IDENTIFYING PREVALENT NEWS TOPICS AND RANK THOSE USING SOCIAL MEDIA FACTORS

VUYYURI DURGA SOWJANYA [#1], D.KANAKA DURGA[#2], L.SOWJANYA[#3]

[#1] MCA Student, Master of Computer Applications,
D.N.R. College, P.G.Courses & Research Center, Bhimavaram, AP, India.

[#2] Assistant Professor, Master of Computer Applications,
D.N.R. College, P.G.Courses & Research Center, Bhimavaram, AP, India.

[#3] Assistant Professor, Master of Computer Applications,
D.N.R. College, P.G.Courses & Research Center, Bhimavaram, AP, India

ABSTRACT

Mass media sources, specifically the news media, have traditionally informed us of daily events. In modern times, social media services 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.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. Our experiments show that SociRank improves the quality and variety of automatically identified news topics.

Key Words:

Services, Socirank, Unsupervised Framework.

I. INTRODUCTION

The mining of valuable information from online sources has become a prominent research area in information technology in recent years. Historically, knowledge that apprises the general public of daily events has been provided by mass media sources, specifically the news media. Many of these news media sources have either abandoned their hardcopy publications and moved to the World Wide Web, or now produce both hard-copy and Internet versions simultaneously.

These news media sources are considered reliable because they are published by professional journalists, who are held accountable for their content. On the other hand, the Internet, being a free and open forum for information exchange, has recently seen a fascinating phenomenon known as social media. In social media, regular, nonjournalist users are able to publish unverified content and express their interest in certain events.

Microblogs have become one of the most popular social media outlets. One microblogging service in particular, Twitter, is used by millions of people around the world, providing enormous amounts of user-generated data. One may assume that this source potentially contains information with equal or greater value than the news media, but one must also assume that because of the unverified nature of the source, much of this content is useless. For social media data to be of any use for topic identification, we must find a way to filter uninformative information and capture only information which, based on its content similarity to the news media, may be considered useful or valuable.

The news media presents professionally verified occurrences or events, while social media presents the interests of the audience in these areas, and may thus provide insight into their popularity. Social media services like Twitter can also provide additional or supporting information to a particular news media topic. In summary, truly valuable information may be thought of as the area in which these two media sources topically intersect. Unfortunately, even after the removal of unimportant content, there is still information overload in the remaining news-related data, which must be prioritized for consumption.

To assist in the prioritization of news information, news must be ranked in order of estimated importance. The temporal prevalence of a particular topic in the news media indicates that it is widely covered by news media sources, making it an important factor when estimating topical relevance. This factor may be referred to as the MF of the topic. The temporal prevalence of the topic in social media, specifically in Twitter, indicates that users are interested in the topic and can provide a basis for the estimation of its popularity. This factor is regarded as the UA of the topic. Likewise, the number of users discussing a topic and the interaction between them also gives insight into topical importance, referred to as the UI. By combining these three factors, we gain insight into topical importance and are then able to rank the news topics accordingly.

II. LITERATURE SURVEY

In this section we will mainly discuss about the background work that is carried out in order to prove the performance of our proposed Method. Now let us discuss about them in detail

MOTIVATION

1. 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.

2. Map Reduce: Simplified Data Processing On Large Clusters

AUTHORS: J. Dean and S. Ghemawat

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 largescale 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.

3. 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%

III. EXISTING METHODOLOGY

To the best of our knowledge, there is no concept like SociRank in the existing system to identify the ranking related news using Social Media Factors .Hence the users cannot able to filterout the information in a proper ranked manner from the social networks. Also there is no concept like clustering the tweets and their re-tweets as per the several social influence categories like :Positive,Negative and Neutral.

LIMITATIONS OF THE EXISTING METHODOLOGY

The following are the limitation of existing system. They are as follows:

1. There is no Information filtering for social computing.
2. There is anonymous topic ranking
3. There is no single method to gather all the individual users interest into one location.

.

IV. PROPOSED METHODOLOGY

In the proposed system, the method proposed an unsupervised method— SociRank— which identifies news topics prevalent in both social media and the news media, and then ranks them by taking

into account .The proposed system can able to gather all the prevalent information from various users in social media and then try to collect their interest and promote the topic for others who search for the same information.

ADVANTAGES OF THE PROPOSED SYSTEM

The following are the advantages of the proposed system. They are as follows:

1. We show that our proposed approach exhibits better results ranking the prevalent information.
2. We can able to find the Percentage of prevalent information based on topic wise
3. Here the users can able to view the re-tweet information in three categories like : Positive,Negative and Neutral.
4. It is very efficient method to collect user reviews and try to recommend the same for multiple users who are in search of the same information.

