A Study on Constructing the Ontology of LIS Journal*

문헌정보학 학술지를 대상으로 한 온톨로지 구축에 관한 연구

Younghee Noh**

ABSTRACT

This study constructed an ontology targeting journal articles and evaluated its performance. Also, the performance of a triple structure ontology was compared with the knowledge base of an inverted index file designed for a simple keyword search engine. The coverage was three years of articles published in the Journal of the Korean Society for Information Management from 2007 to 2009. Protégé was used to construct an ontology, whilst utilizing an inverted index file to compare performance. The concept ontology was manually established, and the bibliography ontology was automatically constructed to produce an OWL concept ontology and an OWL bibliography ontology, respectively. This study compared the performance of the knowledge base of the ontology, using the Jena search engine with the performance of an inverted index file using the Lucene search engine. As a result, The Lucene showed higher precision rate, but Jena showed higher recall rate.

초 록

본 연구에서는 학술지를 대상으로 온톨로지를 구축하고 그 성능을 평가하고자 하였으며, 트리플 구조로 구축된 온톨로지의 성능을 단순 키워드 검색엔진을 위한 도치색인 파일의 지식베이스와 그 성능을 비교하였다. 온톨로지 구축대상은 정보관리학회지 2007년부터 2009년까지의 3년간의 논문기사를 대상으로 하였으며, 구축방법은 온톨로지 구축도구인 프로티지를 이용하였다. 개념온톨로지는 수작업으로 구축하였고, 서지온톨로지는 자동으로 구축하여 각각 OWL 개념온톨로지와 OWL 서지온톨로지를 생성하였다. 성능비교를 위해 각각 제나 검색엔진과 루씬 검색엔진에 의해 검색된 결과를 비교하였다. 루씬은 정확률이 높게 나왔고, 제나는 재현률이 높게 나왔다.

Keywords: ontology, jena, lucene, semantic retrieval, library and information science, semantic web
온톨로지, 제나, 루씬, 시맨틱 검색, 문헌정보학, 시맨틱 웹

---

* This work was supported by a National Research Foundation of Korea Grant funded by the Korean Government(KRF-2009-327-H00016).
** Associate Professor in Department of Library & Information Science, Konkuk University (irs4u@kku.ac.kr)
■ Received : 17 May 2011 ■ Revised : 24 May 2011 ■ Accepted : 3 June 2011
■ Journal of the Korean Society for Information Management, 28(2): 177-193, 2011.
〔DOI:10.3743/KOSIM.2011.28.2.177〕

1. Introduction

1.1 Research Goals and Needs

Today's Internet system is undoubtedly the outcome of many people's efforts; but, in the beginning, it was the effort of one person, Tim Bernes-Lee. He founded the World Wide Web as a tool that could instantly exchange information for collaborative research projects amongst physicists spread across numerous continents and different universities or research labs. It could also be said that when Tim Berners-Lee suggested the Semantic Web, he was aiming to initiate the next stage of the web; it is the most promising next generation web technology that could expedite the 2nd information technology revolution.

The Semantic Web is a new technology created to provide efficient retrieval, integration and re-use of information by constructing the knowledge base from machine-readable definitions among the terms. Constructing Semantic web requires various key technologies and conceptions such as the Uniform Resource Identifier (URI) system that endows unique, fixed identification of the source; data elements that could prevent conflicts of meaning in the elements produced in individual information institutions; Resource Description Framework (RDF) schema that are fundaments for the element relation definition of class that are applicable to such data features; DAML+OIL, a web ontology language that strengthened logical reasoning and expressiveness in the RDF schema; and Web Ontology Language (OWL) that deletes, modifies and supplements constructors.

OWL is the language that defines the warehouse of knowledge in accordance with web ontology. The warehouse of knowledge could be said a set of propositions accumulated from the inference system. The propositions not only include facts of their constituent elements but also include logically inducible facts that were not clearly defined in the ontology in sentence structure format.

