can begin to "TWEET," which is the terminology for sending a message. Users can subscribe to other user's tweets, a process known as FOLLOWING. These subscribers are known as "FOLLOWERS." By default, tweets that a user sends are visible to everyone, however, users can also choose to send tweets specifically to their followers that will not be visible to the public.

Users on Twitter are identified by a user name, and this user name is proceeded by the "@" symbol. When a user identifies another user in their tweet by their user name, it will be visible to the public, and the user that is referenced will be notified by Twitter that they have been "mentioned." [1]

If a user sees a tweet that is interesting and wants to pass the information along, they can "RETWEET" the post, which is similar to forwarding an email message to their followers. Retweets will generally be identified with an "RT" that is embedded in the message.

Lastly, messages can be grouped by topic or type by the use of HASHTAGS (#). A hashtag preceding the topic will allow Twitter users to find tweets related to a particular topic when performing a search. Twitter also has a location function. If users are tweeting from a mobile device, they can choose to turn on their location, and their latitude and longitude will be captured with the tweet.

Tweets can be related to anything, but much of the content on Twitter is related to several key categories. These categories were outlined in research done by Pear Analytics in 2009 on 2,000 tweets [2]. This study found that tweets were primarily related to six categories:

1. Pointless babble – 40%
2. Conversational – 38%
3. Pass along value – 9%
4. Self-promotion – 6%
5. Spam – 4%
6. News – 4%

While these numbers are related to a study that was done two years ago when Twitter was not as widely used as it is now, the general categorization of tweets likely still holds. The areas of interest would be categories 2, 3, and 6, which account for 51% of tweets. So the challenge is to determine how to analyze the tweets in such a way that the important information is separated from the information that is not important.

The data that was captured from Twitter includes the following information:

**TABLE 1.0: TWITTER INFORMATION**

| Caption | Purpose / Explanation |
|---|---|
| **User** | The username that sent the tweet |
| **Tweet** | The content of the tweet |
| **Timestamp** | The date and time the tweet was sent |
| **Tweet ID** | Twitter identification number of the tweet |
| **Geo** | Latitude and longitude of the user |

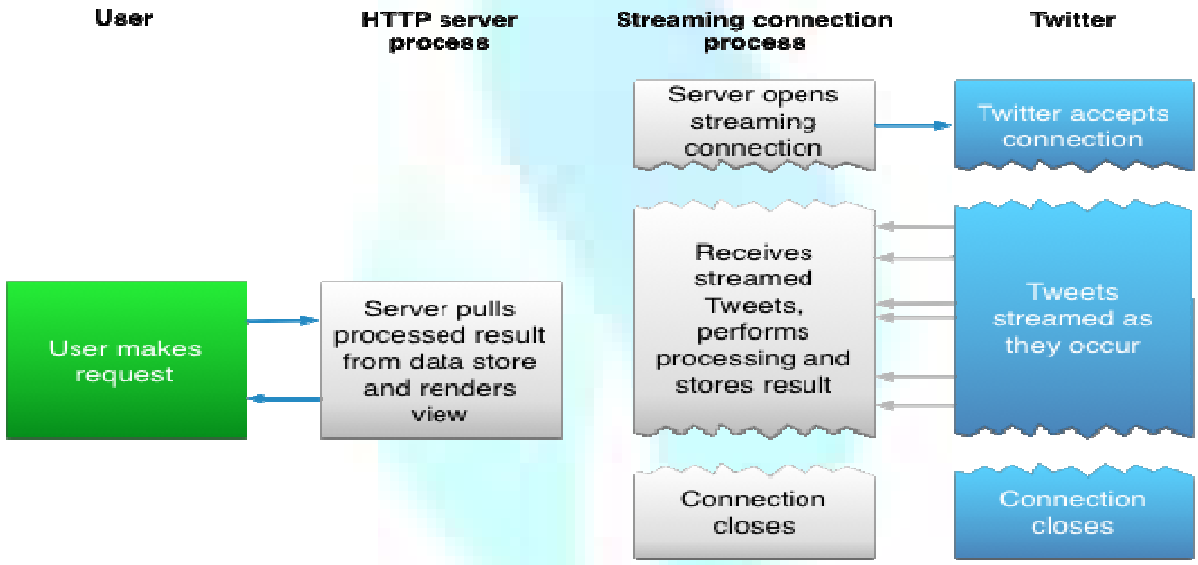
## EXTRACTING TWITTER DATA

Users on Twitter generate over millions of million Tweets each and every day. Some of these Tweets are available to researchers and practitioners through public APIs at no cost. At present Twitter accessibility user needs to pay some amount. But majority of the user don't know about how to access twitter data also why we needed this type of social media data .Whether it is useful to some extent of their human life Yes , all the business functional system was discussed over twitter also this is one of the trendsetter of any event . All the internet users are using this data for their business decision. Later in this paper we discuss about this matter very deeply. Now we will see how to extract the following types of information from Twitter. Usually these are the information can be extracting: Information about a user, Tweets published by a user, Search results on Twitter.

