International Journal of u- and e- Service, Science and Technology
Vol.8, No. 7 (2015), pp.161-172
http://dx.doi.org/10.14257/ijunesst.2015.8.7.16

Web Information Credibility: From Web 1.0 to Web 2.0

Jie Zhao, Xiaoyang Lu, Xueya Wang and Zheng Ma

School of Business, Anhui University
zj_teacher@126.com

Abstract

Web has been one of the major information sources in people's daily life. However, a majority of Web information are regarded as incredible, as most Web information are not well evaluated before they are posted online. This eventually has a substantial impact on Web information utilization. For example, faked Web information will lead to wrong decision in business areas. In addition, the development of Web 2.0 makes people more convenient in producing and spreading information on the Web, which even worsens the information credibility problem on the Web. Thus, in this paper we review the current advances in the researches on Web information credibility, including related systems and methods for Web 1.0 information as well as existing studies on microblog information credibility. Finally, we propose a future research framework focusing on microblog information credibility.

Keywords: *Web, Information credibility, Microblog*

1. Introduction

Web has been one of the major resources for people to obtain news and other useful information. Recently, Web has also been considered to be the source for mining enterprise competitive intelligence [1-3] and other useful information such as location and time of Web objects [4-6]. Those pieces of information are helpful in supporting many applications like business decision making, personalized recommendation, and enterprise crisis analysis [7]. However, to this end, the quality and credibility of Web information becomes a critical issue, as faked or incredible information will probably lead to wrong decisions.

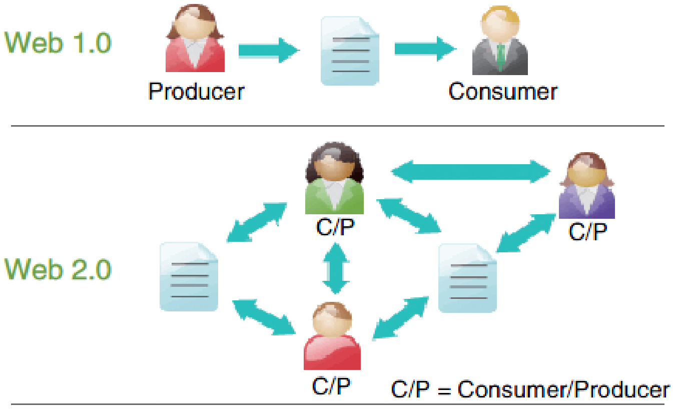

Figure 1. Web 1.0 vs. Web 2.0

ISSN: 2005-4246 IJUNESST

Traditional Web information included in Web pages is termed Web 1.0 information. The term Web 1.0 refers that information can only be produced by specific people such as Web－content managers. People as information consumers are not able to post their own information on the Web. In recent years, with the increasingly development of social network applications such as microblogs and blogs, Web 2.0 has been a hot topic, which allows people to post and spread information freely. Figure 1 shows the difference between the concept of Web 1.0 and Web 2.0.

The typical application of Web 2.0 is microblog, which has been one of the important media for people to express their opinions and spread interested information on the Web. According to a recent report posted by the Chinese Network Information Center, the total number of Chinese microblog users has been increased to over 250 million, and the increasing rate is highly over 290% [8]. As a consequence, microblog has thoroughly changed the traditional information spreading way. On the other hand, microblog information credibility is becoming an important issue in microblog information extraction and mining.

Based on the aforementioned background, in this paper, we first present a review on Web information credibility, including the research advances in Web-1.0 information credibility and in the Web-2.0 context. Then, we propose a future research framework that focuses on microblog information credibility. After an overview of the framework, we exploit the detailed considerations on some fundamental issues connected with the framework.

The remainder of the paper is structured as follows. Section 2 gives the review on Web-1.0 information credibility. Section 3 focuses on microblog information credibility. Section 4 presents future research framework, and finally we concludes the paper in Section 5.

