Integration of Semantic Data Using a Novel Web Based Information Query System

Okkyung Choi, Sangyong Han and Ajith Abraham

Department of Computer Science & Engineering, Chungang University,221,
Huksuk-dong, Dongjak-ku, Seoul, 156-756, Korea
okchoi20@gmail.com, hansy@cau.ac.kr, ajith.abraham@ieee.org

***Abstract*: Semantic search, interoperability, and easy integration between applications are the main advantages of the semantic web. It is considered that the next generation web model can provide solutions to many problems of the current web technology. The main characteristic of the semantic web is to assign elements of ontology to the documents. Therefore, it is necessary to use the rules that are combined with various semantic technologies to extract metadata. These metadata are based on ontology information which is maintained by software agents. In this paper, we propose Semantic Web based Information Query System (SW-IQS) using an Ontology Server. SW-IQS is supposed to enhance the efficiency and accuracy of information retrieval for unstructured and semi-structured documents. For the interoperability and easy integration, we suggest RDF based repository system. A new ranking algorithm is also proposed to measure the similarity between documents with semantic information for fast and accurate information retrieval. It is implemented and compared with the existing models.**

***Keywords*:** Semantic Web, resource description framework, RDF Query Language), Ontology, Information Retrieval, Vector Space Model

I. Introduction

The web was able to make great advancements rapidly because of its convenience and the fact that it is easily accessible to anyone around the world. However, there emerged the problem of having too many search results on a single search for specific information the user needs [2]. The search engine has been developed and new ideas are being tried out in order to solve such a problem. But as long as the search method is based on plain data processing it should be difficult for a user to find the specific material he/she needs on the web because the search is executed fragmentarily based only on words and sentence construction having the semantic contents of the web document left out [2].

The semantic web is a new approach experimenting semantic search, automation, integration and reuse. Regarding this, we propose the SW-IQS (Semantic Web Based Information Query System), which is a combination of intelligent agents and automatic classification techniques upon a substructure of semantic web based techniques. With the suggested system, the lack of precision and recalls in the current search system can be improved. The proposed system is based on the RDF (Resource Description Framework) and Ontology. The efficiency and accuracy of the proposed system is verified through a new method of similarity measurement using semantic metadata.

The study is organized in the following order. In Section 2 presents an overview of the semantic web and its structure and an observation of the cosine similarity technique of the vector space model, which is one of the most widely used similarity measurement models. In Section 3, the architecture of the SW-IQS is explained along with designing method, functions and characteristics of its modules. Next in section 4 a new ranking algorithm technique is introduced and applied for evaluation, comparison and analyzing the current search model and the newly proposed search model. Conclusion and future studies are given in the final section.

II. Related Research

2.1 The Semantic Web

2.1.1 Overview of the Semantic Web and its Structure

The semantic web is a highly advanced intelligent technology, which allows the machine to understand the information. To have the machine understand and process information, the information must be disintegrated into the original data form it had been before being processed and then processed again into another appropriate form. Concerning this, the semantic web expresses the data in a way that can be interpreted by the computer and defines the relation between the different data so that the data that are automated and integrated in the application areas, such as electronic commerce, can be shared.

Figure 1 shows the overall structure of the semantic web. The bottom including Unicode and URI is the basic web resource layer. The second layer composes of XML+NS+xmlschema that assign an arbitrary document structure using the XML, which is easily interpreted by the computer, to indicate the data's sentence construction. In the third layer, RDF+rdfschema enable inference of the data that indicate meaning (semantic). The ontology in the third layer defines the relation between terms and the RDF schema expresses a certain type of ontology that defines the relation between objects and attributes. The top layers, logic, proof and trust, verify and logically express the web that has a certain kind of rule. The digital signature layer is purported for taking care of security matters [9].

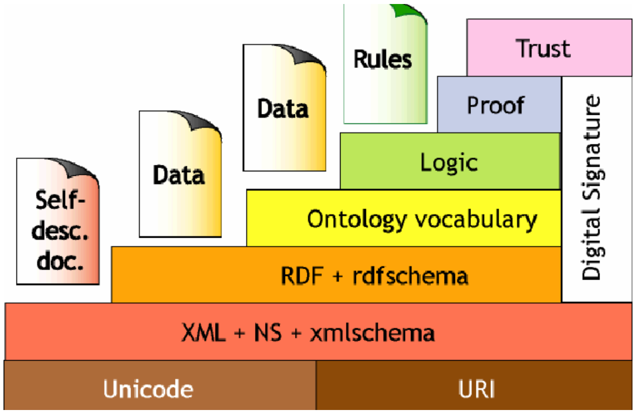


Figure 1. Structure of the semantic Web

2.1.2 RQL (RDF Query Language)

RQL [1] was first introduced and proposed as a query language for RDF and RDF schema. RQL was developed by the European IST project C-Web and was included in the MESMUSES project by FORTH Institute of Computer Science based in Greece. RQL is a function language adopting the syntax of OQL. The results for the RDF schema query are presented in the appropriate RDF code. RQL is defined as a set of core queries or a set of basic filters. RQL is capable of organizing new queries through functional organization and repetition. RDF uses a "select-from-where" expression, similar to that of SQL, to reorganize or filter information. The RQL filter applies the path expression in order to search for graphs from an arbitrary depth.