## TWITTER API

In general an application programming interface (**API**) specifies how some software components should interact with each other. The most procedural languages, an API specifies a set of functions or routines that accomplish a specific task or are allowed to interact with a specific software component. Twitter API data can be classified into two types based on their design and access method: REST APIs are based on the REST architecture now popularly used for designing web APIs. REST API s are having the following resources : Time lines , tweets search, streaming , direct message, friends and followers ,users ,suggested users ,favorites , lists, saved searchers, place and Geo , trends , spam reports, OAuth, help. These APIs use the pull strategy for data retrieval. To collect information a user must explicitly request it. Streaming APIs provides a continuous stream of public information from Twitter. These APIs use the push strategy for data retrieval. Once a request for information is made, the Streaming APIs provide a continuous stream of updates with no further input from the user. They have different capabilities and limitations with respect to what and how much information can be retrieved. The Streaming API has three types of endpoints: Public streams: , User streams, Site streams. Public streams are streams containing the public tweets on Twitter. User streams are single-user streams, with to all the Tweets of a user. Site streams are multi-user streams and intended for applications which access
Tweets from multiple users. As the Public streams API is the most versatile Streaming API, we will use it in all the examples pertaining to Streaming API. There are four layers in the streaming API. [7]
Layer 1: User
Layer 2: HTTP Process
Layer 3: Streaming connection process
Layer 4: Twitter data

