Keyword Extraction from Tweets using Graph-Based Methods

G. Dhana Lakshmi, M. Kanthi Rekha

Abstract: Social media refers to a set of different web sites like Twitter is a microblogging service that generates a huge amount of textual content daily. These methods based on text mining, natural language processing, and information retrieval are usually applied. The text mining approaches, documents are represented using the well-known vector space model, which results in sparse matrices to be dealt with computationally. A technique to extract keywords from collections of Twitter messages based on the representation of texts employing a graph structure, from which it is assigned relevance values to the vertices, based on graph centrality measures. The proposed approach, called TKG, relies on three phases: text pre-processing; graph building and keyword extraction. The first experiment applies TKG to a text from the Time magazine and compares its performance with TFID [1] and KEA[6], having human classifications as benchmarks. Finally, these algorithms are designed to the sets of tweets of increasing size were used and the computational time necessary to run the algorithms was recorded and compared. The results obtained in these experiments showed that building the graph using an all neighbors edging scheme invariably provided superior performance, and assigning weights to the edges based on the weight as the inverse co-occurrence frequency was superior cases. One possible future work is to apply centrality measures TKG showed to be faster for all its variations when compared with TFIDF and KEA, except for the weighting scheme based on the inverse co-occurrence frequency. TKG is a novel and robust proposal to extract keywords from texts, particularly from short messages, such as tweets.

Keywords: Twitter, Text mining, Graph-based text representation, Centrality, Keyword extraction

I. INTRODUCTION

A tremendous measure of information is redressed in an everyday life as a result of which the worldwide economy is being varied to the words which are expressed by Andreas M. Kaplan and Michael Haenlein[3]. Numerous organizations selecting logical specialists to remove the information from different assets like Facebook, Twitter about Marketing, Weather Reports, most recent Trending news, Gadgets these all are assuming an imperative part in gushing the information from various parts of the world [3].

In electronic contraptions to send the information into the servers, we have a few sensors to introduce to remove the information. To comprehend the conclusions of each person and to bring the mindfulness into the general population, we have Social Networking destinations are utilized to facilitate the vast measure of metadata is made and investigated. Recordings are recording with sound and which have an energized substance of 3D visual impacts are for drawing nearer better and for good understanding. These are turning out to be extremely mainstream on YouTube in the early days.

The whole world is focusing more on creating different advancements and devices and investigate the business sector in the early days.

In present occasions, colleges are presented different advances furthermore give heaps of courses about the advances to taught the understudies who need to seek after yet before joining in any college give a few sentiments in any long-range informal communication locales like Twitter by tapping the twitter information and composing a report which is investigated and make a proposal about the information is being recorded in everyday life.

Hadoop Distributed File Systems is utilized to figure capacities and to make new machine calculations because of enhanced and propelled advances keeping in mind the end goal to get the understanding of how the mass information is put away. IT organizations are picking India as a base and creating and elevating their items to the business sector and offering them at the less expensive costs as due to India is a quick creating nation of IT commercial ventures.

II. RELATED WORK

The existing system has collected Twitter data based on Keyword extraction [3] is tasked with the automatic identification of terms that best describe the subject of a document. A Text mining approach is to write the documents and represented using the well-known vector space model [1]. In graph-based text representation, a text can be seen as a collection of interconnected words. The networks are analyzed within an advanced Gephi package, providing several measures- degree centrality, closeness centrality, eccentricity centrality [2]. Twitter is a website under the category of social networking and microblogging. The main feature of Twitter is messaging. In Twitter, the keywords are extracted from a set of tweets represented as a graph of messages posted by using the graph-based and corpus-oriented approach. This text representation model allows the capture of statistical information, such as term co-occurrence frequency, centrality measures, and aggregation of linguistic features.

