SOCIO RANKS : WHICH IDENTIFIES NEWS TOPICS PREVALENT IN BOTH SOCIAL MEDIA AND THE NEWS MEDIA

[1]P.RAGHUVEER, [2]R.MEGHANA, [3]V.BHAYASRI, [4]L.S.V.P.SAI BALAKRISHNA
[1]Associate Professor, CSE Department, NRI INSTITUTE OF TECHNOLOGY, POTHAVARAPPADU
[2,3,4]UG Student, CSE department, NRI INSTITUTE OF TECHNOLOGY, POTHAVARAPPADU

Abstract: Mass media sources, specifically the news media, have traditionally informed us of daily events. In modern times, social media services such as Twitter provide an enormous amount of user-generated data, which have great potential to contain informative news-related content. For these resources to be useful, we must find a way to filter noise and only capture the content that, based on its similarity to the news media, is considered valuable. However, even after noise is removed, information overload may still exist in the remaining data—hence, it is convenient to prioritize it for consumption. To achieve prioritization, information must be ranked in order of estimated importance considering three factors. First, the temporal prevalence of a particular topic in the news media is a factor of importance, and can be considered the media focus (MF) of a topic. Second, the temporal prevalence of the topic in social media indicates its user attention (UA). Last, the interaction between the social media users who mention this topic indicates the strength of the community discussing it, and can be regarded as the user interaction (UI) toward the topic. We propose an unsupervised framework—SociRank—which identifies news topics prevalent in both social media and the news media, and then ranks them by relevance using their degrees of MF, UA, and UI. Our experiments show that SociRank improves the quality and variety

of automatically identified news topics.

Keywords: Information Filtering, Social Computing, Social Network Analysis, Topic Identification, Topic Ranking.

I. INTRODUCTION

Today, online social media such as Twitter have served as tools for organizing and tracking social events. Understanding the triggers and shifts in opinion driven mass social media data can provide useful insights for various applications in academia, industry, and however, there remains a general lack of finding of what causes the hot spots in social media. Typically, the reasons behind the rapid spread of information can be summarized in terms of two categories: exogenous and endogenous factors. Growing factors are the results of information diffusion inside the social network itself, namely, users obtain information primarily from their online social network. In contrast, exogenous factors mean that users get information from outside sources first, for example, traditional news media, and then bring it into their social network. Although previous works have explored both the social media and external news data datasets, few researchers have looked at the endogenous and exogenous factors based on semantical or topical knowledge. They have either sought to identify relevant tweets based on news articles or simply correlated the two data sources through similar patterns in the changing data volume. Still within the same data source, there could be various factors that drive the evolution of information over time. Exogenous factors across multiple datasets make analyzing the evolution and relationship among multiple data streams more difficult.

Watching social media and outside news data streams in a united frame can be a practical way of solving this problem. In this paper, we propose a novel topic model, News and Twitter Interaction Topic model (NTIT), that jointly learns social media topics and news topics and subtly capture the influences between topics. The intuition behind this approach is that before a user posts a message, he/she may be influenced either by opinions from his/her online friends or by articles from news agencies. In our new framework, a word in a tweet can be responsive to the topical influences coming either from endogenous factors (tweets) or from exogenous factors (news).

A straightforward approach for identifying topics from different social and news media sources is the application of topic modeling. Many methods have been proposed in this area, such as latent Dirichlet allocation (LDA) and probabilistic latent semantic analysis (PLSA). Topic modeling is, in essence, the discovery of ―topics‖ in text corpora by clustering together frequently co-occurring words. This approach, however, misses out in the temporal component of prevalent topic detection, that is, it does not take into account how topics change with time.

Furthermore, topic modeling and other topic detection techniques do not rank topics according to their popularity by taking into account their prevalence in both news media and social media. We introduce an unsupervised system—SociRank—which effectively identifies news topics that are prevalent in both social media and the news media, and then ranks them by relevance using their degrees of MF, UA, and UI. Even though this paper focuses on news topics, it can be easily adapted to a wide variety of fields, from science and technology to culture and sports. To the best of our knowledge, no other work attempts to employ the use of either the social media interests of users or their social relationships to aid in the ranking of topics. Moreover, SociRank undergoes an empirical framework, comprising and integrating several techniques, such as keyword extraction, measures of similarity, graph clustering, and social network analysis. The effectiveness of our system is validated by extensive controlled and uncontrolled experiments.

II. EMERGENCE OF TWITTER AS A NEWS MEDIA

Computer science research community has analyzed relevance of on line social media, in particular Twitter, as news disseminating agent. Kwak et al. showed the prominence of Twitter as a news media, they showed that 85% topics discussed on Twitter are related to news. Their work highlighted the relationship between user specific parameters v/s the tweeting activity patterns, like analysis of the number of followers and followees v/s the tweeting (re-tweeting) numbers. Zhao et al. in their work, used unsupervised topic modeling to compare the news topic from Twitter versus New York Times (a traditional news dissemination medium). They showed that Twitter users are relatively less interested in world news, still they are active in spreading news of important world events. Lu et al. showed how tweets related to news event on Twitter can be mapped using energy function. The methods proposed act like novel event detection techniques. The study analyzed 900 news events through 2010-2011. Castillo et al. performed qualitative and quantitative analysis on online social media activity about news articles. They concluded that news articles describing breaking news events have more repetitive social media reactions, than in-depth articles.

III. ANALYZING TWITTER DATA DURING REAL-WORLD EVENTS

The posts and activity on Twitter, impacts and plays a vital role in various real world events. Role of Twitter has been analyzed by computer scientists, psychologists and sociologists for impact in the real-world. Twitter has progressed from being merely a medium to share users' opinions; to an information sharing and dissemination agent; to propagation and coordination of relief and response efforts. Some of the popular case studies analyzed

by computer scientists have been, Twitter activities during elections, natural disasters (like hurricanes, wildfires, floods, etc.), political and social uprisings (like Libya and Egypt crisis) and terrorist attacks (like Mumbai triple bomb blasts). Content and user activity patterns of Twitter during events have been analyzed for both positive and negative aspects. Some of the problems studied that result in bad quality of data, presence of spam and phishing posts, content spreading rumors / fake news, privacy breach of users via the content shared by them and use of Twitter for propagation and instigation of hate among people. Researchers have used machine learning, information retrieval, social network analysis and image and video analysis for the purpose of analyzing and characterizing Twitter usage during real-world events. We introduce some of the research work done in applying user modeling techniques to analyze behavior of users on social networks. Yin et al. modeled user behavior using two factors: the topics related to users' intrinsic interests and the topics related to temporal context.

They created alatent class statistical mixture model, called Dynamic Temporal Context-Aware Mixture model (DTCAM). They evaluated their system on four large-scale social media datasets. The authors demonstrated how user modeling techniques can be effectively used to improve the performance of recommender systems for social networks. Xu et al. introduced a mixed latent topic model to combine various factors to model users' posting behavior on Twitter. The authors assumed that a user's behavior is influenced by three factors: breaking news, posts from social friends and user's interest. They developed and showed that their model outperforms other user models in handling the perplexity of held-out content and the quality of generated latent topics. Abel et al. developed a user modeling framework for news recommendations on Twitter using more than 2 million tweets. The authors proposed different strategies for creating hash tag-based, entity based or topic-based user profiles using semantic enrichment and temporal factors. Their results showed that consideration of temporal profile patterns can improve recommendation quality.