**FIGURE : 1.0: STREAMING API**



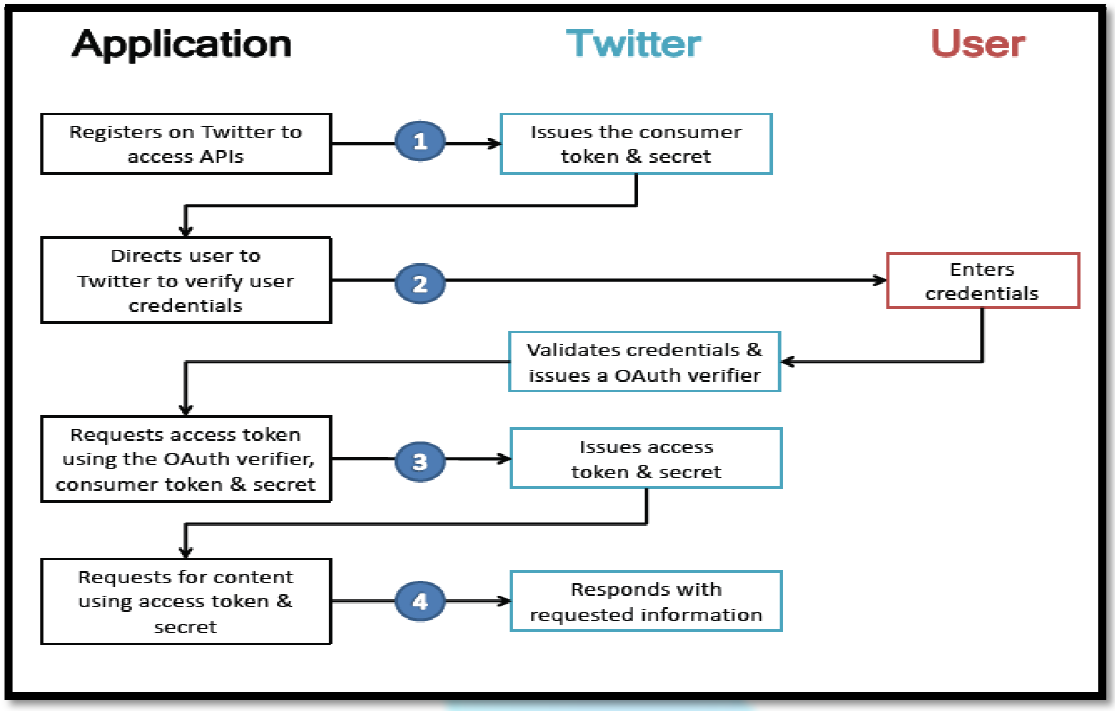

A web application is a type 3 tier layer. Here Layer 1 is a user layer which can sending request and getting response. Layer 2 HTTP Process layer which makes regular process of the Middle layer of the Web system but user requests, makes one or more requests to Twitter's API, then formats and prints the result to the user, as a response to the user's initial request: An app which connects to the Streaming APIs will not be able to establish a connection in response to a user request, as shown in the above example. Instead, the code for maintaining the Streaming connection is typically run in a process separate from the process which handles HTTP requests: The streaming process gets the input Tweets and performs any parsing, filtering, and/or aggregation needed before storing the result to a data store. The HTTP handling process queries the data store for results in response to user requests. While this model is more complex than the first example, the benefits from having a real time stream of Tweet data make the integration worthwhile for many types of apps.
The aforementioned types of information can be collected using both forms of Twitter API. Requests to the APIs contain parameters which can include hash tags, keywords, geographic regions, and Twitter user IDs. Responses from Twitter APIs are in JavaScript Object Notation (JSON) format. JSON is a popular format that is widely used as an object notation on the web. JSON (JavaScript Object Notation) is a lightweight data-interchange format. It is easy for humans to read and write. It is easy for machines to parse and generate. It is based on a subset of the JavaScript Programming Language, Standard ECMA-262 3rd Edition - December 1999. JSON is a text format that is completely language independent but uses conventions that are familiar to programmers of the C-family of languages, including C, C++, C#, Java, JavaScript, Perl, Python, and many others. JSON is built on two structures: A collection of name/value pairs. In various languages, this is realized as an object, record, struct, dictionary, hash table, keyed list, or associative array. An ordered list of values. In most languages, this is realized as an array, vector, list, or sequence. These are universal data structures. Virtually all modern programming languages support them in one form or another. It makes sense that a data format that is interchangeable with programming languages also be based on these structures.[8]
Twitter APIs can be accessed only via authenticated requests. Twitter uses Open Authentication and each request must be signed with valid Twitter user credentials. Access to Twitter APIs is also limited to a specific number of requests within a time window called the rate limit. These limits are applied both at individual user level as well as at the application level. A rate limit window is used to renew the quota of permitted API calls periodically. The size of this window is currently 15 minutes.

## OPEN AUTHENTICATION (OAuth)

Open Authentication (OAuth) is an open standard for authentication, adopted by Twitter to provide access to protected information. Passwords are highly vulnerable to theft and OAuth provides a safer alternative to traditional authentication approaches using a three-way handshake. It also improves the confidence of the user in the application as the user's password for his Twitter account is never shared with third-party applications.[5]

The authentication of API requests on Twitter is carried out using OAuth. The Fig. summarizes the steps involved in using OAuth to access Twitter API. Twitter APIs can only be accessed by applications. Below we detail the steps for making an API call from a Twitter application using OAuth :

**FIGURE 2.0: OPEN AUTHORIZATION (OAUTH) WORKFLOW**




## STEPS FOR WORKFLOW OF OPEN AUTHORIZATION

Step1. Applications are also known as consumers and all applications are required to register themselves with Twitter. Through this process the application is issued a consumer key and secret which the application must use to authenticate itself to Twitter.
Step 2. The application uses the consumer key and secret to create a unique Twitter link to which a user is directed for authentication. The user authorizes the application by authenticating himself to Twitter. Twitter verifies the user's identity and issues a OAuth verifier also called a PIN.
Step 3. The user provides this PIN to the application. The application uses the PIN to request an "Access Token" and "Access Secret" unique to the user. The "Access Token" and "Access Secret" for a user do not change and can be cached by the application for future requests.

## REGISTER OAUTH CREDENTIALS TO TWITTER R SESSION

**Function Name**: registerTwitterOAuth ()
**DESCRIPTION**
This function is used to provide OAuth access tokens to your twitter session. This will enable many bits of functionality as well as allow other commands to provide more options.
**Usage:** getTwitterOAuth (consumer_key, consumer_secret)
registerTwitterOAuth (oauth)
**ARGUMENTS**
**consumer_key** - The consumer key supplied by Twitter
**consumer_secret** - The consumer secret supplied by Twitter
**oauth** - An object of class OAuth
**DETAILS**
The getTwitterOAuth function is a wrapper around the call to OAuthFactory and registerTwitterOAuth, which will return the registered credentials. If your workflow is such that you save the credentials and register them in later R sessions, feel free to do this using registerTwitterOAuth registerTwitterOAuth will store the OAuth argument in an environment which is then accessed throughout the package. When API calls are made, instead of going through RCurl they will go through the ROAuth package. Three URLs will need to be used for the initial OAuth handshake.
**VALUE**
TRUE on success, otherwise an error will be thrown.