The query shown in (1) is an RDF query for finding the property value of the name that matches with "*name" in the class. In RQL, the class variable is defined as $ and the property variable as @ for discrimination.

$$\begin{aligned} &select Y, \$Y \\ &from \{X\} @ P\{Y : \$Y\} \\ &where\ @ Plike"\text{*}name" \end{aligned} \qquad (1)$$

2.2. The Vector Space Model and the Cosine Similarity

The Vector Space Model (VSM) is a way of representing documents through the words that they contain. It is used for information retrieval because it provides a tool that is capable of partial matching. This term weight is used in calculating the similarity between the user's query and each of the documents stored in the system. The vector space model searches for documents that partially match the query terms and lists the searched documents starting from the one with the highest similarity to the one with the lowest. As with the vector space model, a document or a query is regarded as a certain point on the vector space. The vector space is determined by the index that appears in the document collection. Since a query is also regarded as a certain point on the vector space as same as a document, the vector space model does not use the Boolean algebra. Instead, the similarity between two documents is perceived by the cosine similarity, which is the cosine value of the angle formed by the two vectors (Figure 2). The smaller the angle, the greater is the similarity. This is the main idea of the vector space model [13]. In the vector space model, $W_{i,j}$, the weight of the term and document pair (k_i, d_j), has a positive value. The query index also has a weight. If the weight of [k_i, q] is $w_i,q>=0$, the query vector is defined as (w_1,q, w_2,q, ..., w_t,q). Here, t indicates the total number of indexes in the system. The document vector is expressed as (w_1,j, w_2,j, ..., w_t,j).

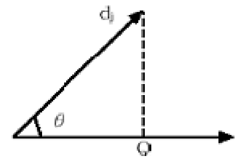


Figure 2. Cosine value applied as sim($\boldsymbol{d_j}$, q)

Document d_j and the user's query q is expressed as shown in (2). In the vector space model the similarity between document d_j and query q can be measured by the correlation between the two vectors. As an example, the similarity can be fixed to a certain value by using the cosine value of the angle formed by the two vectors as shown below.

$$CosSim(d_j, q) = \frac{\vec{d_j} \cdot \vec{q}}{\left|\vec{d_j}\right| \cdot \left|\vec{q}\right|} = \frac{\sum_{i=1}^{t} (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^{t} w_{ij}^2 \cdot \sum_{i=1}^{t} w_{iq}^2}} \qquad (2)$$

III. The Architecture of SW-IQS

This Section presents an overview of the SW-IQS process, the architecture of the system and the function of each module.

3.1 The Necessity

In the beginning stage of the WWW, the development of HTML enabled the development of display-oriented web documents. However, HTML is not capable of structural extension because the structure and content were both expressed in one document. To solve this matter, XML (eXtensible Markup Language), which has the structure and the content separated, was developed to enable the user to arbitrarily produce tags. But being too focused on syntactic metadata, content and meaning could not be combined and semantic connection was impossible. Due to this, the necessity for a semantic web gradually became more apparent.

At this point, a lot of research is underway on developing web technology and standards, and the third generation web is being prepared using semantic web and web services. In the near future metadata will be attached to all the vast amount of data and documents that are given an account on the web. If these metadata are in the form of XML that is interoperable, and are used in the same domain, new information can be drawn out by applying certain rules. So this new information is not meaningless rising out of inaccurate search results but represents accurate knowledge which is based on certain facts [6].

However the search model currently used is poor in efficiency and accuracy, and being unable to draw the specific results requested by the user because the web documents are formed using HTML and XML. Considering such matters, RDF purported for semantic data search and the ontology-based semantic web search model should be adopted as the next generation web prototype and become more and more crucial in the near future.

The following section will deal with the methods and technologies that are necessary for engineering the semantic search model.

(1) Semantic Standardization

Standardization plays an important role in processing semantic heterogeneity that occurs due to multiple data sources. In the metadata standardization process all the contents that belong to the same domain are connected to the same metadata regardless of the source or format. For example, if the article on the JoongAng Daily and the PDF article attached to a web site are both reports on fair election then they must be connected to the same metadata. Semantic standardization is a process of unifying the multiple names that hold the same concept by a common factor. As an example, Lee Chan-jin, the founder of Haansoft is known as "President Lee Chan-jin" within the company but from outside he is known as "the founder of Haansoft." The two titles both indicate the same person but when they are searched on the current search engine the results do not have any connections to each other. This is why the mapping technique is necessary to support the standardization process.

