Conceptual Modeling for Competitive Intelligence Hiding in the Internet

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

***Abstract*—Internet has been a major source to acquire enterprise competitive intelligence. But what types of competitive intelligence are hidden in the Internet? This paper presents an answer to this question. In this article, a conceptual framework to represent the competitive intelligence in the Internet is introduced. The basic idea is to use an entity-oriented method to capture the semantics of the competitive intelligence in the Internet. Traditional approaches regard the competitive intelligence in the Internet as a set of Web pages, which leads to a large amount of manual processing work. Oppositely, we represent them as a set of entities, and further design two types of entities, namely the entities about competitors and those about competition environment. Moreover, the detailed semantics of each entity are discussed. We develop formal approaches to represent these types of competitive intelligence. As a result, a conceptual model for the competitive intelligence in the Internet is proposed, which gives an explicit description for the competitive intelligence in the Internet. Finally, a case study to demonstrate the feasibility of our conceptual model is explained. The results show that it is feasible to use the proposed conceptual model to capture the semantics of the competitive intelligence in the Internet.**

***Index Terms*—competitive intelligence; Internet; conceptual model; entity-based representation**

I. INTRODUCTION

Nowadays, enterprise competitive intelligence has become one of the major factors which determine the core competitive power of companies in the worldwide market [1]. How to efficiently and effectively obtain competitive intelligence is an urgent issue to most companies. With the development and applications of Web technologies, it is possible to obtain a large amount of competitive intelligence from the Web. Previous survey shows that about 90% of competitive intelligence can be acquired from the Internet [2]. This is a challenge and also a chance for enterprises to enhance their competition powers. If we can build a system to acquire competitive intelligence from the Web automatically, it is expected that such a system will bring more and more effective competitive intelligence for enterprises and thus makes it advanced in the market competition.

However, according to the state-of-the-art of the research on competitive intelligence, a lot of issues are needed to be further studied [3]. Currently, most Competitive Intelligence Systems (CIS) are only able to search pieces of information from the Web, typically using a search engine like Google and Yahoo. It is usually inefficient and ineffective because search engines will return much useless information, e.g., advertisings. Recently, the text mining techniques are introduced into CIS and some tools are developed [4, 5]. Unfortunately, while these tools are capable of filtering non-related text blocks in Web pages, it divides a Web page into a set of text blocks. This eventually brings more information processing work into the analysis of competitive intelligence, because the information processing and analysis work is often required to be done manually in current CIS tools.

Thus, many issues must be revised in order to develop a Web-based CIS system, among which the most important thing is to make it clear what types of competitive intelligence we can obtain from the Internet, and what details of those competitive intelligence we can extract, or in other words, to develop an ontology of the competitive intelligence in the Internet [6]. Traditional CIS tools regard the competitive intelligence in the Internet as a set of Web pages, which results in a large amount of information processing work for the generation of competitive intelligence.

This paper aims at developing an ontology and a conceptual model for the competitive intelligence in the Internet. The major contributions of the paper are summarized as follows:

1. We introduce an entity-based representation method to capture the semantics of the competitive intelligence in the Internet, where each type of competitive intelligence is represented as a set of entities. Compared to the traditional Web-page-based viewpoint, our approach is more suitable for real applications, because applications always need a clear and structured competitive intelligence about competitors or competitive environment, but not just a set of Web pages.

doi:10.4304/jsw.5.4.378-386

2. We develop an ER (Entity-Relation)-based conceptual model for the competitive intelligence in the Web. All the competitive intelligence needed for a specific company can be described in such a conceptual model. We describe a case study to demonstrate the feasibility of the model, and the results show it is practical and useful to represent the needed competitive intelligence for enterprises.

The following of the paper is structured as follows. Section 2 briefly introduces the related work. In Section 3, the semantics of the competitive intelligence in the Internet is analyzed. Section 4 presents the framework to represent the competitive intelligence in the Internet, and Section 5 describes the conceptual model for the competitive intelligence in the Internet. Conclusions and future work are in Section 6.

II. RELATED WORK

Competitive intelligence refers to the process that gathering, analyzing and delivering information about competition environment and competitors, and then transforming them into intelligence [1]. Competitive intelligence is acquired, produced and transmitted through competitive intelligence systems (CIS).

Traditionally, people will utilize some publications to acquire competitive intelligence, e.g., 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 [2].

The detailed procedure to produce competitive intelligence from the Web can be described as follows. For example, suppose a company wants to get the competitive intelligence about one of its competitors, e.g., a company C, they will first search the information about the company C through some search engines, typically using some keywords like “C Company”. Then some experts will 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 Web-page-based way has several shortcomings, among which the most serious one is that search engines will return a huge amount of Web pages. For instance, 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 searched results and produce valuable competitive intelligence.