Revised Manuscript Received on March 18, 2020.
  G. Dhana Lakshmi, Department of Computer Science and Engineering, University College of Engineering(A), JNTUK, Kakinada, Andhra Pradesh, India (Email: Dhanalakshmi.gvp@gmail.com)
  M. Kanthi Rekha, Department of Computer Science and Engineering, University College of Engineering(A), JNTUK, Kakinada, Andhra Pradesh, India (Email: Dhanalakshmi.gvp@gmail.com)
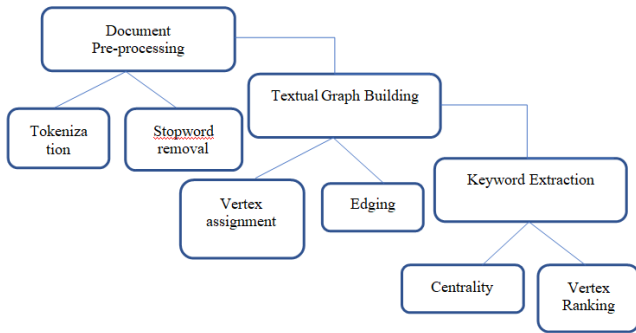
16

Fig 1: Twitter Keyword Graph

III. DEMERITS OF EXISTING SYSTEM

As in the case of an existing system, the problem becomes even worse when the target application involves Twitter messages. Due to short texts (Up to 140 characters), informality, grammatical errors, buzzwords, slangs, and the speed with which real-time content is generated, approximately 250 million messages are posted daily. While assigning keywords to the document manually is a

- Very Costly
- Time Consuming
- Tedious Task and
- The number of digitally available documents are growing

IV. PROBLEM STATEMENT

To overcome the problems in Existing System, the proposed technique to extract key-words[5] from collections of Twitter messages based on the representation of texts employing a graph structure, from which it is assigned relevance values to the vertices based on graph centrality measures. To assess the performance of the proposed approach, three different sets of experiments are performed. The first experiment applies TKG to a text from the Time magazine and compares its performance with centrality. The second set of experiments takes tweets from three different TV shows, applies TKG and compares it with and KEA[6], having human classifications as benchmarks. Finally, these three algorithms are applied to tweets sets of increasing size and their computational running time is measured and compared. The results show that TKG is a novel and robust proposal to extract keywords from texts, particularly from short messages, such as tweets.

V. IMPLEMENTATION PROCESS

This paper aims at extracting tweets from twitter graph-based and corpus-oriented approach by using PYTHON and implement techniques of centrality, degree, closeness, and eccentricity to the extracted data, a technique to extract key-words from collections of Twitter messages based on the representation of texts using a graph structure, from which it is assigned relevance values to the vertices based on graph centrality measures by using Gephi software tool. So, we can eliminate unwanted tweets from twitter.

The proposed work would break down the twitter information for a particular hashtag name exampled as "Mercedes Benz". At that point, the extracted tweets will be

processed with a few measures - degree, closeness, flightiness, Katz centrality by which the advancement of correspondence around a specific idea is evaluated.
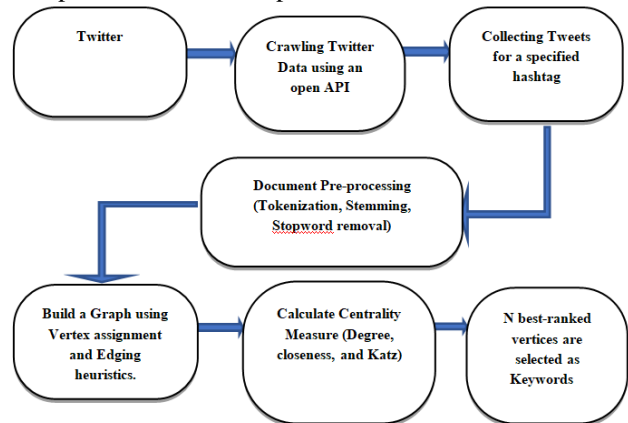

Fig.2: Block Diagram

A. Centrality

Centrality[13] characterizes how essential a hub is inside a system.

