SEMANTIC ASSOCIATION-BASED SEARCH AND VISUALIZATION METHOD ON THE SEMANTIC WEB PORTAL

Myungjin Lee[1], Wooju Kim[1], June Seok Hong[2] and Sangun Park[2]

[1]Dept. of Information and Industrial Engineering, Yonsei University, Seoul, Korea
xml@yonsei.ac.kr and wkim@yonsei.ac.kr
[2]Division of Business Administration, Kyonggi University, Kyonggi, Korea
Junehong@kyonggi.ac.kr and supark@kgu.ac.kr

ABSTRACT

As the information on the web dramatically increases, the existing web reveals more and more limitations in information search because web pages are designed only for human consumption by mixing content with presentation. In order to improve this situation, the Semantic Web based on ontology comes on the stage by W3C, and it will bring a significant advancement in web search. To do this, the Semantic Web must provide novel search and visualization methods which can help users instantly and intuitively understand why and how the results are retrieved. In this paper, we propose a semantic association-based search methodology that consists of how to find relevant information for a given user's query in the ontology, that is, a semantic network of resources and properties, and how to provide proper visualization and navigation methods on the results. From this work, users can search the semantically associated resources with their query and also navigate such associations between resources.

KEYWORDS

Semantic Web, Semantic Web Portal, Ontology, Ontology Retrieval, Semantic Search, Semantic Visualization

1. INTRODUCTION

From the beginning of the Internet, the continuing progress in network technologies and data storage techniques has digitalized huge amounts of documents on the Internet. As the information on the web dramatically increases, the existing web makes it more and more difficult for users to find relevant information because web pages are designed only for human understanding by mixing content with presentation. This problem complicates information search on Internet. For example, in traditional search system, users get search results as web pages that only include the search keyword. Therefore, to searching information they want, search agent needs to understand meaning of resources and navigate through relation between resources. This need for better information retrieval has lead to the creation of the Semantic Web. An extension of the current World Wide Web, Semantic Web is based on the idea of exchanging information with explicit and formal machine-accessible descriptions of meaning [1]. Because of Semantic Web's features, the applications in Semantic Web Portal can obtain an increased accuracy when processing information, and this has the potential of improving the way search engines perform their work.

In order to make and exchange the semantics of information, we have to agree on the way we model information. An ontology, which is a formal explicit description of concepts or classes in a domain of discourse [2], can be used to annotate data with metadata and interrelations in a specific domain. One of the core differences between semantic search and keyword search is how to utilize the interrelations among data which are noted as resources in the Semantic Web.

Semantic search of our approach evaluates the interrelations among resources while traditional keyword search techniques evaluate just resources for a given query. Therefore, the properties which relate resources in the Semantic Web are very important in semantic search. The questions of where should we start and how to span and explore the semantic network which consists of resources and properties are the main issues in our research. Moreover, how to visualize the search results is also an important issue. Keyword search shows just a ranked list as similarities between the user query and documents. However, semantic search should show the relations between the given query and each resource in the result. The relations are important information because they show why and how each resource is related to the query.

The purpose of our research is to propose a semantic search methodology that consists of how to find appropriate information in ontology that is a semantic network of resources and properties and how to provide proper visualization and navigation methods for the results. Moreover, we will propose applied methods using the basic search to find commonly related resources to two keywords of a query. From this work, we provide search results that are connected and ordered relations between search keyword and other resources as link of relation on semantic network.

The paper is structured as follows: section 2 describes some related work about ontology search and visualization. Section 3 shows the architecture of our semantic search system and section 4 discusses how to construct our domain ontology for search system. Section 5 presents an electronic ontology based search and navigation method that this paper suggests on the Semantic Web portal. Section 6 shows the Semantic Web portal based on our approach in a real application. In section 7, we conclude and present further work.

2. RELATED WORK

Recently, a number of semantic search approaches have been published. Their application area and their realization are diverse. However, they are based on a common set of ideas. Christoph Mangold identified and interrelated these ideas in his paper [3]. He presented a categorization scheme that used to classify different approaches for semantic search along several dimensions. In particular, he introduced categories for the following criteria: Architecture, coupling, transparency, user context, query modification, ontology structure and ontology technology. Therefore he selected ten different approaches and project about semantic document retrieval, i.e. Simple HTML Ontology Extensions (SHOE) [4], Inquirus2 [5], TAP [6], etc. He compared the systems by means of the classification criteria and discussed issues that are open to further research and application development. According to his research, our system is classified tight coupling between web pages and the ontology. It means the metadata of documents refer explicitly to concepts of a specific ontology. Therefore our approach is classified graph-based approach that perceives both, ontological concepts and documents as the nodes of a graph.