However, at the moment, language resources－thesauruses, dictionaries, ontologies and so on－for constructing and researching projects in Korea are limited to specific domains (for example: the areas of computing and medicine), and the methodologies are mostly experimental (Lee & Yoon 2011). Also, there is a great need for studies on developing efficient and new language resources that could be applied in various sectors of industry. Thesaurus building and ontology projects are relying on manual execution by numerous experts and specialists; however, the conflicting opinions or ambiguity of concepts found amongst those who are responsible for defining such measures has impeded tangible results in building language resources and information retrieval projects.

Furthermore, due to the specific demands of the industry, such projects tend to require massive budgets, a large pool of specialists, and quite a long development time as well. This is why there is a need to find efficient methodologies that can overcome the difficulties found in the conventional practices of the industry.

This study constructed an ontology, targeting scholarly journal articles, and then evaluated its performance to see whether it would be feasible to expand it and to apply it to other journals for scholarly communication. For this purpose, the current study compared the performance of a triple structure ontology with that of an inverted index file-based indexing system, which is the most widely used at the moment.

1.2 Methodology

Ontology building methodology can be conducted in various ways; semantic retrieval is possible based on a triple structure ontology. The search engine implemented through this study provided knowledge retrieval of desirable results through logical inference. This can be done by expressing procedural relations added to the declarative relations based on the ontology structure. The study methodology and processing stages are detailed below.

First, this study reviewed previous scholarly research on the Semantic Web, ontology, inference schemes and relevant theoretical achievements. This review went beyond simply gathering previous concepts; it performed a comparative analysis of previous research. Second, an actual ontology was constructed. The target was three years of scholarly articles published in the *Journal of the Korean Society for Information Management* from 2007 to 2009. Protégé was used to construct an ontology, whilst utilizing an inverted index file to compare performance. The Jena and Lucene search engines were individually customized to compare performances of the constructed ontology.

2. Previous Studies

There have been numerous and active academic studies worldwide, including studies conducted in Korea, on the Semantic Web and ontology. When ontology was first adopted in Korea, it was mainly applied to the field of medicine (Jeong et al. 2002; Lee & Lee 2003). Lim (2004) compared keyword-based web document searches, which endowed the weight only by using term frequency information, with ontology-based web document searches, which utilized relevant feedback on the information within an ontology. Lim evaluated the performance of two factors: recall rate and precision rate. This was to prove the efficiency of hierarchical relations within an ontology in web document searches.

Lee and others (2005) and Sim (2005) presented electronic catalogue ontology modeling by using predicate logic. Their thesis conceptually expressed and manifested fundamental components for modeling by using Extended Entity Relationship (EER). Their study suggested a methodology that expressed knowledge and additional meanings only in description logic format. However, this thesis did not propose a specified and formalized process for ontology modeling, thus creating difficulties in sharing and understanding information expression.

Oh (2005) emphasized the necessity for the various ontology modeling methodologies for development of the Semantic Web and in the domestic ontology field. However, Oh's study did not include suggesting an ontology modeling methodology. Then, Kim (2008) suggested a web ontology modeling methodology based on predicate logic and rule language to infer and to represent web ontology, which is the key technology in the Semantic Web application. Kim suggested a web ontology modeling hierarchy for building a web ontology as advocated by the Semantic Web, and his recommended modeling methodology expressed information based on TBox and ABox structures and the SWRL of predicate logic according to individual hierarchy level. Kim verified his web ontology performance by utilizing DL Inference of SPARQL and TopBraid.

As for the international studies, Cranefield and others proposed ontology medoling based on Unified Modeling Language (UML). However, there is a great need for a modeling methodology for knowledge inference among ontology properties, which are essential elements for expressing information in web ontology (Cranefield, Haustein, & Purvis 2001).

