A Study on Transforming ICT Research Information Service into Semantic Web Environment

Jong-Cheol Song, Byung-Joo Moon and Hoe-Kyung Jung, *Member, KIMICS*

Abstract—The Research on the ICT(Information & Communication Technology) is proposed the category to IT839 strategy by Government. Government is driving to researching on technology about IT839 Strategy. By transforming this category and research information into Semantic Web environment, it is possible to search function utilizing knowledge base and information object by use of TBox and ABox. In this regard, this study proposes technology for generation of Semantic Web Document about ICT Research Information. The ontology is constructed by using category to IT839 Strategy. The features of framework proposed in this study is to have used a skill to directly map Ontology instance and in case of inability of direct mapping, proposed a skill to establish reliable Semantic Web Document by suggesting indirect mapping skill using mechanical study. In addition, it is possible to establish low cost/high quality Semantic Web Document about ICT research information.

Index Terms—Semantic Web, Ontology, OWL, Metadata

I. INTRODUCTION

Semantic Web was suggested to resolve the limitations of current HTML based web of user centered info exchange by 'style information'. It is to say that Semantic Web is targeting at the technology with which computer or software agent can understand and process web information as people do.

Tim Berners Lee defines Semantic Web as the web that a computer understands and manipulates the meaning of the information [1]. Semantic Web uses ontology, the knowledge representation schemes. Ontology can be said as a concept system which defines the relation between shared concepts [2]. Also, semantic web is more of a next generation web technology which understands, integrate and reuses not only the information representation for people but also information object for agent [3]. Ontology, which is Semantic Web's core technology, has emerged as a means of representing and integrating knowledge in e-commerce, medical information, and bio informatics areas.

However, one of its applications in information service area is semantic web portal has only limited application of semantic web and the relevant research is also limited to ontology. In other words, it only provides the research function on ontology, but fails to deliver the semantic web documentation functions needed to effectively transform the conventional web environment into a semantic web based one.

Therefore, all the extraction of significant information currently provided, construction of ontology containing the category and logics with the capability of processing semantic information within a system, and suggestions of transformation scheme which converts conventional web based information service system into another one which creates and provides semantic web document lowers the barrier to semantic web while encouraging the use of semantic web documents.

In this regard, this paper is aiming at suggesting a system which creates semantic web document for the information resources included in a current system, utilizing constructed ontology by building ontology with significant information with which provided the users in the form of interface in a system managing IT839 strategy and the relevant research information now under active research.

Simple transformation of HTML based web information into a semantic web based information service system does not exclude conventional HTML based web documents from service provision. While the conventional HTML based web documents are continuously on service, information system which can accept semantic web documents would be constructed to effectively undertake information management, distribution and utilization. This is the principle idea of this research.

The system recommended in this research has Dublin Core based Meta Data constructed as DB and restricts the environment to where information object is managed by the documentation unit.

This research consists of the following. The second chapter suggests the scenario which transforms web environment into a semantic web environment. The third chapter designs the initial stage of the generation of semantic web documents. The fourth chapter builds the environment for the simulation of the suggested system and tests it. The fifth chapter seeks the future research

Manuscript received July 4, 2007.

Jong-Cheol Song is with the Department of Information Service Team , IITA, Daejeon, 305-348, Korea (Tel: +82-42-710-1384, Fax: +82-42-710-1389, Email: jcsong@iita.re.kr), Byung-Joo Moon(Email: bjmoon@iita.re.kr), Hoe-Kyung Jung(Email: hkjung@pcu.ac.kr)

direction and reaches a conclusion.

II. THE SCENARIO ON TRANDFORMING WEB INTO SEMANTIC WEB ENVIRONMENT

In this chapter, the paper suggests semantic web based information service transformation scenario to converse conventional information system into a semantic web based one. The suggested transfor-mation scenario is limited to the information service system as following.

First, it has constructed Meta information for information object to be utilized as a search engine. Second, the menu system is well standardized to replace classify-cation system. Third, more direct-ories or categories than category classification exist for information object.

This paper take the premise that the transformation schemes for semantic web environment is the documents level by information object units.

This paper suggests scenario in which factors making up a system is utilized in the conventional information service while trying to add necessary factors of the system to process semantic web documents. It also wants to use Meta information constructed to build or manage conventional web information. Also, the newly constructed web information provides the function of ontology or mapping that the expanded construction of semantic web documents was applied to change the information service environment. This paper's transformation scenario can be found in the Fig.1.