The MultimediaN E-Culture project [7], one of the semantic search system, is to demonstrate how novel Semantic Web and presentation technologies can be deployed to provide better indexing and search support within large virtual collections of cultural heritage resources. To search semantic paths, it checks all RDF literals in the repository for matches on the given keyword and traverses the RDF graph until a resource of interest is found, finally the results are clustered based on the paths from the matching literals to their result. This research has a point of sameness with our approach. However it has some lack of how to assign the weight to the property and decide the threshold to stop the search. It is one of the most important issues because of determinant of how to expand the semantic network. Therefore another difference is that they traverse one direction only: always from the object in the triple to the corresponding subject.

Finally, SemRank [8] is a paper about considering information search method on the Semantic Web. we present an approach that ranks results based on how predictable a result might be for users. It is based on a relevance model SemRank, which is a rich blend of semantic and information-theoretic techniques with heuristics that supports the novel idea of modulative searches, where users may vary their search modes to effect changes in the ordering of results depending on their need. Based on information theory, they built the specificity and the θ-specificity model for measuring the information content of a semantic association by considering the occurrence of edge as an event and RDF properties as its outcomes. We expand the specificity and θ-specificity method to measure an amount of information that is links on semantic network. Therefore we find more valuable and special semantic information in this paper.

3. OVERVIEW OF SYSTEM ARCHITECTURE

The aims of our research are to propose ontology-based semantic search and visualization methods on a Semantic Web portal. Figure 1 shows the components of our association-based search system.

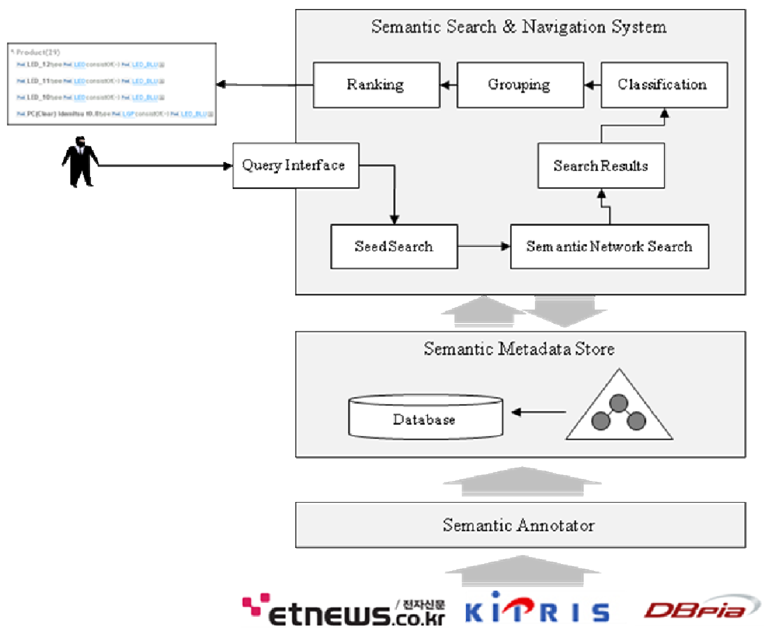

Figure 1. System architecture of Semantic Web portal

Web data is translated into triple structure and stored on semantic metadata store through the semantic annotator. User can find the information relate to query string using the semantic search and navigation system. The semantic search and navigation system consists of querying and navigation task. The goal of the querying task is to find related information which can be concepts and instances over the ontology from the user's given query. Querying procedure consists of finding seed instances relate to user query and spanning of resource through properties. From this search result on querying task they are classified, grouped and sorted on the navigation task. As a result of semantic search, user can search for the relation between information. This step will be discussed more detail on section 5.

