Business Relations in the Web: Semantics and a Case Study

Jie Zhao
School of Management, University of Science and Technology of China, Hefei, China
School of Business Administration, Anhui University, 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

Yanhong Liu
School of Computer Science and Technology, University of Science and Technology of China, Hefei, China

***Abstract*—Web has been one of major sources to acquire competitor intelligence. In this paper, we first present a framework to acquire competitor intelligence from the Web, which consists of profile extraction, events extraction and business relations extraction. Then we investigate the semantics of business relations in detail. A classification of business relations is presented, based on which a conceptual ontology for business relations is proposed. Finally, a case study of extracting business relations from Web pages is studied. We focus on the extraction of position relations from the Web. A structure-based approach is used to recognize the position relations hiding in Web pages. The basic idea as well as the detailed procedures is discussed in the paper. We also conduct an experiment to extract position relations from Web pages. The experimental results show that our approach is effective in the extraction of position relations.**

***Index Terms*—Competitive intelligence; Web; Business relations**

I. INTRODUCTION

How to extract competitor intelligence from the Web has become a hot issue in recent years (Kahaner et al., 1996). Many firms begin to realize that most information about competitors can be found in the Web, and therefore it is possible to build a software system to automatically acquire, analysis, and generate user-defined competitor intelligence from the Web. A previous survey showed that about 90% of competitive intelligence can be acquired from the Web (Thompson and Wing, 2001; Lamar, 2007). This provides enterprises the opportunities to gain competitive values from the Web, in case that they can build a competitive intelligence system extracting intelligences from the Web effectively.

As most enterprises are interested in the competitor intelligence, we concentrate on competitor intelligence extraction issues in this paper. In particular, we will focus on the business relations of a competitor. Business relations are one of important aspects in competitor intelligences. They play important roles in the analysis of competitor intelligence and decision making procedures. Compared with other features in competitor intelligence, such as competitor profiles, business relations are more difficult to be extracted from the Web, since they are usually not explicitly expressed in Web pages.

Another problem of extracting business relations from the Web is what types of business relations we want to extract from the Web. Hence, it is necessary to first study the semantics of business relations and further to construct an ontology for the business relations in competitor intelligence.

This paper mainly discusses semantics and extraction of business relations extraction. It is an extended version of our previous work in the Global Conference on Science and Engineering (GCSE'09) (Zhao and Jin, 2009). We will first present a framework of extracting competitor intelligence from the Web, and then discuss the semantics of business relations. After that, an ontology of business relations is presented. Finally, we will present a case study of extracting business relations, which concentrates on the position relations extraction. The main contributions of the paper can be summarized as follows:

(1) We present a Web-based framework of extracting competitor intelligence. The major components of such a system are analyzed (see Section 3).

(2) The semantics of business relations are studied, and a formal classification on business relations as well as an ontology is presented (see Section 4).

(3) We present a case study to extract business relations from Web pages (see Section 5). Detailed algorithms are developed to realize the case study. And we also conduct an experiment on real Web pages to evaluate the performance of our approach. The experimental results show that our method is effective to extract position relations from the Web.

The following of the paper is structured as follows. In Section 2 we discuss the related work. Section 3 discusses the framework of Web-based competitor intelligence extraction. Section 4 gives the discussion

© 2010 ACADEMY PUBLISHER
doi:10.4304/jsw.5.8.826-833

about the semantics and ontology of business relations. Section 5 presents a case study of extracting business relations. And conclusions and future work are in the Section 6.

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

In the context of information science, ontology usually refers to a set of general items in a specific domain, as well as the relationships among those items (Gruber, 1995; Uschold et al., 1996)]. An ontology for Web-based enterprise competitive intelligence can serve as the foundation of acquiring and representing competitive intelligence in the Web, because it is necessary to make it clear what types of competitive intelligence we can obtain from the Web, and what details of those competitive intelligence we can extract (Li et al., 2006).

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

© 2010 ACADEMY PUBLISHER

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. 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.

