Extraction and Credibility Evaluation of Web-based Competitive Intelligence

Jie Zhao [1, 2]
[1]School of Business Administration, Anhui University, Hefei, China
[2]School of Management, University of Science and Technology of China, Hefei, China
Email: zjpq@mail.hf.ah.cn

Peiquan Jin
School of Computer Science and Technology, University of Science and Technology of China, Hefei, China
Email: jpq@ustc.edu.cn

Abstract— **Web has been one of major information sources for enterprises to acquire competitive intelligence. However, traditional approaches focus on collecting Web pages and fail to generate practical competitive intelligence from Web pages. Another problem in the research on Web-based competitive intelligence is that Web pages may contain a lot of incredible information which will have big influence on the effectiveness of competitive intelligence. Aiming at solving these problems, we propose a framework in this paper for the extraction and credibility evaluation of Web competitive intelligence. We present an entity-based approach to extracting Web competitive intelligence, and a social-network-based method to evaluate the credibility of acquired competitive intelligence. The entity-based extracting approach is based on an ontology of Web competitive intelligence, which represents competitive intelligence as a set of competitor intelligence and competition environment intelligence. Some critical issues about the entity-based approach and the social-network-based method are analyzed in detail. The results show that our system is useful to improve the effectiveness of the extraction and credibility evaluation of Web competitive intelligence.**

Index Terms—**competitive intelligence; Web; entity-based approach; credibility evaluation**

I. INTRODUCTION

Web has been one of major information sources due to its huge data volume and rich semantics. In recent years, many researchers and organizations tried to find competitive intelligence from the Web, because competitive intelligence may be valuable and enhance enterprises' core competitiveness in the worldwide market (Kahaner et al., 1996). A previous survey indicated that about 90% of competitive intelligence can be acquired from the Web (Thompson and Wing, 2001; Lamar, 2007). Thus, many enterprises devoted themselves into building a Web-based competitive intelligence system to acquire, represent, and analyze competitive intelligence from the Web.

Unfortunately, Web-based competitive intelligence system is still an ongoing work in most enterprises. The most important thing is that a lot of fundamental issues are needed to be further studied (Deng and Luo, 2007), among which the extraction and credibility evaluation of Web competitive intelligence receives most attention. Traditional competitive intelligence systems (CIS) simply regard Web competitive intelligence as a set of Web pages, and tried to collect more and more related Web pages through some search engines. This method will result in a large amount of information processing work. For example, if you search the information about a corporation "Microsoft" in Google, you may get a result containing more than six millions of Web pages. It is hard for one to manually get intelligence from so a big data set. Recently, some people introduced the text mining techniques into the intelligence acquiring process (Mikroyannidis et al., 2006). However, while they are capable of filtering the non-related text blocks in a Web page, it divides a Web page into a set of text blocks. This eventually brings more information processing work in order to produce the competitive intelligence.

This paper mainly concentrates on the issue of extracting and evaluating competitive intelligence in the Web. It is an extended version of our previous work in the Second World Summit on Knowledge Society (WSKS'09) (Zhao and Jin, 2009) and IITSI'10 (Zhao and Jin, 2010). The contribution of the paper can be summarized as follows:

(a) We present a system framework to extract and evaluate Web competitive intelligence and discuss the design issues of the system.

(b) We present an entity-based approach to extract Web competitive intelligence. According to this approach, Web competitive intelligence is generated by three steps, namely entity extraction, entity relations extraction, and ontology instantiation. Our approach can generate entity-level competitive intelligence, and is more effective than traditional Web page based approaches.

(c) We propose a social-network-based method to evaluate the credibility of extracted Web competitive intelligence. The new method utilizes the nature of social network, and is based on a social-network-based credibility model. We investigate the details of the credibility model and the basic algorithm to evaluate the credibility of Web competitive intelligence.

doi:10.4304/jsw.6.8.1513-1520

The following of the paper is structured as follows. In Section 2 we discuss the related work. Section 3 gives out the system structure of extracting and evaluating Web competitive intelligence. Section 4 presents an ontology for Web competitive intelligence. Section 5 focuses on the entity-based extraction of Web competitive intelligence. In Section 6, the social-network-based method for the credibility evaluation of Web competitive intelligence is explored. And conclusions and future work are in the Section 7.