4. CONSTRUCTING DOMAIN ONTOLOGY

In order to search the semantic association, we constructed the ontology about information of the target domain. To build ontology on the domain of electronics, we gathered information that was published for one year from different data sources such as an electronics news site [9], a patent service site [10], and a research paper service site [11] in Korea. For example, we collected news related to electronics that was published for a year by an electronics news site. For building ontology, we used semi-automatic semantic annotation. At first we extracted concepts like title, author, keyword, URL, published date, etc. from each news, and then made triples that is relation between concepts. Ontology that we made from different sites consists of 7 categories - News, Person, Article, Product, Patent, and Technology, but data sources are different. It means that the information of each source was separated from each other, so it was impossible to connect information between sources while independent search was possible at each source. The users who work in electronics industry, especially the companies which produce electronic parts, are very sensitive to new technology, trends, and parts producers. They read newspaper related to electronics and search for information about patents and technology from each site because data sources are different. This problem can be solved a search system using ontology that consists of semantic network because this information has semantically related over several fields. For example, a patent is related to a trend or technology. Also, the patent is owned by a company. Therefore, we transformed and combined the information of the sources into ontology so that they are connected to each other. This is one of the important improvements of the semantic search system.

An ontology that uses on Semantic Web portal provides an understanding of the domain of the data set. The ontology on the Semantic Web involves publishing in languages specifically designed for Web Ontology Language (OWL) [12] based on Resource Description Framework (RDF) [13]. The RDF data model is based upon the idea of making statements about Web resources in the form of subject-predicate-object expressions, called triples in RDF terminology. All information is semantically related and linked in ontology, so that the user can search for the relation between information. Therefore search results can represent semantic association paths connected between resources.

Figure 2 shows semantic annotation result from news web site. We gathered title, writer, date, source URL, etc. information from each web site like left side of Figure 2 and built OWL ontology that shows right side of Figure 2.


```
<News rdf:ID="NW006">
    <dc_source rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        http://www.etnews.co.kr/news/detail.html?id=200704230077
    </dc_source>
    <dc_title>
        <Title rdf:ID="Title_22">
            <koreanTitle rdf:datatype="&xsd;string">
                루미마이크로, 조명 및 플래시용 고휘도 LED 개발
            </koreanTitle>
        </Title>
    </dc_title>
    <dc_creator>
        <Reporter rdf:ID="Reporter_uhyeongjun">
            <name rdf:datatype="&xsd;string">유형준</name>
        </Reporter>
    </dc_creator>
    <dc_date rdf:datatype="&xsd;dateTime">
        2007-04-24T00:00:00
    </dc_date>
    <articleCode rdf:datatype="&xsd;string">
        NW006
    </articleCode>
</News>
```


Figure 2. Semantic annotation from news web site

5. Semantic Search and Navigation

We will explain our semantic search approach in two steps which are querying and navigation. First, we will explain how to find related resources over the ontology from the given query in querying step. In next step which is navigation, we will explain the visualization part of our system which shows the results in appropriate ways.

5.1. Querying

Ontology is a mechanism for representing formal and shared domain knowledge. Ontology consists of a hierarchy of concepts, role relations among concepts, and instances attached to it. The goal of the querying task is to find related information which can be concepts and instances over the ontology from the user's given query. Our querying procedure consists of two main steps. The first step is seed search with user's query and the second step is semantic spanning of resources through properties. Seed search is very similar to the traditional keyword search. The results of seed search are concepts and instances containing the words of the given query string. Therefore, the first step is not the important issue of our approach. The second step, which is the main issue of our approach, is semantic spanning of resources. The procedure finds related concepts and instances through properties over the ontology from the seeds by spanning properties. In this step, we need our own policy to decide which properties should be spanned.

We give the definition of knowledge base [14]. A knowledge base can be viewed as three tuples:

$$KB = (I, C, P)$$

where C denotes the set of concepts; P denotes the set of property; I denotes the instance set of all concepts. Specially, let $c_i \in C$ denotes a concept, $p_i \in P$ denotes a property and $i_i \in I$ denotes an instance of concept c_i.

In the first step of querying task, we look through the seed candidates with the keyword the user submitted from the ontology. We now define a set of seed instance that is a start point of semantic relation search. A set of seed instances is:

$$Seed_q = I_q \cup \{ i \mid i \in \leq C_q \}$$

where $Seed_q$ denotes a set of seed instances, q denotes the user's query string, I_q denotes $I_q = \{i_i \mid \text{name of } i_i \text{ contains } q\}$, C_q denotes $C_q = \{c_i \mid \text{name of } c_i \text{ contains } q\}$, and $\leq C$ denotes a partial order on C.