IV. LITERATURE SURVEY

A. Analysis Of Key-Exchange Protocols And Their Use For Building Secure Channels

AUTHORS: R. Canetti and H. Krawczyk

We present a formalism for the analysis of key-exchange protocols that combines previous definitional approaches and results in a definition of security that enjoys some important analytical benefits: (i) any key-exchange protocol that satisfies the security definition can be composed with symmetric encryption and authentication functions to provide provably secure communication channels (as defined here); and (ii) the definition allows for simple modular proofs of security: one can design and prove security of key-exchange protocols in an idealized model where the communication links are perfectly authenticated, and then translate them using general tools to obtain security in the realistic setting of adversary-controlled links. We exemplify the usability of our results by applying them to obtain the proof of two classes of key- exchange protocols, Diffie-Hellman and key-transport, authenticated via symmetric or asymmetric techniques.

B. Map Reduce: Simplified Data Processing On Large Clusters

Map Reduce is a programming model and an associated implementation for processing and generating large datasets that is amenable to a broad variety of real-world tasks. Users specify the computation in terms of a map and a reduce function, and the underlying runtime system automatically parallelizes the computation across large-scale clusters of machines, handles machine failures, and schedules inter-machine communication to make efficient use of the network and disks. Programmers find the system easy to use: more than ten thousand distinct Map Reduce programs have been implemented internally at Google over the past four years, and an average of one hundred thousand Map Reduce jobs are executed on Google's clusters every day, processing a total of more than twenty petabytes of data per day.

C. Scalable Security For Petascale Parallel File Systems

AUTHORS: A.W. Leung, E.L. Miller, and S. Jones

Petascale, high-performance file systems often hold sensitive data and thus require security, but authentication and authorization can dramatically reduce performance. Existing security solutions perform poorly in these environments because they cannot scale with the number of nodes, highly distributed data, and demanding workloads. To address these issues, we developed Maat, a security protocol designed to provide strong, scalable security to these systems. Maat introduces three new techniques. Extended capabilities limit the number of capabilities needed by allowing a capability to authorize I/O for any number of client-file pairs. Automatic Revocation uses short capability lifetimes to allow capability expiration to act as global revocation, while supporting non-revoked capability renewal. Secure Delegation allows clients to securely act on behalf of a group to open files and distribute access, facilitating secure joint computations. Experiments on the Maat prototype in the Ceph petascale file system show an overhead as little as 6--7%.

D. Scalable Performance Of The Panasas Parallel File System

AUTHORS: B. Welch, M. Unangst, Z. Abbasi, G.A. Gibson, B. Mueller, J. Small, J. Zelenka, and B. Zhou

The Panasas file system uses parallel and redundant access to object storage devices (OSDs), per-file RAID, distributed metadata management, consistent client caching, file locking services, and internal cluster management to provide a scalable, fault tolerant, high performance distributed file system. The clustered design of the storage system and the use of client-driven RAID provide scalable performance to many concurrent file system clients through parallel access to file data that is striped across OSD storage nodes. RAID recovery is performed in parallel by the cluster of metadata managers, and declustered data placement yields scalable RAID rebuild rates as the storage system grows larger. This paper presents performance measures of I/O, metadata, and recovery operations for storage clusters that range in size from 10 to 120 storage nodes, 1 to 12 metadata nodes, and with file system client counts ranging from 1 to 100 compute nodes. Production installations are as large as 500 storage nodes, 50 metadata managers, and 5000 clients.

E. Scale And Performance In A Distributed File System

AUTHORS: J.H. Howard, M.L. Kazar, S.G. Menees, D.A. Nichols, M. Satyanarayanan, R.N. Sidebotham, and M.J. West

The Andrew File System is a location-transparent distributed tile system that will eventually span more than 5000 workstations at Carnegie Mellon University. Large scale affects performance and complicates system operation. In this paper we present observations of a prototype implementation, motivate changes in the areas of cache validation, server process structure, name translation, and low-level storage representation, and quantitatively demonstrate Andrews ability to scale gracefully. We establish the importance of whole-file transfer and caching in Andrew by comparing its performance with that of Sun Microsystems NFS tile system. We also show how the aggregation of files into volumes improves the operability of the system.

V. SOCIRANK FRAMEWORK

The goal of our method—SociRank—is to identify, consolidate and rank the most prevalent topics discussed in both news media and social media during a specific period of time. The system framework can be visualized in Fig. 1. To achieve its goal, the system must undergo four main stages.

- **Preprocessing:** Key terms are extracted and filtered from news and social data corresponding to a particular period of time.
- **Key Term Graph Construction:** A graph is constructed from the previously extracted key term set, whose vertices represent the key terms and edges represent the co-occurrence similarity between them. The graph, after processing and pruning, contains slightly joint clusters of topics popular in both news media and social media.
- **Graph Clustering:** The graph is clustered in order to obtain well-defined and disjoint TCs.
- **Content Selection and Ranking:** The TCs from the graph are selected and ranked using the three relevance factors (MF, UA, and UI).

Initially, news and tweets data are crawled from the Internet and stored in a database. News articles are obtained from specific news websites via their RSS feeds and tweets are crawled from the Twitter public timeline [41]. A user then requests an output of the top k ranked news topics for a specified period of time between date d1 (start) and date d2 (end).

VI. SYSTEM ANALYSIS

A. Existing System

- Many existing NLP procedures vigorously depend on phonetic highlights, for example, POS labels of the encompassing words, word upper casing, trigger words (e.g.,Mr.,Dr.),and gazetteers. These phonetic highlights, together with compelling managed learning calculations (e.g., concealed markov demonstrate (HMM) and contingent arbitrary field (CRF)), accomplish great execution on formal content corpus. In any case, these strategies encounter extreme execution decay on tweets as a result of the loud and short nature of the last mentioned.
- In Existing System, to enhance POS labeling on tweets, Ritter et al. prepare a POS tagger by utilizing CRF demonstrate with ordinary and tweet-particular highlights. Dark colored grouping is connected in their work to manage the badly shaped words.

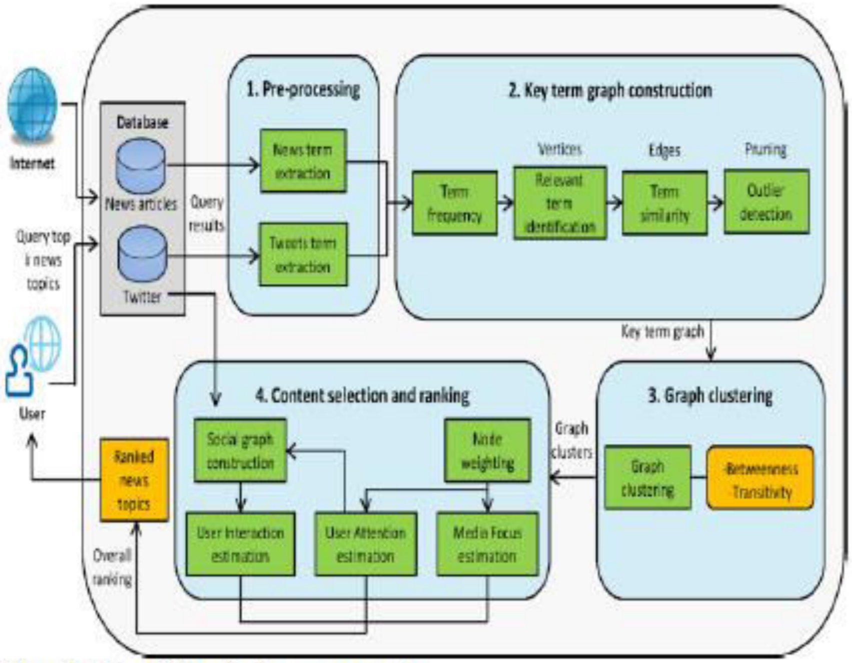


Fig. 1. SociRank framework.

Drawbacks of Existing System:

 Given the constrained length of a tweet (i.e., 140 characters) and no limitations on its composition styles, tweets frequently contain syntactic blunders, incorrect spellings, and casual shortened forms.

 The mistake inclined and short nature of tweets frequently makes the word- level dialect models for tweets less dependable.