Recently, researchers introduced Web-based text mining approach into CIS. This approach aims at finding implicit knowledge from a lot of text data [4, 5] by using some fundamental technologies, such as computing linguistics, statistical analysis, machine learning, and information retrieval. Some specific methods have been proposed to process Web pages so far, including extracting text blocks from Web pages [7] and detecting changes of Web pages [8]. Through 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, this method will consequently produce a large number of text blocks for each Web page, which will bring much more work to experts. Furthermore, if the text blocks are clustered under specific rules, the information about competitors or competition environment will spread among different clusters and introduce additional work for information analysis.

Competitive intelligence serves companies and people, so in order to make competitive intelligence systems more effective, first we should study what competitive intelligence companies need. As a survey reported [9], most people prefer to look up information by competitor. When we further ask one more question: “What is the competitive intelligence about competitors?”, most companies will present 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 contains the needed information, but note they are NOT competitive intelligence. CIS is expected to produce competitive intelligence about competitors or competition environment from a large set of Web pages, but not just deliver 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, CIS should output competitive intelligence about entities such as competitors (or sub-entities such as the products of a specific competitor), rather than just the Web pages that contain the basic information.

Extracting entities from Web pages is one of the hottest issues in Web information extraction and retrieval [10]. The term is called Named Entity Recognition. Named entity recognition was first introduced as a sub-task in the Message Understanding Conference (MUC) [11]. 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, namely time entities, number entities, and organization entities [12]. The major methods used in named entity extraction include rule-based approaches, statistical methods, as well as hybrid methods [10]. However, although there are some existing work in named entity extraction from text (or Web pages), little has been done for competitive intelligence acquiring.

In this paper, we focus on the entity-based competitive intelligence acquiring from the Internet and present an entity-based representation framework for competitive intelligence. Fig.1 shows a high-level architecture of the entity-based competitive intelligence acquiring in the Internet. As shown in Fig.1, the entity-based representation framework is applied in this system to extract and produce different granularities of entities, which are corresponding to different types of competitive intelligence.

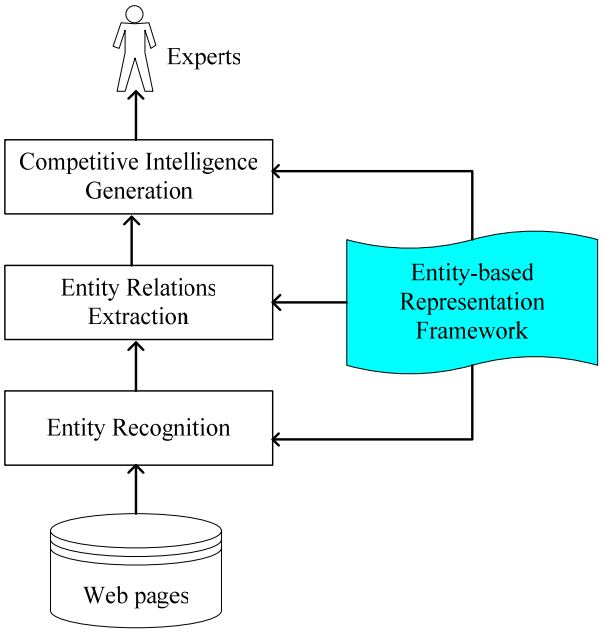


Figure 1. The architecture of the entity-based competitive intelligence acquiring in the Internet

III. SEMANTICS OF THE COMPETITIVE INTELLIGENCE IN THE INTERNET

Compared to the traditional sources of competitive intelligence, the Web has a larger volume of data, and meanwhile the Web data is updated frequently, which means they usually contain fresh information. According to the time requirement of competitive intelligence, more fresh means more valuable. The disadvantages of Web data source should also be mentioned here. First, its data volume is too large to be processed efficiently. Second, due to the security reason, some internal information is usually not posted in the Web. And third, the data in the Web are usually less confidential than printed documents since some one (e.g. some competitor) may post false information in the Web. Hence, we need a filtering mechanism to distinguish true information from those false ones [13]. However, this topic is out of the scope of this article.

According to the traditional theories about competitive intelligence, there are three types of competitive intelligence, which are related with competitors, competition environment, and competition strategies. Since the competition strategies are usually not posted in the Web due to the security reason, only the first two types of competitive intelligence may be available in the Web.

A. Competitors

In business areas, competitors refer to other companies that may contest with you in some business fields. Competitive information about competitors contains three aspects, as shown in Table 1.