If the user input "*BLU*" – stands for Back Light Unit – as a keyword for semantic search, we look through all the candidates from the ontology because we don't know the exact search intention of the user. The candidates include all instances that have the word "*BLU*" in their character strings and instances of classes (include subclasses of it) that have the query string. Even if the candidates which is found at this step is not seed instances that user wants, it doesn't care because seed instances will be ordered by similarity of user's intention in the next navigation task.

Once we found seeds from the ontology, the next step is to find related concepts with the seeds. The main difference between semantic search and keyword search is searching for related classes and instances that are linked in the semantic network. All concepts are connected in ontology that has a simple graph structure. Figure 3 shows the pseudo code for finding related concepts.

```
Calculate Path Decision Threshold Theta(p) for all property p
FOR all instance s_i in Seed_q do
        Set r with seed instance s_i
        WHILE depth of path < maximum depth
                FOR all triple(r, p, t) in Semantic Graph do
                        IF Theta(r, p) < Theta(p) THEN
                                Expand path into Semantic Graph
                        ENDIF
                ENDFOR
                Set r with the last resource of the path
        ENDWHILE
ENDFOR
```

Figure 3. Algorithm to trace a semantic network

This algorithm evaluates outgoing properties from a resource. In the beginning of the semantic spanning, it evaluates properties connected to seeds. For one seed, it evaluates every properties outgoing from the seed to other resources and decides properties for spanning. If a property is chosen, the resources connected to the seed through the property are added to the spanned resources. At the next loop, the algorithm evaluates properties of one resource of the spanned resources.

As shown in Figure 3, *depth of path* means the count of nodes which are connected to the seed, and path decision threshold θ denotes a measure for spanning links connected to the seed. In our research, *maximum depth* is limited to 3. Therefore, the search results include only the nodes which are connected to the seeds within 3 links. The number of spanned links will affect the search results a lot. We will research to deal with this issue in our future research.

The issue of this algorithm is how to decide the properties that should be spanned. If we span all links connected to the seed or middle concepts, maybe the search results are huge and useless. Therefore it needs to span links that are more important. In information theory, the amount of information contained in an event is measured by the negative logarithm of the probability of occurrence of the event [15]. It means that information occupied by a few people is more valuable than those occupied by many people. According to the information theory, the search method that we use expands with properties that are less frequently used, higher information gain, in connections as semantic network. To this search method, we need thresholds for every property and *path decision threshold* is the factor for deciding spanning properties.

In Kemafor's research [8], authors used the θ-specificity for measuring the amount of information. The θ-specificity of a property is a measure of its uniqueness relative to all other properties in the description base whose domain and range belong to the same Representative Ontology Class (ROC)'s respectively. Intuitively, the threshold is the ratio of the number of connections where the property is actually used to the number of connections where the property can be used. In other words, it can be explained as the average degree of property usage for all resources in connections.

We use the θ-specificity [8] for measuring path decision threshold as follows:

$$\theta(p) = \frac{|(*, p, *)|}{|(x, *, y)|}$$

where p denotes a property, x denotes $x \in p.domain$, and y denotes $y \in p.range$. It means that a property of low frequency has more valuable than a property of high frequency. However, it considered only uniqueness of a property, not resources. Therefore, we devise a new type of θ-specificity for a resource and a property as follows:

$$\theta(r, p) = \frac{|(r, p, *)| + |(*, p, r)|}{|(r, *, y)| + |(x, *, r)|}$$

where *r* denotes a resource and *p* denotes a property. Intuitively, this measure indicates the ratio of the number of connections where the resource *r* and the property *p* are used to the number of connections where the resource *r* is used. It means the average degree of property usage for the given resource *r*.

If $\theta(r, p)$, θ-specificity of resource *r* and property *p*, is smaller than path decision threshold, $\theta(p)$, then we span to other resources connected to the property *p*. It means that if the number of usage of property *p* for the resource *r* is smaller than the average usage of property *p* for all resources, the property *p* is valuable enough to span.