2. Web-1.0 Information Credibility

Most of previous studies on information credibility focused on the Web 1.0 context. Information credibility refers to the believability of some information and/or its source [9]. It not only refers to the objective evaluation on information quality and precision, but also refers to the measurement on information sources. Recently, Web information credibility has been a hot topic and some works have been conducted. The earliest research on this area can be found in [10], in which the authors present a new method considering some trust mechanism in society to measure the information credibility. However, most works in Web information credibility were published after 2005.

2.1 Related Systems

There are some prototypes in Web information credibility evaluation, among which the most famous ones are *WISDOM* [11] and *Honto*?*Search* [12].

WISDOM extracts information from Web pages and clusters them according to senders and opinions. It is designed as a computer-aided tool to help users to determine the credibility of querying topics. *Honto*?*Search* is a Web Q/A system. It allows users to input a query about some fact and delivers the clustered analysis on the given fact. It is also a computer-aided system to help users evaluate information credibility. Besides, *HONCode* [13] and *MedPICS* [14] are two prototypes in the medical domain which also support Web information credibility evaluation. *HONCode* is built by the NGO Health on the Net Foundation. It can help users to find the credible medical websites, which are trusted by some third-party authoritative organization. The third-party-based evaluation method is very common is some specific areas, such as electronic commerce. *MedPICS* allows website owner to add some trust tags in the Web pages. And then users can filter Web information based on the trust tags in Web pages. For example, they can require that only the information whose trust tags are higher that a certain value be returned to them.

Previous works usually focus on different contents in the Web. Most researchers paid attention to the Web news credibility, searched results credibility, and products information credibility. Those works are generally based on Web pages and try to compute the credibility of Web pages. For example, *Google News* [15] uses the trustiness of news posters to evaluate the news credibility. Some news websites adopt a vote-based approach to measure the news credibility, such as www.126.com and www.sohu.com. There are also some works concerning Web information quality and the credibility of Web information sources. A lot of people also make investigation on Web information credibility. For example, a survey in 2004, which is focused in the electronic commerce area, shown that about 26% American posted comments on products in the Web [16], which indicates that users' comments is a key factor in the information credibility evaluation.

2.2. Related Methods

The basic methods used in Web information credibility evaluation can be divided into four types [9], which are the *Checklist* method, the *Cognitive Authority* method, the *Iterative Model*, and the *Credibility Seal Programs*.

(1) *Checklist*

This method uses a checklist to perform a user survey and then to determine the information credibility. This method is usually not practical in real applications. For example, some checklists contain too many questions that will consume too much time of users.

(2) *Cognitive Authority*

The *Cognitive Authority* method pays much attention to the authority of information. It is similar with the checklist method, except that it usually utilizes some automatic tools. For example, it suggests users use the *Whois*, *Traceoute*, and other tools to evaluate the authority of the information senders and websites.

(3) *Iterative Model*

The *Iterative Model* evaluates information credibility through three steps. First, it checks the appearance of the website. Second, it measures some detailed factors, including the professional level, the trustable level, the freshness, precise, and relevance to users' needs. Finally, users are required to mark the evaluated results. The similarity between the iterative model and the *Checklist* method is that both of them provide some criteria for users to mark the information credibility. The difference between them is that the iterative model pays more attention to the importance of the information receiver in the evaluation process.

(3) *Credibility Seal Programs*

This method is much different from other three methods. It provides some credibility seal program to help users to find credible sources in the Web. For example, the *HONCode* can help users to find trusted medical websites [13]. However, this method is usually restricted in certain areas due to the huge amount of Web information.

3. Microblogging Information Credibility

Microblog is a representative of Web 2.0 applications. Microblogging services such as Twitter and *Sina Weibo* have been widely used all over the world. It has been demonstrated that microblog usually has a big impact on a lot of social and political events. Therefore, recently research on microblog data has been a hot topic. While microblogging services can continuously produce a huge amount of data, how to effectively use those data is still a challenging problem. The hardest issue is the credibility of microblog data, which makes people difficult to utilize microblog data directly in business applications. In this section, we summarize recent studies in microblog information credibility.

3.1. Microblog Credibility

Microblog platforms produce a huge amount of data every day, due to its great number of users. A previous study [17] showed that microblogs were considered more incredible than other kinds of Web information such as news portal and blogs. On the other side, there are many microblog-based new applications, which indicate that it is urgent to study microblog information credibility.