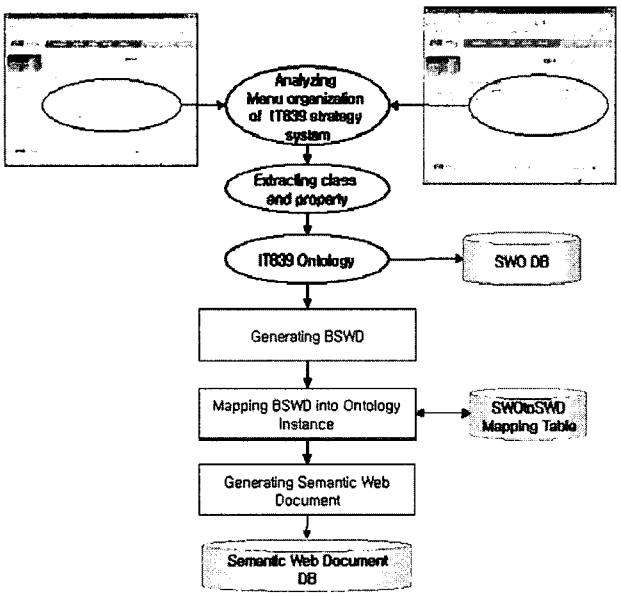


Fig. 1 The Scenario on Transforming Web Environment into Semantic Web Environment.

The transformation scenario in the Fig. 1 first extracts and analyzes hierarchy, inclusion, equivalent, and the reverse by analyzing the menu system and classification system relation in the information system. Second, the extracted factors utilize ontology technology language such as OWL to build reference Semantic Web Ontology, or SWO. Third, the constructed Meta information is used to create Basic Semantic Web Document, BSWD. Fourth, those created BSWD is mapped with SWO to create Semantic Web Document, SWD. As OWL Lite contains class classification system and simple restriction phrase for effective transformation, OWL Lite is used to build SWO.

III. THE DESIGN FOR SEMANTIC WEB DOCUMENT GENERATOR

The generation stage of semantic web document suggested in this paper was developed by using the algorism which creates semantic web document with Meta information built on the information service system and applicable ontology.

The generation stage of the semantic web document is made up of semantic web document generation module which creates semantic web document in a constructed database, and the other semantic web generation module for newly registered information object.

The Fig. 2 shows the semantic web generation process for the newly registered information. Registered Meta information through CMS and the original text is stored in a Meta DB. Stored Meta DB is used to create BSWD document and begins to map SWO which built reference ontology. BSWD is a documents written by the use of XML and RDF, and it also contains Dublin Core based Meta data.

The mapping of BSWD and SWO chooses SWO domain and provides the input person with the instances which could be mapped with BSWD among other class instance or property instance, so that BSWD and SWO could be mapped.

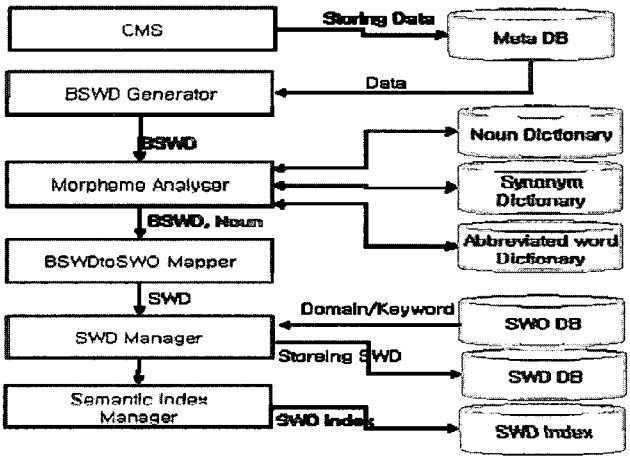


Fig. 2 Generating Semantic Web Document about input information

The semantic web document transformation device is seen in the Fig 3. The device transforms established Meta DB into BSWD in information

service system and creates semantic web documents in alliance with reference ontology. Semantic web document transformation device can only be used when classification system is well established and the category and ontology class and characteristics can be well matched.

The semantic web document transformation device for Meta DB as in the Fig. 3 has the functions as below. To create BSWD from Meta DB, DB schemato DC mapping device, DC based element generator, and triple construct generator are used. The reference ontology and BSWD element's Tag Set are mapped with the characteristics and class of ontology using the category operated in the information service system. Mapping information and BSWD are used to create semantic web documents.