5.2. Navigation

The navigation task arranges the search results and provides the method for users to navigate them. This step aims to find what they want among the results. The search results from the previous step are semantic associations from seeds to other resources. There are two issues in this step. Firstly, how can we arrange those different paths with the same seed and resource? Each path that is association paths from a seed to a resource has a different reason why the resource is related to the seed, and the reason is important for the user. Therefore, we should show every path from the seed to the resource even if the final resource of the path is the same. Secondly, we need grouping and ranking policy for the search results. In general cases, the number of search results of semantic search is much larger than the number of keyword search results because they contain related concepts and instances in addition to the resources containing the words of the query. Therefore, it is very important to decide how to group and rank the results.

For the visualization of query results, the navigation task consists of three steps. The first step is classification in categories that are determined according to their information sources. The search results are categorized into 7 main categories – Company, News, Person, Article, Product, Patent, and Technology – as class of a seed instance, so that the user can easily find their search interest.

The second step is grouping association paths per internally disjoint [16] from query results in each domain. For instance, two paths are internally disjoint if there exist two semantic paths e_1, P_1, e_2, P_2, e_3, …, e_{n-1}, P_{n-1}, e_n and f_1, Q_1, f_2, Q_2, f_3, …, f_{n-1}, Q_{n-1}, f_n connecting e_1 with *en* and f_1 with f_n, respectively, and both of the following conditions holds: e_1 equals f_1 and e_n equals f_n where n is $1 < n \leq$ *maximum depth*. Therefore, the user can grasp the point that is relations between the seed instance and other nodes.

The final step is ranking. The ranking step consists of the ranking of groups and ranking of semantic paths in a group. Each group and semantic path in groups is ranked as follows:

$$rank(sim(q, e_1), distance(e_1, e_n))$$

where q denotes the user's query string, e_1 denotes the seed instance, and e_n denotes the end node in association path. We use $distance(e_1, e_n)$ to denote the number of the links between e_1 and e_n, and $sim(q, e_1)$ to denote word similarity between the user's query string and the seed instance. The most important factor of our ranking algorithm is word similarity between seed instance and the keyword. To measure a word similarity, we use the term frequency [17] that is simply the number of times a given term appears in document for similarity measure. According to this method, semantic paths are lined up in the rank which is the higher similarity between the query string and the seed instance. The next factor is the number of the links between the seed and the final node of the resulting path. If the number of the links is zero, it means the node is seed themselves. They have the highest priority. The next is the node that is directly connected to the seed. The nodes with the same value of word similarity are ordered with the number of links. Therefore, query results is firstly ranked by $sim(q, e_1)$, and then by $distance(e_1, e_n)$.

6. Application for Semantic Visualization

We have discussed how to retrieve information related to the seeds which contain the user's keyword. Now, the main issue here is how to visualize the retrieved information. We especially focused on devising the visualization method by which the user can easily identify the retrieved resources and recognize how they are associated with the keyword. To accomplish this goal, we provide a series of visualization methods and the first is a layout of our semantic web portal, which is depicted in Figure 4. Figure 4 actually shows a screen shot of our semantic web portal, which was developed with practical purpose to help small and medium sized companies in the field of Korean electronics industry under the support of Korea e-business association (KOEB). As shown in the figure, overall layout consists of four parts.

Figure 4. A user interface of Semantic Web portal

The first part appears on the upper area of the screen and is input box for the user's query. The left side of the screen is the second part and shows the matched seed instances with the given keyword grouped by each belonging class. If the user clicks a seed instance of the left panel, she or he can narrow the retrieved results only into the ones related to a designated seed. The center of the screen shows the third part and provides the main result of the query, which is grouped in terms of seven dominant classes in our case - News, Person, Article, Product, Patent, and Technology. In each category, the results are further grouped by the destination resource of the retrieved path from the seed. As a result, what we can see in this main panel is the representative path, that is, the highest ranked path in each specific group. In addition, the user can browse more results having the same destination for a given seed by clicking the plus button attached to each list item in the end of paths. We will discuss this visualization issue in more details later. Finally, the right panel of the layout is reserved to track history of user's query string all this while.

Now, let's get into more detailed discussion on the main panel to present the retrieved results. Figure 5 shows only search results for the 'Product' category in case of *LED_BLU* keyword.