(2) Semantic Connection

By establishing the relation between data the search engine will be able to determine whether A and B have any connection to each other and provide the user with fast and accurate information. The ontology technology must be used here in order to benefit from such semantic connection processes.

Let's take an example of a search for "major league Chanho Park" for a better understanding of the semantic technique. With the current search engines, the frequency and arrangement of words and the reliability, importance and popularity of the web site are referred to along with category information to restrict the result to a certain category when searching for pages containing "major league Chanho Park," "major league and Chanho Park" and "major league or Chanho Park." This method does not go through semantic analysis of what specific meaning "major league Chanho Park" has to the user and simply suggests the "official Chanho Park site" or the "Chanho Park Fans' Site." But when the semantic method is applied, "Chanho Park" is not just considered as neighboring words next to "major league" but is conceived to refer to a famous baseball player. This is the main difference between the current search model and the semantic search model. Through the semantic method Chanho Park is conceived as a baseball player and also a famous celebrity. In the case where a document contains the name Chanho Park within a context on baseball, the related metadata would include a tag or baseball course that could be referred to or inferred in the text. In short, information can be searched more efficiently and accurately by applying the semantic connection method to establish relations between data.

3.2. An Overview of Information Flow Process

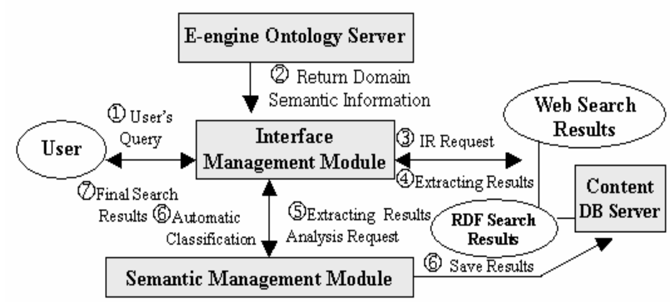


Figure 3. An overview of the SW-IQS search process

Figure 3 shows the information flow process of an SW-IQS. When the user inputs the desired search information, the Interface Management Module searches for related pages on the web using the Domain Semantics information returned by the E-engine Ontology Server. Then, the Semantic Management Module automatically classifies and ranks the searched pages and the Content DB Server's RDF documents to provide the final search results to the user.

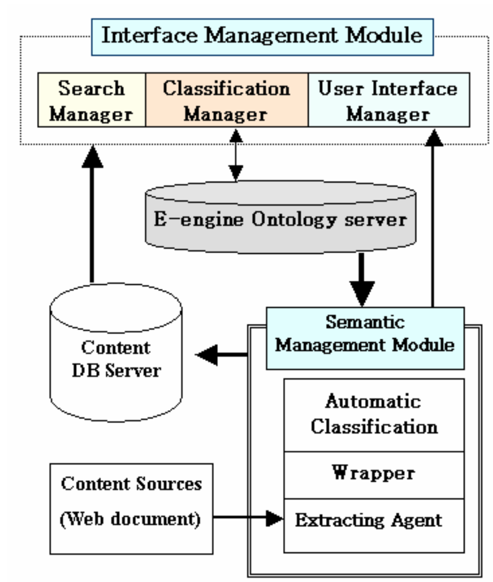


Figure 4. System architecture

3.3 The System Architecture and the Function of Each Module

As shown in Figure 4, the architecture of the SW-IQS is composed of Interface Management Module, E-engine Ontology Server, Content DB Server and Semantic Management Module. This section will describe the functions and characteristics of its modules in detail.

3.3.1 E-engine Ontology Server

E-engine Ontology Server [2] is a systematic method of expression that can improve the present condition where information is processed simply as data and the semantic context must be provided by man, and allow information to have value as knowledge.

In other words, two different data can possess the same meaning and syntax if they are observed semantically. For example, the tag for the data "movie" can be expressed either as <movie>, <cinema> or <screen>. So the three tags actually have the same meaning and they must be integrated into a single abstract schema or a semantic connection technique must be applied. Here the World Map is in charge of such standard rules and ontology. It is a key factor of the semantic web for it is the foundation that enables each individual factor to have a common understanding through artificial intelligence knowledge expression. The ontology can be classified into the domain ontology, the metadata ontology and the conceptual ontology. The domain ontology is an ontology restricted to a certain field and the metadata ontology is an ontology that is used for describing the contents of the online information resource. Meanwhile the conceptual ontology, like the frame ontology, is a meta-level ontology that expresses the object-oriented concepts such as frame and slot that are used in expressing concepts in an object-oriented way. This ontology is used for expressing concepts but not concepts from a certain field but rather different concepts.