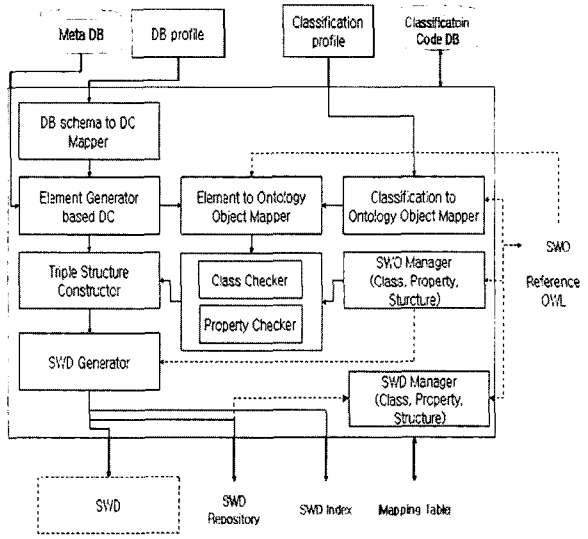


Fig. 3 Generating Semantic Web Document Using metaDB

BSWDtoSWO mapping device primarily uses direct mapping technique and secondly uses indirect mapping techniques using mechanical study toward information object and ontology instance as in the Fig. 4. This double structure provides the channel which can correctly link information object to ontology instance.

BSWDtoSWO mapping device divides the Meta DB into a noun unit to create key words, and uses those key wards in a thesaurus or abbreviation dictionary to map with SWO so that BSWD can be mapped with SWO. The Meta DB based mapping method starts from the premise as below to build a model. The terminology representing a certain category is already been selected and included in the ontology when it was constructed. Also, the representative domain of the registered information is already chosen.

Under the assumption, the Meta DB registered in the mapping device uses morpheme analyzer to analyze and extract nouns. Extracted nouns SWO DB and the hit rates' threshold is applied to recommend ontology term to the register person. The class instance of SWO DB is selected by the register person to map the register information with ontology.

When the SWO and BSWD are not mapped, the mechanical study is utilized to undertake the mapping of BSWD and SWO. Mechanical study uses Baysian model[4] which has already been frequently used in many studies for document classification. The Baysian model uses the term shows up in a granted Meta DB for a single information resource to calculate the rates that each ontology class and characteristics is in sync. When the biggest rates come up, the ontology of the class and characteristics has the Meta DB allocated and mapped with the information resource.

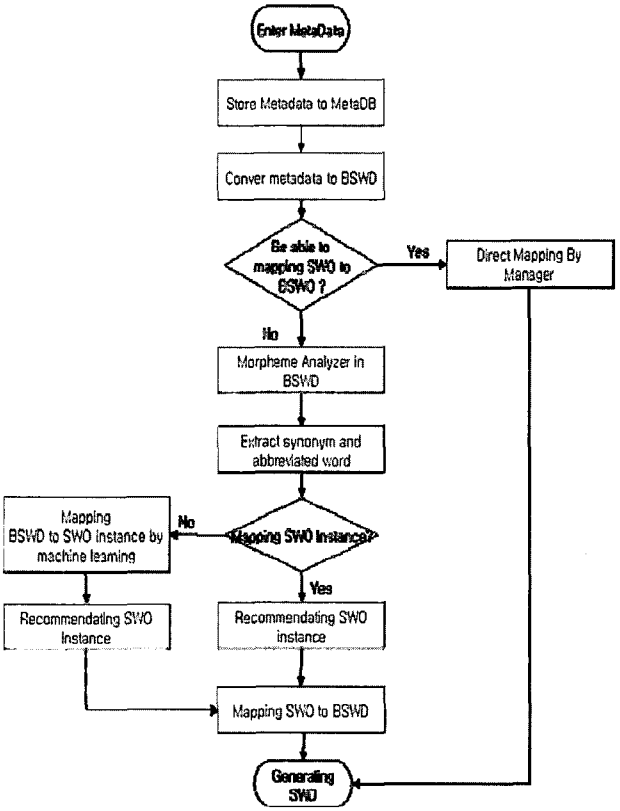


Fig. 4 Algorithm on the SWD Generation

Fig. 5 uses the classification system called IT839 strategy provided in a metadb.net and shows the example of constructing ontology. When the class structure for IT839 strategy ontology has been completed, it consists of object property with information type and service provider. In other words, hasPattern is the object property which has the information type of information object and hasOffer has the object property toward the service provider held by information object. The value of hasPattern is paper, technology report, market report, patent, and standard. hasOffer is lab, university, software company or hardware companies.

