

Mining Social Response to Crisis via Electronic Media

Juan Yang, Xiqian Zhao, and Qing Li
Southwestern University of Finance & Economics
Chengdu, China
Email: liq_t@swufe.edu.cn

Abstract—In this paper, we are interested in using text mining to analyze textual data obtained from electronic social media in order to find out social concerns on crisis. In particular, we explore the temporal theme patterns in the natural catastrophe, i.e. the Yushu Earthquake, and compare the foci of the Chinese public, Chinese government and the overseas. By incorporating sources of information with a variety of aspects, we can distill not only snapshots of social concerns at given times but also the evolution of these concerns over time. Our intension is to assist the government in publicizing appropriate information and making intelligent decisions.

Keywords: text mining, electronic social media, Yushu Earthquake, response.

I. INTRODUCTION

When a crisis happens, the government should take measures to solve the problem promptly. Our research focuses on people's concerns with respect to the crisis in order to find some useful knowledge that can help the government to adjust its strategy in time. The goal is to meet the needs of the people, to solve problems in crisis situations, and to maintain social stability.

News is an important source of information reflecting the society. The media give the public the most timely messages including the government's policy with respect to the event and life coverage. With the development of the information technology, electronic media provides us a comprehensive view via network news. Note that the perspectives from news consumers home and abroad can vary considerably. In the past twenty years, the Web has evolved from a framework of information dissemination to a social interaction facilitator for its users. Such technological advancement has fertilized vibrant creation, sharing, and collaboration among the users. Internet-based forums usually express the public concerns to a great extent. Yet, these rapidly growing on-line textual data necessitate automated information processing.

Here, we extract the themes of news articles and forum discussions using text mining technologies to find what the social concerns are for further analysis. This article takes the Yushu Earthquake as a case study. We focus on textual data taken from traditional news Web sites and Internet-based discussion and forums. In particular, we gather text contents from 1) official Chinese media, 2) overseas media, and 3) domestic Chinese forums. The reason we selected these categories of news media is that they reflect different views of different people on the crisis. By incorporating sources

of information with a variety of aspects, we can distill not only snapshots of social concerns at given times but also the evolution and difference of these concerns over time. Our intension is to assist the government in publicizing appropriate information and making intelligent decisions.

In this paper, we are interested in using text mining to analyze textual data obtained from electronic social media in order to find out the social concerns. The rest of this article is organized as follows. First in Section 2, we review research related to this work. We describe the design of our framework and explain the functions of the modules in Section 3. Next, we conduct experimental studies by analyzing the texts from three forms of electronic social media (Section 4). In Section 5, we conclude this article with discussion on some fundamental assumptions made in this work and speculation on future extensions of our work.

II. RELATED WORK

In a broader context, research on social concerns have attracted scholars of different disciplines for years. In social science, studies focus on the different targets of public concerns. For example, Murray et al. [5] study the evolution of public response to energy crisis. Iyengar and Simon [4] explore three types of media to legitimize the administration's perspective on the Gulf Crisis. Dean [2] investigates the effects of company reputation for social responsibility related with a corporate crisis. Wang et al. [13] study the utilization of social concerns for social media recommendation.

Most of these studies are based on the pre-collected data including the Continuous National Survey carried by National Opinion Research Center [5] and a corporate database [2]. We argue that social media should be utilized to understand the public interests due to its responsiveness and extensive coverage. In this article, we take a different route using text mining on Web information. Text mining is the process of analyzing collections of textual materials in order to capture key concepts, themes, and to uncover hidden relationships and trends. Recent researches study the advanced text mining techniques to trace the topic trends of textual information. For instance, Kontostathis et al. [6] compare several commercial products for topic tracking. Montes-y-Gómez et al. [8] explore a simple statistical measures to uncover news trends, and associates their relations. Clifton [1] utilizes clustering techniques to assist topic identification. Morinaga and Yamanishi [11]