It is composed of three layers, the Content Manager, Schema Manager and the Thesaurus Manager. The Content Manager applies the definition of the semantic metadata and the definition of the classification model for semantic data search along with succession and equivalence to define the relation between different metadata. The Schema Manager defines the standard data type and format of the Content Manager's standard classification model and the Thesaurus Manager's semantic integration model. The Thesaurus Manager is like an encyclopedia. It defines the identification and property standards in accordance with the international standards for electronic commerce. The Thesaurus Manager integrates schema and unifies and reorganizes similar terms. In other words, it is in charge of integrating terms that are semantically the same.

3.3.2 Interface Management Module

The current web search engine is not able to discriminate whether the search word "OEM" stands for "original equipment manufacturer," "Macromedia Korea's Partner : Partner Program for B2B partnership" or the "Object Embedding Method". This is because the current web search engine does not consider the semantic connection of the query. For this reason, the metadata factors that connect the semantic notes with the content information must be defined.

(1) Search Manager

The Search Manager induces accurate search results by bringing the related domain semantic information from the ontology server based on the query input by the user and requesting the user for a second query. In other words, when a multiple number of domain semantic information has been found the manager suggests the subject words and descriptions of the semantic information for the user to select.

(2) Classification Manager

The Classification Manager's method of searching by subject is quite distinctive compared to the current layered structure method. Searching by subject determines the relations between terms based on the RDF documents and enables more precise and efficient search for documents by applying a flexible network structure.

(3) User Interface Manger

The User Interface Manager provides various users' search input screens, ontology information selecting screens and final information search results screens.

3.3.3 Semantic Management Module

As with the Semantic Management Module, the information-extracting agent is used for extracting related web pages, and then the wrapper is used to return XML documents based on the material. After this, the Automatic Classification Module is used to automatically classify the pages and then the results are stored in the Content DB Server. Here, the similarity of the web pages must be measured in order to automatically classify and rank them. In this study, the ***term relationship variable***, a measurement of the synonymous relation between terms(*j*), and the ***Semantic Distance variable***, the measurement of the relationship or proximity according to distance between terms are used. The ***term relationship variable*** uses the similarity level (the range of similarity) of each term to measure the similarity. The range is 1 to 9, those leveling close to 1 have high similarity while those leveling closer to 9 have low similarity. The factors that discriminate the levels are the search word, terms contained in the extracted documents, synonyms of the search word extracted through ontology and the synonyms of terms contained in the documents extracted through ontology. The definition of ***term relationship*** is as follows.

Definition 1: ***term relationship*** (synonymous relations between each term)

$$R_j = \frac{f_{ij}}{t_r}$$

f_{ij} : The number of occurrence of term*(j)* in document*(i)*

t_r : Measurement variable for similarity between terms*(j)* = *level*(t_r)

The ***Semantic Distance Variable*** uses the proximity between each horizontal node ($\boldsymbol{H_p}$) and the proximity between each vertical node ($\boldsymbol{V_p}$) of each of the documents' structures. Figures 5 and 6 illustrate the comparison of the ***Semantic Distance*** measured for the XML documents and the RDF documents respectively.