 The Existing framework doesn't bolster the point score esteem and rating according to current framework date and time this is the significant disadvantage in the past framework.

 There is no Monitoring idea in the Entire application from administrator level this one affected on the clients.

B. Proposed System In this paper, we centre around the undertaking of tweet division. The objective of this errand is to part a tweet into a grouping of sequential n-grams, every one of which is known as a portion. A section can be a named substance (e.g., a film title "discovering nemo"), a semantically significant data unit (e.g., "formally discharged"), or some other kinds of expressions which seem "more than by shot" as shown in Fig.2.

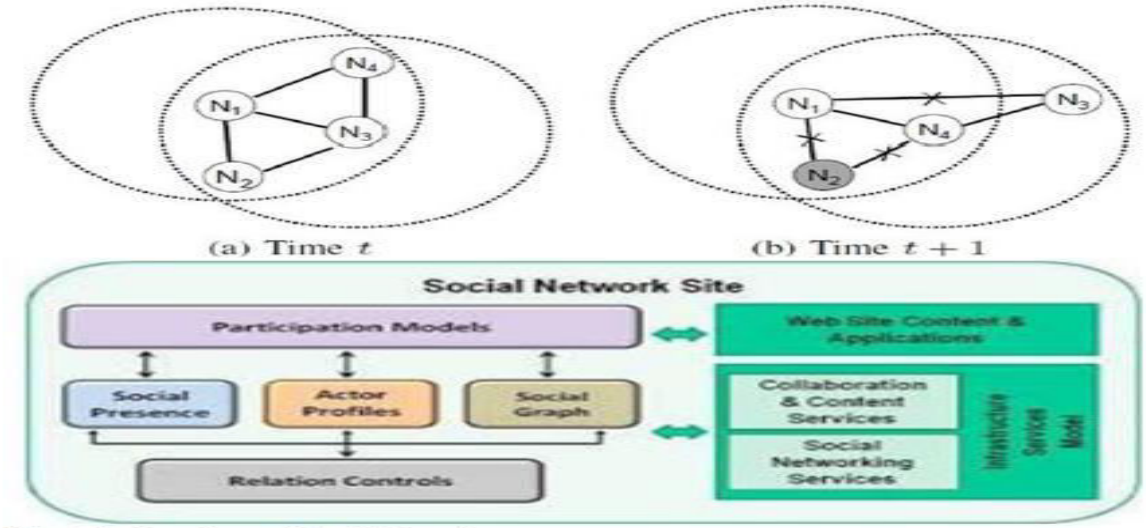
 To accomplish top notch tweet division, we propose a nonexclusive tweet division system, named HybridSeg. HybridSeg gains from both worldwide and nearby settings, and has the capacity of gaining from pseudo input.

 Global setting. Tweets are posted for data sharing and correspondence. The named elements and semantic expressions are very much protected in tweets.

- **Local setting.** Tweets are profoundly time-delicate with the goal that numerous developing expressions like "She Dancin" can't be found in outer information bases. In any case, thinking about an extensive number of tweets distributed inside a brief timeframe period (e.g., multi day) containing the expression, it isn't hard to remember "She Dancin" as a substantial and significant portion. We in this manner examine two neighborhood settings, specifically nearby etymological highlights and nearby collocation.

Focal points of Proposed System:

- Our work is likewise identified with substance connecting (EL). EL is to recognize the specify of a named substance and connection it to a section in an information base like Wikipedia.
- Through our system, we show that nearby semantic highlights are more solid than term-reliance in managing the division procedure. This discovering opens open doors for devices created for formal content to be connected to tweets which are accepted to be considerably more loud than formal content.
- Helps in safeguarding Semantic importance of tweets.
- It screens the whole framework from administrator side and it bolsters graphical portrayal.
- The Entire framework is spoken to utilizing two variables one is time and another is subject.


N4
N1
N3
N2
N1
N3
N4
N2
Social Network Site
Participation Models
Web Site Content & Applications
Social Presence
Actor Profiles
Social Graph
Collaboration & Content Services
Social Networking Services
Infrastructure Services Model
Relation Controls

(a) Time t

(b) Time $t + 1$

Fig.2. System Architecture.

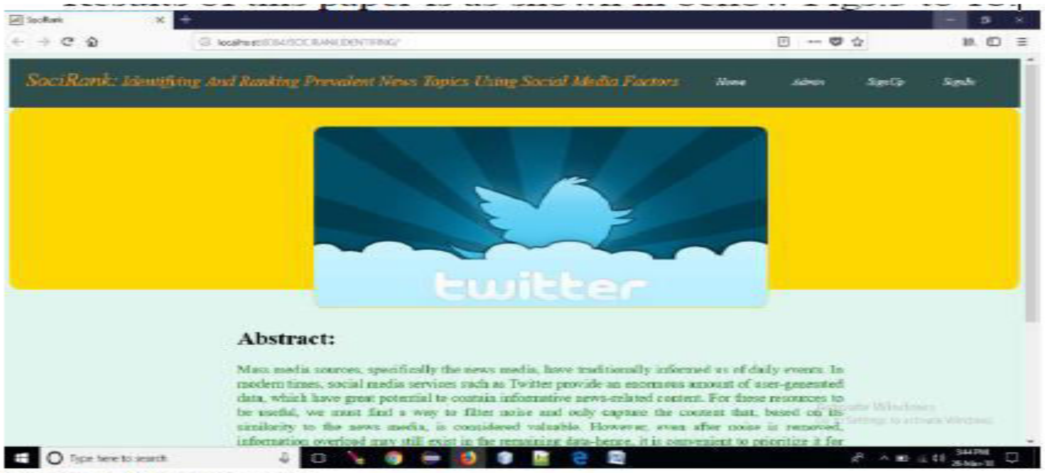
VII. RELATED WORK

The primary research regions connected in this paper include: subject identification, theme positioning social, organize investigation, catchphrase extraction, co-event comparability measures, and chart clustering. Broad work has been led in the greater part of these territories. All the more as of late, inquire about has been led in recognizing points and occasions from online networking information, considering fleeting data. Cataldi et al. [7] proposed a subject detection system that recovers constant developing themes from Twitter. Their technique utilizes the arrangement of terms from tweets and model their life cycle as indicated by a novel maturing hypothesis. Moreover, they consider social connections—all the more specifically, the specialist of the clients in the system—to deflect mine the significance of the subjects. Zhao et al. [8] did comparable work by building up a Twitter-LDA display intended to recognize subjects in tweets. Their work, in any case, just thinks about the individual interests of clients, and not pervasive points at a worldwide scale. Another significant idea that is fused into this paper is theme positioning. There are a few means by which this errand can be refined, generally being finished by evaluating how oftentimes and as of late a theme has been accounted for by broad communications.

The primary motivation behind chart bunching in this paper is to recognize and isolate TCs, as done in Warden and Brussels work [4]. Wanaka and Tanaka -Ishii [37] additionally proposed a technique that bunches a co-event diagram in view of a chart measure known as transitivity. The essential thought of transitivity is that in a connection between three components, if the relationship holds between the first and second components and between the second and third components, it likewise holds between the first and third components. They recommended that each out-put group is relied upon to have no equivocalness, and this is just accomplished when the edges of a diagram (speaking to co-event relations) are transitive.

VIII. RESULTS

Results of this paper is as shown in bellow Figs.3 to 18.



SociRank: Identifying And Ranking Prevalent News Topics Using Social Media Factors

Abstract:

Mass media sources, specifically the news media, have traditionally informed us of daily events. In modern times, social media services such as Twitter provide an enormous amount of user-generated data, which have great potential to contain informative news-related content. For these resources to be useful, we must find a way to filter noise and only capture the content that, based on its similarity to the news media, is considered valuable. However, even after noise is removed, information overload may still exist in the remaining data-hence, it is convenient to prioritize it for

Fig.3. Homepage.