So far, most studies in microblog information credibility concentrated on Twitter and English texts, but there are also some studies towards other languages such as Arabic and Chinese.

In [18], Al-Eidan et al. classified microblog credibility into three levels, namely credible, incredible, and questionable. They conducted a feature-based approach to determine the credibility of microblogs. The features used included the following ones:

1. Similarity between the concerned microblog and known credible texts.
2. Appearance of specific inappropriate words.
3. URLs within the microblog that point to trusted sources.
4. Certification of users who posted the concerned microblog.
5. Evaluation score from TwitterGrader.com.
6.

Table 1. Features for Evaluating Microblog Information Credibility in *SPOT* [19]

No.	Feature
1	Number of Followers
2	Number of Followees
3	Reputation of User
4	Frequency of user’s postings
5	Average URLs in a microblog
6	Frequency of a microblog being replied
7	Distance among microblogs

In [19], Perez, *et al.,* presented *SPOT* for evaluating microblog information credibility. *SPOT* detected the URLs embedded in a microblog to measure information credibility and detect questionable users on Twitter. Because of the 140-character limitation on text length, many people tend to user short URLs in microblogs. However, some bad URLs may be embedded in microblogs by malicious users. The *SPOT* system first obtains necessary information by using *Twitter* APIs. These pieces of information include users’ personal data, microblog contents, and the entities in microblogs. Then, it extracts features from the collected microblog data and classifies users, *e.g.,* using the *SVM* approach, into trusted and untrusted ones. Typical features are shown in Table 1.

In [20], Suzuki, *et al.,* evaluated microblog information credibility by analyzing the forward behaviors of microblogs. This approach is based on the assumption that a credible microblog is very likely to be kept unchanged when being forwarded. On the contrary, an incredible microblog is usually modified by users when it is forwarded.

In [21], Castillo, *et al.,* analyzed the credibility of topics on microblogging platforms. They assumed that there must be some clues to evaluate microblog information credibility in a social network environment. Particularly, they first extracted events and topics from microblogs and then evaluated credibility. Microblog topics were detected first by *TwitterMonitor* [22] and further by feature-based classification models. Different features were investigated in [21]. As a result, they used features shown in Table 2 and different classification models including *SVM*, *Decision Tree*, and *Naïve Bayes Model* to determine microblog information credibility.

Table 2. Features for Evaluating Microblog Information Credibility in [21]

No.	Type	Feature
1	User Features	Registration Date
		Number of Followers
		Number of Followees
		Certification Status
2	Message Features	Length of Microblog
		Question and Exclamation Mark
		Number of URLs
3	Propagation Features	Fan-out of the Propagation Tree
		Number of Microblogs in Sub-Trees
		Levels of the Propagation Tree
4	Topic Features	Number of Microblogs in a Topic
		Average Length of Microblogs in a Topic
		Percentage of Microblogs containing URLs

In [23], Kang, *et al.,* studied the credibility of microblogs about specific topics. They classified microblog credibility into two types: textual credibility and social credibility. Textual credibility refers to the credibility of microblogs within a specific topic, and social credibility is the credibility of users that is determined by social network data. They employed different models including social model, content model, and hybrid model to evaluate microblog information credibility.

In [24], Gupta, *et al.,* studied the credibility of events in microblogging platforms. They used a *PageRank*-similar approach to model the correlations among users, microblogs, and events, which is shown in Figure 2. Then, they proposed an algorithm called *EventOptCA* to iteratively compute the correlation scores and determine the credibility score.