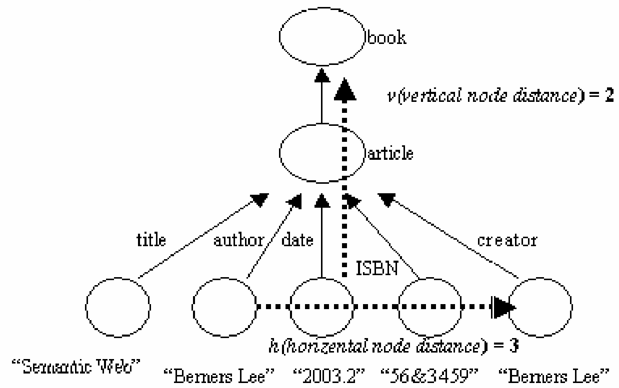


Figure 5. Semantic distance using XML document

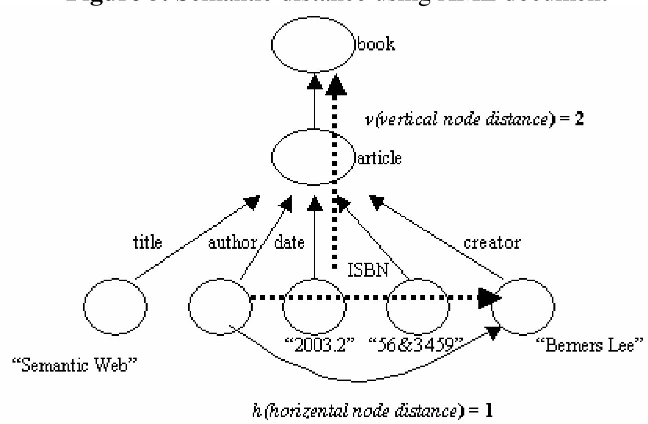


Figure 6. Semantic distance using RDF document

The reason why the *Semantic Distance Variable* differs between the XML and RDF documents is, because the XML document is based on a tree structured hierarchical method while the RDF document is based on the graph structured method. When the user is looking for "a book whose author and publisher are both Berners Lee," the horizontal node distance between "author" and "creator" in the case of RDF documents is "1," which means they are very closely related. Meanwhile, the same distance is "3" in the case of XML documents, which shows that "author" and "creator" are less related than in the former case. Different modeling methods are required for accessing and handling XML and RDF documents, which are formed in different structures. The XML document having a tree based hierarchical structure; it requires DOM (Document Object Model) to handle XML data. However, for the RDF document, a new modeling method is required that is capable of intensive analysis and expression of the attributes or relations between attributes and the relation between classes.

The definition of the ***Semantic Distance*** is as follows.

Definition 2: ***Semantic Distance*** (the relationship between terms)

$$D_j = H_p * V_p$$

$$H_p = \frac{1}{C^h}, h = |k - j|, C = \frac{level(i_j)}{\max V(i)}$$

H_p : The horizontal proximity between each node

h : Horizontal proximity between each term

C : The measurement variable for the level of each tree

$level(i_j)$: The level value of the term *(j)*'s location in document *(i)*

max V(i) : The highest-level value in document *(i)*

$$V_p = \frac{1}{F^v}$$

V_p : The vertical proximity between each node

F : The vertical proximity determining factor ($0<F<1$)

$V : level(i_k) - level(i_j)$: The vertical node distance between each term

And finally, the proportionally applied value of weight ($\boldsymbol{k_j}$) for automatic classification and ranking is as follows.

Definition 3: Proportionally applied value of weight ($\boldsymbol{k_j}$)

$$k_j = \frac{R_j}{D_j}$$

$\boldsymbol{R_j}$: term relationship variable measuring the synonymous relation between terms *(i)*

$\boldsymbol{D_j}$: Semantic Distance variable measuring relationship between terms

In the next section, a new ranking algorithm using the proportionally applied value of weight ($\boldsymbol{k_j}$) is suggested.

IV. Evaluation of Performances

4.1 The Ranking Algorithm

Former method [2] exclude semi-structural and non-structural documents, which possibly may contain meaningful information, and only refers to whether semantic information (RDF documents) are included or not to determine the similarity and ranking. This method rather lowers the preciseness of the ranking. Regarding this, this study attempts to rank documents by the ranking algorithm given in (3), which applies the proportionally applied value of weight ($\boldsymbol{k_j}$) and the cosine similarity of the vector model, and does not classify and extract the documents.

$$sim(d_j q) = k_j \times \frac{\vec{d}_j . \vec{q}}{|\vec{d}_j| \times |\vec{q}|} k_j = \frac{R_j}{D_j}. \qquad (3)$$

4.2 Comparison and Analysis

For evaluation of performance, "the book which the author is Burners Lee" was searched on the current search engine Google [14] and the documents listed from one to ten were selected to test the proposed model. The URL's of the selected documents are as listed in Table 1 and XML. RDF documents are also added to the test documents. In the left box, we added the document numbers to explain our test results.

No	URL
0	http://ec.cse.cau.ac.kr/okchoi/rdf.xml
1	http://ec.cse.cau.ac.kr/okchoi/person_book.rdf
2	http://static.userland.com/userLandDiscussArchive/msg011396.html
3	http://www.amazon.co.uk/exec/obidos/ASIN/1587990180/202-4832593-9905459
4	http://www.w3.org/People/Berners-Lee/1996/ppf.html
5	http://lists.w3.org/Archives/Public/www-talk/1992MarApr/0030.html
6	http://www.aarp.org/computers-books/Articles/a2002-06-26-weaving
7	http://www.bookbrowse.com/dyn_/title/titleID/125.htm
8	http://www.businessweek.com/bwplus/books/bookauth1.htm
9	http://www.freelists.org/archives/hacknot/07-2002/msg00004.html
10	http://www.addall.com/Browse/Author/2688303-1
11	http://search.barnesandnoble.com/booksearch/

Table 1. Web pages for testing

The comparison of performance is practiced according to a new step-by-step searching method (Figure 7) and the procedures are described below.