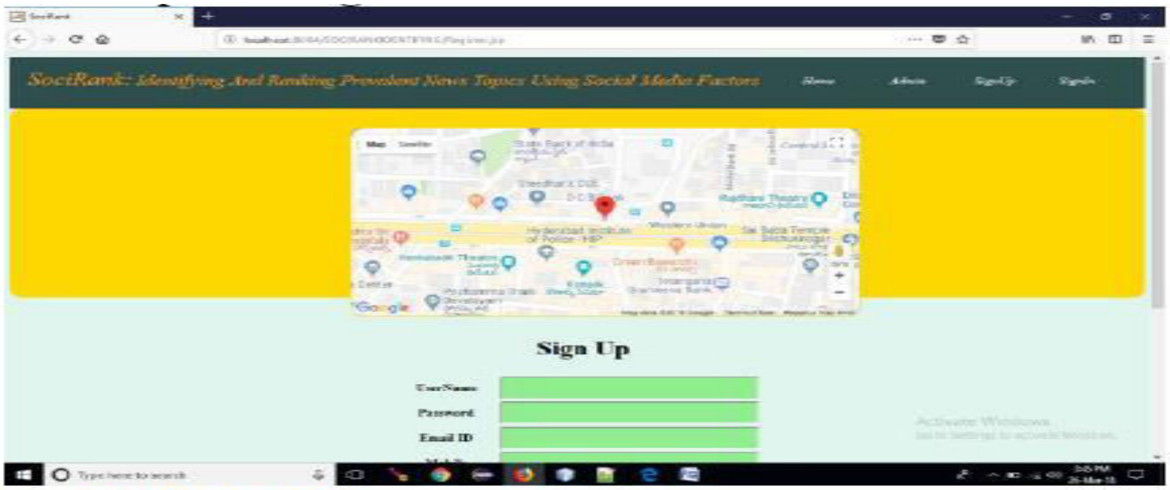

SociRank: Identifying And Ranking Prevalent News Topics Using Social Media Factors
Home
Admin
SignUp
SignIn
Google
Sign Up
UserName
Password
Email ID

Fig.4. Registration.

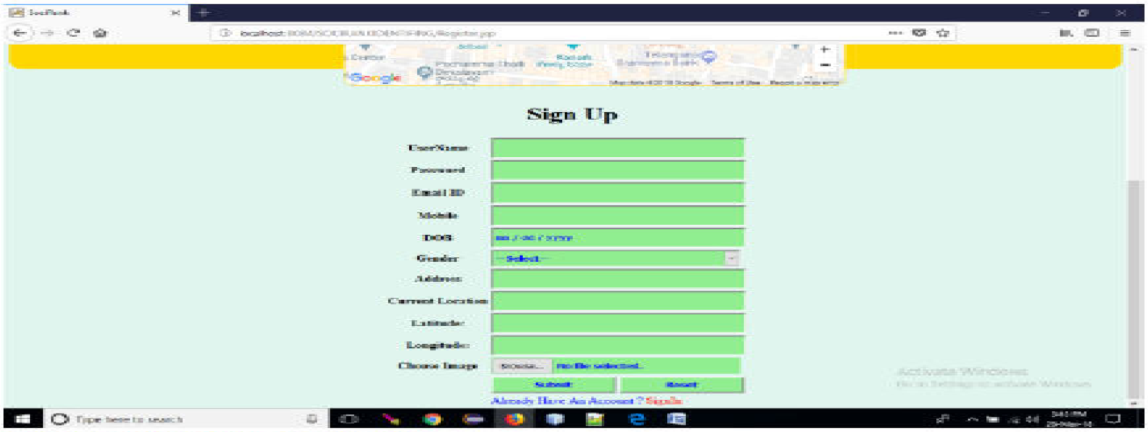
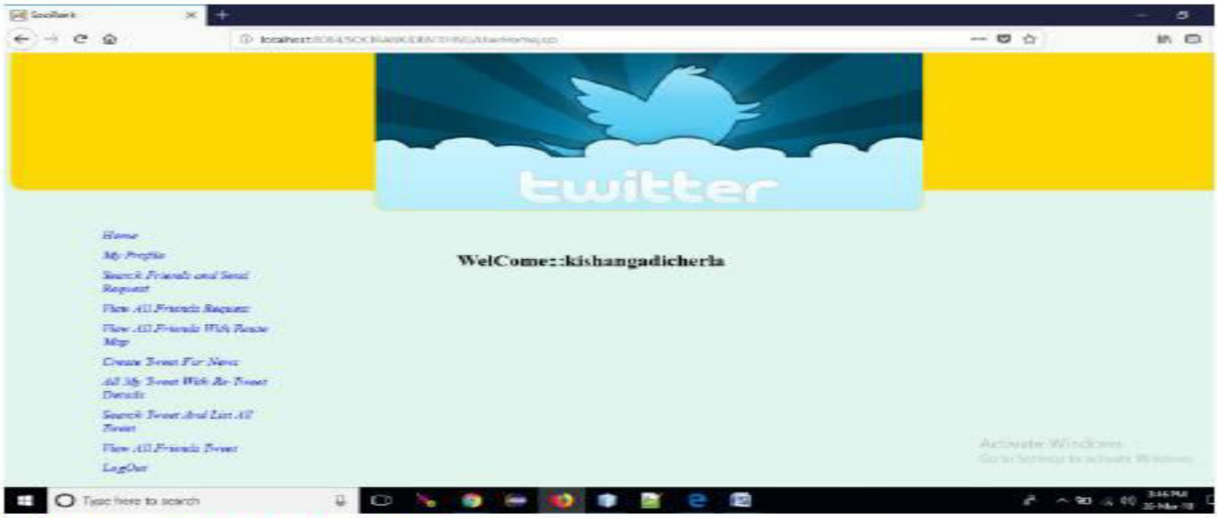
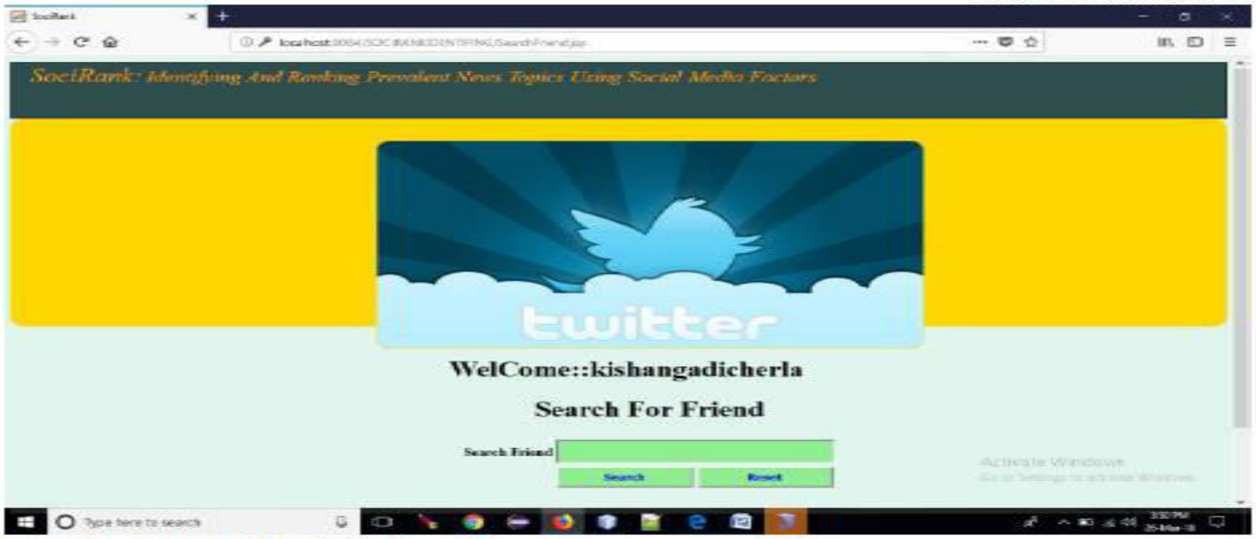


Fig.5. Userhome page.


twitter
Home
My Profile
Search Friends and Send Request
View All Friends Request
View All Friends With Route Map
Create Tweet For News
All My Tweet With Re-Tweet Details
Search Tweet And List All Tweet
View All Friends Tweet
LogOut

WelCome::kishangadicherla

Fig.6. Search friend.