II. Related Work

A. Competitive Intelligence Extraction

Competitive intelligence refers to the process that gathering, analyzing and delivering the information about the competition environment as well as the capabilities and intensions of the competitors, and then transforming them into intelligence (Kahaner, 1996). Competitive intelligence is acquired, produced and transmitted through the competitive intelligence systems (CIS).

Traditionally, people usually utilize some publications to acquire competitive intelligence, such as news paper, magazines, or other industry reports. With the rapid development of the Web, people can search any information in a real-time way, thus it has become an important way to obtain competitive intelligence from the Web (Thompson and Wing, 2001).

The detailed procedure of producing competitive intelligence from the Web can be described as follows. For example, suppose the company wants to get the competitive intelligence about one of its competitors, namely, the company C, they will first search the information about the company C through some search engines, e.g. Google, typically using some keywords like "C Company". Then the experts analyze the gathered Web pages to make out a report about the company C. In this paper, we call this type of intelligence acquiring "Web-page-based competitive intelligence acquiring". The disadvantages of the Web-page-based way are obvious. Since the search engine will usually return a huge amount of Web pages, e.g. when you search in Google using the keywords "Microsoft Office 2008" you will get billions of Web pages, it is ultimately not feasible for experts to analyze all the searching results and produce valuable competitive intelligence.

Recently, researchers introduced the Web text mining approach into the CIS. The Web text mining aims at finding implicit knowledge from a huge amount of text data (Mikroyannidis, 2006). It depends on some fundamental technologies, including the computing linguistics, statistical analysis, machine learning, and information retrieval. So far, re-searchers have proposed some approaches to processing Web pages, such as extracting text from Web pages (Hotho et al., 2005) and detecting changes of Web pages (Khoury et al., 2007). According to the text-mining-based approaches, the noisy data in Web pages can be eliminated, and a set of text blocks are obtained and even clustered in some rules. However, since a Web page typically contains a lot of text blocks, this method will consequently produce a large number of text blocks which is much more than the number of Web pages. Besides, if the text blocks are clustered under specific rules, the information about competitors and competition environment will spread among different clusters and bring too much work for information analysis.

Competitive intelligence serves for companies and people, so in order to make the competitive intelligence systems more effective, first we should study what competitive intelligence companies need. As a survey indicated (Lamar, 2007), most people prefer to look up information by competitor. When we further ask one more question: "What is the competitive intelligence about the competitors?", most companies will give out the answer: "We want to know everything about our competitors, their history, products, employees, managers, and so on." Are these information only Web pages? The answer is definitely "no". Web pages are only the media that contain the needed in-formation, but note they are NOT competitive intelligence. The CIS is expected to produce competitive intelligence about competitors or competition environment from a large set of Web pages, but not just deliver the Web pages or the text blocks in them. This means we should transfer the Web-page-based viewpoint into an entity-based viewpoint. In other words, the CIS should deliver competitive intelligence about the entities such as the competitors (or sub-entities such as the products of a specific competitor), rather than just deliver the Web pages that surly contain the basic information.

B. Ontology

Although there are no standards to construct a domain ontology, it has been widely accepted that constructing an ontology should obey some methodology.

Gruber presented five rules of constructing an ontology in 1995 (Gruber, 1995), which are:

(1) Clearness and Objectivity. An ontology should describe the meanings of terms clearly, and the definitions of terms should be objective and independent on some specific background.

(2) Consistence. The concepts inducted from an ontology should be consistent with the terms included in the ontology.

(3) Extensibility. Nothing is needed to be revised when new concepts are added into an ontology.

(4) Minimal Deviation of Representation. An ontology should not depend on some specific representing method, i.e., we can use different representing methods to depict an ontology while keeping the meanings of the ontology unchanged.

(5) Minimal Constraints. The constraints on an ontology should be minimized. If an ontology is able to represent the requirements on knowledge sharing, we should use the minimal constraints in modeling the concepts and relationships in the ontology.