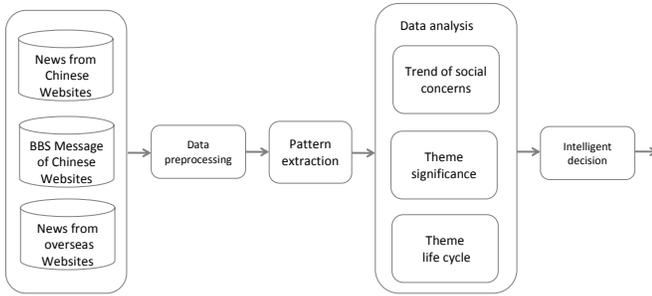


Fig. 1. Overview of mining social response

propose a finite mixture model to capture the dynamics of topic evolution. Mei and Zhai [10] study the language model to discover evolutionary theme patterns in news. Compared to the previous approaches on temporal theme pattern analysis, Mei et al. [9] take a further step by exploring the spatiotemporal patterns in blogosphere. Such studies provide a strong technique support to track public concerns on a crisis via news and Internet-based forums.

In this paper, we obtain and analyze the social response using text mining techniques. In particular, we explore the temporal theme patterns in the natural catastrophe, i.e. the Yushu Earthquake, and compare the foci of the Chinese public, Chinese government and the overseas.

III. RESEARCH DESIGN

In this article, we take a recent crisis, i.e. Yushu Earthquake, as a case study. On April 14, 2010, a 7.1-magnitude earthquake struck Yushu County of Qinghai Province in China. According to the Xinhua News Agency, 2,698 people have been confirmed dead, 270 missing, and 12,135 injured of which 1,434 are severely injured. The epicenter was located in Rima village, upper Laxiu township of Yushu County, in remote and rugged terrain, near the border of Tibet Autonomous Region. Right after the tragedy, Chinese government promptly took a series of actions to deal with this crisis. This crisis was cared by lots of people inland and abroad. In this article, we study social responses to this crisis by mining various kinds of Web information. We believe the responsiveness and extensiveness of Internet-based media provide a sufficient source for social response analysis. With the power of text mining, the trends and differences of various social entities can be dug out.

The outline of our research design is described in Figure 1. The whole procedure to analyze the social responses via Internet-based media consists three modules:

- **Data preparation** : it collects a large volume of online news information from various types of Web sites. These accumulated textual information is cleaned up for further analysis.
- **Data processing** : text mining techniques are applied to extract themes/topics in the Internet-based media.
- **Data analysis** : data analysis is further utilized to extract useful knowledge to assist the government in tackling the

crisis. In particular, we track the topic developing trend, and find topic life cycles and topic significance.

The details of these steps are explained in the following sections.

A. Data Preparation

In this article, we study people's concerns with respect to the crisis to find some useful knowledge that can help the government to adjust its strategy in time. We collect and analyze three kinds of textual information related with Yushu Earthquake from various types of Web sites. In particular, we gather text contents from 1) official Chinese media, 2) overseas media, and 3) domestic Chinese forums.

To obtain the relevant news from official Chinese and overseas media, we query a domestic (BAIDU.com) or global search engine (Google.com) with "Yushu earthquake" in Chinese and English, respectively, and collect relevant news articles published from April 14 to May 5 in 2010. We also collect 30,000 groups of discussion on the subject of the Yushu Earthquake from Chinese Internet forums. Similar to the news, the discussion messages are posted within 22 days after Yushu earthquake happened.

Each document (news or discussion message) is treated as a term vector where term weight is valued by the standard TF*IDF definition [12]. In particular, We assume that each term is associated with an "inverted document frequency (IDF)", denoted by $I(t) = \log \frac{N}{n(t)}$, where N is the corpus size and $n(t)$ is the number of documents in corpus containing term t . We use a function $f(t, d)$ to denote the number of occurrences of term t in document d , i.e. "term frequency (TF)". The final term weight is the product of TF and IDF. With such transformation, we can apply several data mining techniques to analyze these textual documents.