TABLE I. TYPES OF COMPETITIVE INTELLIGENCE ABOUT COMPETITORS

	Profile	**Events**	**Business Relations**
Definition	the basic information about a company	time-related facts about a company	relations with other organizations
Examples	company name, telephone number, address, products set, managers' names, etc.	establishment of the company, release of new products, staff reduction, Being listed stock, etc.	suppliers of the company, investors, customers served, etc.

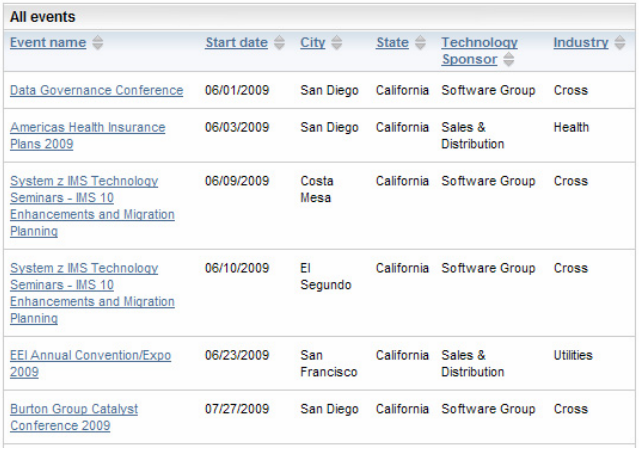
1. *Profile*

The profile is the general information about competitor. Many websites such as *Wikipedia* provide some general information about companies, such as names, employee counts, managers' names, etc. Fig.2 shows an example, which is the general information of the *Microsoft Corporation* in *Wikipedia*.

Figure 2. Example of profile intelligence extracted from the Web

2. *Events*

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, we are able to know the recent development of competitors. Typical events are the establishment of a competitor company, the listed-in-stock, the progress of some specific project, etc. Fig.3 is an example about IBM's recent events, which are extracted from its Website.



All events

Event name	Start date	City	State	Technology Sponsor	Industry
Data Governance Conference	06/01/2009	San Diego	California	Software Group	Cross
Americas Health Insurance Plans 2009	06/03/2009	San Diego	California	Sales & Distribution	Health
System z IMS Technology Seminars - IMS 10 Enhancements and Migration Planning	06/09/2009	Costa Mesa	California	Software Group	Cross
System z IMS Technology Seminars - IMS 10 Enhancements and Migration Planning	06/10/2009	El Segundo	California	Software Group	Cross
EEI Annual Convention/Expo 2009	06/23/2009	San Francisco	California	Sales & Distribution	Utilities
Burton Group Catalyst Conference 2009	07/27/2009	San Diego	California	Software Group	Cross

Figure 3. Example of events in the Web

3. *Business relations*

Differing from profile and events, business relations are usually more implicit. This is because most companies do not want their 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 some 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 perform an intelligent analysis on the contents of Web pages. For example, the Web page in Fig.4 shows the companies acquired by Oracle Corporation during the last three years, from which we get to know that *Oracle Corporation* is the major shareholder of the *Virtual Iron Software Corporation*.

Acquisition date	Company	Business
	2009	
May 2009	Virtual Iron Software	Server Virtualization Management Software
April 2009	Sun Microsystems (announced, but not yet completed)	Computers, software and IT services
March 2009	Relsys International	Drug Safety and Risk Management
	2008	
October 2008	Advanced Visual Technology	Retail Space Planning
October 2008	Primavera	Project Portfolio Management
January 2008	BEA Systems	Enterprise Software
	2007	
September 2007	Bridgestream	Enterprise Role Management software
July 2007	Bharosa, Inc	Online Identity Theft and Fraud Detection
May 2007	Agile Software Corporation	Product Lifecycle Management
March 2007	Hyperion Corporation	Enterprise Performance Management
March 2007	Tangosol Inc	Datagrid Software

Figure 4. Example of business relations in the Web

B. Competition Environment

The competition environment refers to many aspects of surrounding information that may have impacts on the business development of the company, e.g. the policies issued by the government, the habits about the resistance in a specific area or city. Table 2 lists the two types of competition environment as well as their explanations that are possible to be extracted in the Web.

TABLE II. TYPES OF COMPETITIVE INTELLIGENCE ABOUT COMPETITION ENVIRONMENT

	Macro Environment	Business Environment
Definition	the environmental information about the politics, economics, cultures, laws, society, science & technologies, and nature.	the business or industry state that a company belongs to
Examples	government laws, economic plans, local policies, state-of-the-art of specific technology, traditions of the nation, etc.	the companies set, suppliers, customers, alternatives of products, potential new comers in a specific area.