There was a study that provided an Ontology Definition Metamodel (ODM) based on an Model Driven Architecture (MDA) to connect ontology modeling and the Semantic Web (Djuric, Gasevic, & Devedzic 2005). They proposed four stages in ontology modeling: the meta-metamodel stage, metamodel stage, model stage and instance stage. They expressed knowledge by utilizing UML at the meta-metamodel stage and the model stage; and the result was then mapped with knowledge of OWL through ODM. However, their study showed knowledge while focusing on UML as Cranefield did in his study, lacking a provision for knowledge inference between properties, which is inconvenient because one has to readopt ODM-like language to express the knowledge.

Posada and others (2005) proposed an industry-standardized ontology modeling design by utilizing Protégé, an ontology editing tool, to a visualize mass-scale model; however, what they suggested was not actually an ontology modeling tool, and they only proved convenience of the mass-scale model's visualization upon adopting an ontology. Cui and others (2004) suggested an ontology methodology on the Web service's Semantic synthesis. However, their study insisted on an overall process of modeling methodology that was necessary to provide Web service rather than proposing a modeling methodology for an ontology.

As for studies on relations between concepts in a thesaurus system, Han Sang-kil's (1999, 2000) study can be cited. Han analyzed basic and additional relations between 20 thesauruses found in Korea and worldwide. With these, Han overcame the limits found amongst relations among thesauruses and concentrated on expanding such relations for the online search environment.

Kim (2001) performed an experimental study developing a thesaurus by importing term definitions. Nam (2004) expanded equivalence relation and synonym relation, which previous thesauruses had failed to capture accurately. Kim tried to further construct in-depth meaning relations among descriptors in thesauruses.

Jeong (2003) concentrated on the correlation between thesauruses and ontology, and established part of the Art and Architecture Thesaurus (AAT) in Web Ontology Language (OWL), which is facet-type thesaurus. Jeong's study examined ways to define new relations between concepts that were not yet scrutinized through previously existing online thesauruses by utilizing axioms of ontology.

Cho and Nam (2004) verified the compatibility between thesauruses and ontology by adopting OWL Lite's vocabulary for the thesaurus's basic relations. This study also suggested how to build a multilingual thesaurus. While most of the studies produced in Korea could not deal with the inference system of ontology, Soergel and colleagues (2004) were able to analyze relations between thesaurus concepts, proposing new relation types between concepts and, at the same time, proposing new discourses in the potential of inference rule applications.

The following are examples of ontology that are being renewed or were already established:

Mikrokosmos is an information-based machine-translation system developed by New Mexico State University in the United States, sponsored by the US Department of Defense. This system aims at massive scale but practical machine translation. This system allows high quality meaning analysis using five thousand of concepts and seven thousand Spanish articles (Mahesh 1996).

HowNet is a Chinese ontology built to develop a Chinese-English machine-translation system. This system includes a total of 53,000 terms from a Chinese dictionary and 57,000 terms from an English dictionary, but its shortcoming is that this system is Chinese reliant (Dong 1999).

Cyc adopted artificial intelligence technology, as a collection of commonsense database, which was built 10 years ago by The Microelectronics and Computer Technology Corporation (MCC) to provide quasi-human or human-like inference. Thesuperordinate Cyc ontology holds 3,000 general concepts, which are then connected to large volumes of subordinate facts. This system has approximately 1 billion concepts and 1 million facts or rules (Lent 1995).

WordNet is widely used in many sectors of the natural language processing and information retrieval industry and uses English language vocabulary data based on word-relations. Based on achievements of psycholinguistic research on human vocabulary knowledge, the Princeton University Cognitive Science Laboratory has been building this system since 1985. Unlike a dictionary that enumerates words by words, WordNet features a network system in which the meaning of the word constitutes the network. At the moment, this system has a vocabulary of 1.4 billion, and there have been attempts to make this system multilingual (Miller 1990).