B. Theme Extraction

Each document (news or discussion message) contains one or more topics. The goal of theme extraction is to find out the topics of each document. With symbolizing each text in term vector, we can extract representative terms (keywords) as the concepts in our corpus. Please note that different terms with similar or the same semanteme composite a concept. In addition, each term can be classified into a proper pre-defined group including products, organizations, or people with priori-knowledge. With the knowledge of concept and group information, each document is associated with several tags revealing its topics and group. Thus, we synthesize the groups and concepts into a theme (or topic) and associate them with each document. Each theme reflects what concern a social entity cares about.

C. Data analysis

We analyze the textual information through three different aspects. Specifically, we track the developing trends of extracted topics, find their life cycles, and evaluate their significance.

TABLE I
THE NUMBER OF NEWS ARTICLES EVERYDAY

Date	Frequency		Date	Frequency		Date	Frequency	
	C	F		C	F		C	F
1	537	13	9	230	3	17	23	0
2	720	12	10	150	3	18	10	1
3	602	9	11	90	2	19	17	0
4	464	5	12	50	3	20	10	0
5	457	4	13	98	0	21	24	0
6	418	6	14	44	0	22	9	0
7	287	2	15	29	1			
8	346	9	16	28	0			

The topic developing trend is measured by the number of news reports or message posts. t_i is defined as the i th day after the crisis happened. it starts from 1 to 22. n_i denotes the number of the news or discussions per day. The larger the number of reported news or posts n_i , the stronger the cares of the entire society. Note that three social entities are considered, i.e., the Chinese public, Chinese government and the overseas. Knowing the associated documents of each theme from different social entities, we can easily track the concerns of each social entities in the crisis.

The intensity of concerns is represented by the topic developing trend. We also apply the life cycle of each theme to evaluate its duration. Theme life cycle is the lasting days after the first appearance of the reported news or posts.

In order to evaluate the importance of each theme in the crisis, we introduce a variable h_j , the number of texts contains theme j . We define s_{ij} is the ratio between h_j and n_i on the i th day ($s_{ij}=h_j/n_i$). Observing the change of this value, the topic significance can be easily tracked.

IV. EXPERIMENTAL STUDY

A. The trends of social concerns

In order to study the trends of social concerns, we calculate the number of news articles related with "Yushu" earthquake within 22 days after the earthquake happened.

The number of news articles are summarized in Table 1. Here, C denotes the news articles are obtained from Chinese Web sites, and F denotes that the articles are from overseas Web sites.

To well understand the trends of social concerns, The trends of social concerns in two different social entities are shown in Figure 2 and Figure 3, respectively. In a nutshell, it is observed that social concerns on the Yushu earthquake from both Chinese official and overseas media decrease as time goes on. This is quite understandable that the effect of a crisis tends to be gradually attenuated with the time passing by. However, there are three peaks in the curve of social concern from Chinese official media.

- The first peak occurs on April 5th, the first day after the earthquake happened. Right after the crisis, more and more news agencies knew this tragedy and reported it online. It usually takes one night to reach the peak. On the second day of the earthquake, the number of news

