

ONTOLOGY DESIGN FOR THE EFFICIENT CUSTOMER INFORMATION RETRIEVAL

Mi Sug Gu, Jeong Hee Hwang, Keun Ho Ryu
Database Laboratory, Chungbuk National University
Cheongju Chungbuk, 361-763, Korea
{gumisug,jhhwang,khryu}@dmlab.chungbuk.ac.kr

ABSTRACT:

Because the current web search engine estimates the similarity of documents, using the frequency of words, many documents irrespective of the user query are provided. To solve these kinds of problems, the semantic web is appearing as a future web. It is possible to provide the service based on the semantic web through ontology which specifies the knowledge in a special domain and defines the concepts of knowledge and the relationships between concepts. In this paper to search the information of potential customers for home-delivery marketing, we model the specific domain for generating the ontology. And we research how to retrieve the information, using the ontology. Therefore, in this paper, we generate the ontology to define the domain about potential customers and develop the search robot which collects the information of customers.

KEY WORDS: Semantic Web, Ontology, XTM (XML Topic Maps)

1. INTRODUCTION

The current web search engine estimates the similarity of the documents, using the frequency of words. Therefore the users spend too much time finding the useful information. To solve this problem, the semantic web is emerging. The semantic web defines the meaning of the information, specifies the relations of the concepts, and gives the information of the meaning to the web documents. And the software agent retrieves and provides the information to the users automatically[1].

The ontology is needed to provide the semantic web based services. It not only specifies the domain of knowledge, but also formalizes the concept of knowledge and the relationships between the concepts. The ontology describes the words in the specific domain by the hierarchical structure, and contains the reasoning rules to expand it. The ontology is used when the different databases use the different words or identifiers about the same concepts. Therefore when we compare and integrate the information of the two different databases, the ontology can help to compare the differences of them. Therefore the ontology is essential and indispensable to the semantic web based information retrieval.

In this paper, to retrieve the customer information of the home delivery service, we design the specific domain for generating the ontology. And also we propose a technique of the information retrieval using the ontology. Accordingly, to retrieve intelligently the information for the home delivery service, we implemented the retrieval robot[2] which can collect the information of the internet shopping sites on the web, and the information of the customers. And we describe the method about how to implement the retrieval system based on the ontology.

This paper comprises as follows. Chapter 2 will describe the technique of the information retrieval based

on the ontology. And chapter 3 will explain the domain and the design of the ontology to search for the customer information. In chapter 4, we will describe the design of the automatic ontology generation of the retrieval system. And we will not only explain the design of the agent and the whole framework of the ontology based information retrieval system but also describe the implementation of the ontology based retrieval system and the result of the retrieval. And then chapter 5 will conclude this paper.

2. SEMANTIC WEB BASED CUSTOMER INFORMATION RETRIEVAL SYSTEM

Semantic web is the extended one of the existing web. We give the well defined meaning to the information on the web. Therefore it has the paradigm that people can work with the computers co-operatively[1, 2].

2.1 Ontology based information retrieval

As the semantic web gives the well defined meaning to the information on the web, not only people but also computer can interpret and understand the meaning of the documents. Therefore it is proposed to retrieve, interpret, and integrate the information automatically using computer. In general, the semantic web contains the ontology for the definition of the knowledge, and the agent for collecting the information automatically[3, 4].

The ontology is the specification of conceptualization. That is, it is used to define the concepts of the specific domain and the relationships of the concepts[5].

The semantic web retrieval system combines the semantic web techniques to the current web. As the role of the ontology is increasing, it searches the important information fast. The ontology based information retrieval system utilizes the concept and the rules defined in the ontology. Because it uses the reasoning rules to develop the exactness of the retrieval, it not only can provide the

documents consistent to the user query, but also can query and modify the query and the related information properly according to the relations defined in the ontology. [6] suggested the structure of the semantic web based retrieval system. This system has subsystems such as the search engine and the ontology system.

To share the knowledge among people and machines, it is necessary to define the terms which can understand one another. It is necessary to design the ontology which contains the specific knowledge about the specific domain. Therefore in this chapter we will design the knowledge in a specific domain which is the base of this system.