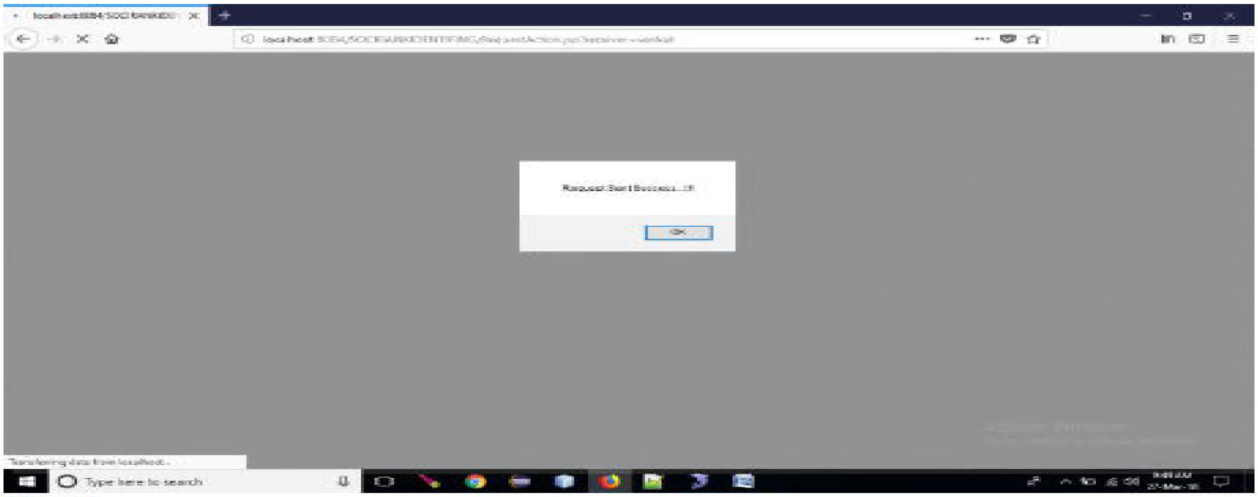
SociRank: Identifying And Ranking Prevalent News Topics Using Social Media Factors

WelCome::kishangadicherla

Search For Friend

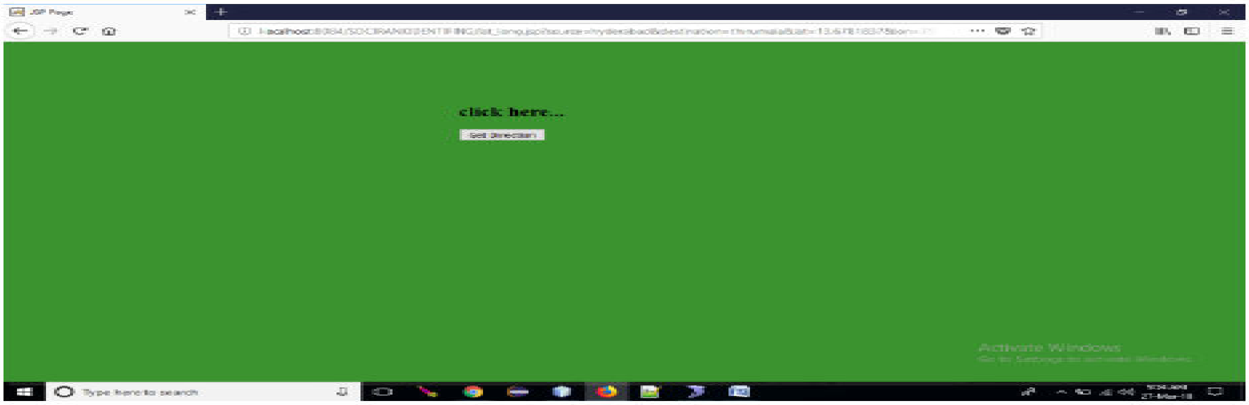
Search Friend

Fig.7. Searched friend.



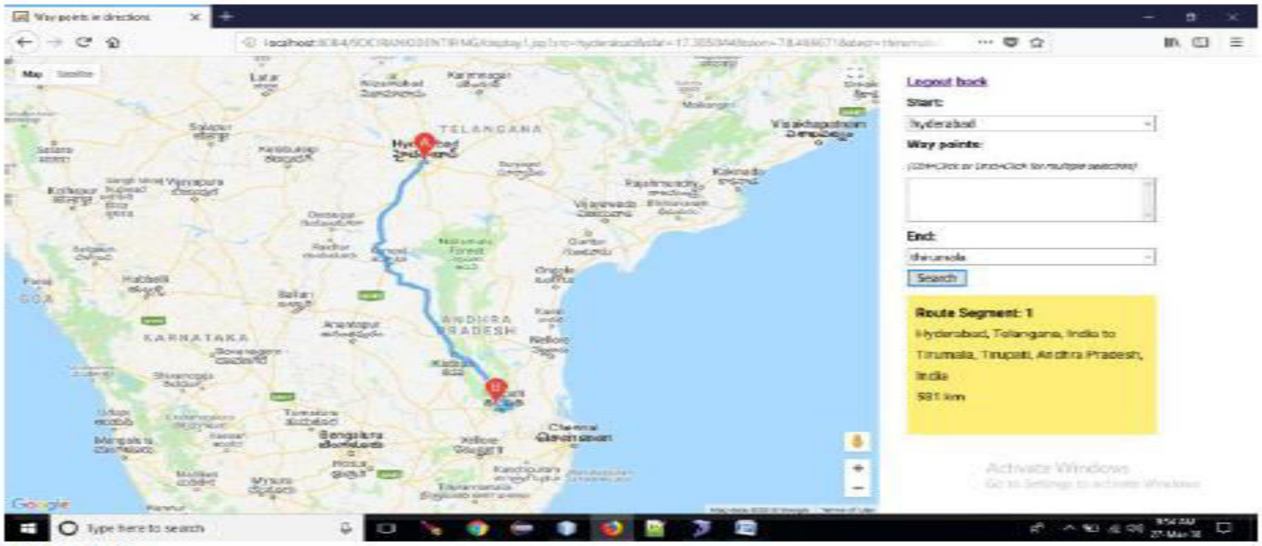
Request Sent Success...!!

Fig.8.



click here...

Fig.9.