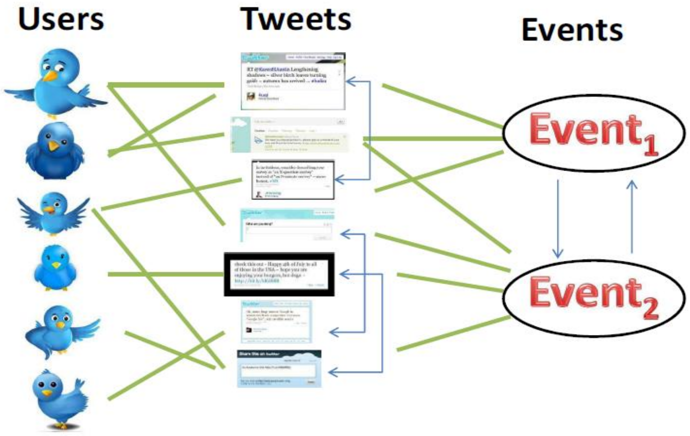

Figure 2. Modeling the Correlations among Users, Tweets, and Events [24]

3.2. Microblog Spam Detection

Spam has been one of the most influential factors that affect the credibility and usability of microblog information. Spams are usually seen in commercial advertisement, computer virus propagation, and news push. Spammers in microblogging platforms are driven by several goals, such as spreading advertisement to promote sales and compromising system reputation.

Spamming may jeopardize the trust of users on the system. Furthermore, it wastes users' time and energy to filter out spam messages. To this end, it is highly desirable to devise techniques and methods for identifying spammers and their behaviors in on-line social networks.

Spam detection has been observed in various social network systems, including YouTube, Twitter, Facebook, and Myspace. And many spamming or spammers detection methods have been proposed in literatures.

Feature-based spammer detection is a widely-used method. For example, in [25] the researchers proposed many features and adopted classical approaches in machine learning to identify spammers. In [26], a Naïve Bayes to detect spam names was introduced. Lee, *et al.*, in [27] introduced social honeypots to attract spammers, statistical of the properties of these spam profiles and creating spam classifiers to actively filter out existing and new spammers.

Lin, *et al.*, in [28] found three respective users, however the classifier is based on classical features that cannot tackle the new escaping mechanism [29]. Proposed an unsupervised method for automatic identification of spammers with the method link structure of the network in order to derive a legitimacy score for each account.

Another mainstream of anti-spam researches is focused on social network or graph based detection algorithms. Lee, *et al.*, [30] considered the correlated redirect chains of URLs in a number of tweets and trained a statistical classifier with features derived from correlated URLs. However, this work was not suitable for tweets that do not contain URLs [31]. Proposed to build a sender and receiver based graph, and to find the spammers according to the distance between senders and receivers. However, it cannot handle the spammers who seldom send post to others. Zhu, *et al.*, [32] proposed a supervised matrix factorization method with social regularization (*SMFSR*) for spammer detection in social networks, which exploited both social activities and users' social relationships in an innovative and highly scalable manner.

4. Future Research Framework

Figure 3 shows the future research framework we propose to evaluate microblog information credibility and to detect trusted events. We propose to first detect microblog spammers and remove the microblogs posted by spammers. Then, we evaluate the remained microblogs in terms of social network features. After that, we have filtered spammers and incredible microblogs. Finally, we propose to perform event detection and clustering on credible microblogs to output trusted events for further studies like business decision making.

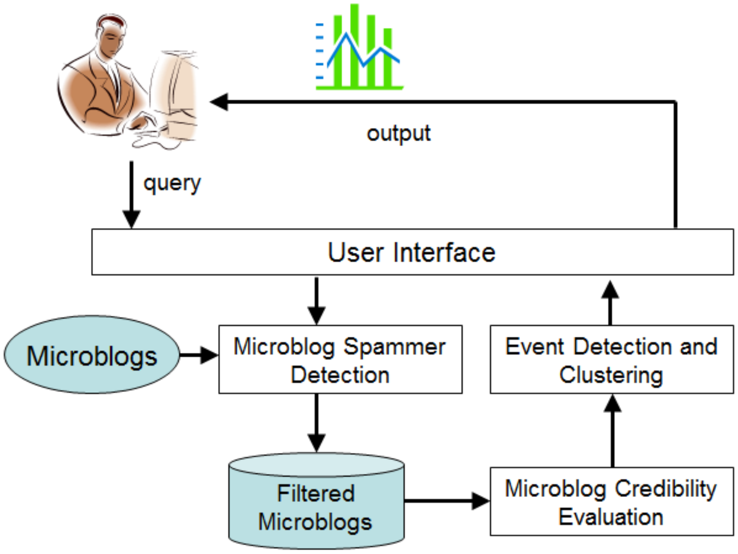
output
query
User Interface
Microblogs
Microblog Spammer Detection
Event Detection and Clustering
Filtered Microblogs
Microblog Credibility Evaluation