IV. ENTITY-BASED REPRESENTATION OF THE COMPETITIVE INTELLIGENCE IN THE INTERNET

The general idea of the entity-based representation of the competitive intelligence in the Internet is to define different sets of entities for the elements listed in Table 1 and Table 2. This is inspired by the fact that users need competitive intelligence about entities, e.g., a specific competitor or a specific business area. Thus the representation of competitive intelligence should fulfill such requirements because it acts as the foundation of the extraction of competitive intelligence (see Fig.1).

Fig.5 shows the high three levels of entities for competitive intelligence. An upper-level entity consists of all the sub-levels of entities. There is also 3rd level or even 4th level of entities if necessary. In the following sections, 4.1 and 4.2, we will discuss the detailed sub-levels of entities about the 2nd level of entities.

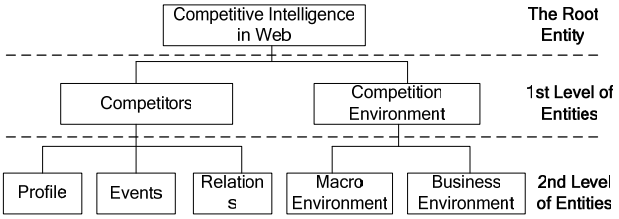


Figure 5. The high levels of entities for competitive intelligence

A. Entities for Competitors

A competitor consists of three sub-levels of entities, namely *profile*, *events*, and *relations*.

1. *Profile*

The profile of a competitor is represented as the entity "*Profile*". It is structured as follows, where the *AREA* indicates the business area of the company *C*.

PROFILE (*C*) :: = {*NAME*, *ADDR*, *TEL*, *FAX*, *EMAIL*, *PRODUCTS*, *AREA*}

2. *Events*

The events of a company is represented as the entity "*EVENTS*". It is structured as follows, where the *TOPIC* is the event's topic, the *LOCATION* is the site that the event takes place, and the *TIME* is the happened instant or the lasting time period of the event.

EVENTS (*C*) :: = {*TOPIC*, *LOCATION*, *TIME*}

3. *Relations*

The business relations of a company consist of a sub-level of entities, as shown in Fig.6. The binary relation *COOPERATION*(*C, D*) refers to the historical cooperation of the company with another one. If a *COOPERATION* relation exists between *C* and *D*, it is very possible for *C* to cooperation with these companies in the future. The binary relation *INVEST*(*C, D*) refers to that the company *D* is one of the investors of the company *C*. Thus *D* can be looked as one friend of *C*'s. The binary relation *SALES*(*C, D*) indicates that the company *D* is one of the customers of the company *C*, e.g., the *BMW* corporation is one of the *IBM*'s customers. This relation is very important in business competition, since the customers of a company usually determine its incomes. The binary relation *SUPPLY*(*C, D*) means that the company *D* is a supplier of the company *C*.

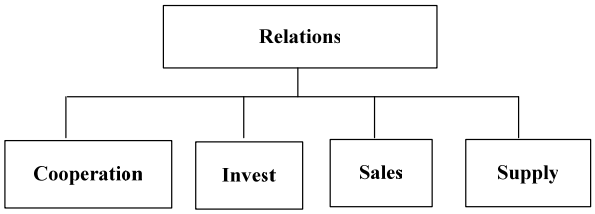


Figure 6. The sub-entities for the relations entity

These relations are defined as follows.

COOPERATION (*C, D*) :: = { *C, D, PERIOD, AREA, AMOUNT, TYPE*}

Here, *PERIOD* refers to the time duration of the cooperation, *AREA* means the business field, *AMOUNT* is the total contracted currency amount, whereas *TYPE* is the cooperation type.

INVEST (*C, D*) :: = { *C, D, PERIOD, AREA, AMOUNT*}

Here, the attributes have similar meanings as which are in the relation *COOPERATION*.

SALES (*C, D*) :: = { *C, D, PERIOD* }

SUPPLY (*C, D*) :: = { *C, D, PERIOD, AREA, ITEMS*}

Here, the *ITEM* is the set of the supplied item names.

B. Entities for Competition Environment

The entity of competition environment consists of two sub-leveled entities, the *macro environment* entity and the *business environment* entity.

1. *Macro Environment*

The macro environment contains many aspects of factors. It is difficult to obtain a complete description of the macro environment, because many laws or policies are not available in the Web. However, it is still possible to obtain some information about the changes of the macro environment. For example, new laws or policies usually appear in news reports, which enable us to extract at least the name and area of the new laws or policies.

Thus in our representation framework, due to the unfeasibility of constructing a complete description of the macro environment, we concentrate on representing and extracting the new coming information about the macro environment. There are also some sub-entities of the macro environment, which are shown in Fig.7.