III. A FRAMEWORK FOR COMPETITOR INTELLIGENCE EXTRACTION FROM THE WEB

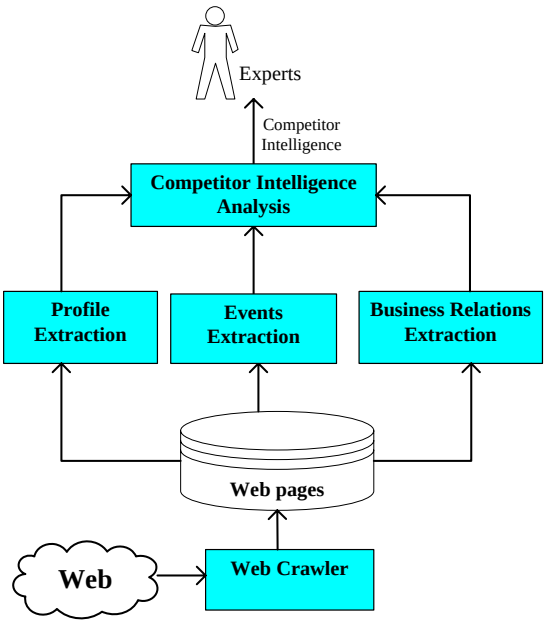

Figure 1. Architecture of competitor intelligence extraction from the Web

The architecture of competitor intelligence extraction from the Web is shown in Fig.1. There are five modules in the system. The Web Crawler module performs the traditional tasks of a spider. It collects different Web pages from the Web. These Web pages will be further processed by other three modules, as shown in Fig.1. The Profile Extraction module is designed to extract some basic information about a company, such as name, address, managers, employees, telephone numbers, and so on. This type of information is called *profile* of a company in this paper. Profile is usually easier to be extracted from the Web, compared with the other two types of information, event and business relations, because many websites, such as Yellow Pages and Wikipedia, offer detailed information about a company. The Event Extraction module acquires events related with a given company. An event represents a specific activity which is related to the interested company. Typical events are creation of a company, bankrupt, being stock listed, and so on. Events are often hided in the content of a Web page, or must be extracted through a lot of Web pages. The Business Relations Extraction module extracts business relations related to a given company. Since there are many types of business relations in the real world, we have to first classify the types of business relations and then conduct effective ways to extract business relations from Web pages. In this paper, we will concentrate on the business relations. In particular, we will study the semantics of business relations and build an ontology for business relations, which forms the foundation of the Web-based competitor intelligence extracting system.

Based on a systematic view, we give out the following description of the requirements on competitor extraction from the Web (see Table I).

TABLE I. Different aspects of competitor intelligence in the Web

Type	Description
Profile	Basic information about competitor, e.g. company name, telephone number, address, products set, managers' names, etc.
Events	Events related with competitors. A typical event consists of a topic, a location, and a time element. Examples of events are establishment of a new company, release of new products, staff reduction, Being listed stock, etc.
Business Relations	Relations between a competitor and its internal employees or other objects, or relations between a competitor and other companies, e.g. suppliers of the company, investors, customers served, etc.

A. *Profile*

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, etc. Fig.2 shows the extracted general information of the TOSHIBA Corporation.

B. *Events*

Events about competitor usually refer to the news about it. Many websites provide news which is updated

© 2010 ACADEMY PUBLISHER

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. Fig.3 shows some recent events about IBM.

C. Business Relations

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. For example, from the Web page shown in Fig.4, we get to know that IBM has 98 partners in Franfurt, Germany.

TOSHIBA

Type	Corporation TYO: 6502, (LSE: TOS)
Founded	1939 (merger)
Headquarters	Minato, Tokyo, Japan
Key people	Hisashige Tanaka, Founder of Shibaura Engineering Works Ichisuke Fujioka and Shoichi Miyoshi, Founders of Tokyo Electric Atsutoshi Nishida, CEO
Products	Digital products, Electronic devices & components, Social Infrastructure, Home appliances, and others
Revenue	▼¥6,654 billion (Fiscal year ended March 31, 2009)[1]
Net income	▼¥-343.6 billion (Fiscal year ended March 31, 2009)
Website	Toshiba Worldwide

Figure 2. Example of profile extracted from the Web

All events

Event name	Start date	City	State	Technology Sponsor	Industry
Waters USA 2009	12/07/2009	New York	New York	Software Group	Cross
Cohesive Information Integration) Blending ETL, Data Quality, MDM to Unify the Enterprise Information View	12/08/2009	Online	Online	Software Group	Cross
IBM Cognos Performance 2009 - Dallas	12/08/2009	Fort Worth	Texas	Software Group	Cross
IBM Pulse Comes to You Virtual Forum	12/08/2009	Online	Online	Software Group	Cross
NYC Metro Oracle Users Group Day	12/08/2009	New York	New York	Software Group	Cross