Figure 3. The Framework for Future Research

4.1. Microblog Spammer Detection

In recent years, social media services have grown to become important media for information sharing and communication. For example, *Sina Weibo* is currently the dominant microblogging service provider in China. Microblogging platforms allow users to share their personal information and opinions with friends. A microblog message may contain up to 140 characters of text and URL links. The posts published by a user are shared on the news feeds of the user's followers.

Spammers on microblogging platforms usually focus on advertising unsolicited messages instead of share information with other people, which will lower the information quality and credibility of microblogs. Detecting spammers aims to find out those users that have spamming behaviors. We assume that a spammer will very probably post spams, thus all the microblogs posted by spammers should be removed from the collected dataset.

Different kinds of spammers usually have differing behaviors and objectives. Therefore, it is necessary to first define spammers. Based on observations, we give the following properties of spammers:

(1) Spammers usually use feeds of third-party content to update and maintain accounts under the names of those third parties.

(2) Spammers usually issue many invitations to other users.

(3) Spammers usually publish or link to malicious contents.

Basically, we can classify spammers on microblogging platforms into two types: advertising intention users and following intention users. Advertising intention spammers are users whose goal is to publish production advertisement or promote other activity. Following intention spammers are users whose goal is to following unrelated accounts randomly.

4.2. Microblog Credibility Evaluation

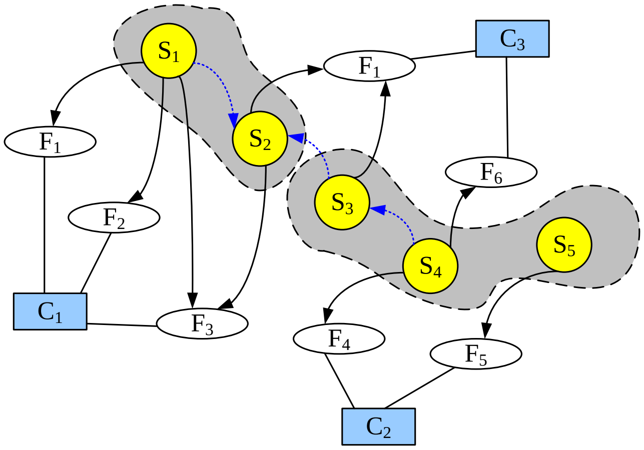

Figure 4. Social Network Modeling for Microblogs

Microblogging platforms provide a social network for users. Therefore, social-network features can be regarded as the most important and unique ones of microblogs, which should be considered with priority to evaluate microblog credibility. According the definition in Wikipedia, social network refers to the network of personal relationships. People are basically located in one or more social networks. Moreover, if we want to evaluate some people, it is reasonable to refer to the evaluation from the social network in which he or she is involved. Based on this viewpoint, we present the social-network-based method to perform the credibility evaluation of microblogs, which is shown in Figure 4.

As Figure 4 shows, the model consists of three types of nodes, *i.e.*, *S* node, *F* node, and *C* node. The circle *S* nodes represent original microblogs. The *C* nodes represent specific entities that appear in original microblogs. They can be extracted using existing named-entity recognition tools. The *F* nodes, represented by ellipses, indicate facts that appearing in microblogs. There are three types of edges between the nodes. The solid directed edge between an *S*-type node and an *F*-type node represents that a fact stems from a certain microblog. The dash directed edge between two *S*-type nodes represents the linking relationship between two microblogs. The solid undirected edge between an *F*-type node and a *C*-type node represents that an entity involves some facts.

Based on the model in Figure 4, we can computer the credibility of each node in the model. We assume that the credibility of a node can be influenced by its connected nodes. So we first compute the initial credibility of microblogs, and then compute the fact credibility in one social network (the gray area in Figure 4), and finally compute the credibility of each entity by combining the credibility of related facts.