The most well-known example of an ontology built in Korea is the Korean vocabulary database, an ETRI noun concept map, which connects various relations between concepts expressed by Korean nouns. This was built by the Electronics and Telecommunications Research Institute (ETRI) and has approximately 10,000 general nouns and 15,000 economy nouns. Researchers at ETRI are currently working to expand this system to include a verb concept map.

Corenet is built by KAIST based on a Word

List set from the National Institute for Japanese Language and Linguistics. This system connected meanings and concepts of words by setting up a concept system composed of lexical, semantic property systems, and connected significations and concepts of words. There are total of 2,938 hierarchical concepts and a total of 92,448 lexical and semantics (Choi 2001).

Kim's (2001) study is relevant to information science subjects. Kim established a simple ontology by utilizing previous ontology semantic information of research materials found at universities. Then, Kim studied whether and how relevant ontology could be mapped with Resource Description Framework (RDF) schema. However, Kim's study is concerned with RDF schema mapping methodology of ontology constructing materials, rather than focusing on the ontology itself, failing to provide in-depth analysis and insights for ontology construction.

On the other hand, Song (2002) established an ontology specifically for the field of information science and analyzed whether there is any difference in the efficiency of the semantic information of Song's own ontology compared to that of a thesaurus. Kwak (2005) proposed a basic model for establishing an ontology based on an information structure that is commonly shared by human. For this, Kwak scrutinized categories and relations of concepts as reported by humans. This study concluded that there was a difference between KDC or DDC structure and students' information structure, and that students' academic performance does influence information structure. These research studies provide information for building an ontology for the information science field.

3. Implementing Ontology

3.1 Tool for Constructing Ontology

This study utilized the Protégé 4.1 Beta version (Protégé 2011) to implement an ontology. Protégé was developed by the Stanford Center for Medical Informatics Research at the Stanford University School of Medicine after 15 years of studies compiling information-based structure systems, which is supported by grant LM007885 from the United States National Library of Medicine; the National Science Foundation (NSF) and the Defense Advanced Research Projects Agency (DARPA) have also provided support for the Protégé project.

A key tendency in the latest development in ontology is expanding scopes of applicability so the technology can be used by a wider audience in a variety of industries. There are approximately one thousand projects that are using Protégé. Protégé is a free software, in accordance with the Mozilla Public License and has broad compatibility with other knowledge representation languages.

3.2 Ontology Constructing Process

The target of ontology for this study is 189 scholarly articles published through the *Journal of the Korean Society for Information Management*, extracted from the DBpia of Nuri Media. The database includes a bibliography of articles, contents, abstracts and author information in a meta-data format and also includes the original document information. This study constructed meta-data from all the meta-data except the original document information. To enhance convenience in building the ontology, Chinese characters were converted to Korean characters.

This study adopted Methontology, which is the most widely and generally used methodology when building an ontology. This methodology has sub-categories of ontology building stages including: specification, conceptualization, formalization, implication and maintenance. Among them, specification and conceptualization will be examined in below.

3.2.1 Specification Stage

This stage is a preparation stage before building the ontology itself. This stage includes defining ontology usage and aim, scope, standardization stage and potential users.

1) Ontology Usage and Aim

The Library and Information Science (LIS) field is complex and interdisciplinary, combining both characteristics of the science of a) Classification, an academic field that is developing relatively slowly and steadily, and b) Information Science or Information Services, an academic field that is evolving relatively quickly. In this respect, building an ontology for the LIS field would enable us to test whether this ontology could be adapted to other academic fields. In other words, the trials and errors found or encountered in constructing the ontology for this specific yet complex field would inform others' attempts to build ontologies to be used for other complicated academic sectors. Furthermore, no one has yet implemented an ontology for the LIS field in Korea or abroad. Therefore, this study will critically guide later studies and will support further development in the information science industry.

2) Ontology Construction Scope

This study targeted 189 scholarly articles published in the *Journal of the Korean Society for Information Management*. The database includes the articles' bibliography, contents, abstract and information on authors.