Other researchers also proposed some advanced rules. However, no rules have been accepted as a standard in the research on ontology construction. Since different ontologies aim at solving different problems of diverse

applications, we can define some basic rules in ontology construction, and refine and explain the details of those rules.

In order to solve the problems in ontology construction, many researchers used ontology engineering methods to develop different ontologies. For example, M. Uschold and King suggested the Skeletal Approach in 1996 (Uschold et al., 1995), Gruninger et al. presented the TOVE method to model enterprises (Gruninge and Fox, 1995), and Gaily et al. proposed a new representation method for the REA ontology (Gailly and Poels, 2008). However, most of these methods are towards a specific domain and can not suit the requirements from different application. For instance, the approach proposed in (Gruber, 1995) was used in constructing a news ontology, but it is difficult for one to use it in other domains.

Many methods were used to represent an ontology, including natural language, frame, logical language, and so on. The natural language is usually used in early stages of constructing an ontology. The frame method is effectively when it is used to represent concepts, attributes, and relationships. A concept in the ontology is represented as a frame, in which the attributes of the concept as well as its relationships with other concepts are described by the slots of the frame. The logical language uses predicate logic to describe an ontology.

C.S. Lee proposed a four-layered framework to represent an ontology of news (Lee et al., 2005). This framework consists of four components:

(1) Domain Layer. This layer represents the domain name of the ontology, which is composed with some different categories defined by domain experts.

(2) Category Layer. This layer contains a set of events. Each category may have several events.

(3) Event Layer. This layer contains some concepts which have some relationships among them. An event can be contained in one or more categories.

(4) Extended Concept Layer. This layer contains several concepts and relationships. Each concept can be event concept or object concept, and is represented as a category. Each extended concept contains attributes and operations. Each operation represents some specific action of a concept.

C. Credibility Evaluation of Competitive Intelligence

According to our knowledge, there are few works focused on the credibility evaluation of competitive intelligence. Most of previous related works concentrated on information credibility. Basically, competitive intelligence stems from information. But competitive intelligence credibility is different from information credibility. There is some relationship between those two types of credibility, which is still an unrevealed issue in the research on competitive intelligence.

Information credibility refers to the believability of some information and/or its source (Metzger, 2007). It not only refers to the objective evaluation on information quality and precise, but also refers to the measurement on information source. 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 (Alfarez and Hailes, 1999), 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.

There are some prototypes in Web information credibility evaluation, among which the most famous ones are WISDOM (Akamine et al., 2009) and Honto?Search (Yamamoto and Tanaka, 2009). 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 (Fritch, 2003) and MedPICS (Eysenbach, 2000) 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 (Google, 2005) 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 (Rainie and Hitlin, 2007), which indicates that users' comments is a key factor in the information credibility evaluation.

The basic methods used in Web information credibility evaluation can be divided into four types, which are the Checklist method, the cognitive authority method, the iterative model, and the credibility seal programs. The Checklist 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 (Metzger, 2007). 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. 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. The credibility seal program 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 (Fritch, 2003). However, this method is usually restricted in certain areas due to the huge amount of Web information.

III. A FRAMEWORK FOR THE EXTRACTION AND CREDIBILITY EVALUATION OF WEB COMPETITIVE INTELLIGENCE

According to practical applications, competitive intelligence must be connected to some specific domain. For example, many competitive intelligence softwares provide intelligence analysis towards a given domain. Based on this assumption, we propose a domain-constrained system to extract and evaluate Web competitive intelligence (as shown in Fig.1). The system consists of four modules, which are the competitive intelligence extraction module, the credibility evaluation module, the query processing module, and the user interface. The basic running process of the system is as follows. First we use a focused crawler to collect Web pages about a specific application domain. Then we extract competitive intelligence from Web pages according to some rules, the competitive intelligence ontology, and the domain dictionary. After that, the credibility of the extracted competitive intelligence is evaluated based on a defined social-network-based credibility model. When users submit queries about competitive intelligence through the user interface, the query processing model will retrieve appropriate results from the competitive intelligence database.