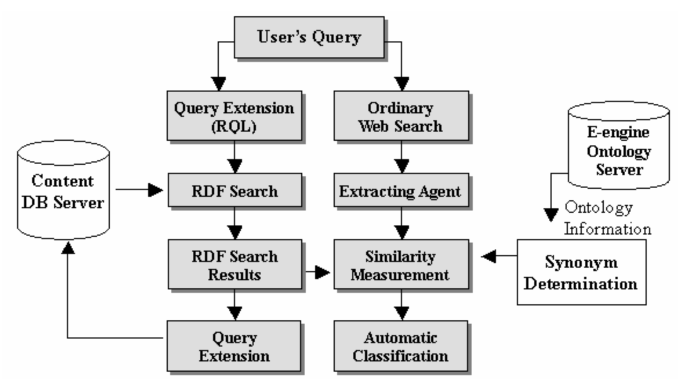


Figure 7. Step-by-step search method using ranking algorithm

Step 1: The top ten documents are searched through the ordinary search engine with the search word "the book which the author is Burners Lee". The unnecessary web pages are extracted from the searched pages by the extraction agent. Here, document numbers 3, 9 and 11 are removed from the web page list because there are broken links.

Step 2: The RDF document is searched by using RQL. As a result, a person_book.rdf document 4, Appendix A] is searched. Figure 8 shows how to find the RDF documents which are satisfied with the conditions. From this method, we could find the exact result.

Step 3: The vector based cosine similarity is calculated for the seven web documents and two XML and RDF documents, which were obtained through the extraction agent.

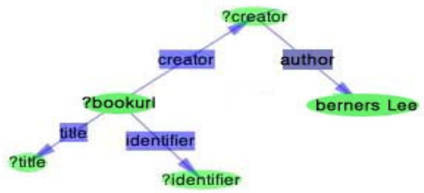
Step 4: The HTML documents are converted to XML documents by using the HTMLtoXML Wrapper in order to measure the relationship between terms.

Step 5: Then using the new ranking algorithm, the similarity cosine is measured for the XML documents and the RDF document.

As observed in Table 2, the ranking was different from the ranking results using the cosine similarity. Based on the difference, the ranking for each document is readjusted and the documents are automatically classified.

The currently used search engine does not take the weight of words and synonyms and the relationships into account at all. Furthermore, as observed in Table 1, the RDF document was ranked at the 4th, which is relatively lower than the rankings of ordinary web documents, when the vector based cosine similarity was used. This is because in the case when the vector model's cosine similarity is used, no application are made of the ***term relationship variable***, which measures the synonymous closeness between terms (***i***), and the ***Semantic Distance variable***, which measures the relationship between terms, that is the proximity measured by the distance between two terms. Meanwhile, when the new ranking algorithm, an improvement of the current vector model, is applied for measuring similarity, the RDF document that had been ranked at 4 rose to rank 1 and document number 0, which had been ranked at 1 with the vector model, was ranked lower at 3. The new ranking algorithm performed higher preciseness and recalls, which can be relied more by the user.

```
select ?bookurl, ?title
  where
  (dc::title ?bookurl ?title)
  (dc::creator ?bookurl ?creator)
  (dc::identifier ?bookurl ?uri)
  (foaf::author ?creator ?author)
  (foaf::author ?creator berners Lee)
  using dc for http://dublincore.org/2002/08/13/dces
  foaf for http://xmlns.com/foaf/0.1/
```



Bookurl	Title
Person_book.rdf	**Spinning the Semantic Web**
Person_book.rdf	**Weaving the Web**

Figure 8. Step-by-step RDF document searching method using RQL

Table 2. Similarity Measurement Results

Document No.	Term Weight				Cosine Similarity	Rank	k_j	**Similarity (Proposed method)**	**Final Rank (Ranking Algorithm)**
	Book	Author	Berners Lee	Book Author Berners Lee					
0	0	0	0	0	0	**8**	0	0	**8**
1	0	0.00117450	0.00033550	0.001510110	0.00075505	**4**	0.002064926	0.001409910	**1**
3	0.00695923	0.00046600	0.00232970	0.009755110	0.00487755	**0**	0.001553185	0.003215370	**0**
5	0.00067344	0.00050500	0.00018930	0.001367778	0.00068388	**5**	0.000399741	0.000541815	**6**
6	0.00153377	0.00089855	0.00089850	0.003330889	0.00166544	**1**	0.000994426	0.001329935	**3**
7	0.00054144	0.00008000	0.00015220	0.000769778	0.00038488	**6**	0.001152593	0.001329935	**5**
8	0.00165611	0.00058988	0.00052777	0.002773778	0.00138688	**3**	0.000910704	0.000768741	**1**
9	0.00004300	0.00004000	0.00004000	0.000122000	0.00006100	**7**	0.000053000	0.000057000	**7**
11	0.00037555	0.00223922	0.00006890	0.003295778	0.00149486	**2**	0.001205704	0.001350284	**2**

V. Conclusion and Further studies

The present study has suggested the SW-IQS composed of Semantic Management Module, Content DB Server, E-engine Ontology Server and Interface Management Module. As a solution to the problematic matters of the current search model, an integrated information searching system is suggested to enhance efficiency and preciseness of searching by information extraction and automatic classification, and to maximize the processing of both semi-structured and unstructured documents. This integrated information searching system uses the semantic web factors, which are emerging as the next generation web. With the system, semantic data search and integration are possible through establishment of ontology, data standardization, data integration and semantic connection methods. Through the establishment of the ontology, data standardization, data integration and the semantic connection method, the semantic web search model is capable of semantic data search and integration.