2.2 Definition and design of the ontology domain

In this paper, we used the Topic Maps to construct the ontology. RDF[7,8,9,10,11,12,13] in W3C usually focuses on the resources on the web, and links among the web page. On the other hand, Topic Map has the characteristics that can represent not only objects such as web page, picture, and electronic documents, but also abstract concepts such as psychology, history, ethics, etc. Topic in a specific document consists of words which represent the writer's intention. From now on, I will explain three kinds of topic models[11,12].

First, I will explain topic. In this paper, we constructed the ontology to retrieve the information of the customers of the home shopping sites for marketing. Accordingly, the ontology has three topic types. First, the region is the area which the home shopping sites are located on. Second, the items are what the home shopping sites are dealing with. And third, the company information means the basic information about the company which operates the home shopping sites. They are phone numbers, URL, e-mail and so on.

The region has a domain that home shopping sites are located on, and it is represented hierarchically. The root of the topic is the region. And the region has sub nodes that represent each region and the name of the city in a hierarchical structure. And the whole item is a root node. And there are also sub nodes which have items in a hierarchical structure. And the company information contains the address, the phone number, the items and URL of the home shopping sites. Therefore, we can make up the associations with other topic types such as region and items.

Second, I will talk about occurrence. Each topic is linked to the resources which it refers to. For example, to investigate the company information we can refer to URL, <http://dblab.chungbuk.ac.kr/~information>. The information about linking is called occurrence. And then we store the information which occurrence refers to into the database.

Third, another element of the Topic Maps is the association. The standard specification about Topic Map provides the associations among topics. The association of topic defines not the hierarchical structure but the semantic association among topics. There are some examples as follows.

"Home Shopping Site of Computer is located in Seoul.",
 "Home Shopping Site of Television has the information of the company."

In the above example, "Computer" and "Seoul" are topics, and they have the association such as "is located in" each other. Topics in the topic map not only are independent objects, but also have links of special relation. Just as topics are classified into topic types, the associations are classified into association types. That is, in the above example, "is located in" and "has a member of" are the association types. Topic type and association type are important functions of the topic map for representing, classifying, and structuring the knowledge and the information.

3. DESIGN OF THE RETRIEVAL SYSTEM BASED ON THE ONTOLOGY

In this chapter we will describe the semantic web based customer information retrieval system. And also we will explain the automatic ontology generation and the retrieval agent.

To generate the ontology automatically, we need to design the ontology about the related domain. Figure 1 shows the flow diagram for the ontology construction. Based on the basic information about the shopping sites on the web, we define the region and the item information for the ontology.

The information of the company in shopping sites is classified into the region information and the item information. Based on it, the topic map can be generated. Figure 2 shows the knowledge map based on the topic map. This information can provide the exact knowledge while the user buys and retrieves the specific products.

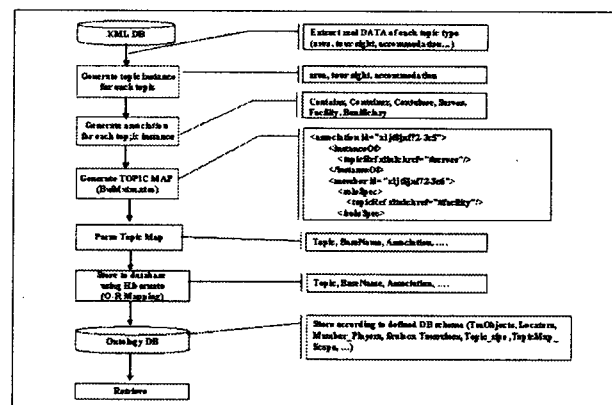


Fig. 1. Ontology generation procedure

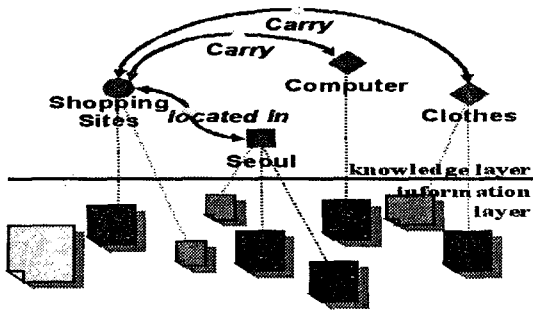


Fig. 2. Knowledge Map of the Shopping Sites

After defining the domain for the ontology construction, we specify the basic elements of the Topic Map such as topic, association, and occurrence. The specification of the topic association was written using the XTM, XML Topic Map. We verified the syntax errors about the source of XTM for the ontology by Omnigator of Ontopia[14].