The main function of the competitive intelligence extraction module is to extract domain-constrained competitive intelligence from Web pages and further to deliver them to the credibility evaluation module. We use an entity-based approach in this module to extract competitive intelligence, which will be discussed in the next section.

The credibility evaluation module adopts the social-network-based method to evaluate competitive intelligence credibility. The credibility of competitive intelligence is influenced by a lot of factors. These factors are classified into two types in our paper, which are the inner-site factors and inter-site factors. Then we use different algorithms to evaluate the competitive intelligence credibility according each type of factors. Finally we will integrate the both results and make a comprehensive evaluation on the competitive intelligence credibility.

The user interface supports keyword-based queries on competitive intelligence. Users are allowed to input topics, time, or locations as query conditions.

The query processing module aims at returning competitive intelligence related with given topics or other conditions. Competitive intelligence workers can further process the returned results and produce integrated competitive intelligence. This module contains two procedures. The first one is a database retrieval procedure, and the second is clustered visualization of the results. The system provides several ways of clustered visualization, including time-based clustering, location-based clustering, and topic-based clustering.

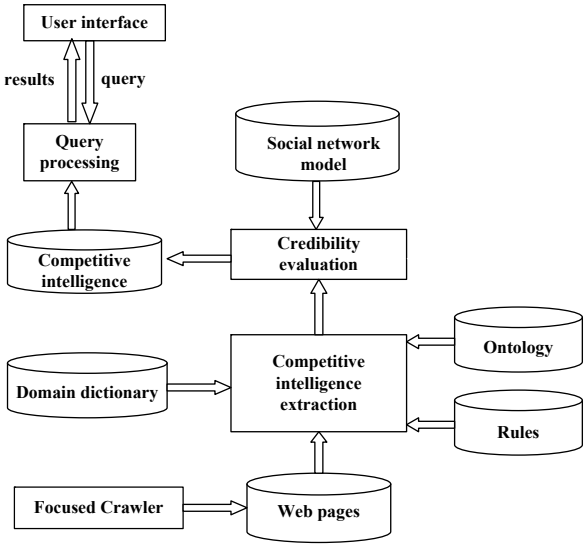

Figure 1. The system framework for the extraction and credibility evaluation of Web competitive intelligence

IV. ONTOLOGICAL FOUNDATION FOR WEB COMPETITIVE INTELLIGENCE

The ontology of Web competitive intelligence provides a systematic definition on the concepts and composition of Web competitive intelligence. It aims at answering the following questions: (a) What types of competitive intelligence is hided in the Web? (b) What types of competitive intelligence is possible to be extracted from the Web?. After studying the characteristics of Web pages, we define a layered ontology for Web competitive intelligence.

As shown in Fig.2, Web competitive intelligence contains two types of entities, competitor and competition environment, each of which consists of some sub-layered entities. The competitor intelligences is composed with three sub-entities, which are profile, events, and business relations (as shown in Table I). The profile intelligence is the general information about competitor. Many web-sites such as Wikipedia (http://www.wikipedia.org)

provide some general information about companies, such as names, employee counts, managers' names. Events about competitor usually refer to the news about it. Many websites provide news which is updated frequently. Through the events expressed in the news, people are able to know the recent development of the competitors. Typical events are the establishment of the competitor, the listed-in-stock of the competitor, the progress of some specific project, etc. Compared with profile and events, the business relations are usually more implicit. This is because most companies do not want that the competitors know their suppliers or customers. However, this type of competitive intelligence may be more useful than others. For example, if you know exactly the suppliers of your competitor, you may have some countermeasures to control those suppliers so as to leave the competitor in a passive situation. To obtain the business relations about competitor, we must per-form an intelligent analysis on the contents of Web pages.