4.3. Event Detection and Clustering

After evaluating microblog information credibility, we get trusted microblogs which are used as credible sources for further event detection. We aim to construct the evolution of events, *i.e.*, to detect evolutional events from microblogs. Generally, evolutional events consist of two kinds of information, which are event information and evolutional information. For this reason, we conduct an object-oriented approach in this paper to establish the representation of evolutional events. An evolutional event is modeled as a triple: *E = {EID, AD, DD}*, where *EID* is the identifier of the event, *AD (Attribute Descriptor)* describes the static properties of the event, *i.e.*, those properties that are not changing with time, *DD(Development Descriptor)* represents the dynamic evolutional properties of the event.

According to this representation, the dynamic properties of an evolutional event can be regarded as the temporal changes of the event. On the other hand, most evolutional events are related with locations. So we can model an evolutional event as a spatiotemporal object, and construct a spatiotemporal model to represent the dynamic evolutional properties of the evolutional events. The spatial dimension in evolutional events usually refers to a location in a large scale, *e.g.*, Beijing or Shanghai. Besides, the spatiotemporal changes in the evolutional event model are also different from those in traditional spatiotemporal data model. That is, traditional spatiotemporal changes generally refer to the split, merge, or shape change of a geographical object, while the evolution of event usually refers to the beginning, development, end, and influence of an event.

For the algorithms of event detection, we propose to use a fine-grained approach [33, 34], which is shown in Figure 5. As microblogs are very short, we directly use the POS (Part of Speech) tagging and grammar analysis to extract the description of WHO/WHY/HOW/WHAT. For time and location extraction, we can use previous works on time and location extraction to obtain the information about WHEN/WHERE [4, 5, 35, 36].

After extracting the events, we first cluster the events according to topic words, and further organize them along a time line. The critical issue is how to cluster the events. We propose to employ a time-based clustering approach [37] to compute the time-based similarity among events and finally get the clusters.

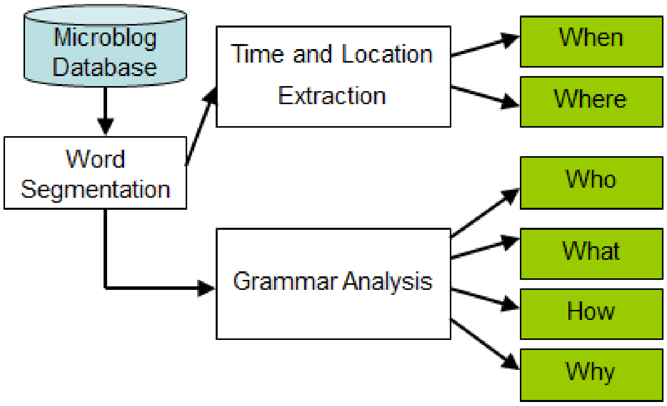

Figure 5. Extracting Events from Microblogs

5. Conclusions

Web information credibility has been a hot research topic in recent years, with the increasingly development on big data management and analysis. Traditional studies in Web information credibility focused on Web pages, which was identified as Web-1.0 information credibility. At present, Web 2.0 information attracts more attention from both academia and industries. Microblog, as a representative of Web 2.0 technologies, has become a new trend of Web information studies. Thus, Web-2.0 information credibility has also become a focus. In this paper, we summarized recent advances in both Web-1.0 and Web-2.0 information credibility. We discussed some detailed approaches regarding Web-1.0 information credibility systems and methods, as well as microblog credibility and spam detection. Finally, we propose a future research framework with a focus on microblog spammer detection, microblog credibility evaluation, and microblog event detection and clustering.

Acknowledgements

This work is supported by the National Science Foundation of China (no. 71273010), the Doctor Start-up Fund of Anhui University, and the Academic Innovative Research Program of Anhui University (no.yfc100274 and no.yfc100280).