XTM source which has no syntax errors is parsed by the TM4J[15], Topic Map engine. During parsing the source code, to guarantee the integrity of the data we investigate the constraint condition in advance. Accordingly, we refer to the schema of Topic Map to verify the constraint condition of the association cardinality. And also we verify the occurrence type and the association role.

The automatic ontology generation means that we design and define the concepts and the associations of the concepts and then make programming them. In this paper, we classified the data from the web by the agent, and then construct the ontology with the data automatically. And we define the knowledge resource information from the web as topic, association, and occurrence.

The information retrieval engine which uses the agent is a web site which supports the information retrieval as an on-line database[2]. Therefore the retrieval engine is a database which stores the linking information. The robot agent travels around the web sites, and then collects the information, following the links. And then it classifies, indexes the information and stores it into the database.

It starts retrieving after inputting the address of the web site as an initiative value. To filter the shopping sites, it checks the words or the documents related to the word, "shopping". It examines the words and the documents which contain items, and then it determines whether it is a shopping site or not. From the retrieved shopping sites, it extracts the URL of the shopping sites, the URL of the linked sites, the titles, the items, the address of the shopping sites, and so on

Here we will describe the information retrieval system based on the ontology to extract the information of the shopping sites. Figure 3 shows the framework of the information retrieval system. This system extracts the information from the web by the agent, and then retrieves the exact information from the extracted information using the ontology. Figure 4 shows the specific modules. They are classified into two kinds of modules; the module

for the web retrieval agent and the module for generating and managing the ontology.

The ontology based retrieval system is composed of the user interface, the ontology storage, the ontology generation engine, the web server and so on. In this paper we used TM4J engine to generate the ontology automatically.

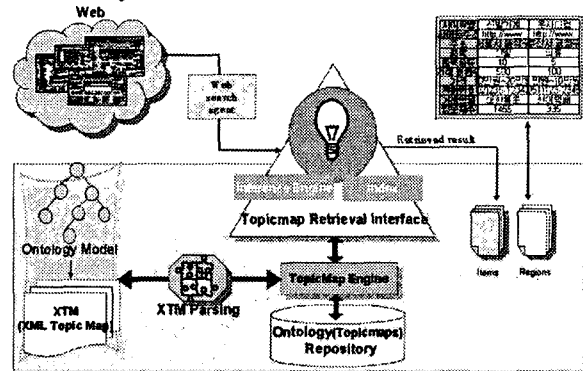


Fig. 3. Framework of the Retrieval System

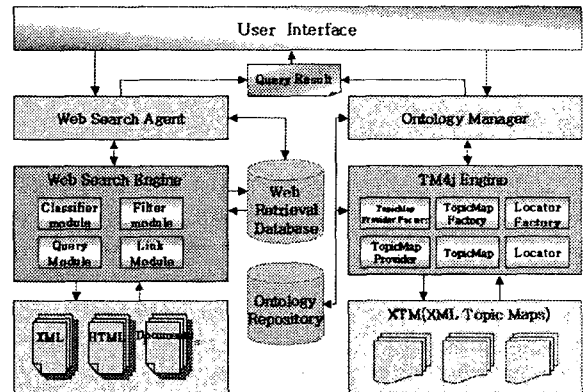


Fig. 4. Structure of the specific module

In this system, the user can query the information which he wants to know through the user interface. Therefore the user can be provided the information from the storage, where the ontology engine already stored the information which extracted by the agent.

4. IMPLEMENTATION OF THE INFORMATION RETRIEVAL SYSTEM

Here we will explain the automatic ontology construction. Topic Map information is stored into the table of the object relational database according to Hibernate of TM4J[16].

Here we can see the URL of the shopping sites on the right, and also the phone number, the title and the items of the special site on the left. If we select a special site on the right, we can see the special information about it. And "INPUT URL" is used to determine the initiatives of the site which the user wants to retrieve. If we input the address of the site, the agent travels and searches a lot of sites which are linked to. Figure 5 shows the retrieved results by the region ontology.