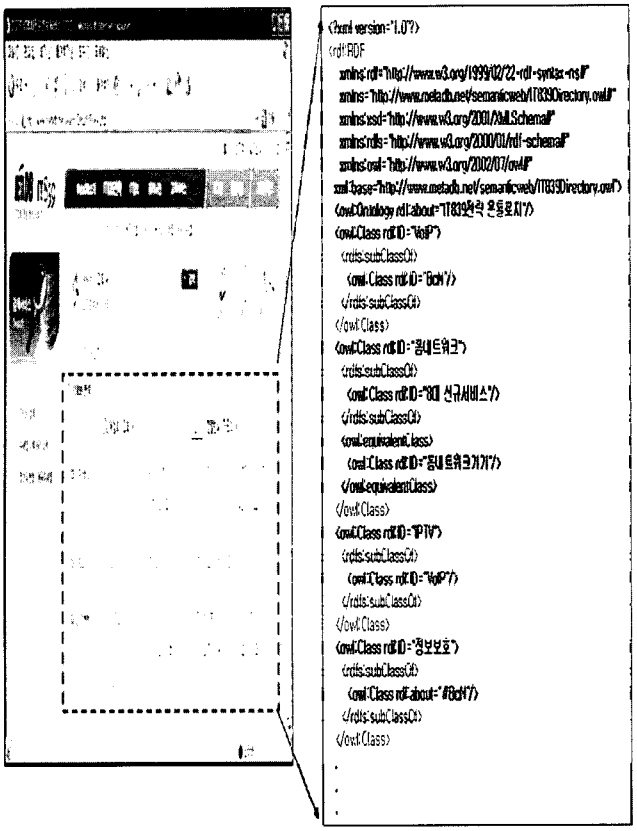


Fig. 5 IT839 Ontology

IT839 strategy ontology has the characteristics information of information type and service provider to identify the information type and service provider of IT839 strategy information object. Fig. 5 is the case of object property of IT839 strategy ontology.

IV. SIMULATION

IT839 strategy ontology uses Protégé to build IT839 strategy classification system, also using IT technology factors. Java language was used along with Jena 2.0[5,6] for the storage, management and reasoning of ontology.

Table 1 System development and simulation environment

Development Tool	Description
System	IBM PC Compatibility
OS	MS Windows XP
Development Language	Java
Ontology Generator	Protégé 3.1 OWL Plugin 2.1
Development Tool Kit	Jena 2.0
DataBase	MySQL

Semantic web document generator analyzes morpheme from the Meta DB field which was newly registered information and includes the algorism which matches the extracted noun to the ontology instance, Fig. 6 shows the algorism included in the semantic web document generator.

```
1: SWDTransform(MetaInformation md){
2: BSWD tripledoc = new BSWD();
3: tripledoc.generate(md);
4:MorphologicalAnaly mar = new
MorphologicalAnaly(tripledoc) ;
5:Machinelearning ml= new Machinelearning();
6: tripledoc.inputnoun(mar.noun);
7: tripledoc.inputsynonym(mar.synonym);
8: tripledoc.inputabbreviate(mar.abbreviate);
9:SWOManager swom = new SWOManager(swo);
10:BSWDtoSWOMap map = new
BSWDtoSWOMap(tirpledoc, swo);
11: if(map.mapping()){
12:     SWD swdm = new SWD(map.extract());
13:     swdm.print();
14:     while(!swdm.num)
15:       if(swdm.yes()) swdm.store();
16:       swdm.print();   }
17:  else{
18:        ml.match(tripedoc);
19:        ml.print();
20:        while(!ml.yes())   mal.print();
21:  swdm.store();  }   }
```

Fig. 6 Algorithm on the Semantic Web Document Generation

Semantic web document generator is composed of BSWD generator, BSWDtoSWO mapping device, SWD manager, and semantic index manager, but it is also connected to morpheme analyzer. The Fig. 6 and semantic web document generator's makeup factors can be applied in the explanation. BSWD generator is the module creating basic semantic web document and corresponds to line number two and three.