Logout back

Start:

hyderabad

Way points:

(Ctrl+Click or Cmd+Click for multiple selection)

End:

thirumala

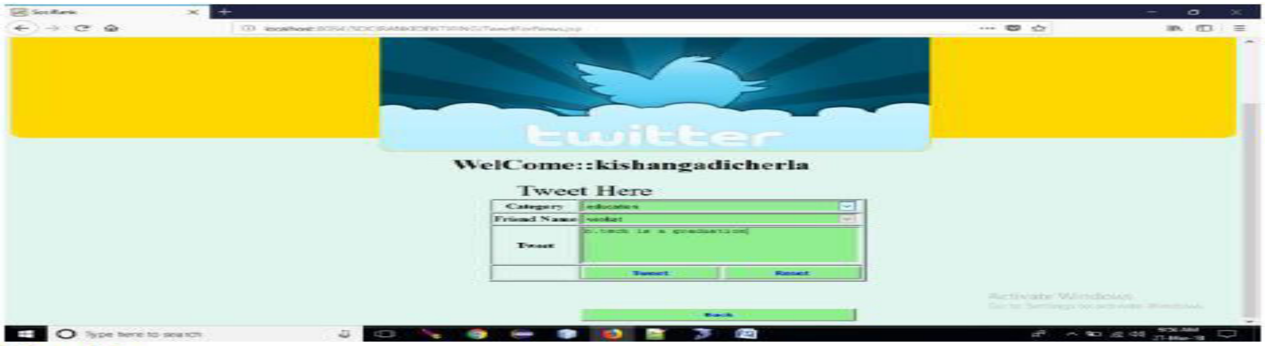
Search

Route Segment: 1

Hyderabad, Telangana, India to Tirumala, Tirupati, Andhra Pradesh, India

581 km

Fig.10.



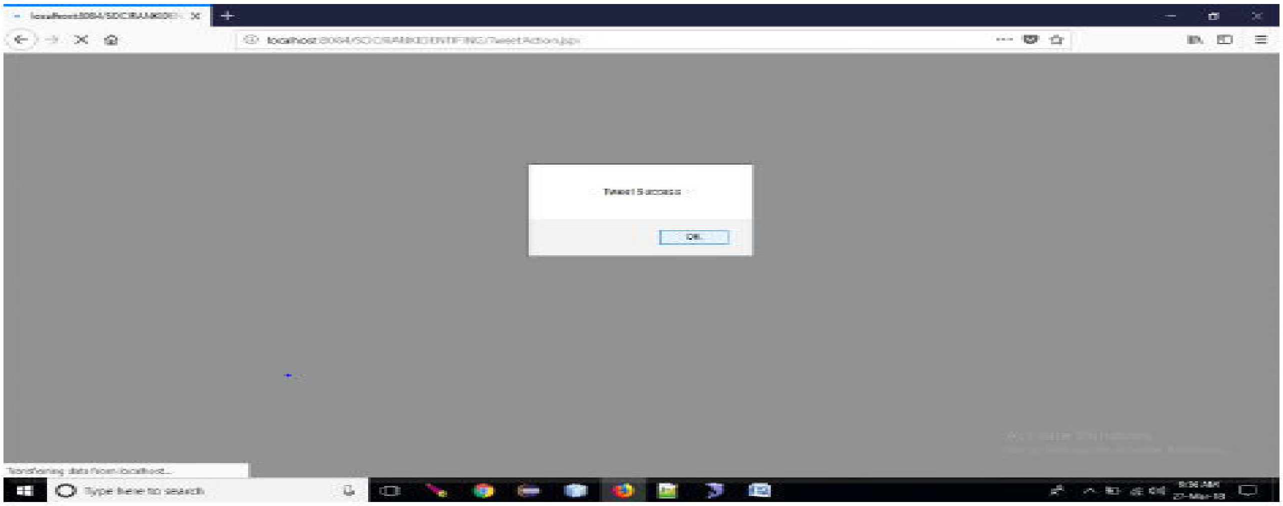
WelCome::kishangadicherla

Tweet Here

Category	education
Friend Name	venkat
Tweet	b.tech is a graduation
	Tweet Reset

Back

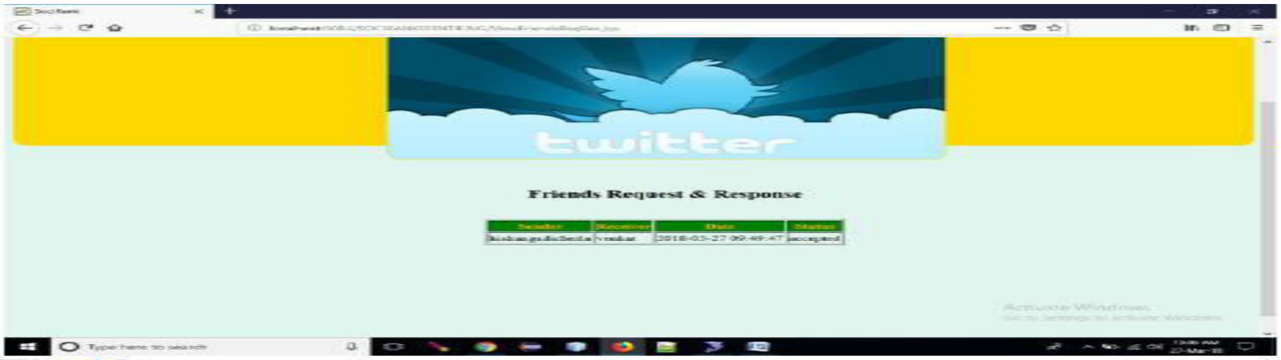
Fig.11.



Tweet Success

OK

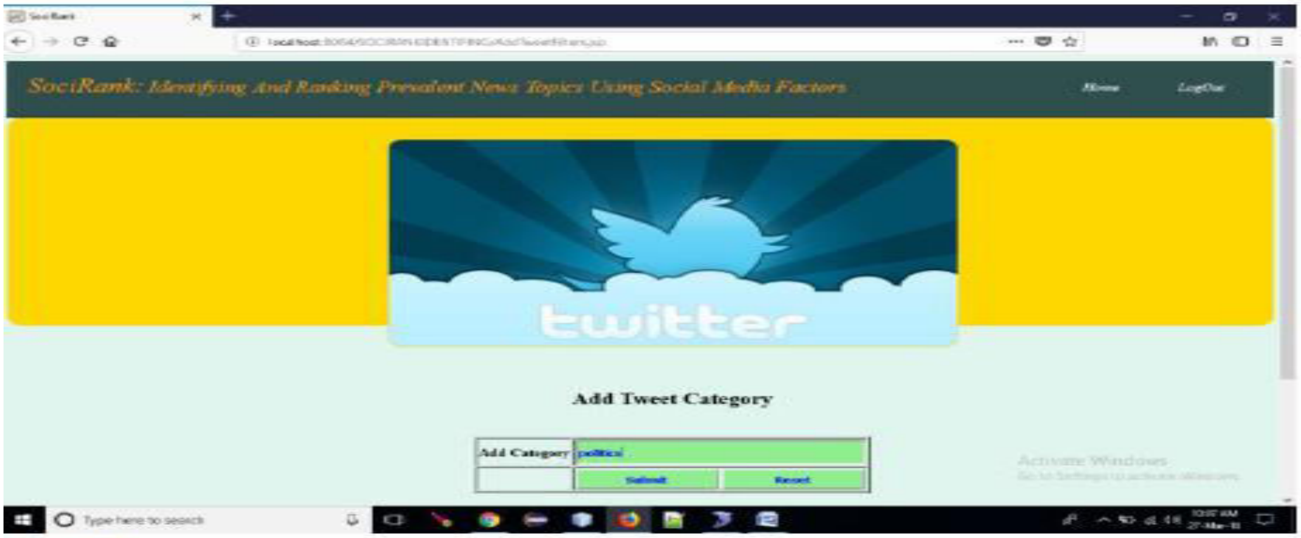
Fig.12.



Friends Request & Response

Sender	Receiver	Date	Status
kishangadicherla	venkat	2018-03-27 09:49:47	accepted

Fig.13.