B. Degree Centrality

Degree centrality [13] exchanges the same thought into a measure. The degree centrality measure positions hubs with more associations higher as far as centrality.

The degree centrality C_D for the hub v_i in an undirected diagram is

$$C_{D(v_i)} = d_i$$

Where d_i is the degree (number of adjoining edges) of v_i a hub In coordinated diagrams, we can either utilize the in-degree, the out-degree, or the blend as the degree centrality esteem:

$$C_{D(v_i)} = d_i^{in}$$
$$C_{D(v_i)} = d_i^{out}$$

C. Closeness Centrality

In closeness centrality [13], the more focal hubs are, the all the more rapidly they can achieve different hubs. Formally, these hubs ought to have a littler normal most limited way length to different hubs. Closeness centrality is characterized as

$$C_C v_i = \frac{1}{l_{vi}}$$

Where $l_{vi} = \frac{1}{n-1} \sum_{vj \neq vi} l_{i,j}$ is node v_i's normal most brief way length to different hubs. The littler the normal most limited way length, the higher the centrality for the hub.

D. Eigenvector Centrality

Eigenvector centrality (also called eigen centrality) is a measure of the influence of a node in a network. It assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes. Google's PageRank is a variant of the eigenvector centrality measure[28].

VI.   RESULT AND DISCUSSION

We utilized Twitter's open API to gather approximately 10,000 tweets crosswise over five times that we consequently channel down to 3,000 messages because of the catchphrases found in the tweets. The messages have a normal length of 18 tokens, and the corpus vocabulary includes 4,68,000 special words.

Pre-processing just a little portion of Twitter messages are pertinent to the objective removing assignment. Straightforwardly handling the crude unfiltered stream would restrictively increment computational expenses and make adapting more troublesome because of the clamor characteristic in the information.

A. Katz Centrality

A noteworthy issue with eigenvector centrality emerges when it considers coordinated diagrams. Centrality is just passed on when we have active edges, and in exceptional cases, for example, when a hub is in a coordinated non-cyclic diagram, centrality gets to be zero, even though this hub can have numerous edges associated with it. The inclination term β is added to the centrality values for all hubs regardless of how they are arranged in the system. The subsequent centrality measure is known as the Katz Centrality and is detailed as takes after:

$$C_{Katz}(\upsilon_i) = \alpha \sum_{j=1}^{n} A_{j,i}\ C_{Katz}(\upsilon_i) + \beta \ \dots\text{(i)}$$

The main term is like the eigenvector centrality, and its impact is controlled by the steady α. The second term β, is the inclination term that keeps away from zero centrality values. We can rework the above condition in a vector structure as,

$$C_{Katz} = \alpha\ A^T\ C_{Katz} + \beta 1 \qquad \dots\text{(ii)}$$

Where 1 is a vector of each of the 1's. taking the main term to one side hand side and considering C_{Katz}

$$C_{Katz} = \beta(I - \alpha\ A^T)\text{-}1.1 \qquad \dots\text{(iii)}$$

Getting Twitter Data: Gathering twitter information OAuth 2.0 library to get to live stream, and has built up a twitter application as depicted in the accompanying strides.
1. Create a twitter account.
2. Go to https://dev.twitter.com/applications and log in with your twitter certifications.
3. Click "Make New App".
4. Fill out the structure and consent to the terms.
5. On the following page, tap the "Programming interface Keys" tab along the top, then look over the distance down until you see the segment "Your Access Token".
6. Click the catch "Make My Access Token".
7. You will now duplicate four qualities into twitterstream.py. These qualities are your "Programming interface Key", your "Programming interface Secret", your "Entrance Token" and your "Entrance Token Secret " every one of the four should now be obvious on the API Keys page.