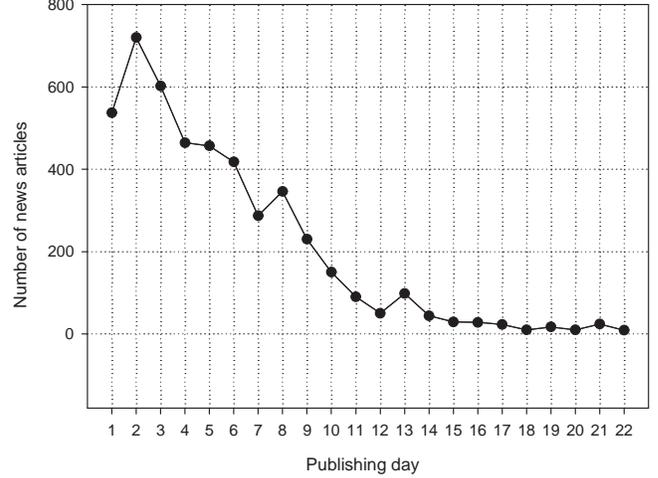


Fig. 2. Trends of social concerns in Chinese Web sites

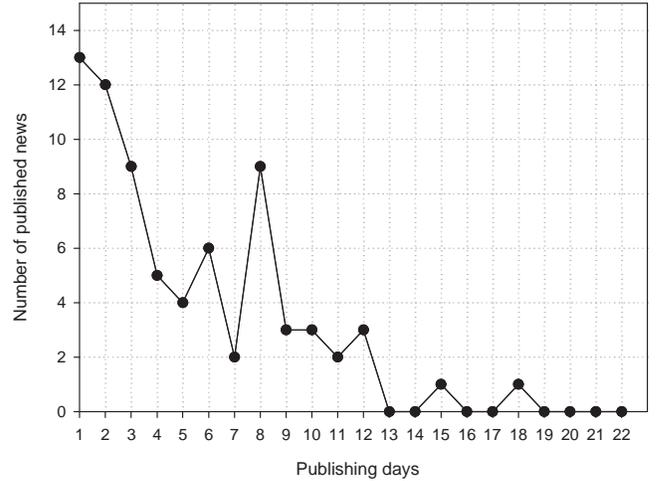


Fig. 3. Trends of social concerns in overseas Web sites

articles on this crisis reaches the maximum, and decreases gradually with the attenuation of the crisis.

- The second peak occurs on the 8th day of the earthquake happened. This is a special day for Chinese people to memory the dead. In Chinese traditional culture, people always hold a ritual to pray for the dead on the 7th day of their death. Even though the impact of the crisis attenuates with the time passing by, such kind of ritual bring eyeballs of the society on this crisis again. Therefore, they are more news articles on the crisis than several days before.
- The third peak occurs on the 13th day after the earthquake. Since it is a long time after the tragedy, the rescue activity is almost to the ends, and the affairs of disaster recovery has just been officially started on that day. This peak reflects people concern on the disaster recovery.

In Figure 3, similar conclusion can be drawn. This actually

reveals that most overseas media on Yushu earthquake was affected by the news reports from the inland media.

B. Themes analysis

TABLE II
THE THEMES EXTRACTED FROM NEWS ARTICLES AND INTERNET FORUM

Chinese News	emergency	disaster	mainland
	Tibet	quake	drug
	rescue	Yushu	Qinghai
	building	centre	department
Overseas' news	organization	relief	
	Tibet	regions	area
	recovery	quake	provinces
	inhabitants	government	building
Chinese forum	Asia		
	friends	supplies	school
	recovery	quake	provinces
	bless	relief	building
	Asia	earth	peer
	compatriots	people	

In our theme extraction module (Section III-B, we extract out the main topics of the relevant news articles or posts. such themes indeed represent the foci of social concerns on the crisis.

Table 2 shows the themes extracted from all relevant news articles both inland and overseas, and messages from Chinese Internet-based forums. From official Chinese news, we capture 14 themes in total. Each theme carries a strong information revealing social concerns on the crisis. For example, Theme "emergency" reveals people concern on the rescue of victims in the earthquake. Theme "mainland" contains the information that many foreign leaders and Chinese people abroad expressed their sympathy and solicitude for the casualty. From the overseas news articles, 10 themes are extracted to reflect the concerns of social entities abroad. It is interesting to find out that overseas media has different opinions on the same event with Chinese media. For instance, Theme "inhabitants" extracted from overseas news carries the message of foreign social entities' care on Chinese local policy, especially the rules to the Tibetans. Some news articles within this theme complain that Chinese government did not permit them to rescue victims because the disaster is happened in Tibetan-inhabited area. They pointed out that Chinese government is too sensitive to the Tibet affairs.

Even though reported the same event, inland and overseas news articles may have different opinions. Take theme "Tibet" as an example. News articles associated with theme "Tibet" in Chinese media describe the situation of the earthquake, post-quake relief and rescue work in Tibetan-inhabited areas. It emphasizes that one in trouble, all to help. In contrast, news articles in overseas media focus on the Dalai Lama, Tibetan Buddhist monks and the funerals of Tibetans who died in the earthquake. Theme "building" is another good example. Relevant articles from Chinese media mainly refer to the temples which have been destroyed in the disaster. However, news from overseas focus on the evidence that most school buildings are more weak than others to resist the earthquake.