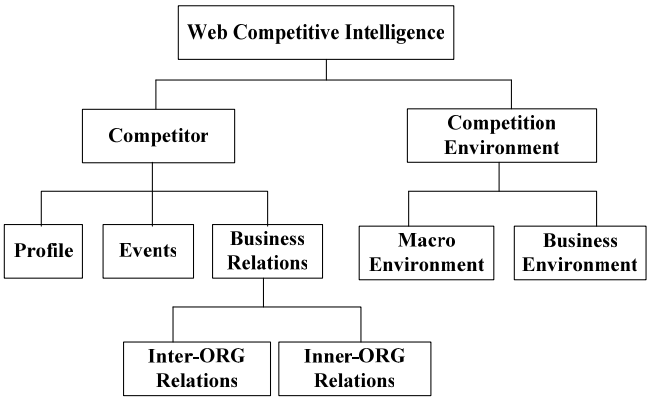

Figure 2. The layered ontology of Web competitive intelligence

TABLE I. Details about competitor intelligence in the Web

Type	Description
Profile Intelligence	This type of intelligence refers to the basic information about competitor, e.g. company name, telephone number, address, products set, managers' names, etc.
Events Intelligence	This type of intelligence refers to the news usually co-rellated with time and location, e.g. establishment of the company, release of new products, staff reduction, Being listed stock, etc.
Business Relations Intelligence	This type of intelligence refers to the business relations between competitor and other companies, e.g. suppliers of the company, investors, customers served, etc.

Table II shows the description of competition environment.

TABLE II. Details about competition environment in the Web

Type	Description
Macro Environment	The environmental information about the politics, economics, cultures, laws, society, science & technologies, and nature.
Business Environment	The business or industry state that a company belongs to, e.g., the companies set, suppliers, customers, alternatives of products, potential new comers in a specific area.

V. ENTITY-BASED EXTRACTION OF WEB COMPETITIVE INTELLIGENCE

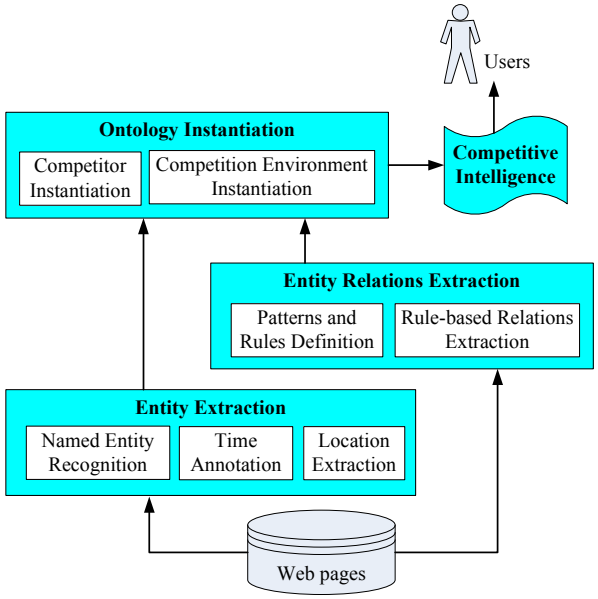

Figure 3. Architecture of Entity-based Extraction of Web competitive intelligence

Fig.3 shows the architecture of entity-based extraction of Web competitive intelligence. Our approach is based on an entity extraction step, an entity relations extraction step, and an ontology instantiation step. The entity extraction step is used to detect different types of concerned entities in Web pages, including named entities, time entities, and location entities. The entity relations extraction step is then used to build relations among entities according to some patterns and rules. After that we will get some facts like "Kaifu Li is CEO of Google" or "Microsoft is located in Redmond". Basic relations among entities are inner-ORG relations and Inter-ORG relations, as shown in Fig.2. The ontology instantiation step will generate Web competitor intelligence and competition environment intelligence based on the predefined ontology and deliver the result to users.

A. Entity Extraction

The entity extraction module uses three approaches to detect concerned entities from Web pages, which are

named entity recognition, time annotation, and location extraction.

Named-entity recognition is a hot research field in Web information extraction and retrieval (Whitelaw et al., 2008). It was first introduced as a sub-task in the MUC (Message Under-standing Conference) conference (Sundheim, 1995). Its main task is to recognize and classify the specific names and meaningful numeric words from the given texts. Typical named entities are company names, person names, addresses, times, etc. Most of the previous research in this field focused on three types of named entities: time entities, number entities, and organization entities (Khalid et al., 2008). In our system, several types of named-entities are needed to be studied. However, different methods are also required for different named-entities. For example, we use a hierarchy method to recognize the addresses from Web pages, i.e. "China [*country*] → Beijing [*city*] → Chaoyang District [*district*] → Peace Road [*street*] No.128 [*number*]", while for email extraction we use another approaches such as pattern matching, i.e. strings like "[*strings*]@[*strings*].[*strings*]".