Open twitterstream.py and set the variable comparing to the programming interface key, programming interface mystery, access token, access token mystery. You will see code like the beneath.

api_key = "<Enter your programming interface key>"
api_secret = "<Enter your programming interface secret>"

api_token = "<Enter your entrance token key>"
api_token_secret= "<Enter your entrance token mystery key>"
8. Run the accompanying and ensure you see information streaming and that no blunders happen. $ python twitterstream.py>output.txt
9. If you wish, change the record to utilize the twitter hunt API to look down indicated terms. For instance, to hunt down the expressions "Microsoft", you can pass the accompanying URL to the twitter demand capacity: https://api.twitter.com/1.1/seek/tweets.json?q=microsoft

With the application created above, we have gathered around 10MB to 15MB live streams tweets. Presently the gathered tweets have been separated to pick tweets identified with courses, colleges physically. The tweets in light of the system and content corpus.

Stopwords List:

	Stopwords List		
	Terms	Terms.1	Freq
1	amg	amg	202
2	amp	amg	1
3	automotive	amg	14
4	benz	amg	117
5	black	amg	15
6	buy car	amg	0

Fig 3: Network generated with the concept of closeness centrality containing the hashtag: #cars
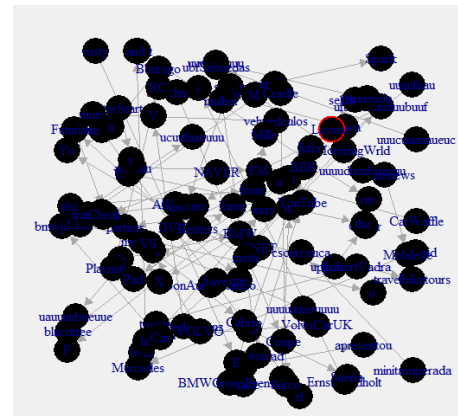

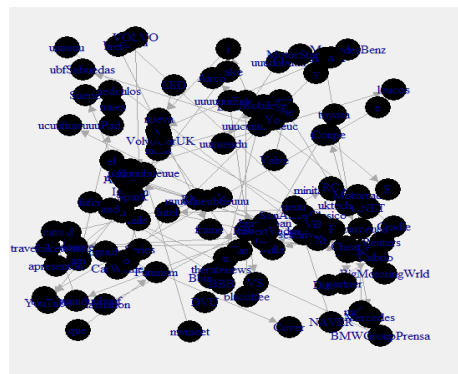**Fig 4: Network generated with the concept of Katz centrality containing the hashtags: #cars**

18

Table 1: Katz Centrality for the keywords of the nearest neighbor (TKG1) contained with the hashtag: #cars

Words in the Tweets	Katz Centrality		
	TKG1W1	TKG1WF	TKG1W1/F
Volvo	0.297	0.310	0.321
Bmw	0.365	0.374	0.369
Autonomous	0.354	0.364	0.370
Cars	0.239	0.232	0.240
Mercedesbenz	0.480	0.480	0.481
Technology	0.171	0.169	0.175
Intel	0.308	0.300	0.310
Roadshows	0.056	0.057	0.058
Mobileye	0.480	0.480	0.481
Online	0.239	0.232	0.240
Product	0.480	0.480	0.481
Price	0.171	0.169	0.175
Automotive	0.308	0.300	0.310
Plata	0.056	0.057	0.058
Drive	0.480	0.480	0.481
Radio	0.308	0.300	0.310
Secreted	0.056	0.057	0.058
Dos	0.480	0.480	0.481

The above table described that the comparison between the tweets of TGK1W[1], TKG1W[F], and TKG1W[1/F] with TFIDF and with the values of Katz centrality. Here the centrality value of TKG1W[1/F] gives the superior performance when compared with TGK1W[1] and TKG1W[F].