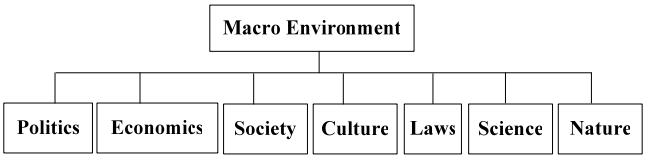


Figure 7. The sub-entities for the macro environment entity

The politics environment of a given business field *A* in a specific geographical area *L* is represented as the *POLITICS* entity, which is defined as follows.

POLITICS (*A, L*) :: = { *TOPIC, TYPE, DATE, DEPT* }

Here, the *TOPIC* refers to the detected new topic of the political information, such as the name of a political meeting. The *TYPE* is the type of the information. The *DATE* is the issued date of the topic. The *DEPT* is the department name who issues the new topic.

The definitions of the other six entities are much like that of the *POLITICS* entity. Here we do not repeat their definitions.

2. *Business Environment*

The business environment entity consists of six sub-entities (see Fig.8). The business environment refers to the whole business situation about a specific business field in a given geographical area. The *COMPANIES* entity refers to a set of company entities. The *CUSTOMERS* entity is the types of different customers in the given business filed and local area. The *SUPPLIERS* entity is the set of the supplier entities. The *PRODUCTS* entity is the set of different products. The *ALTERNATIVES* entity is the set of alternatives of products, which represent the potential products that may enter into the business environment. The *NEWCOMERS* represents those companies that may enter into the business field. These companies are the possible new competitors in the business environment.

The *COMPANIES* entity of a given business field *A* in a specific geographical area *L* is defined as follows, where the *COMPETITOR* refers to a competitor entity (see Fig.5).

COMPANIES (*A, L*) :: = { *COMPETITOR* }

The *SUPLLIERS* and *NEWCOMERS* entity has the same definition as the *COMPANIES* entity.

The *CUSTOMERS* entity is defined as follows. Here, the *TYPE* is the different customer types, while the *AMOUNT* is the number of customers corresponding to a given *TYPE*.

CUSTOMERS (*A, L*) :: = { *TYPE, AMOUNT*}

The *PRODUCTS* entity and the ALTERNATIVES entity have the same representation structure. The

PRODUCT_NAME and *TYPE* denote the product name and product type respectively.

PRODUCTS (*A, L*) :: = { *PRODUCT_NAME, TYPE*}

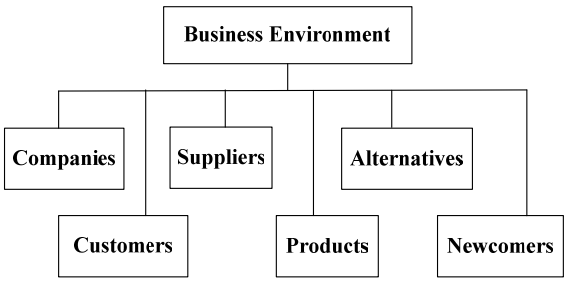


Figure 8. The sub-entities for the business environment entity

V. CONCEPTUAL MODELLING OF THE COMPETITIVE INTELLIGENCE IN THE INTERNET

In this section, we introduce a conceptual model for the competitive intelligence in the Internet. The conceptual model is based on the representation framework defined in Section 4 and focuses on the graphical representation of the competitive intelligence. We first define the notations, and then present the conceptual model. In order to demonstrate the feasibility of the model, a case study is discussed at the end of this section.

A. Methods and Notations

We use the *ER* (*Entity-Relation*) method [14] to build the conceptual model for the competitive intelligence. The ER method is widely used in data engineering to conceptually model the application data in the real world. Many *CASE* (*Computer Assisted Software Engineering*) tools support the ER-based conceptual modeling procedure. In this paper, we use the *Sybase PowerDesigner* to build the ER-based conceptual model for competitive intelligence. The basic notations used in the model are described in Fig.9.

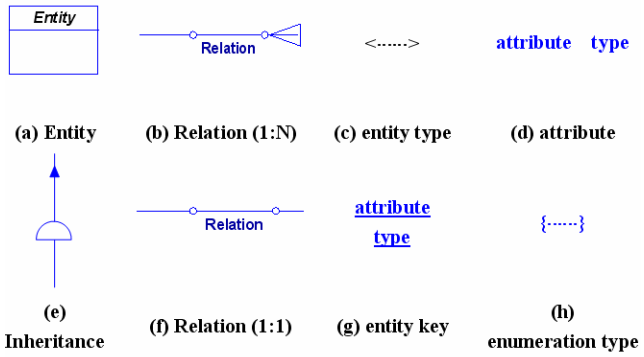


Figure 9. Notations used in the conceptual model

B. The Conceptual Model

The conceptual model for the competitive intelligence in the Internet is shown in Fig.10. The entity "*Competitive Intelligence*" represents the root entity in Fig.5, and it consists of a set of "*Competitor*" entities and a set of "*Environment*" entities. Note an entity "*Competitor*" has exact one "*Profile*" sub-entity, while has a set of "*Event*" entities and a set of "*Relation*" entities. The "*Relation*" entity and "*Environment*" entity both have several sub-entities, as shown in the figure.