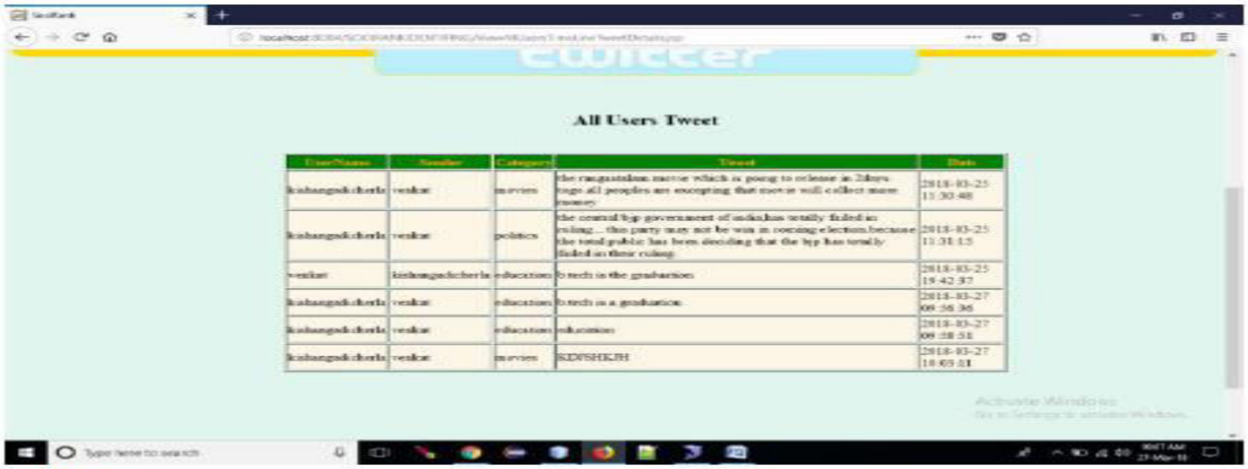


SociRank: Identifying And Ranking Prevalent News Topics Using Social Media Factors

Add Tweet Category

Add Category	politics
	Submit Reset

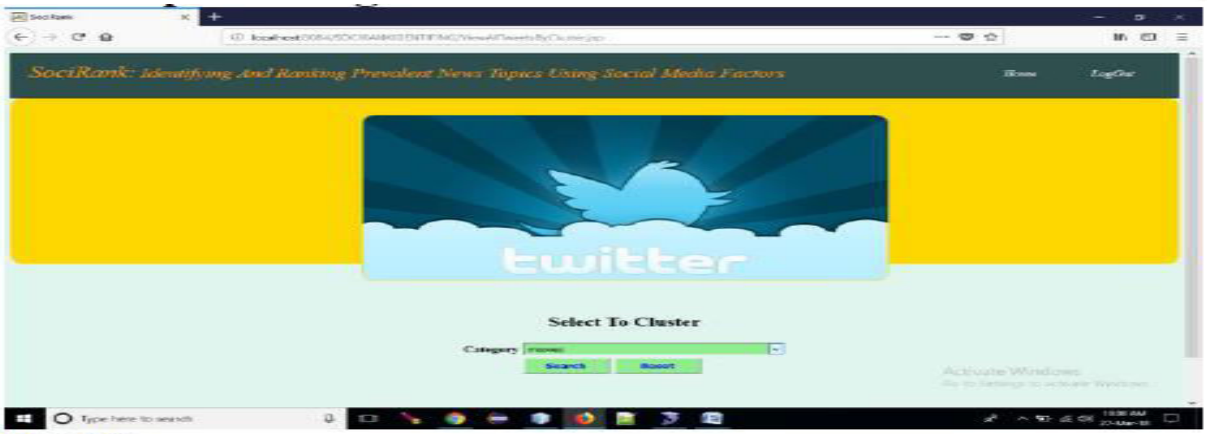
Fig.14.



All Users Tweet

UserName	Sender	Category	Tweet	Date
kishangadicherla	venkat	movies	the rangasthalam movie which is going to release in 2days togo all peoples are excepting that movie will collect more money	2018-03-25 11:30:48
kishangadicherla	venkat	politics	the central bjp government of india,has totally failed in ruling... this party may not be win in coming election,because the total public has been deciding that the bjp has totally failed in their ruling	2018-03-25 11:31:15
venkat	kishangadicherla	education	b tech is the graduation	2018-03-25 19:42:37
kishangadicherla	venkat	education	b.tech is a graduation	2018-03-27 09:38:36
kishangadicherla	venkat	education	education	2018-03-27 09:38:51
kishangadicherla	venkat	movies	KDFSHKJH	2018-03-27 10:05:11

Fig.15.

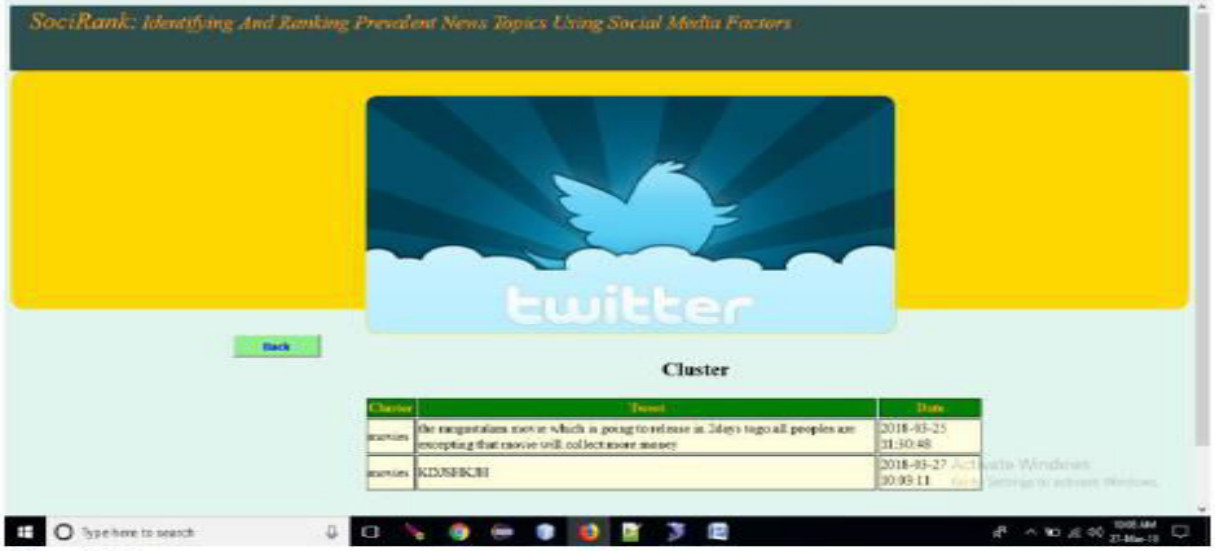


SociRank: Identifying And Ranking Prevalent News Topics Using Social Media Factors
Home
LogOut
twitter

Select To Cluster

Category: movies

Fig.16.

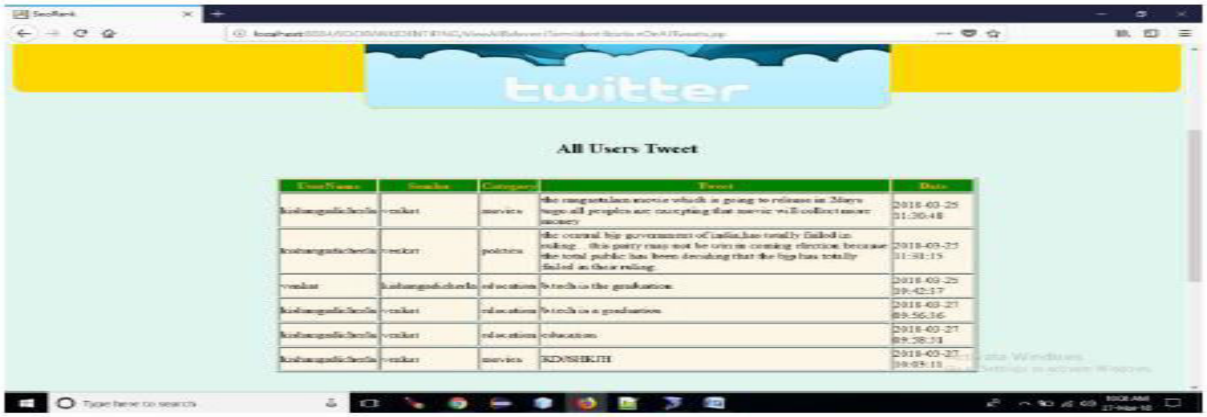


SociRank: Identifying And Ranking Prevalent News Topics Using Social Media Factors
twitter
Back

Cluster

Cluster	Tweet	Date
movies	the rangastalam movie which is going to release in 2days togo all peoples are excepting that movie will collect more money	2018-03-25 11:30:48
movies	KDJSHKJH	2018-03-27 10:09:11

Fig.17.