In order to evaluate the performance of the proposed integrated searching system, the formerly suggested method [2] was improved and a new performance-evaluating algorithm has been suggested to verify the efficiency and preciseness of the system.

The suggested system was tested and analyzed by applying the new ranking algorithm. Results showed that the ranking of the documents extracted from the web and documents stored in the Content DB Server is more efficient and precise compared to the former method.

For further studies, the proposed system's Content DB Server will be modified and implemented in the electronic transaction framework for e-business to enable semantic data integration and to activate information services.

Acknowledgment

This research was supported by the MIC (Ministry of Information and Communication), Korea, under the Chung-Ang University HNRC-ITRC (Home Network Research Center) support program supervised by the IITA-Institute of Information Technology Assessment.

References

[1] The RDF Query Language (RQL), http://139.91.183.30:9090/RDF/RQL/.

[2] Okkyung Choi, Seokhyun Yoon, Myeongeun Oh. Sanygoung Han, "Semantic web Search Model for information retrieval of the semantic data", HSI 2003, pp 588-593, June 2003

[3] Vagan Terziyan, Oleksandr Kononenko, "Semantic Web Enabled Web Services:State-of-Art and Industrial Challenges" Int. Conference on Web Services Europe (ICWS-Europe'03), Erfurt, Germany, Sept.23-25, 2003

[4] RDF document,

http://ec.cse.cau.ac.kr/okchoi/person_book.rdf
[5] Tim Berners-Lee, "Why RDF model is different from the XML model", http://www.ilrt.bris.ac.uk/discovery/rdf/resources/#sec-examples , October 1998,
[6] Lee Gang-chan's Standard Research Team, "W3C's Strategies for the Next Generation Web", ITFIND Weekly Technology Magazine, 2002
[7] Kim Yeong-cheon et al, "Research on Enhancing Information Search by Re-giving Term Weight", Korea Fuzzy Logic and Intelligence Systems Society, Vol. 11 No. 9, pp. 811-816, 2001
[8] "The Semantic Web: An Invisible Innovation of Space", *MicroSoftware* feature article, April 2002
[9] Forum on the Standardization of Electronic Commerce, "Electronic Commerce Standard Operation System Applying the Semantic Web Technology", ECIF, 2001
[10] Lassila, O.; van Harmelen, F.; Horrocks, I.; Hendler, J.; McGuinness, D.L. , "The semantic Web and its languages", IEEE Intelligent Systems , Volume: 15 Issue: 6 , pp. 67 –73, Nov.-Dec. 2000
[11] Decker, S.; Mitra, P.; Melnik, S., "Framework for the semantic Web:an RDF tutorial", IEEE Internet Computing , Volume: 4 Issue: 6 , pp. 68 –73, Nov.-Dec. 2000
[12] Decker, S.; Melnik, S.; van Harmelen, F.; Fensel, D.; Klein, M.; Broekstra, J.; Erdmann, M.; Horrocks, I., "The Semantic Web:the roles of XML and RDF", IEEE Internet Computing , Volume: 4 Issue: 5 , pp. 63 –73, Sept.-Oct. 2000
[13] Kim Yeong-cheon et al, "Research on Enhancing Information Search by Regiving Term Weight", Korea Fuzzy Logic and Intelligence Systems Society, Vol. 11 No. 9, pp. 811-816, 2001
[14] Google, http://www.google.com/en
[15] A Toolkit for Creating Vector Models from Text Documents, http://www.cs.utexas.edu/users/jfan/dm/index.html
[16] FOAF: the 'friend of a friend' vocabulary, http://xmlns.com/foaf/0.1/
[17] M. Paolucci, T. Kawamura, T. Payne, and K. Sycara. "Semantic matching of web services capabilities", In *Proc. Of the 1st International Semantic Web Conference (ISWC)*, Lecture Notes In Computer Science; Vol. 2342, pp.333-347, 2002.
[18] Lei Li and Ian Horrocks, "A Software Framework For Matchmaking Based on Semantic Web Technology", Proceedings of the 12th international conference on World Wide Web, Budapest, Hungary, pp.331-339, 2003
[19] A. Ankolekar, M. Burstein, J. R. Hobbs, O. Lassila, D. L. Martin, S. A. McIlraith, S. Narayanan, M. Paolucci, T. Payne, K. Sycara, and H. Zeng. Daml-s: Semantic markup for web services. In Proc. of the International Semantic Web Workshop, 2001
[20] Amit Sheth, Clemens Bertram, David Avant, Brian Hammond, Krysztof Kchut, Yashodhan Warke, "Managing Semantic Content for the Web", IEEE Internet Computing, Vol. 6, No. 4, pp.80-87, July/August 2001
[21] XOL Ontology Exchange Language, http://www.ai.sri.com/pkarp/xol/
[22] SHOE, http://www.cs.umd.edu/projects/plus/SHOE/
[23] OML, http://www.ontologos.org/OML/
[24] Ontology Inference Layer OIL, http://www.ontoknowledge.org/oil/
[25] Daniel A.Menasce, "QoS Issues in Web Services", IEEE Internet Computing, Vol. 6, Issue 6, pp.72–75, Nov. 2002
[26] M. Nodine, W. Bohrer, and A. Ngu, "Semantic multibrokering over dynamic heterogeneous data sources in InfoSleuth.", Proceedings of the Fifteenth International Conference on Data Engineering. Sydney, IEEE Computer Society Press, pp. 358-365, 1999.
[27] I. Horrocks, U. Sattler, and S. Tobies, "Practical reasoning for expressive description logics. In H. Ganzinger", D. McAllester, and A. Voronkov, editors, LPAR'99, Springer-Verlag, pp.161–180, 1999.
[28] M. Paolucci, J. Soudry, N. Srinivasan, and K. Sycara, "A Broker for OWL-S Web Services", in Cavedon, Maamar, Martin, Benatallah, (eds) Extending Web Services Technologies: the use of Multi-Agent Approaches, Kluwer, 2005.