IV. IMPLEMENTATION STAGE

Implementation Stage is where the hypothetical structure is changed over into automatically way. In this stage we will partition the application into various modules and afterward coded for arrangement. The application is separated essentially into following 2 modules. They are as follows:

1) Admin Module
2) User Module

Now let us discuss about each and every module in detail as follows:

1) Admin Module

In this module, the Admin has to login by using valid user name and password. After login successful he can perform some operations such as Authorizing users, Login ,View all users and authorize, give click option to view all users locations using Multiple Markers ,View all Friend Request and Response ,View all users time line tweet details with Soci rank, rating and give tweet ,View all tweets by clustering based on tweet name and show tweeted details,SociRank,rating and View all Relevant Term Identification on all tweets and group together(similar tweeted details for each and every created tweet) ,View all users outlier detection tweet with its tweeted details,Soci_Rank,rating and View all term frequency on all tweets count(Display the tweets which is getting tweet regularly) based on tweet name, View all tweet news Socirank in chart and View all tweet term frequency count in chart based on date and time, View all tweets tweeted socirank in chart

2) User Module

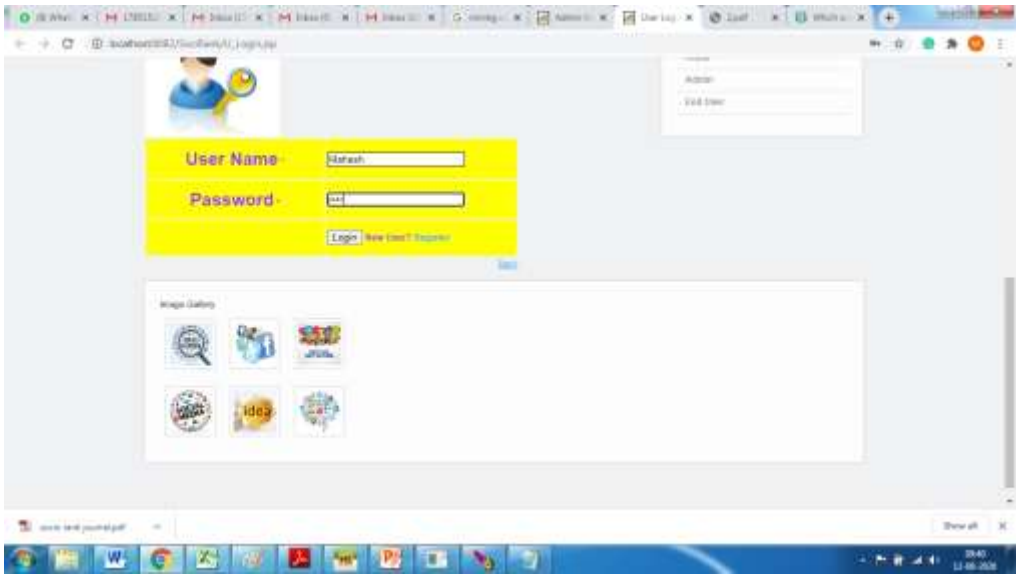
In this module, there are n numbers of users are present. User should register before performing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user can perform some operations like Register with Location and Login, View Your Profile with location ,Search Friend and Find Friend Request, View all Your Friends Details and Location Route path from Your Location, View all your time line tweets with Soci rank, rating and give tweet, Create tweet for News like Tweet name, tweet uses, Tweet desc(enc),tweet image and View all your tweet with re tweet details,Socirank,rating,Search tweet and list all Tweets and view its details and give re tweet, give rank by hyper link and View all your friends Tweets and give Tweet

V. EXPERIMENTAL REPORTS

Admin Views the Socio Rank

Figure. Represents the Socio Rank

USER LOGIN



USER MAIN PAGE

Figure . Represents the User Login and Home Page

VIEW ALL FRIENDS TWEETS AND RE-TWEET

Figure . Represents the All Friends Tweets

VI. CONCLUSION

We presented a valuable 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. Our experiments show that SociRank improves the quality and variety of automatically identified news topics.

VII. REFERENCES

[1] M. R. Morris, J. Teevan, and K. Panovich. A Comparison of Information Seeking Using Search Engines and Social Networks. In In Proc. of ICWSM, 2010.

[2] M. R. Morris, J. Teevan, and K. Panovich. What do People Ask Their Social Networks, and Why?: A Survey Study of Status Message Q&A Behavior. In Proc. of CHI, 2010.

[3] Z. Gyongyi, G. Koutrika, J. Pedersen, and H. Garcia-Molina. Questioning Yahoo! Answers. In Proc. of QAWeb, 2008.

[4] Yahoo!Answers Team. Yahoo! Answers BLOG. http://yahooanswers.tumblr.com/, [Accessed on 10/20/2014].

[5] B. Li and I. King. Routing Questions to Appropriate Answerers in Community Question Answering Services. In Proc. of CIKM, 2010.

[6] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In Proc. of WWW, 2008.

[7] G. Drosatos, P. Efraimidis, A. Arampatzis, G. Stamatelatos, and I. Athanasiadis. Pythia: A privacy-enhanced personalized contextual suggestion system for tourism. In COMPSAC, 2015.

[8] S. Li, Q. Jin, X. Jiang, and J. Park. Frontier and Future Development of Information Technology in Medicine and Education: ITME 2013. Springer Science & Business Media, 2013.

[9] A. Mtibaa, M. May, C. Diot, and M. Ammar. Peoplerank: Social Opportunistic Forwarding. In Proc. of Infocom, 2010.

[10] E. Pennisi. How Did Cooperative Behavior Evolve? Science, 2005.

[11] H. Shen, Z. Li, G. Liu, and J. Li. Sos: A distributed mobile q&a systembased on social networks. TPDS, 2014.

[12] A. Spagnolli and L. Gamberini. Interacting via sms: Practices of social closeness and reciprocation. British Journal of Social Psychology, 2007.

[13] M. L. Radford, C. Shah, L. Mon, and R. Gazan. Stepping Stones to Synergy: Social Q&A and Virtual Reference. Proceedings of the American Society for Information Science and Technology, 2011.