Figure 3. Example of events in the Web

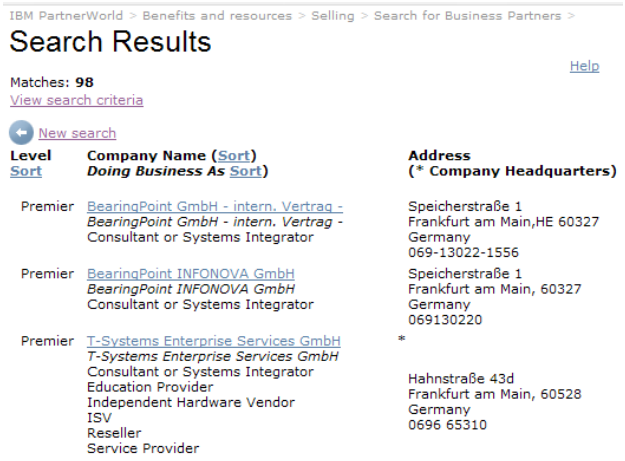

Figure 4. Example of business relations in the Web

IV. SEMANTICS OF BUSINESS RELATIONS

A. Relations Defined by ACE

TABLE II. The relations defined by ACE

Type	Subtypes
Phisical	Located, Near
Part-Whole	Geographical, Subsidiary
Personal-Social	Business, Family, Lasting-Personal
ORG-Affiliation	Employment, Ownership, Founder, Student-Alum, Sports-Affiliation, Investor-Shareholder, Membership
Agent-Artifact	User-Owner-Inventor-Manufacturer
Gen-Affiliation	Citizen-Resident-Religion-Ethnicity

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), which are listed in Table II. However, those relations are not defined for competitor intelligence. The only interested types in ACE are the Person-Social relation and ORG-Affiliation relation. But these relations are too rough for business relations intelligence extraction.

B. Types of Business Relations

We classify the business 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.

(1) Inner-ORG relations

The Inner-ORG relations refer to the business relations among the entities of the same organization. A lot of information about a company can be extracted from the

© 2010 ACADEMY PUBLISHER

Web, e.g., name, address, email. This task is somehow easy to perform, because many methods have been proposed to extract different named entities (Whitelaw et al., 2008). 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). According to the context of competitor intelligence extraction, several types of named-entities are needed to be studied. However, we can use previous approaches to extract the named entities needed in the extraction of Inner-ORG relations. We further classify the Inner-ORG relations into four types, which are ORG-person relations, ORG-location relations, ORG-time relations, and ORG-statistics relations.

- ORG-person relations

The ORG-person relations refer to the business relation between a company and one of its employees. Due to the fact that there are many types of positions in a company, we have to determine many types of position relations for a company. For example, who is the general manager of Lenovo? Is John an employee of Lenovo? Those relations all involve a person and a company.

- ORG-location relations

The ORG-location relations refer to the business relations between a company and some location. Location information plays a very important role in the decision making process. Enterprises usually make different market policies for different areas or cities. Typical ORG-location relations include the city of a company located and the sales area of a company.

- ORG-time relations

The ORG-time relations are the business relations between a company and a time value. For example, the founding time of a company, the date being stock lised, the bankrupt date of a company, and so on. Different types of time values may exist in this type of relations. For example, the founding time a company may be a calendar day, while the duration of a company locating in a city may be a time period.

- ORG-statistics relations

This type of Inner-ORG relations refer to the relations between a company and some numeric value. For example, how many employees does Lenovo hire? Or what is the total market value of Lenovo?

(2) Inter-ORG relations

The Inter-ORG relations refer to the business relations between two companies. With the development of virtual enterprises and enterprise union, the relationships among different companies become more and more important in the market competition. Therefore, it is very important to recognize the competitors' business relations with other companies. Typical Inter-ORG relations are the relations among the companies who are contained in the same supply chain. For example, who are the suppliers of Lenovo?

We classify the Inter-ORG relation into four types of relations, which are cooperation relation, invest relation, sales relation, and supply relation.

- Cooperation relations

The cooperation relations refer to the contracted cooperation between two companies. Normally, these types of relations appear when two companies are working together for the same project.

- Invest relations

The invest relations usually exist between a stock listed company and another organization or person who has its stocks. Many companies will buy some stocks of other companies as a future investment.

- Sales relations

The sales relations refer to the customers of a company. Customers may be persons or other companies. So this type of relation indicates the users of a company.

- Supply relations

The supply relations give the suppliers of a company. For example, who are suppliers of KFC in China?

C. An Ontology for Business Relations

Based on the semantics we analyzed in Section 4, we formally construct an ontology for business relations. Such an ontology describes the concepts in business relations, as well as the relationships between different concepts. In this paper, we use the UML-model to formally describe the ontology. Fig.2 shows the UML-model-based ontology for business relations.