Time annotation stems from the traditional research on natural language processing (NLP) (Wong et al., 2005)]. Although there is some previous work on the time annotation on text, rare work has been done for that on Web pages. The time annotation on text is usually based on the two standards: TIMEX2 (http://timex2.mitre.org) and TimeML (http://www.timeml.org/site/index.html). The most important difference between Web page and text is that a Web page has some tags. So in this paper, we first eliminate the tags in a Web page, and then apply the traditional time annotation approaches to obtain the time information in the Web page.

For the location detection in a Web page, we conduct a pattern-based approach to detect location information in a Web page. First, we construct a hierarchy dictionary to store the hierarchy location names according to their geographical relationships, e.g. "[*country*]→[*city*] → [*district*] → [*street*] → [*number*]". We have found that there are common patterns in the expression of locations and addresses. Thus a set of patterns are defined and then used to detect locations from Web pages.

B. Entity Relations Extraction

Business relations are very important for companies. Generally, there are several types of business relations. The ACE (Automatic Content Extraction) has defined six types of relations in English texts (ACE, 2008). However, those relations are not defined for competitive intelligence. In this paper, we classify the entity relations into two types: Inner-ORG relations and Inter-ORG relations. The Inner-ORG (ORG is the abbreviation of the word "organization") relations refer to the business relations between a company and its components, e.g. company-manager, company-employee, and so on. The Inter-ORG relations are relations among different companies. Examples of the Inter-ORG relations are company-investor, company-supplier, company-partner, etc.

VI. SOCIAL-NETWORK-BASED CREDIBILITY EVALUATION OF WEB COMPETITIVE INTELLIGENCE

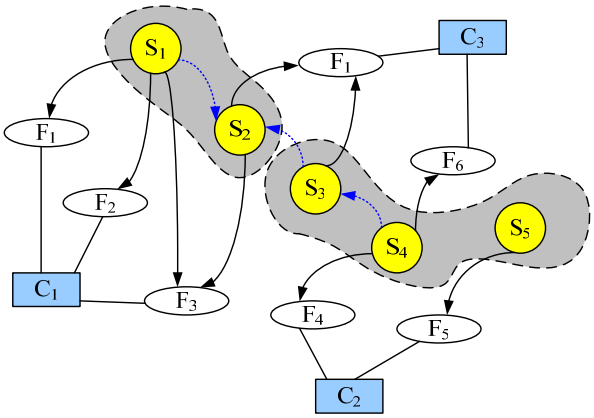

Figure 4. Social-network-based credibility evaluation of Web competitive intelligence

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 Web competitive intelligence.

Fig.4 shows the basic idea of the social-network-based evaluation model. The model consists of three types of nodes, i.e., S node, F node, and C node. The C node is represented as a rectangle in Fig.4. It represents a specific entity of competitive intelligence, which is from the layered ontology model in Fig.2. The F node is represented as an ellipse, which indicates a fact. The circle S node represents Web sites. 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 comes from a certain website. The dash directed edge between two S-type nodes represents the linking relationship between two websites. The solid undirected edge between a F-type node and a C-type node represents that a competitive intelligence element is composed with some facts.

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

One of the key issues in Fig.4 is how to find social networks related with some given fact. In our system, we focus on four types of social communication networks

and construct social network finding algorithms respectively.

(a) BBS system: We can build a graph model to modeling the BBS nature. The users in a BBS system can be looked as nodes, the replies to a post can be regarded as edges, and the weight of each edge can be determined by the number of replies. After that, we can use some statistical model to find the social network related with some topic in a BBS system.

(b) Online friends system: Typical online friends system such as Facebook provides a list of friends for each user. Thus we can utilize the friends list to find the social network.

(c) Online social community: Online social community usually provides a lot of services including personal homepage, blog, BBS, et al. Since each user in an online social community has their own websites, we can use the linking relationships in user's website as well as the IP geographic information to find social network.