Author Biographies

Okkyung Choi is currently a Ph.D. Student in the Department of Computer Engineering, ChungAng University, Seoul, Korea. She received Master of Science degree at ChungAng University. From 1996-98, she worked as a Computer Programmer at Samsung Electronics, and wrote software for the Switchboard Interface System using MFC (VC++ 5.0) and constructed the intranet system named Technical Information Services.

Her research focuses on the design and implementation of Semantic Web Services System, e-Commerce and Auction System. In these fields she has published more than 30 papers in journals and conference proceedings.

Sangyong Han is a professor of the school of computer science and engineering, ChungAng University, Seoul, Korea. He received Bachelor of Engineering in College of Engineering from Seoul National University in 1975, and the Ph.D. degree in College of Engineering, University of Minnesota in 1984. From 1984 to 1995, he worked at Poughkeepsie Lab and Waston Research Center in IBM, USA.

His research interests include Web Technologies, Web Services, Semantic Web, Information Retrieval and Multimedia.

Ajith Abraham currently works as a Distinguished Visiting Professor under the South Korean Government's Institute of Information Technology Assessment (IITA) Professorship programme at Chung-Ang University, Korea. His primary research interests are in computational intelligence with a focus on using evolutionary computation techniques for designing intelligent paradigms. Application areas include several real world knowledge-mining applications like Web services, information security, Web intelligence, financial modeling, multi criteria decision-making etc. He has authored/co-authored over 200 research publications in peer reviewed reputed journals, book chapters and conference proceedings of which three have won 'best paper' awards.

He is serving the Editorial board of over a dozen International Journals and has also guest edited 12 special issues for reputed International Journals. He received PhD degree from Monash University, Australia. More information at: http://www.softcomputing.net

Appendix A

-person_book.rdf

```
<rdf:RDF

xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:foaf="http://xmlns.com/foaf/0.1/">
<foaf:Person>
<foaf:author>Edd Dumbill</foaf:author>
<foaf:mbox rdf:resource="mailto:edd@xml.com" />
<foaf:title> O'Reilly Network</foaf:title>
</foaf:Person>

<foaf:Person>
<foaf:author>Simon St.Laurent</foaf:author>
<foaf:mbox rdf:resource="mailto:simon@xmlhack.com" />
<foaf:title> Programming Web Services with XML-RPC
</foaf:title>
</foaf:Person>

<foaf:Person>
<foaf:author>Berners Lee</foaf:author>
<foaf:mbox rdf:resource="mailto:eric@xmlhack.com" />
<foaf:title>Weaving The Web</foaf:title>
<foaf:subtitle>The Original Design and Ultimate Destiny of
the World Wide Web by Its Inventor</foaf:subtitle>
</foaf:Person>

<foaf:Person>
<foaf:author>Berners Lee</foaf:name>
<foaf:mbox rdf:resource="mailto:eric@xmlhack.com" />
<foaf:title>>Spinning the Semantic Web</foaf:title>
<foaf:subtitle>Bringing the World Wide Web to Its Full
Potential</foaf:subtitle>
</foaf:Person>

</rdf:RDF>
```