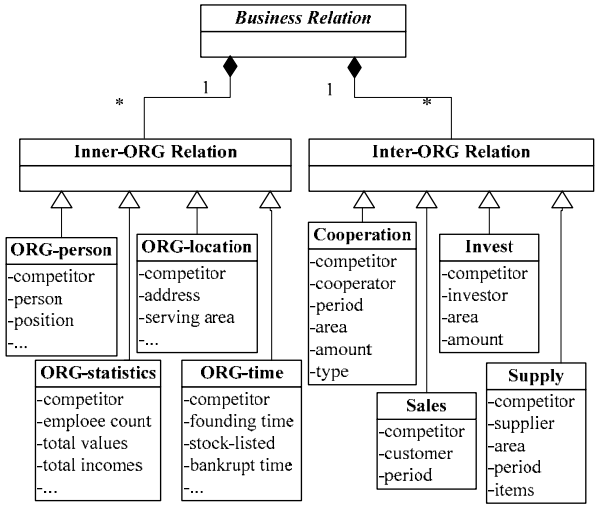

Figure 5. The ontology for business relations

V. A CASE STUDY: EXTRACTING POSITION RELATIONS FROM THE WEB

In this section, we discuss the extraction of position relations from the Web.

© 2010 ACADEMY PUBLISHER

A. *Extracting Position Relations*

Position relation describes the fact that a person holds a position in a specific organization. A position relation typically contains three elements, which can be formalized as a triple {O, P, R}, where O, P and R stand respectively for organization name, position name and person name. Position relation extraction aims at obtaining such position triples from natural language text or Web pages.

The detailed algorithm to extract position relations from Web pages is illustrated in Fig.6. We first determine the structural parts of a Web page, and then focus on these parts and extract position relations using some templates.

To extract structured file segments, we first find some structural sentences in Web pages. A position relation candidate consists of a person name, a position name, and a separator. By defining a static set of position names and separators, we are able to find all the structural sentences which may contain position relation candidates. For example, if we have already defined a position name called "General Manager" and a separator ":", then we can determine all the position relation candidates from Web pages, which have the form as "XXX : General Manager", where "XXX" is a person name. Table III shows the separators used in our algorithm.

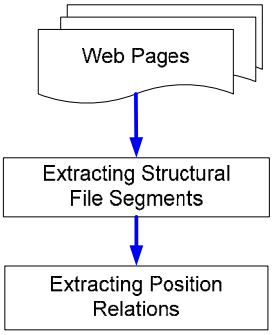

Figure 6. Extracting position relations from Web pages

TABLE III. Some general separators in Web pages

Symbol	Meaning	Example
	blank	杨宁(*Yang Ning*) 空中网总裁(*the president of Kong Zhong Website*)
:	colon	千橡集团总裁(the president of Qian Xiang Group)：陈一舟(Chen Yizhou)
—	dash	大贺集团董事长(the board chairman of Da He Group)—贺超兵(He Chaobin)
<td> ...</td>	column tag	**<td>**中华广告网董事长(the board chairman of Zhong Hua Advertisement Website)**</td><td>**姜杉(Jiang Bing)**</td>**
(...)	parenthesis	雷军(Lei Jun)（金山总裁(the president of Jin Shan)）

After determining the structural file segments in a Web page, we then concentrate on these structural file segments to extract position relations. In this process, we first add tags to the file segments, and then we generate position relation candidates.

(1) *Tagging of Person Names*

Here we use ICTCLAS (ICTCLAS, 2009) as the tool for the lexical segmentation and tagging of person names. However, the ICICLAS tool can not recognize some special person names in Web pages, such as "杨(Yang)皓(Hao)", where there is a blank character between surname and given name. In this paper, we add two rules into the ICTCLAS to tackle with these special cases and call this process as person name complement.

Rule 1: After lexical segmentation and POS tagging, if there is a substring looking like "A/*nr1* # B/*x*" or "A/*nr1* B/*x* - ", "A" and "B" will be combined and tagged "/*r*" to get "AB/*r*"

Rule 2: After lexical segmentation and POS tagging, if there is a substring looking like "A/*nr1* BC/*x* - " or "A/*nr1* B/*x* C/*x* - ", "A", "B" and "C" will be combined and tagged "/***r***" to get "ABC/***r***".

In the rules, "A", "B" or "C" denotes one Chinese character respectively, the tag "*nr1*" denotes "A" is a Chinese surname. The tag "*x*" represents any POS tag. The symbol "#" denotes a blank character while " - " represents separators we described above. The tag "*r*" denotes a complete Chinese person name.