(d) Functional community: Functional community contains the websites that have special functions, such as Flickr, Digg, YouTube, CiteULike. Those websites are usually designed for some special purpose, e.g., to collect videos, pictures, music, or digests. Users in such community usually have a friend list and some bookmarks, which can be used to find social networks.

VII. Conclusions and Future Work

In this paper, we briefly analyzed the state-of-the-art of Web-based competitive intelligence, especially of the extraction and credibility evaluation issues of Web-based competitive intelligence. Based on the analysis, we presented a system framework to first extract competitive intelligence from Web and then evaluate competitive intelligence credibility to produce credible competitive intelligence from users. For the extraction of Web competitive intelligence, we investigated an entity-based approach and discussed the detailed procedures to extract entities and relations from Web pages, as well as the ontology instantiation issue. The extraction is based on predefined layered ontology for Web competitive intelligence. For the credibility evaluation of Web competitive intelligence, we propounded a social-network-based method. As social network becomes more and more popular in both Web and real lives, it is helpful to combine the nature of social network into the credibility evaluation of Web competitive intelligence. We constructed a social-network-based credibility model in the paper, and discussed some related issues, e.g., how to find social network in Web.

However, the paper has not considered the implementation details and experiments. In our future plan, we will concentrate on the implementation of the system discussed in this paper, and conduct comprehensive experiments on real Web dataset to demonstrate the effectiveness and performance of our design.

Acknowledgement

This work is supported by the Open Projects Program of National Laboratory of Pattern Recognition, the Key Laboratory of Advanced Information Science and Network Technology of Beijing, and the National Science Foundation of China under the grant no. 60776801 and 70803001.

References

[1] ACE (Automatic Content Extraction) English Annotation Guidelines for Relations, Version 6.2, (2008) *Linguistic Data Consortium*, In: http://www.ldc.upenn.edu/Projects/ACE/

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

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

[4] Deng, D., Luo, L., (2007) An Exploratory Discuss of New Ways for Competitive Intelligence on WEB 2.0, In W. Wang (ed.), *Proc. of IFIP International Conference on e-Business, e-Services, and e-Society*, Volume 2, Springer, Boston, pp. 597-604

[5] DOM, http://www.w3.org/DOM/

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

[7] Fritch, J.W., (2003) Heuristics, Tools, and Systems for Evaluating Internet Information: Helping Users Assess a Tangled Web, *Online Information Review*, Vol.27(5), pp. 321-327

[8] Gailly, F., Poels, G., (2008) Ontology-Driven Business Modeling: Improving the Conceptual Representation of the REA Ontology, *Proc. of ER*, LNCS 4801, pp. 407-422

[9] Google, (2005) Google News Patent Application, In: http://www.webpronews.com/news/ebusinessnews/wpn-45-20050503GoogleNewsPatentApplicationFullText.html. Accessed in 2010.3

[10] Gruber, T., (1995) Towards principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, (43), pp.907–928

[11] Gruninge, M., Fox, M., (1995) *The Logic of Enterprise Modeling, Modeling and Methodologies for Enterprise Integration*. Bernus, P.; Nemes , L. (Eds.), Cornwall, Great Britain: Chapman and Hall, pp. 83-85

[12] Hotho, A., Nürnberger, A., Paass, G., (2005) A Brief Survey of Text Mining. *LDV Forum* (LDVF), Vol.20(1), pp.19-62

[13] Kahaner, L., (1996) *Competitive Intelligence*, New York: Simon & Schuster

[14] Khalid, M., Jijkoun, V., Rijke, M., (2008) The Impact of Named Entity Normalization on Information Retrieval for Question Answering, In *Proc. of ECIR'08*, pp.705-710

[15] Khoury, I., El-Mawas, R., El-Rawas, O., et al., (2007) An Efficient Web Page Change Detection System Based on an Optimized Hungarian Algorithm. *IEEE Transaction on Knowledge Data Engineering* (TKDE), Vol.19(5), pp.99-613