Generated BSWD is put in the morpheme analyzer to extract nouns and the relevant synonyms and abbreviated words to the extracted nouns are also extracted. Also extracted nouns and synonyms, and abbreviated words are included in BSWD. This corresponds with line number 4 to 8. In the stage when the extracted nouns are used to match ontology instance number 9 and 10 are correspondent. When the extracted nouns and ontology instance are matched, the final mapping decision was reached by the manager. In the next stage that the mapping information is stored in the SWD DB by SWD manager is correspondent to line number 11 to 16. When the extracted nouns do not match the ontology instance, mechanical study is used for the mapping recommendation to the manager for the mapping selection and to store SWD DB. In this stage, line numbers from 17 to 21 are correspondent.

V. CONCLUSIONS

This paper suggested a transformation system from average web system into the semantic web environment.

The goal of semantic web technology application in the information service system is to more effectively search and distribute information using semantic information x. In terms of semantic web, the ontology construction and search engine function might be at the core of the effective development, but in terms of information service system, it is only one other functions which helps ontology information service to undertake more effectively. Rather, in regard to the information service system, the ontology in the semantic web and semantic web information processing capabilities should be loaded in the systems to more effectively transform information in conventional systems into semantic one, which in turn is to build ontology based information search functions along with semantic web document generation and distribution functions.

Accordingly, this paper suggested ways to build highly reliable semantic web document by direct or indirect mapping techniques on information object and ontology instance to generate and transform semantic web document, thus to suggest an alternative to build low cost and high quality semantic web document.

REFERENCES

[1] Tim Berners-Lee, James Hendler and Ora Lassila, "Scientific American article : The Semantic Web", Scientific American Issue, 2001.

[2] Gruber T.R, "A Translation Approach to Portabl Ontology Specifications", Knowledge Acquisition, vol 5 Num 2, pp.199-220, 2004.

[3] Joong-in Choi, "The overview and the research environment of Semantic web", Information Technology Academy, Vol. 21 Num. 3, pp.4-9, 2003.

[4] Mehran Sahami, "Learning Limited Dependence Bayesian Classifiers", Proc. Of the 2nd International Conference on KDD'96, 1996.

[5] Stanford Medical Informatics, "Protégé-Frames and Protégé-OWL", from http://protege.stanford.edu/.

[6] HP, "Jena-A Semantic Web Framework for Java", from http://jena.sourceforge.net/.

[7] Li Ding, rong Pan, Tim Finin, Anupam Joshi, Yun Pen, and Pranam Kolari, "Finding and Ranking Knowledge on the Semantic Web" from http://ebiquity.umbc.edu/papers/, 2005.

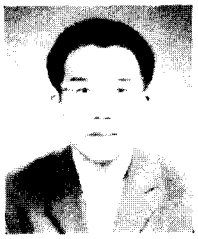
[8] Ebiquity, "Swoogle semantic web search 2006", from http://swoogle.umbc.edu.

[9] W3C, "Resource Description Framework (RDF)", from http://www.w3.org/RDF.

[10] W3C, "RDF Vocabulary Description Language 1.0:RDF Schema" from http://www.w3.org/TR/rdf-schema.

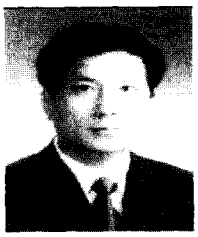
Jong-Cheol Song

He Received B.S degree in the Department of Computer Engineering from Kwangwoon University, Korea, in 1999, and he received Ph. D. degree in the Department of Computer Engineering from Paichai University, Korea, in 2007. During 1999-2003, he worked for ETRI as Researcher. Since 2003, he has worked in Information Serve Team at Institute for Information Technology Advancement(IITA), where he now works as a Senior Researcher. His current research interests Semantic Web and Information Retrieval.



Byung-Joo Moon

He Received B.S degree in the Department of Computer Science from Chungbuk University, Korea, in 2000. During 1991-2003, he worked for ETRI as Senior Researcher. Since 2003, he has worked in Information Serve Team at Institute for Information Technology Advancement(IITA), where he now works as s Team Leader. His current research interests Web 2.0, Information Retrieval, Information Clustering.



Hoe-Kyung Jung

He Received B.S degree in 1987, and Ph. D. degree in 1993, in the Department of Computer Engineering from Kwangwoon University, Korea. During 1994-2005, he worked for ETRI as Researcher. Since 1994, he has worked in the department of Computer Engineering at Paichai University where he now works as a Professor. His current research interests Multimedia Document Architecture Modeling, Information Processing, Web Services, Semantic Web, USN and MPEG-21.