**FIGURE 3.0: GENERATING OAUTH FOR A USER**

| Generating OAuth Token for a User |
|---|
| >library(twitter) |
| >library(ROAuth) |
| >library(plyr) |
| >library(stringr) |
| >reqURL <-"https://api.twitter.com/oauth/request_token" |
| >accessURL <-"http://api.twitter.com/oauth/access_token" |
| >authURL <-"http://api.twitter.com/oauth/authorize" |
| >consumerKey <-"xxxxxxxxxxxxxxxxxxxx" |
| >consumerSecret <-"xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx" |
| >twitCred<-OAuthFactory$new |
| (consumerKey=consumerKey,consumerSecret=consumerSecret,requestURL=reqURL,accessURL=accessURL,authURL=authURL) |
| >library(RCurl) |
| >options(RCurlOptions = list(cainfo = system.file("CurlSSL","cacert.pem",package="RCurl"))) |
| >twitCred$handshake() |

## TWITTER DOMAIN (REQUIREMENTS)

Twitter messages have certain unique attributes that make it complex to analyze and classify while training the data sets. Some of them are as follows:

**LENGTH:** The maximum length of a Twitter message is 140 characters, thus the limited length of tweet, might comprise of one or two sentences. Thus researcher task is the simple breakdown of the tweet to extract the polarity from it.

**AVAILABLE DATA:** The twitter API supports searching tweets pertaining to a query thus we can obtain a large training set using the twitter API.

**LANGUAGE MODEL.** The language under consideration is English. Basic problem faced is the limited length which encourages the user to use slang and acronyms. Moreover users post tweet from different devices including mobile phones, thus increasing the rate of misspelling words.

**NEGATIVE SENTENCES:** Many people would write their tweets with negation before the adjective or verb, which complicates the training set..

**CONFUSING POLARITY:** For certain tweets there will be a confusion or disagreement for the polarity to be assigned.

**DEALING WITH EMOTICONS:** The training data should contain clean labels. The emoticons serve as a noisy label. There are some cases in which the emoticon label would normally not make sense to a human evaluator. For example user ayakyl tweeted; hurrah :) won't have class today!!!! If we remove the emoticon from this phrase, it becomes hurrah won't have class today!!!! in which a human evaluator would normally assess as negative.

**CASUAL LANGUAGE:** Tweets contain very casual language. For example, a user may want to right the word happy as: happpppyyy happpiieee happy hap-e besides showing that people are happy; this emphasizes the casual nature of Twitter and the disregard for correct spelling.

**USERNAMES:** Users often include usernames in their tweets, in order to address messages to particular users. A de facto standard is to include the @ symbol before the username

After receiving the consumer key and consumer secret key, the user could register the secret keys with the Twitter. As stated in Listing1.1, we could store the secret keys in variables and using "twitCred$handshake()" function, the consumer key and consumer secret key is registered to Twitter, requesting for access into the public tweets in the Twitter.

**NOTE**

For Mac OS: The registration process could be done only once.

For Windows OS: The registration process must be repeated every time the user tries to access the tweets.

**FIGURE 4.0: REGISTRATION WITH TWITTER TO ACCESS THE PUBLIC TWEETS**

```
Registration with Twitter to access the public tweets

To enable the connection, please direct your web browser to:
http://api.twitter.com/oauth/authorize?oauth_token=xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
When complete, record the PIN given to you and provide it here: xxxxxxx          ##7 Digit PIN Number
>registerTwitterOAuth(twitCred)
##returns TRUE on success
```

When the handshake with the Twitter DB is over, it returns a URL to where the Web browser must be directed. When the user accesses the URL, an OAuth Token is returned which is a 7-digit number. This number is again entered into the R console. The command "registerTwitterOAuth(twitCred)" is used to complete the registration process, which enables the user to access the Twitter API for fetching users' tweets. It returns 'TRUE' to indicate the connection between the R and Twitter API.