Table 2: Katz Centrality of the keywords of all neighbor (TKG2) contained with the hashtags: #cars

Words in the Tweets	Katz Centrality		
	TKG2W1	TKG2WF	TKG2W1/F
Volvo	0.477	0.410	0.519
Bmw	0.173	0.326	0.446
Autonomous	0.124	0.293	0.343
Cars	0.096	0.186	0.208
Mercedesbenz	0.094	0.214	0.391
Technology	0.229	0.138	0.355
Intel	0.221	0.178	0.357
Roadshows	0.292	0.098	0.340
Mobileye	0.219	0.184	0.232
Online	0.153	0.087	0.225
Product	0.114	0.093	0.215
Price	0.185	0.125	0.279
Automotive	0.517	0.419	0.523
Plata	0.620	0.622	0.618
Drive	4	0.517	0.516
Radio	0.594	0.518	0.515
Secreted	0.415	0.487	0.411
Dos	0.467	0.387	0.415

The above table described that the comparison between the tweets of TGK2W[1], TKG2W[F], and TKG2W[1/F] with TFIDF and with the values of Katz centrality. Here the centrality value of TKG1W[1/F] gives the superior performance when compared with TGK2W[1] and TKG2W[F].

Table 3: TFIDF

Term	Frequency
Volvo	0.2625
BMW	0.1957
Patch	0.1555
Car	0.1522
Technology	0.1304
Plata	0.0981
Product	0.1037
Price	0.0917
Lily	0.0864
Online	0.0778
Mobileye	0.0702
Roadshow	0.0693
Carl	0.0585
Secreted	0.0585
Radio	0.0456
Scene	0.0582
Drive	0.0549
Dos	0.0533
Ted	0.0524

In the hashtags, Volvo, BMW, and Mercedes Benz, to check off 145 edges by TKG1 and 940 edges by TKG2 for both the graphs, the edges may calculate the weights as W1, WF, and W1/F which has been removed by 3000 tweets gathered from twitter information.

4.2 Performance Evaluation

For Prediction issues, the outcomes are regularly introduced as a disarray lattice by utilizing Precision, Recall, F-measure. Accuracy is a measure of result importance, while the review is a measure of what number of genuinely applicable results is returned. A high zone under the bend speaks to both high review and high exactness, where high accuracy identifies with a low false.

The well-known Precision, Recall, F-measure from information retrieval were used as evaluation metrics:

$$\text{Precision (Pr)} = \frac{|\{Relavant\} \cap \{Retrieved\}|}{|\{Retrieved\}|} \quad \ldots\ldots(i)$$

$$\text{Recall (Re)} = \frac{|\{Relavant\} \cap \{Retrieved\}|}{|\{Retrieved\}|} \quad \ldots\ldots\ldots(ii)$$

$$\text{F-measure} = F = 2*\frac{Pr*Re}{(Pr+Re)} \quad \ldots\ldots\ldots\ldots(iii)$$

As the number of catchphrases recovered by every technique was characterized as 10, to process Pr the number of important archives recovered was equivalent to the number of watchwords proposed by the strategy that shows up in no less than one of the human records. By difference, to process Re the number of important reports was considered to the number of watchwords in the convergence set of the three human records.

To evaluate the computational versatility of the calculations; that is, how their computational time increments in connection to the dataset size, another arrangement of analyses were performed. A dataset with 50,000 tweets identified with the project "Volvo" was taken and inspected as takes after: {0.1, 1, 10, 20, 50 k} tweets. For each of these five datasets sizes, the running time of the calculations was figured and put away for examination. The calculations were kept running in a PC, Win 7, Core i7 2.2 GHz, 8 GB RAM. TFIDF, and the chart era of TKG kept running in Java SE 1.7, and the centrality measures of TKG were run utilizing Graph-tool.