Figure 5. Search result of *LED_BLU* in product domain

The first line of the search results in Figure 5 shows the best ranked path from *LED_BLU* (a seed) to *LED_12* (a destination resource). We deployed destination resources to the front of each line because the user is more interested them than seeds. The path consists of 2 triples in the form of subject-predicate-object expressions. For example, the first triple is (*LED_12 type LED*) which means *LED_12*'s type is *LED*. The second triple is (*LED consistOf(-) LED_BLU*). The *(-)* symbol of the *consistOf* means that *LED* is connected to *LED_BLU* with inverse property of *consistOf*. We can read the second triple as *LED_BLU* consists of *LED*. With the interpretations of triples, the user can find out how and why the destination is related to the user's query. Figure 6 shows the data model of the first result using RDF Validator [18].

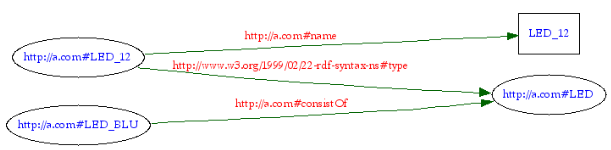

Figure 6. Semantic network of the first result

The plus (+) button in the rear of the first line means that it has other paths from *LED_BLU* to *LED_12*. We already discussed the reason why showing multiple paths from the same keyword to the same destination resource is important in navigation section. If the user clicks the (+) button, it draws out all the paths from *LED_BLU* to *LED_12* into a box as shown in Figure 7 so that the user can confirm the other paths and acknowledge more reasons why *LED_12* is related to *LED_BLU*.

Figure 7. Various result in a group

Our semantic search system provides advanced search method, AND operator, in addition to the basic search function. If a user wants to find concepts that are related to two keywords, she or he can use AND operator with the two keywords. If user searches including two keywords in traditional keyword search site, it shows contents which include both keywords. However, the results are not what user wants. In our semantic search, we find the concepts that are connected to the both seeds of keywords through properties even though they might not include any keywords.

The search results of AND operator with two keywords are two semantic paths that are connected to both keywords in the semantic network of resources and properties. In other words, they are intersecting resources over two independent search results with each keyword. Figure 8 shows the search results of AND operator with *Samsung* and *BLU* keywords.

Figure 8. Search result of AND operator

The first line of the search results in the patent domain shows that *Samsung*(삼성전자) is the right holder of one patent and the patent is connected to *BLU* at the same time via the *hasKeyword* property, the *LGP* resource, and the inverse property of *consistOf*. From this result, the user can find out any concept or instance that is related to both of *Samsung* and *BLU* in the semantic network.

Semantic visualization is an important issue in Semantic Web technologies because it requires whole different way of search and arrangement from the traditional keyword search. Moreover, it is tightly coupled with the semantic search method. Therefore, the semantic visualization should be carefully designed by considering semantic search procedure and the characteristics of the search results.

7. CONCLUSIONS

In this paper, we proposed a semantic search and navigation methods for semantic visualization in the domain of electronic parts. The procedure consists of two main steps - querying and navigation. In the querying step, we used interrelations between concepts which are represented with properties in ontology to find out related resources to the keyword while the traditional keyword search focuses on just resources. Therefore we find semantic paths that have more valuable and important on semantic network. In the navigation step, we grouped the search results into 7 categories based on the information sources where the results come from. Also, we ranked the results in each category with two measures, similarity and distance. In order to adopt and test our semantic visualization procedure, we designed and implemented a Semantic Web portal which provides semantic navigation services. We expect that our research has shown a practical implementation of semantic search and visualization on the Semantic Web portal.

As some of the aspects of our future work, we are planning to refine our metrics for measuring the semantic relation between two resources in the semantic graph, and to apply our methodology on social networks such as freebase [19].

ACKNOWLEDGEMENTS

This research was financially supported by the Ministry of Knowledge Economy(MKE) and Korea Industrial Technology Foundation (KOTEF) through the Human Resource Training Project for Strategic Technology.

REFERENCES

[1] Tim, B.L., James, H. & Ora, L., (2001) "The Semantic Web", *Scientific American Magazine*.