It tries to imply that there is corruption in the construction of these school buildings in China.

Themes extracted from Posts in the Chinese forum represent the public concerns on the crisis. There are totally 14 themes extracted from this information source. Most of them are similar to these themes obtained from Chinese official news media. However, there are some exceptions reflecting the opinions from the public. Theme "friends", "bless" and "compatriots" are good examples of such exceptions. Posts of these themes convey the public's love and care for the victims and their satisfaction on the response of the government to the crisis. In addition, many messages on finding lost people in the earthquake remind that the government should set up an official information exchange platform to assist victims in finding their friends and relatives.

From the above, we find that Chinese media and the overseas media have different perspectives based on their interests even though they intend to report the same event. And the concerns of the public is a good metric to evaluate the work of the government and remind them the necessary things but neglected.

C. Life cycle analysis

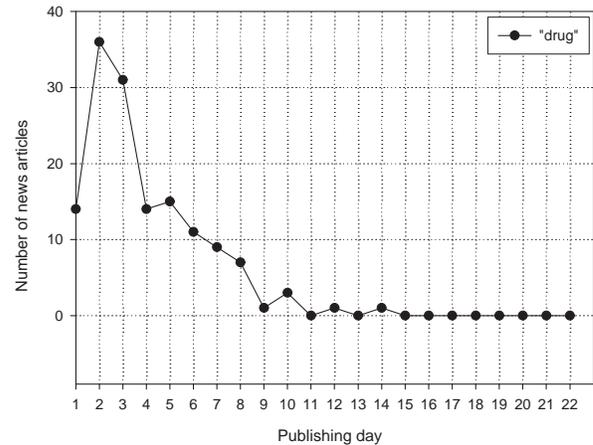


Fig. 4. Life Cycle of "drug"

Different themes have different life cycles. Theme life cycle tells us how long people focus on a particular event and how this concern changes as time goes by. For example, due to the shortage of medicines to save victims and control the post-disaster epidemic, there were a strong social concern on the provision of medical goods. With the strengthening of the rescue, more and more relief supplies were shipped into the disaster area. Such concern on "drug" was attenuated as time passing by. this trend is actually described by the life cycle curve in Figure 4.

In addition, we can easily find out the different opinions of inland and overseas entities by comparing event life cycles. For example, the trends of theme "quake" in Chinese official news and overseas news are described in Figure 5. It can be

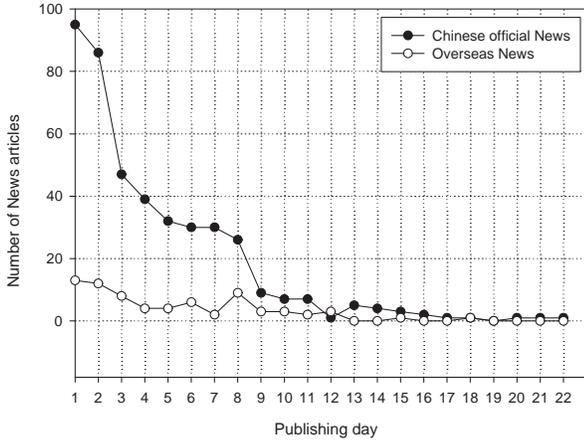


Fig. 5. Life cycle of "quake"

observed that the concern on earthquake event from inland is stronger and last longer than that of overseas.