Table 4: Comparison of TFIDF with TKG of Tweets with Precision, Recall and F-measure

S. No	TKG1W1 TKG1WF/ TKG1W1/ F	TKG2W1/TKG2WF TKG2W1/F	TFIDF
1	**Volvo**	**Volvo**	**Volvo**
2	**B**		**MW** **BMW** **BMW**
3	Autonomous	Creta	Patch
4	**Cars**	**Cars**	**Car**
5	Mercedes-Benz	Velocidad	**Technology**
6	**Technology**	**Technology**	Plata
7	Intel	Intel	Product
8	**Roadshows**	**Roadshows**	**Price**
9	**Mobileye**	**Mobileye**	Lily
10	Online	Online	Online
11	Product	Exhaust	**Mobileye**
12	**Price**	Product	**Roadshow**
13	Automotive	Automotive	Carl
14	Plata	**Price**	Secreted
15	Drive	Drive	Radio
16	Radio	Todo	Scene
Precision	79%	80%	79%
Recall	60%	58%	58%
F-measure	39%	40%	31%

Table 5: Comparison of TFIDF with TKG of Tweets for Katz centrality with Precision, Recall and F-measure

S. No	TKG1W1/ TKG1WF/ TKG1W1/ F	TKG2W1/ TKG2WF/ TKG2W1/ F	TFIDF
1	0.309	0.468	0.195
2	0.369	0.315	0.155
3	0.362	0.253	0.152
4	0.237	0.163	0.13
5	0.48	0.233	0.098
6	0.306	0.24	0.103
7	0.057	0.252	0.098
8	0.48	0.243	0.103
9	0.237	0.211	0.091
10	0.481	0.155	0.086
11	0.171	0.14	0.077
12	0.306	0.196	0.07
13	0.057	0.486	0.069
14	0.48	0.62	0.058
15	0.362	0.677	0.045
16	0.057	0.542	0.035
Precision	79%	80%	79%
Recall	60%	58%	58%
F-measure	39%	40%	31%

To evaluate the execution of the calculation, it was at first connected to a single text and performs the algorithms to access the results. After this experiment, a new set of tests were performed, TKG was compared with TFIDF approach. The benchmark was a human keyword extraction performed by three different tweets.

Table 6: Comparison between Closeness centrality and Katz centrality with Precision, Recall and F-measure of TKG and TFIDF

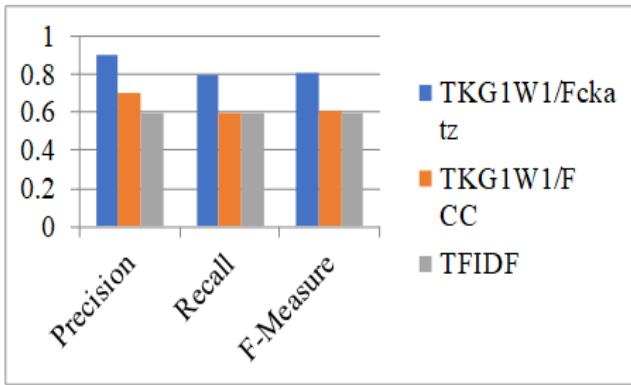| Tweets	$TKG^1W^{1/F}$ C_C	$TKG^1W^{1/F}C_{katz}$	TFIDF
bmw	0.199	0.398	0.195
volvo	0.189	0.369	0.162
car	0.189	0.325	0.152
nissan	0.188	0.321	0.155
mercedesbenz	0.175	0.240	0.130
technology	0.163	0.239	0.098
intel	0.155	0.228	0.103
mobileye	0.135	0.175	0.086
online	0.134	0.168	0.077
roadshow	0	0.058	0.091
PRECISION	0.70	0.82	0.60
RECALL	0.60	0.80	0.60
F-MEASURE	0.61	0.81	0.60

Fig. 5: Demonstrates the graph with tweets and frequency of TKGW1 contrasted with TFIDF

Table 7: Comparison between Closeness centrality and Katz centrality with Precision, Recall and F-measure of TKG and TFIDF