Item ontology is composed of the hierarchy of the items which sell in the shopping mall. If we select the suitable

item, we can see the information of the shopping site which sells the item.

To retrieve the information for the home delivery marketing, we generated three ontologies automatically. They are the company information ontology, the region ontology, and the item ontology. And we design and implement the web search agent to implement the retrieval engine using the ontology. And also we show the retrieved results by the ontology.

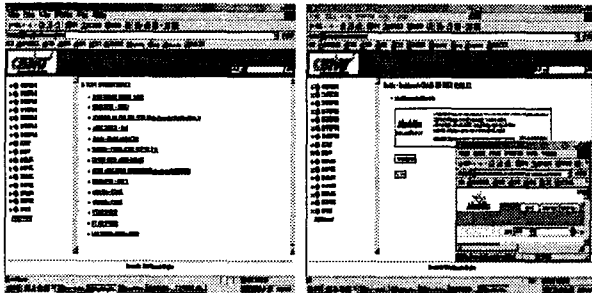


Fig. 5. Result of the retrieval

5. CONCLUSIONS

The current web search engine estimates the similarity of the documents, using the frequency of words, and then gives the grades. Therefore the users waste too much time finding the useful information.

To solve this problem, the program or the software agent extracts the meaning from the documents automatically. To do this, the semantic web adds the semantic information and meta data to the web documents, and also the software extracts the information automatically. Therefore, it is possible not only to extract the information automatically, but also to expand or share the information.

Therefore, to retrieve the information for the semantic web based home delivery marketing, the ontology is needed. Accordingly, in this paper, we define the knowledge domain for retrieving the information of the customers and formalize the knowledge. Therefore, we design and implement the agent to extract the customer information from the web.

In this paper, we generated the ontology and implemented the information retrieval system for the home delivery marketing. This is based on the semantic web. Because the ontology shares the knowledge and considers the consistency of the domain knowledge, it is helpful for the product recommendation, the marketing, and the home delivery service.

We are going to experiment the performance, the precision and the recall of our system compared with the existing retrieval engine.

Acknowledgements

This work was supported by the Regional Research Centers Program of Ministry of Education & Human Resources Development in Korea.

References

- [1] T. Berners-Lee, J. Hendler, O. Lassila, "The Semantic Web," *Scientific American*, 2001.
- [2] J. Hendler, "Agents and the Semantic Web," *IEEE Intelligent Systems*, Vol.16, No.2, March/April, 2001.
- [3] S. Decker, S. Melink, F. van Harmelen, D. Fensel, M. Klein, J. Broekstra, M. Erdmann, I. Horrocks, "The Semantic Web; the roles of XML and RDF," *IEEE Internet Computing*, Vol. 4, No. 5, pp.63-73, 2000.
- [4] M. Klein, "XML, RDF, and Relatives," *IEEE Intelligent System*, vol.16, no.2. March/April. 2001.
- [5] A. Gomez-Perez, O. Corcho, "Ontology Languages for the Semantic Web," *IEEE Intelligent Systems*, Vol. 17, No. 1, 2002.
- [6] Choi Ho Seop, Ok Cheol Young, "Information Retrieval System and Ontology," *KISS Journal* 22, 4, 2004.4.
- [7] W3C. RDF website. <http://www.w3c.org/RDF>
- [8] W3C. RDFS website. <http://www.w3c.org/TR/rdf-schema>
- [9] International Organization for Standardization, ISO/IEC 13250, Information Technology SGML Applications-Topic Map (ISO, Geneva 2000).
- [10] M. Biesunski, M. Bryan, S. Newcomb, ISO/IEC 13250 TopicMaps.
- [11] S. Pepper, G. Moore, "XML Topic Maps(XTM) 1.0", TopicMpas.org.
- [12] Steve Pepper, "The TAO of Topic Maps," XML Conference & Exposition, 2000.
- [13] Mi Sug Gu, Jeong Hee Hwang, Keun Ho Ryu, "Designing the Ontology of XML Documents Semi-automatically", ICIC2005, Part I, LNCS 3644, pp. 818-827, 2005
- [14] <http://www.ontopia.net>
- [15] <http://www.tm4j.org>
- [16] <http://www.hibernate.org>