All Users Tweet

UserName	Sender	Category	Tweet	Date
kishangadicherla	venkat	movies	the rangastalam movie which is going to release in 2days togo all peoples are excepting that movie will collect more money	2018-03-25 11:30:48
kishangadicherla	venkat	politics	the central bjp government of india,has totally failled in ruling .. this party may not be win in coming election because the total public has been deciding that the bjp has totally failled in their ruling	2018-03-25 11:31:15
venkat	kishangadicherla	education	b.tech is the graduation	2018-03-25 11:42:17
kishangadicherla	venkat	education	b.tech is a graduation	2018-03-27 09:56:16
kishangadicherla	venkat	education	education	2018-03-27 09:58:31
kishangadicherla	venkat	movies	KDJSHKJH	2018-03-27 10:09:11

Fig.18.

IX. CONCLUSION

In this paper, we proposed an unsupervised technique SociRank which distinguishes news points pervasive in both web based life and the news media, and after that positions them by considering their media centre, client consideration and client communication as pertinence factors. The transient pervasiveness of a specific theme in the news media is viewed as the media centre of a theme, which gives us knowledge into its broad communications fame. The transient commonness of the subject in web based life, particularly Twitter, demonstrates client intrigue, also, is viewed as its client consideration. At long last, the collaboration between the online networking clients who say the point demonstrates the quality of the network talking about it, and is viewed as the client communication. To the best of our insight, no other work has endeavoured to utilize the utilization of either the interests of online networking clients or then again their social connections to help in the positioning of subjects.

X. REFERENCESS

[1]D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet allocation,” J. Mach. Learn. Res., vol. 3, pp. 993–1022, Jan. 2003.

[2]T. Hofmann, “Probabilistic latent semantic analysis,” in Proc. 15th Conf. Uncertainty Artif. Intell., 1999, pp. 289–296.

[3]T. Hofmann, “Probabilistic latent semantic indexing,” in Proc. 22nd Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, Berkeley, CA, USA, 1999, pp. 50–57.

[4]C. Wartena and R. Brussee, “Topic detection by clustering keywords,” in Proc. 19th Int. Workshop Database Expert Syst. Appl. (DEXA), Turin, Italy, 2008, pp. 54–58.

[5]F. Archetti, P. Campanelli, E. Fersini, and E. Messina, “A hierarchicaldocument clustering environment based on the induced bisecting k-means,” in Proc. 7th Int. Conf. Flexible Query Answering Syst., Milan, Italy, 2006pp.257–269. [Online].Available: http://dx.doi.org/ 10.1007/11766254_22.

[6]C. D. Manning and H. Schütze, Foundations of Statistical Natural Language Processing. Cambridge, MA, USA: MIT Press, 1999.

[7]M. Cataldi, L. Di Caro, and C. Schifanella, “Emerging topic detection on Twitter based on temporal and social terms evaluation,” in Proc. 10th Int. Workshop Multimedia Data Min. (MDMKDD), Washington, DC, USA, 2010, Art. no. 4. [Online]. Available: http://doi.acm.org/ 10.1145/ 1814 245.1814249.

[8]W. X. Zhao et al., “Comparing Twitter and traditional media using topic models,” in Advances in Information Retrieval. Heidelberg, Germany: Springer Berlin Heidelberg, 2011, pp. 338–349.

[9]Q. Diao, J. Jiang, F. Zhu, and E.-P. Lim, “Finding bursty topics from microblogs,” in Proc. 50th Annu. Meeting Assoc. Comput. Linguist. Long Papers, vol. 1. 2012, pp. 536–544.

[10]H. Yin, B. Cui, H. Lu, Y. Huang, and J. Yao, “A unified model for stable and temporal topic detection from social media data,” in Proc. IEEE 29th Int. Conf. Data

Eng. (ICDE), Brisbane, QLD, Australia, 2013, pp. 661–672.

[11]C. Wang, M. Zhang, L. Ru, and S. Ma, “Automatic online news topic ranking using media focus and user attention based on aging theory,” in Proc. 17th Conf. Inf. Knowl. Manag., Napa County, CA, USA, 2008, pp. 1033–1042.

12] C. C. Chen, Y.-T. Chen, Y. Sun, and M. C. Chen, “Life cycle modeling of news events using aging theory,” in Machine Learning: ECML 2003. Heidelberg, Germany: Springer Berlin Heidelberg, 2003, pp. 47–59.

[13]J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling, “TwitterStand: News in tweets,” in Proc. 17th ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst., Seattle, WA, USA, 2009, pp. 42–51.

[14]O. Phelan, K. McCarthy, and B. Smyth, “Using Twitter to recommend real-time topical news,” in Proc. 3rd Conf. Recommender Syst., New York, NY, USA, 2009, pp. 385–388.

[15]K. Shubhankar, A. P. Singh, and V. Pudi, “An efficient algorithm for topic ranking and modeling topic evolution,” in Database Expert Syst. Appl., Toulouse, France, 2011, pp. 320– 330.

[16]S. Brin and L. Page, “Reprint of: The anatomy of a large-scale hypertextual web search engine,” Comput. Netw., vol. 56, no. 18, pp. 3825–3833, 2012.

[17]E. Kwan, P.-L. Hsu, J.-H. Liang, and Y.-S. Chen, “Event identification for social streams using keyword-based evolving graph sequences,” in Proc. IEEE/ACM Int. Conf. Adv. Soc. Netw. Anal. Min., Niagara Falls, ON, Canada, 2013, pp. 450–457.

[18]K. Kireyev, “Semantic-based estimation of term informativeness,” in Proc. Human Language Technol. Annu. Conf. North Amer. Chapter Assoc. Comput. Linguist., 2009, pp. 530– 538.

[19]G. Salton, C.-S. Yang, and C. T. Yu, “A theory of term importance in automatic text analysis,” J. Amer. Soc. Inf. Sci., vol. 26, no. 1, pp. 33–44, 1975.

[20]H. P. Luhn, “A statistical approach to mechanized encoding and searching of literary information,” IBM J. Res. Develop., vol. 1, no. 4, pp. 309–317, 1957.

[21]J. D. Cohen, “Highlights: Language- and domain-independent automatic indexing terms for abstracting,” J. Amer. Soc. Inf. Sci., vol. 46, no. 3, pp. 162–174, 1995.

[22]Y. Matsuo and M. Ishizuka, “Keyword extraction from a single document using word co- occurrence statistical information,” Int. J. Artif. Intell. Tools, vol. 13, no. 1, pp. 157–169, 2004.

[23]R. Mihalcea and P. Tarau, “TextRank: Bringing order into texts,” in Proc. EMNLP, vol. 4. Barcelona, Spain, 2004.

[24]I. H. Witten, G. W. Paynter, E. Frank, C. Gutwin, and C. G. Nevill-Manning, “KEA: Practical automatic keyphrase extraction,” in Proc. 4th ACM Conf. Digit. Libr., Berkeley, CA, USA, 1999, pp. 254–255.