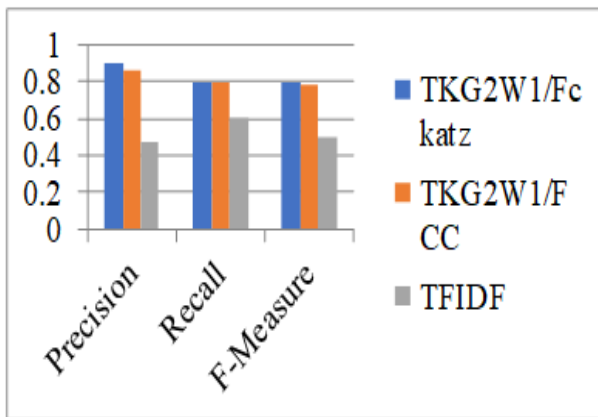| Tweets	$TKG^2W^{1/F}C_C$	$TKG^2W^{1/F}C_{katz}$	TFIDF
Car	0.599	0.519	0.152
Bmw	0.589	0.446	0.195
Volvo	0.585	0.343	0.162
Nissan	0.573	0.308	0.155
Mercedesbenz	0.552	0.291	0.130
Technology	0.463	0.255	0.098
Intel	0.458	0.157	0.103
Online	0.435	0.140	0.077
Mobileye	0.363	0.132	0.086
Roadshow	0	0.125	0.091
PRECISION	0.80	0.85	0.47
RECALL	0.69	0.71	0.60
F- MEASURE	0.79	0.80	0.50

Fig. 6: Demonstrates the graph with tweets and frequency of TKGW1 contrasted with TFIDF

VII. CONCLUSION

The experiment's done on the twitter information, which has "#volvo", "#mercedes Benz", "#bmw" hashtags. Because of the outcomes from the investigations, it has been observed that Katz's centrality is superior to the closeness centrality.

The existing approach called the Vector Space Model (VSM) is the well-known model for representing documents as vectors of identifiers, which results in a sparse matrix to be dealt with computationally. However, it has some difficulties and limitations such as scalability, sparsity and so on. It motivates the proposal of an alternative approach called TKG, relies on three phases: text pre-processing; graph building; and keyword extraction. The text pre-processing methods used are standard ones from the text-mining literature, such as tokenization, stemming, and stopword removal. Graph building takes each token as a vertex and edging can be performed in one of two ways: using a nearest-neighbor approach, or using an all neighbors approach. Finally, keywords are extracted from the graph by applying, in cascade, some graph centrality measures.

To assess the performance of both of these algorithms, it was initially applied to the three different sets of experiments having human classifications as benchmarks are performed.

The results obtained in these experiments showed that building the graph using an all neighbors edging scheme invariably provided 66% of superior performance, and assigning weights to the edges based on the weight as the inverse co-occurrence frequency was superior in two-thirds of the cases. Katz centrality was the keyword determination strategy chosen by the preliminary experiment and was utilized as a part of one of the varieties of the second arrangement of analysis. It also eliminates the nodes which have zero values. TKG showed to be faster for all its variations of 80% Precision and 62% of Recall when compared with TFIDF of 68% Precision and 58% of Recall, except for the weighting scheme based on the inverse co-occurrence frequency. Its scalability is achieved due to how the edging processes receive a vector of tokens and assign vertices and edges to build the graph, which requires local updates in the structure.

REFERENCES

1. A. Anagnostopoulos, R. Kumar, and M. Mahdian, "Influence and correlation in social networks", Proceedings of the 14th ACM SIGKDD international conference on Knowledge Discovery and Data Mining, ACM, pp. 7–15, 2008.
2. Abhishanga Upadhyay, Luis Mao, Malavika Goda Krishna, "Mining Twitter Data", Packt Publishing 246 pages May 2016.
3. Andreas M. Kaplan, Michael Haenlein, "User of the world, unite! The challenges and opportunities of Social Media", Science Direct, Business Horizons 53, pp. 59-68, ESCP Europe, 2010.
4. C. Zhang, H. Wang, Y. Liu, Y. Wu, Y. Liao, B. Wang, "Automatic keyword extraction from documents using conditional random fields", J. Comput. Inf. Syst., pp. 1169–1180, 2008.
5. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schutze, "Introduction to Information Retrieval", Cambridge University Press, Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo, pp. 100-123, 200