[2] Thomas, G., (1993) "A translation approach to portable ontology specifications", *Knowledge Acquisition*, Vol. 5, No. 2, pp199-220.

[3] Christoph, M., (2007) "A survey and classification of semantic search approaches", *International Journal of Metadata, Semantics and Ontologies*, Vol. 2, No. 1, pp23-34.

[4] Heflin, J. & Hendler, J., (2000) "Searching the web with SHOE", *AAAI Workshop on Artificial Intelligence for Web Search*, pp35-40.

[5] Glover, E.J., Lawrence, S., Gordon, M.D., Birmingham, W.P. & Giles, C.L., (2001) "Web search – your way", *Communications of the ACM*, Vol. 44, No. 12, pp97-102.

[6] Guha, R., McCool, R. & Miller, E., (2003) "Semantic search", *Proceedings of the Twelfth International Conference on World Wide Web*, pp700-709.

[7] Guus, S., Alia, A., Mark, A., Victor, B., Lynda, H., Michiel, H., Laura, H., Zhisheng, H., Janneke, K., Marco, N., Borys, O., Jacco, O., Ronny, S., Jos, T., Jan, W. & Bob, W., (2006) "MultimediaN E-Culture Demonstrator", *The Semantic Web Challenge at the Fifth International Semantic Web Conference*, pp951-958.

[8] Kemafor, A., Angela, M. & Amit, S., (2005) "SemRank: Ranking Complex Relationship Search Results on the Semantic Web", *The 14th international conference on World Wide Web*, ACM, pp117-127.

[9] Electronic Times Internet. http://www.etnews.co.kr/.

[10] Korea Intellectual Property Rights Information Service (KIPRIS). http://eng.kipris.or.kr/.

[11] Smart Academic Knowledge Base DBpia. http://www.dbpia.co.kr/.

[12] Deborah, L.M. & Frank, v.H., (2004) "OWL Web Ontology Language Overview", *W3C Recommendation*.

[13] Frank, M. & Eric, M., (2004) "RDF Primer", *W3C Recommendation*.

[14] Liang, B., Tang, J. & Li, J., (2005) "Association search in semantic web: search + inference", *The 14th international conference on World Wide Web*, ACM, pp992-993.

[15] David, J.C.M., (2003) *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press.

[16] West, D.B., (2001) *Introduction to Graph Theory 2/E*, Prentice Hall.

[17] Gerard, S. & Michael, J.M., (1983) *Introduction to Modern Information Retrieval*, McGraw-Hill.

[18] RDF Validator. http://www.w3.org/RDF/Validator/.

[19] Freebase. http://www.freebase.com/.

Authors

Myungjin Lee is a PhD candidate of Information and Industrial Engineering at Yonsei University in Korea. He has experiences about XML, XML Web Services, and Semantic Web. Therefore, He has been developing SMART Semantic Web Framework which provides a programmatic environment for ontology reasoning and management. His primary research interests are in Semantic Web, Decision Support Agent, Semantic Web Mining, Semantic Information Retrieval, combining Semantic Web with Machine Learning. He has published a contributed volume, Novel XML & XML Web Services and Novel JSP, in Korea.

Wooju Kim is a professor of Information and Industrial Engineering at Yonsei University in Korea. He received a BBA degree from Yonsei University in 1987, and a PhD in Management Science from KAIST in 1994. He has published many papers related to the issues including Semantic Web, Web Services, e-Business, Expert Systems, S/W Engineering, and Managerial Forecasting. His current research areas are decision support systems on the semantic web environment, semantic web mining, knowledge management and intelligent web services.

June S. Hong received the B.S. in Business Administration from Seoul National University, Korea, in 1989, and the M.S. degree and Ph.D. degree in Management Science from KAIST, Korea, in 1991 and 1997, respectively. Since 2003, he has been an Associate Professor in Department of Management Information Systems at the Kyonggi University. His current research interests include semantic web, intelligent agent, and various business application systems.

Sangun Park is a professor in the Division of Business Administration at Kyonggi University. He has a Ph.D. in management engineering from the Graduate School of Management at the Korea Advanced Institute of Science and Technology (KAIST). He has developed intelligent information systems and e-commerce applications with various industrial partners. His current research interests are in the fields of Semantic Web, ontology matching, ontology reasoning, intelligent information systems, and electronic commerce.