References

[1] J. Zhao and P. Jin, "Towards the Extraction of Intelligence about Competitor from the Web, Second World Summit on the Knowledge Society (WSKS'09)", LNAI 5736, Chania, Crete, Greece, (**2009**), pp. 118-127

[2] J. Zhao and P. Jin, "Business Relations in the Web: Semantics and a Case Study", Journal of Software, vol. 5, no. 8, **(2010)**, pp. 826-833.

[3] P. Jin, J. Yang, J. Zhao and Y. Liu, "A Structural Approach to Extracting Chinese Position Relations from Web Pages", Journal of Web Engineering, vol. 12, no. 5, **(2013)**, pp. 363-382.

[4] Q. Zhang, P. Jin and L. Yue, "Extracting Focused Locations for Web Pages", First International Workshop on Web-based Geographic Information Management (WGIM) (in conjunction with WAIM), LNCS, Wuhan China, vol. 7142, **(2011)**, pp. 76-89.

[5] J. Zhao, P. Jin, Q. Zhang and R. Wen, "Exploiting Location Information for Web Search", Computer in Human Behaviors, vol. 30, **(2014)**, pp. 378-388.

[6] X. Zhao, P. Jin and L. Yue, "Automatic Temporal Expression Normalization with Dynamic Reference Time", The 23rd International Conference on Computational Linguistics (COLING), Beijing, China, **(2010)**, pp. 1498-1506.

[7] J. Zhao and P. Jin, "Extraction and Credibility Evaluation of Web-based Competitive Intelligence", Journal of Software, vol. 6, no. 8, **(2011)**, pp. 1513-1520.

[8] "CNNIC, The 29th Report of the Internet Development in China", **(2014)**, In: http://www.cnnic.cn/dtygg/dtgg/201201/W020120116337628870651.pdf.

[9] M. Metzger, "Making Sense of Credibility on the Web: Models for Evaluating Online Information and Recommendations for Future Research", Journal of the American Society of Information Science and Technology, vol. 58, no. 13, **(2007)**, pp. 2078-2091.

[10] A. Alfarez and S. Hailes, "Relying On Trust to Find Reliable Information", Proc. of International Symposium on Database, Web and Cooperative Systems (DWA-COS), **(1999)**.

[11] S. Akamine, D. Kawahara, Y. Kato, *et al.,* "WISDOM: A Web Information Credibility Analysis System", Proc. of the ACL-IJCNLP 2009 Software Demonstrations, Suntec, Singapore, **(2009)**, pp. 1-4.

[12] Y. Yamamoto and K. Tanaka, "Finding Comparative Facts and Aspects for Judging the Credibility of Uncertain Facts", Proc. Of WISE, LNCS, vol. 5802, **(2009)**, pp. 291-305.

[13] J. Fritch, "Heuristics, Tools and Systems for Evaluating Internet Information: Helping Users Assess a Tangled Web", Online Information Review, vol. 27, no. 5, **(2003)**, pp. 321-327.

[14] G. Eysenbach, "Consumer Health Informatics", British Medical Journal, vol. 320, **(2000)**, pp. 1713-1716.

[15] "Google, Google News Patent Application", **(2014)**, In: http://www.webpronews.com/news/ebusinessnews/wpn-45-20050503GoogleNewsPatentApplicationFullText.html.

[16] L. Rainie and P. Hitlin, “The Use of Online Reputation and Rating Systems”, **(2014)**, In: http://www.pewinternet.org/ PPF/r/140/report_display.asp.

[17] J. Schwarz and M. Morris, “Augmenting web pages and search results to support credibility assessment”, Proceedings of the 2011 annual conference on Human factors in computing systems. New York, NY, USA: ACM, **(2011)**, pp. 1245-1254.

[18] R. Al-Eidan, H. Al-Khalifa and A. Al-Salman, “Measuring the credibility of Arabic text content in Twitter”, Proc. of Digital Information Management (ICDIM), Washington, DC, USA: IEEE Computer Society, **(2010)**, pp. 285-291.

[19] C. Perez, M. Lemercier, B. Birregah, *et al.*, “SPOT 1.0: Scoring Suspicious Profiles on Twitter”, Proc. of Advances in Social Networks Analysis and Mining (ASONAM), Washington, DC, USA: IEEE Computer Society, **(2011)**, pp. 377-381.