As shown in Fig.10, we use fourteen entities to conceptually model the competitive intelligence in the Internet. Each entity has an entity name and some attributes describing its characteristics. Each attribute is composed with a name and a domain. There are three types of domains in the conceptual model, as listed in Table 3.

TABLE III. DOMAINS IN THE CONCEPTUAL MODEL

Domain	Description	Example
General types	common types to represent an atomic value	*string*, *integer*, *real*, *date*, etc.
Entity references	a reference to a specific entity	<*competitor*>, <*customer*>, etc.
Enumeration	a finite enumeration of some values	{*P*, *E*, *S*, *C*, *L*, *T*, *N*}, representing the seven types of macro environment in Fig.7.
Entity set	a set of entities.	*Set of* <*event*>, *Set of* <*relation*>, etc.

C. A Case Study

In this section, we present a case study to demonstrate our representation framework. Suppose a company wants to find the competitive intelligence about one of its competitors, e.g., the *Oracle Corporation*. It runs the whole process shown in Fig.1 to generate the competitive intelligence about Oracle. The intelligence extraction and generation approaches are out of the scope, and we will not discuss them here. Now we just want to know whether our presented conceptual model is able to represent the extracted competitive intelligence about Oracle.

As shown in Fig.10, the competitive intelligence about competitors consists of three sub-entities, namely the *Profile*, *Events*, and *Relations*, each of which is a set of entities. Hence, the competitive intelligence about Oracle consists of a set of <*Profile*> entities, a set of <*Event*> entities, as well as a set of <*Relation*> entities. Fig.11 shows the single <*Profile*> entity about Oracle, and Fig.12 shows the set of <*Event*> entities about Oracle.