D. Theme significance analysis

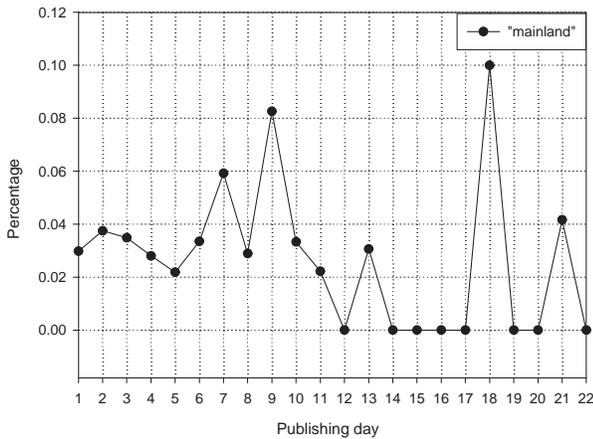


Fig. 6. Theme significance of "mainland"

Since there are several themes in daily news articles, we draw a theme significance curve to evaluate its relative importance among all daily themes. Therefore, we can easily find out how much people pay attention to this theme. Figure 6 illustrates the significance change of theme "mainland" within 22 days after the earthquake. This topic has the greatest impact on May 1st, 2010 on which about 10% news articles report about the topic of "mainland".

By comparing theme significance in different media sources, we can also judge concern variance of different social entities while they report news on the same event. As shown in Figure 7, the discussion on "quake" topic only takes a small portion of all topics covered by Chinese media. Besides the earthquake itself, Chinese people also paid attention to other

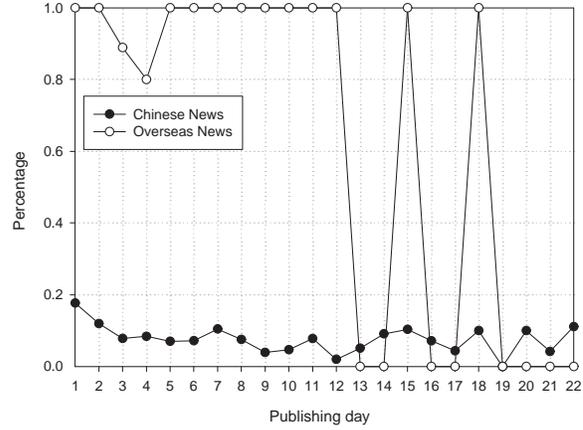


Fig. 7. Significance Comparison of Theme "quake" in Chinese and overseas media

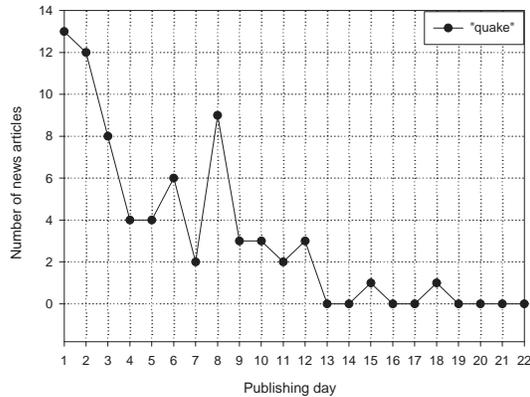
things including victim rescue, post-disaster epidemic control and reconstruction. In contrast, overseas media paid a lot of attention to the "quake" topic. Most news reports of overseas media were related with this topic. Obviously, Chinese and overseas media may value the same topic differently for news-making.

In addition, theme significance is a good indicator to deeply understand social concerns, especially when it is utilized with theme life cycle. For example, we may feel theme "quake" is no more cared by the overseas social entities from Figure 8(a). However, as shown in Figure 8(b), we can find that even though the total number of news articles on "quake" is reduced gradually, news articles on this issue still take a large portion of all relevant daily news.

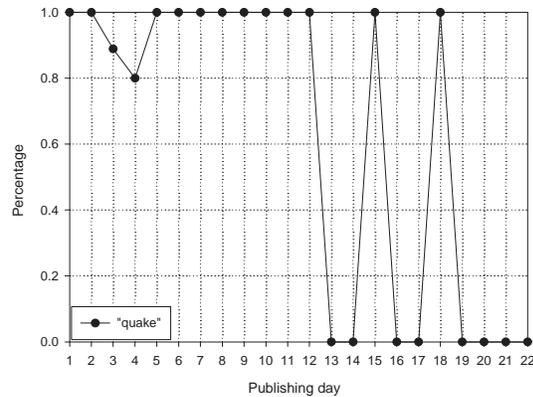
V. CONCLUSION

This research studies on the utilization of text mining on electronic media to find out social concerns on crisis. Our intension is to assist the government in publicizing appropriate information and making intelligent decisions. In particular, we explore theme patterns of news reports related with Yushu Earthquake, and compare the foci of different media including Chinese government, Chinese public and overseas societies. There are several interesting findings summarized as follows in this case study:

- Chinese official media covers a more widely topics on the crisis and has more news articles on the Yushu earthquake than overseas media.
- Most reports from overseas media is derived from the Chinese news articles. However, overseas media has different opinions with Chinese media on the same event because of political reasons.
- The public of China shows a positive attitude to several actions of Chinese government including rescuing victims, post-disaster reconstruction.
- It is necessary for the government to set up an official information exchange platform for fining lost people right



(a) Life cycle of "quake"



(b) Theme significance of "quake"

Fig. 8. Life cycle versus theme significance

after the tragedy.

This study can be extended in a few interesting ways. For example, More online information sources can be considered. For example, we can enrich the information sources for the public by mining messages from Blog and twitter. In addition, A more sophisticated text mining technique is expected to be applied for more accurate theme detection and tracking.

ACKNOWLEDGMENT

This research is supported by National Natural Science Foundation of China (Grant No.60803106), the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, and the Fok Ying-Tong Education Foundation for Young Teachers in the Higher Education Institutions of China.

REFERENCES

- [1] Clifton, C. and Cooley, R. *TopCat: data mining for topic identification in a text corpus*, Principles of Data Mining and Knowledge Discovery,174-183, 1999.
- [2] Dean, D.H. *Consumer reaction to negative publicity: Effects of corporate reputation, response, and responsibility for a crisis event*, Journal of Business Communication,Vol.41, No.2,192-211, 2004.
- [3] Glance, N. and Hurst, M. and Tomokiyo, T. *BlogPulse: Automated trend discovery for weblogs*, WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, Vol.2004,185-221, 2004
- [4] Iyengar, Shanto and Simon, Adam. *News Coverage of the Gulf Crisis and Public Opinion, A Study of Agenda-Setting, Priming, and Framing*, Communication Research, Vol. 20, No.3, 365-383, 1993.
- [5] Murray, J.R. and Minor, M.J. and Bradburn, N.M. and Cotterman, R.F. and Frankel, M. and Pisarski, A.E. *Evolution of public response to the energy crisis*, Science, Vol.184, No. 4134, 257-263, 1974.
- [6] Kontostathis, A. and Galitsky, L. and Pottenger, W.M. and Roy, S. and Phelps, D.J. *A survey of emerging trend detection in textual data mining*, Survey of Text Mining: Clustering, Classification, and Retrieval,Vol.1,185-221, 2003
- [7] Li, Z. and Wang, B. and Li, M. and Ma, W.Y. *A probabilistic model for retrospective news event detection*, Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval,106-113, 2005
- [8] Montes-y-Gómez, M. and Gelbukh, A. and López-López, A. *Mining the News: Trends, Associations, and Deviations*, Computación y Sistemas,Vol.5, No.1,14-24, 2001.

- [9] Mei, Q. and Liu, C. and Su, H. and Zhai, C.X. *A probabilistic approach to spatiotemporal theme pattern mining on weblogs*, Proceedings of the 15th international conference on World Wide Web, 533-542, 2006
- [10] Mei, Q. and Zhai, C.X. *Discovering evolutionary theme patterns from text: an exploration of temporal text mining*, Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, Vol.2004,198-207, 2005
- [11] Morinaga, S. and Yamanishi, K. *Tracking dynamics of topic trends using a finite mixture model*, Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Vol.2004,811-816, 2004
- [12] Ricardo, B. and Berthier, R. *Modern information retrieval*, Addison Wesley Longman Publishing Co. Inc., 1999.
- [13] Wang, J. and Li, Q. and Chen, Y.Z. and Lin, Z.X. *Recommendation in Internet Forums and Blogs*, Proceedings of 48th Annual Meeting of the Association for Computational Linguistics (ACL), 2010.