6. Cody Buntain, Jimmy Lin, Jennifer Golbeck, "Discovering Key Moments in Social Media Streams", 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), 2016.
7. Dr. Ovidiu Vermesan, Dr. Peter Friess, "Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems", River Publishers, Denmark, 2013.
8. Erik Westra, "Modular Programming with Python", Packt Publishing 246 pages, May 2016.
9. J.H. Kietzmann, K. Hermkens, I.P. McCarthy, B.S. Silvestre, "Social media? Get serious! Understanding the functional building blocks of social media", Bus. Horizons 54 (3), pp. 241–251, 2007.
10. Matthew Denny, "Social Network Analysis", Institute for Social Science Research, Friday 2 6th September, 2 0 1 4
11. Muller and Christian, "Information Company on its way toward Information Quality: Keynote", DGQI, 2007.
12. N.A. Christakis and J.H. Fowler, "The spread of obesity in a large social network over 32 years"., New England Journal of Medicine 357, no. 4, 370–379, 2007.
13. Reza Zafarani, Mohammad Ali Abbasi, Huan Liu, "Social Media Mining An Introduction", Cambridge University Press, pp. 74-282 Draft version: April 20, 2014.
14. Rocio Abascal-Mena, Rose Lema; Florence Sedes," From tweet to graph: Social network analysis for semantic information extraction", IEEE Eighth International Conference on Research Challenges in Information Science (RCIS), pp 1-10, Marrakech, 28-30 May 2014.
15. Shamanth Kumar, Fred Morstatter, Huan Liu, "Twitter Data Analytics", Springer, pp. 5-19, August 19, 2013.
16. T. La Fond and J. Neville, "Randomization tests for distinguishing social influence and homophily e_ects", Proceedings of the 19th international conference on the World Wide Web, ACM, pp. 601–610, 2010
17. W. Jin, R.K. Srihari, "Graph-based text representation and knowledge discovery", in Proceedings of the 2007 ACM Symposium on Applied, Computing, pp. 807–811, vol. 7, 2007.
18. Willyan D. Abilhoa, Leandro N. de Castro , "A keyword extraction method from twitter messages represented as graphs", Applied Mathematics and Computation, Volume 240, Pages 308-325, 2014.

AUTHORS PROFILE

Mrs. Gorle Dhana Lakshmi is currently as Assistant Professor in the Department of Computer Science and Engineering in Jawaharlal Nehru Technological University Kakinada since 2018. She has completed her B. Tech in Information Technology from Gayatri Vidya Parishad College of Engineering for Women. She has an M. Tech degree in Computer Science and Engineering from Gayatri Vidya Parishad College of Engineering (A). She has attended AICTE-NPTEL Faculty Development Programme on Introduction to Machine Learning. Her area of specialization including Big Data Analyzing, Machine Learning, and Deep Learning. She has a total experience of 4 years which includes teaching experience of 3 years and one year of industrial experience.

Mrs. Miriyala Kanthi Rekha is currently as Assistant Professor in the Department of Computer Science and Engineering in Jawaharlal Nehru Technological University Kakinada since 2018. She has completed her B. Tech in Computer Science and Engineering from GMRIT. She has an M. Tech degree in Computer Science and Engineering from JNTUK. She has attended AICTE-NPTEL Faculty Development Programme on Artificial Intelligence and Deep Learning. Ratified as an Assistant Professor from UCEK, JNTUK. GATE qualified in the year 2012. Her area of specialization including Data Mining, Artificial Intelligence and Deep Learning. She has a total experience of 5 years teaching experience.