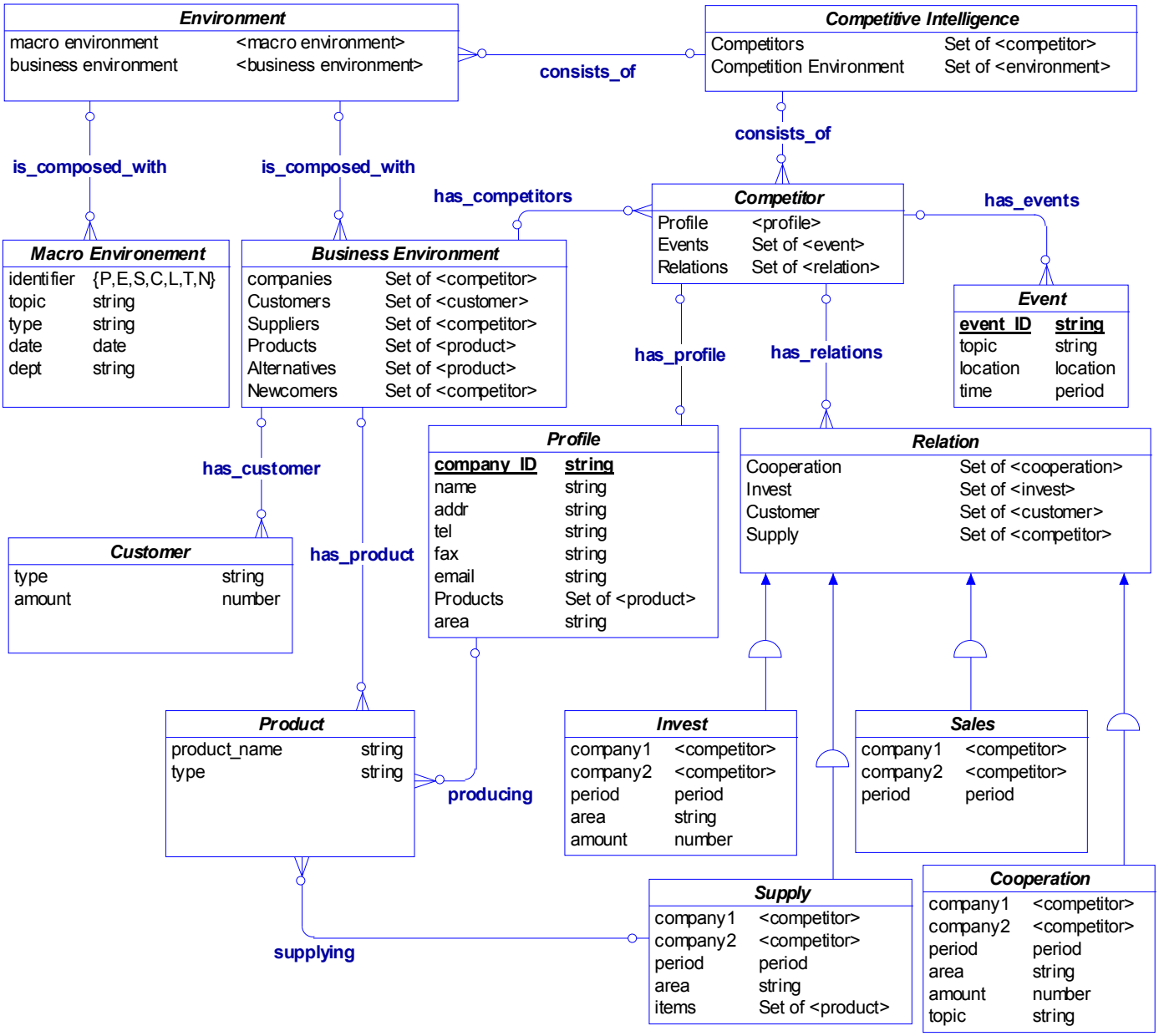


Figure 10. The conceptual model for the competitive intelligence in the Internet

Profile	
company_ID	CMP100
name	Oracle Corporation
addr	500 Oracle Parkway Redwood Shores California 94065, USA
tel	+1.650.506.7000
fax	+1.800.223.1711
email	support@oracle.com
products	Set of <product>
area	Information Technology

Product	
product_name	***type***
Oracel Database	Software
Berkeley Database	Software
Secure Enterprise Search	Software
Database Gateway	Software
Express Server	Software
CODASYL DBMS	Software
WebLogic	Software
Tuxedo	Software
Oracle Business Intelligence	Software
.....	.....

Figure 11. The <*Profile*> entity about Oracle

Event			
event_ID	topic	location	time
1	Sponsor the 2009 IRMCO conference	Cambridge, USA	19-Apr-09
2	Oracle at RSA Conference 2009	Anaheim, USA	20-Apr-09
3	Oracle Customer Services Day Chicago	Chicago, USA	20-Apr-09
4	German Oracle User Group - DOAG Regional meetin	Dreieich, Germany	20-Apr-09
5	Oracle Business Intelligence Symposium	Stockholm, Switcherlan	20-Apr-09
6	Oracle E-Business Suite Release 12.1 Launch	Reston, USA	20-Apr-09
……	……	……	……

Figure 12. The set of <*Event*> entities about Oracle

Cooperation					
company1	company2	period	area	amount	topic
Oracle	Relsys International	[2009, NOW]	Drug	NA	NA
Oracle	BEA System	[2008, NOW]	Software	850 million	Enterprise Software
Oracle	Hyperion Corporation	[2007, NOW]	Software	330 million	Performance Management
Oracle	Tangosol Inc.	[2007, NOW]	Software	NA	Datagrid
Oracle	Siebel Systems	[2006, NOW]	Software	585 million	CRM
Oracle	Sunopsis	[2006, NOW]	Software	NA	ETL
……	……	……	……	……	……

Figure 13. The set of <*Cooperation*> entities about Oracle

Supply				
company1	company2	period	area	items
Oracle	Toplink	[2002, NOW]	Software	{Toplink}
Oracle	Reliaty	[2003, NOW]	Software	{Data protector}
Oracle	Collaxa	[2004, NOW]	Software	{BPM}
Oracle	Oblix	[2005, NOW]	Software	{Identity Manager}
Oracle	Telephony	[2006, NOW]	Communication	{IP center}
Oracle	Captovation	[2008, NOW]	Software	{Document Capturer}
……	……	……	……	……

Figure 14. The set of <*Sales*> entities about Oracle

Sales		
company1	company2	period
Oracle	WorleyParsons	[2005, NOW]
Oracle	ADP INC	[2008, NOW]
Oracle	Genworth Financial	[2007, NOW]
Oracle	Grupo Mexicana	[2008, NOW]
Oracle	Iron Mountain	[2009, NOW]
Oracle	InnerWireless	[2008, NOW]
……	……	……

Figure 15. The set of <*Sales*> entities about Oracle

Invest				
company1	company2	period	area	amount
Oracle	Arete Research	[1996, NOW]	Financial Services	10,000 million
Oracle	Goldman Sachs	[1999, NOW]	Bank	20,000 million
Oracle	Morgan Stanley	[1995, NOW]	Financial Services	8,000 million
Oracle	UBS	[2006, NOW]	Bank	6,000 million
Oracle	Deutsche Bank	[2001, NOW]	Bank	30,000 million
Oracle	Oppenheimer	[1999, NOW]	Bank	26,000 million
……	……	……	……	……

Figure 16. The set of <Invest> entities about Oracle

The <*Relation*> entity consists of four sub-entities, which are the <*Invest*> entity, the <*Cooperation*> entity, the <*Sales*> entity, and the <*Supply*> entity. The acquired intelligence about the <*Sales*> entity is shown in Fig.13, which represents the investors to the Oracle Corporation. The other three entities are shown in Fig.14 to Fig.16.