[16] LaMar, J., (2007) Competitive Intelligence Survey Report, In: http://joshlamar.com/documents/CIT Survey Report.pdf

[17] Lee, C., Chen, Y., Jian, Z., (2005) Ontology-based fuzzy event extraction agent for Chinese-news summarizations, *Expert Systems with Applications*. (25), pp. 431-447

[18] Li, J., Huang, M., Zhu, X., (2006) An Ontology-Based Mining System for Competitive Intelligence in Neuroscience, In N. Zhong et al. (ed.), *Proc. Of WimBI 200*6, LNCS 4845, Springer, Heidelberg, pp. 291-304

[19] Metzger, M., (2007) 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(13), pp. 2078-2091

[20] Mikroyannidis, A., Theodoulidis, B., Persidis, A., (2006) PARMENIDES: Towards Business Intelligence Discovery from Web Data, In *Proc. Of WI*, pp.1057-1060

[21] Rainie, L., Hitlin, P., (2007) The Use of Online Reputation and Rating Systems, In: http://www.pewinternet.org/ PPF/r/140/report_display.asp, Accessed in 2010.3

[22] Sun, B., Mitra, P., Giles, C., et al., (2007) Topic segmentation with shared topic detection and alignment of multiple documents. In *Proc. of SIGIR*, pp. 199-206

[23] Sundheim, M., (1995) Named Entity Task Definition-Version 2.1, In *Proc. Of MUC*, pp.319-332

[24] Thompson, S., Wing, C., (2001) Assessing the Impact of Using the Internet for Competitive Intel-ligence, *Information & Management*, Vol.39(1), pp.67-83

[25] TimeML. In http://www.timeml.org/site/index.html

[26] TIMEX2. In http://timex2.mitre.org/

[27] Uschold, M., et al., (1996) Ontologies: Principles, Approaches and Applications, *Knowledge Engineering Review*, (11), 93- 155

[28] Whitelaw, C., Kehlenbeck, A., Petrovic, N., et al., (2008) Web-scale Named Entity Recognition, In *Proc. of CIKM'08*, pp.123-132

[29] Wikipedia, In http://www.wikipedia.com

[30] Wong, K., Xia, Y., Li, W., et al., (2005) An Overview of Temporal Information Extraction, *Int. Journal of Computer Processing of Oriental Languages*, vol.18(2), pp.137-152

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

[32] Zhao, J., and Jin, P., (2010) A Framework for Credibility Evaluation of Web-Based Competitive Intelligence, *Proc. Of IITSI'10*, Finland Academic Publisher, Jingangshan, China

[33] Zhao, J., and Jin, P., (2009) Towards the Extraction of Intelligence about Competitor from the Web, *Proc. Of WSKS'09*, LNAI 5736, Chania, Greece, pp.118-127

Jie Zhao was born in Hefei, China, in October, 1974. She received her master degree in management science and engineering from Hefei University of Technology, Hefei, China, in 2003. Now she is a Ph.D. candidate in School of Management, University of Science and Technology of China. She is currently an associate professor in School of Business Administration, Anhui University, China. Her research interests include Web intelligence, information retrieval, and information extraction. Assoc. Prof. Zhao is a member of SCIP (Society of Competitive Intelligence Professionals), and serves as a PC member of several international conferences, such as ICCIT'09, NCM'09, and IMS'09.

Peiquan Jin was born in Zhejiang, China, in August, 1975. He received his Ph.D. degree in computer science from University of Science and Technology of China, Hefei, China, in 2003. Before that, he received his master and bachelor degree in management science both from Tianjin University of Finance and Economics, Tianjin, China, in 2000 and 1997, respectively. He is currently an associate professor in School of Computer Science and Technology, University of Science and Technology of China, China. His research interests include Web intelligence, knowledge management, and databases. Dr. Jin is a member of ACM, ACM SIGMOD, IEEE, and IEEE ComSoc, and is an editor of International Journal of Advancements in Computing Technology, Journal of Convergence Information Technology, and International Journal of Digital Content Technology and its Applications. He serves as a PC member of many international conferences, including DEXA'09-10, NCM'09-08, ICCIT'09-08, NISS'09, and NDBC'09.