For example, the substring " 杨(Yang)/*nr1* # 皓(Hao)/*ng*" will be converted to " 杨皓(Yang Hao)/*r*" according to **Rule 1** and the substring "蒋(Jiang)/*nr1* 天龙(Tianlong)/*nz* - " will be converted to "蒋天龙(Jiang Tianlong)/***r*** - " according to **Rule 2**.

(2) *Tagging of Position Names*

Position name tagging is conducted through a position dictionary which is constructed manually. Note that the dictionary only contains simplified position names. A simplified position name only contains core words of a complete position name. For example, "总裁(president)" is simplified position name while " 副总裁(vice president)" is not a simplified position name. The tag of a position name is "/***p***".

(3) *Tagging of Potential Organization Names*

A potential organization name in a sentence is tagged by the *OFW* (*organization feature word*). If a sentence contains a person name, a position name and an *OFW* simultaneously and these elements appear in a specific structure, then it is much possible that there is an organization name in the *OFW* position. We first assume there is an organization name in the *OFW* position and then filter out the sentences which contain an illegal organization name in the subsequent stages. In this paper, we tag two most frequent *OFW*: "公司(corporation)" and

© 2010 ACADEMY PUBLISHER

"集团(group)", both with the tag "/***o***". Table IV shows a summary about the tagging symbols used in our algorithm, and Fig.3 shows an example of the position relation candidate generated.

TABLE IV. Tagging symbols in position relation candidates

Symbol	Meaning
/***r***	a person name
/***p***	a position name
/***o***	an *OFW* (organization feature word) tag, i.e., a potential organization name

Sentence: 天极公司总裁(the president of Tian Ji corporation) 一李志高(Li Zhigao)
Position Relation Candidate: 天极(Tian Ji) 公司(corporation)/***o*** 总 裁 (president)/***p*** 一 李 志 高 (Li Zhigao)/***r***

Figure 7. An example of position relation candidate

B. Experimental Results

The Web pages in the experiment are downloaded from famous Chinese search engine Baidu. The keywords look like "position name + Chinese surname" such as "总裁(president)+张(Zhang)丨王(Wang)丨李(Li)丨赵(Zhao)丨刘(Liu)". The method increases the probability that a Web page contains position relation instances. The Web pages are amount to 6028 and we choose five kinds of position relations (president, manager, engineer, CEO, board chairman) to conduct experiments. The position name dictionary contains 66 simplified Chinese position names which are prepared manually.

Table V shows the extracting results of five kinds of position relations over the 1425 structural file segments. The average recall is over 87%, whereas the precision of our approach is much high. The reason why our approach gains high precision is that our approach is based on structural feature of position relations on Web pages.

TABLE V. The experiment result of five position relations

	Recalled number	*Recall*	*Precision*	*F-measure*
总裁 (president)	4991	88.2%	97.0%	95.1%
经理 (manager)	12490	84.3%	95.1%	92.7%
工程师 (engineer)	2547	88.1%	90.1%	89.7%
董事长(board chairman)	9027	90.1%	98.7%	96.9%
首席执行官 (CEO)	613	84.8%	95.4%	93.1%
Sum/Average	29668	87.2%	96.1%	94.2%

VI. CONCLUSIONS

Web has played important roles in competitor intelligence systems. In this paper we present a framework of extracting competitor intelligence from the Web and further develop an ontology to represent the semantics of business relations about a competitor. We studied the classification of business relations, and further develop an effective approach to extract position relations from Web pages.

Our future work will concentrate on the design and implementation of the algorithms to extract other types of business relations and conduct experiments to demonstrate the performance of the algorithms and the whole system.

ACKNOWLEDGEMENT

This work was supported in part by the National Natural Science Foundation of China under the grant no. 70803001 and 60776801, and the Science Research Fund of MOE-Microsoft Key Laboratory of Multimedia Computing and Communication (grant no. 06120804).

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

[3] 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

[4] 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

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

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

[7] 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

[8] 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

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

[10] 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 2006*, LNCS 4845, Springer, Heidelberg, pp. 291-304

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

[12] 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

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

© 2010 ACADEMY PUBLISHER

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

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

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

[17] ICTCLAS, Available at http://www.ictclas.org (accessed in May 18, 2009)

[18] Jie Zhao, Peiquan Jin, Extracting Business Relations from the Web: An Ontological Foundation, in *Proc. Of Global Conference on Science and Engineering (GCSE)*, 2009

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.

© 2010 ACADEMY PUBLISHER