[20] Y. Suzuki, “A credibility assessment for message streams on microblogs”, Proc. of P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), **(2010)**, pp. 527-530.

[21] C. Castillo, M. Mendoza and B. Poblete, “Information credibility on twitter”, Proceedings of the 20th international conference on World Wide Web (WWW), New York, NY, USA: ACM, **(2011)**, pp. 675-684.

[22] M. Mathioudakis and N. Koudas, “Twittermonitor: trend detection over the twitter stream”, Proceedings of the 2010 International Conference on Management of Data (SIGMOD), **(2010)**, pp. 1155-1158.

[23] B. Kang, J. O'Donovan and T. Höllerer, “Modeling topic specific credibility on twitter”, Proceedings of the 2012 ACM international conference on Intelligent User Interfaces, New York, NY, USA: ACM, **(2012)**, pp. 179-188.

[24] M. Gupta, P. Zhao and J. Han, “Evaluating event credibility on twitter”, Proc. of SDM, **(2012)**, pp. 153-164.

[25] M. McCord and M. Chuah, “Spam detection on twitter using traditional classifiers”, Proc. of the 8th International Conference on Autonomic and Trusted Computing (ATC), **(2011)**, pp. 175-186.

[26] D. M. Freeman, “Using naive bayes to detect spammy names in social networks”, Proceedings of the 2013 ACM Workshop on Artificial Intelligence and Security, Co-located with CCS, Berlin, Germany, **(2013)**, pp. 3-12.

[27] K. Lee, J. Caverlee and S. Webb, “Uncovering social spammers: social honeypots+ machine learning”, Proceedings of the 33rd international ACM SIGIR conference on Research and Development in Information Retrieval, ACM, **(2010)**, pp. 435-442.

[28] C. Lin, J. He, Y. Zhou, X. Yang, K. Chen and L. Song, “Analysis and identification of spamming behaviors in sina weibo microblog”, Proc. of SNA-KDD, **(2013)**, p. 5.

[29] M. Bouguessa, “An unsupervised approach for identifying spammers in social networks”, IEEE 23rd International Conference on Tools with Artificial Intelligence (ICTAI), Boca Raton, FL, USA, **(2011)**, pp. 832-840.

[30] S. Lee and J. Kim, “Warning Bird: A Near Real-Time Detection System for Suspicious URLs in Twitter Stream”, IEEE Transactions on Dependable Security and Computing, vol. 10, no. 3, **(2013)**, pp. 183-195.

[31] J. Song, S. Lee and J. Kim, “Spam filtering in twitter using sender-receiver relationship”, Proc. of RAID, **(2011)**, pp. 301-317.

[32] Y. Zhu, X. Wang, E. Zhong, N. N. Liu, H. Li and Q. Yang, “Discovering spammers in social networks”, Proc. of AAAI, **(2012)**.

[33] L. Zheng, P. Jin, J. Zhao and L. Yue, “A Fine-Grained Approach for Extracting Events on Microblogs”, the 25th International Conference on Database and Expert Systems Applications (DEXA), LNCS 8644, Munich, Germany, **(2014)**, pp. 275-283.

[34] J. Zhao, X. Wang and Z. Ma, “Towards Events Detection from Microblog Messages”, International Journal of Hybrid Information Technology, vol. 7, no. 1, **(2014)**, pp. 201-210.

[35] S. Lin, P. Jin, X. Zhao and L. Yue, “Exploiting Temporal Information in Web Search”, Expert Systems with Applications, Elsevier, vol. 41, no. 2, **(2014)**, pp. 331-341.

[36] S. Lin, P. Jin, X. Zhao and L. Yue, “Extracting Focused Time for Web Pages”, The 13th International Conference on Web-Age Information Management (WAIM), Springer, LNCS 7418, Harbin, China, **(2012)**, pp. 266-271.

[37] J. Zhao, X. Li and P. Jin, “A Time-Enhanced Topic Clustering Approach for News Web Search”, International Journal of Database Theory and Application, vol. 5, no. 4, **(2012)**, pp. 1-10.