VI. CONCLUSIONS

It is the future trend to build competitive intelligence systems to automatically acquire competitive intelligence from the Web. Before the design of such systems, one of the important things is to make it clear what types of competitive intelligence are hided in the Internet. In this paper, we have presented an answer to this question. We studied the semantics of competitive intelligence in the Internet, and proposed a framework to represent the competitive intelligence in the Internet. Moreover, a conceptual model for the competitive intelligence was built, which makes it easy to represent the competitive intelligence about competitors as well as competition environment. We also presented a case study for the proposed conceptual model. The results show that it is feasible and practical to represent competitive intelligence acquired from the Web based on our proposed approach.

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] L. Kahaner, (eds.), *Competitive Intelligence*, New York: Simon & Schuster, 1996.
[2] S. Thompson, C.Y. Wing, "Assessing the Impact of Using the Internet for Competitive Intelligence", *Information & Management*, Vol.39 (1), pp. 67-83, 2001.
[3] D. Deng, L. Luo, "An Exploratory Discuss of New Ways for Competitive Intelligence on WEB 2.0". In W. Wang (ed.), *Seventh IFIP International Conference on e-Business, e-Services, and e-Society 2007: Integration and Innovation Orient to E-Society*, Volume 2, Springer, Boston, pp. 597-604, 2007.
[4] J. Froelich, S. Ananyan, D. L. Olson, "Business Intelligence through Text Mining", *Business Intelligence Journal*, Vo.10 (1), pp. 43-50, 2005.
[5] A. Mikroyannidis, B. Theodoulidis, A. Persidis, "PARMENIDES: Towards Business Intelligence Discovery from Web Data", In J. Liu et al. (ed.), *WI 2006: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, IEEE Computer Society, Washington DC, pp. 1057-1060, 2006.
[6] J. Li, M. Huang, X. Zhu, "An Ontology-Based Mining System for Competitive Intelligence in Neuroscience", In N. Zhong et al. (ed.), *WimBI 2006: Proceedings of the First WICI International Workshop on Web Intelligence Meets Brain Informatics*, Lecture Notes in Computer Science 4845, Springer, Heidelberg, pp. 291-304, 2006.
[7] A. Hotho, A. Nürnberger, G. Paass, "A Brief Survey of Text Mining", *LDV Foru*m, Vol.20 (1), pp.19-62, 2005.
[8] I. Khoury, R.M. El-Mawas, O. El-Rawas, et al., "An Efficient Web Page Change Detection System Based on an Optimized Hungarian Algorithm", *IEEE Transaction on Knowledge Data Engineering*, Vol.19(5), pp. 99-613, 2007.
[9] J. LaMar, "Competitive Intelligence Survey Report", Available at: http://joshlamar.com/documents/CIT_SurveyReport.pdf , accessed in 11 May, 2009.
[10] C. Whitelaw, A. Kehlenbeck, N. Petrovic, et al., "Web-scale Named Entity Recognition", In J. G. Shanahan, et al., (ed.), *CIKM 2008: Proceedings of the 17th ACM Conference on Information and Knowledge Management*, ACM, Napa Valley, California, pp.123-132, 2008.
[11] M. Sundheim, "Named Entity Task Definition-Version 2.1", In *MUC 2006: Proceedings of the Sixth Message Understanding Conference*, Morgan Kaufmann, San Francisco, California, pp.319-332, 1995.
[12] M. Khalid, V. Jijkoun, M. Rijke, "The Impact of Named Entity Normalization on Information Retrieval for Question Answering", In C. Macdonald, et al., (ed.), *ECIR 2008: Proceedings of the 30th European Conference on IR Research*, Lecture Notes in Computer Science 4956, Springer, Heidelberg, pp.705-710, 2008.
[13] X. Zhou, Y. Li, P. Bruza, et al., "Using Information Filtering in Web Data Mining Process", In Y. Yao, et al., (ed.), *WI 2007: Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence*, IEEE Computer Society, Washington DC, pp.163-169, 2007.
[14] P. P. Chen, "The Entity Relationship Model - Towards a Unified View of Data", *ACM Transactions on Database Systems*, Vol.1 (1), pp. 9-36, 1976.

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.