## SCIENTIFIC FUNDAMENTALS AND SENTIMENT ANALYSIS

As people are free to give their opinions on anything, example they buy a brand and they express their views on brands in various social networking sites like Twitter, Facebook, Discussion forums, and blogs. Sentiment analysis, also called Opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics and their attributes. In Conversation mining the additional task is to determine the nature of opinion: whether it is positive or neutral in general. Opinion mining is a type of Natural language processing for tracking the attitudes, feelings or appraisals of the public about particular topic, product or services. Sentiment analysis can be useful in several ways

## CONCLUSION

This research introduce the basic of introductory idea about the Twitter , TWITTER API , Streaming API , Open Authorization (OAuth) Workflow, Register OAuth credentials to twitter R **session.** . With the information overload, Web sentiment mining is a new and promising research issue to help users in gaining insight into overwhelming information on the web social media. In this paper, we present a preliminary discussion about crawling data from twitter. There still remain many areas for further research, such as the design of efficient algorithms for opinion mining or sentiment mining for large document collections, and so on. We can do research in Some new techniques can provide the user with the opportunity to analyze the twitter log and determine the sentiment of the text, whether it is positive or negative, which is extended to strength of polarity.

## REFERENCES


1. A.Pappu Rajan ,(2013)," A Study on Security Threat Awareness among Students Using Social Networking Sites, by Applying Data Mining Techniques", International Journal Of Research In Commerce, IT & Management, Vol. No. 3 , Issue No. 09 : ISSN 2231-5756
2. Bissattini C., Christodoulou K.(2013),"Web sentiment analysis for revealing public opinions, trends and making good financial decisions", *Journal of Advanced Research in Computer Science and Software Engineering.*
3. David Alfred Ostrowski, (2010), "System Analytics Research and Innovation Center Ford Motor Company, Sentiment Mining within Social Media for Topic Identification", paper presented at IEEE International Conference on Semantic Computing,Malaysia.
4. http://www.json.org Viewed on 28 March, 2014
5. https://dev.twitter.com/docs/api/1.1 Viewed on 20 March, 2014
6. Manjaly J.S. (2013),"Twitter based sentiment analysis for subject identification" International *Journal of Advanced Research in Computer and Communication Engineering*.226-229
7. Mikut, R. and Reischl, M., (2011)," Advanced review: Data Mining Tools", John Wiley and Sons Inc.
8. Nakov P.,Kozareva Z., Ritter A.(2013),"*Sentiment analysis in Twitter*. Second Joint Conference on Lexical and Computational Semantics". Atlanta, Georgia.
9. Newman B., (2013, March 25),"The importance of social media analytics", *Fairfield County Business Journal*.
10. Savage N. (2011, March)," Twitter as medium and message", Communications of the ACM *Society*, Issue: 3, Vol.54. pp: 18-20.
11. Younggue B.,Hongchul L.(2012),"Sentiment analysis of Twitter audience: Measuring the positive or negative influence", *Journal of the American Society for Information Science and Technology*.

# REQUEST FOR FEEDBACK

**Dear Readers**

At the very outset, International Journal of Research in Computer Application & Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you tosupply your critical comments and suggestions about the material published in this issue as well as on the journal as a whole, on our E-mail**infoijrcm@gmail.com** for further improvements in the interest of research.

If youhave any queries please feel free to contact us on our E-mail infoijrcm@gmail.com.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward an appropriate consideration.

With sincere regards

Thanking you profoundly

**Academically yours**

Sd/-
**Co-ordinator**

# DISCLAIMER

The information and opinions presented in the Journal reflect the views of the authors and not of the Journal or its Editorial Board or the Publishers/Editors. Publication does not constitute endorsement by the journal. Neither the Journal nor its publishers/Editors/Editorial Board nor anyone else involved in creating, producing or delivering the journal or the materials contained therein, assumes any liability or responsibility for the accuracy, completeness, or usefulness of any information provided in the journal, nor shall they be liable for any direct, indirect, incidental, special, consequential or punitive damages arising out of the use of information/material contained in the journal. The journal, nor its publishers/Editors/ Editorial Board, nor any other party involved in the preparation of material contained in the journal represents or warrants that the information contained herein is in every respect accurate or complete, and they are not responsible for any errors or omissions or for the results obtained from the use of such material. Readers are encouraged to confirm the information contained herein with other sources. The responsibility of the contents and the opinions expressed in this journal is exclusively of the author (s) concerned.

# ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

## *Our Other Journals*