3) Ontology standardization stage and potential users

There are four stages of detailing the conceptualization, and all instances belong to a class, and the lowest class should have an instance, and the name of the peripheral class can be same as the name of the instance. The potential users are researchers, professors, or students who study LIS field and librarians as well.

3.2.2 Conceptualization Stage

Conceptualization in Ontology means extracting comprehensive concepts that could represent common properties of certain phenomena or an object. Conceptualization is something of a process rather than end product. There are 11 subdivisions in the conceptualization stage of Methontology, but one doesn't have to go through every substage. This study went through 6 substages:

1) Constructing the Glossary

The glossary includes all types and forms found in and belong to the domain, including any concept, instance, property, verb and any forms. This study enumerates terms/words found in the LIS field that builds up the ontology, excluding meaningless conjunctions, suffixes or special symbols.

2) Concept Classification Tree

This is to define temporary relations included in the glossary. The relations between concepts are expressed by using subclass relations, relative prime relations and others (See Table 1).

3) Building the Concept Lexicon

This process aims to express all concepts included in the domain to be built, including instances of class, concepts of class, property of instances, relations between concepts, synonyms, and disjoint paths and forms. This study included the name of concepts, synonyms, acronyms, relative primes and forms.

4) Building the Class Property Table

This table concretely describs class properties expressed in the concept lexicon. The property of a class is described as the concept itself rather than as the instance of the concept. Therefore, the result includes name of concepts, possible value formats, numerical value evaluation units and precision of values.

5) Instance Property Table

This is to further concretely describe the properties of instances expressed in the concept lexicon. The table declares the properties of instances toward the concept and the property of instances themselves, whilst also defining the value of the properties. In other words, the instance property table includes the name of the instance, name of the relevant classes, the name of the same instance (homonym) and the name of the instance in relative prime relation.

6) Relations table

This stage concretely describes the relation between class and instance. The relations table's instance property in the LIS field table shows the name of relation, relevant class names, subclass name(s), mandatory and multiple choices. Please see below for an example of relation name.

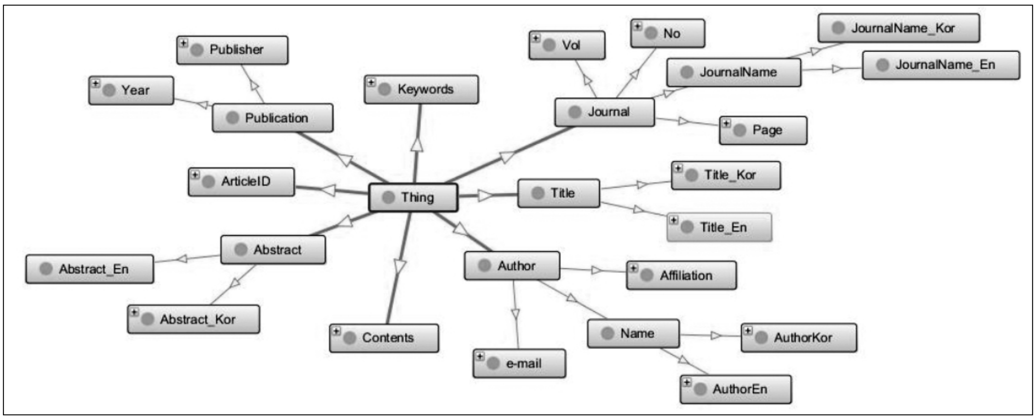
〈Table 1〉 Relation Table(Meanings of bibliography and concept's object property)

Object Property	Domain	Co-domain	Constraint	Meaning
hasKeyword	Article ID	Keywords	some	Article ID Class has Keywords class as its constituent
hasContents	Article ID	Content	some	Article ID Class has Content class as its constituent
hasJournal	Article ID	Journal	exactly 1	Article ID Class has only one Journal class as its constituent

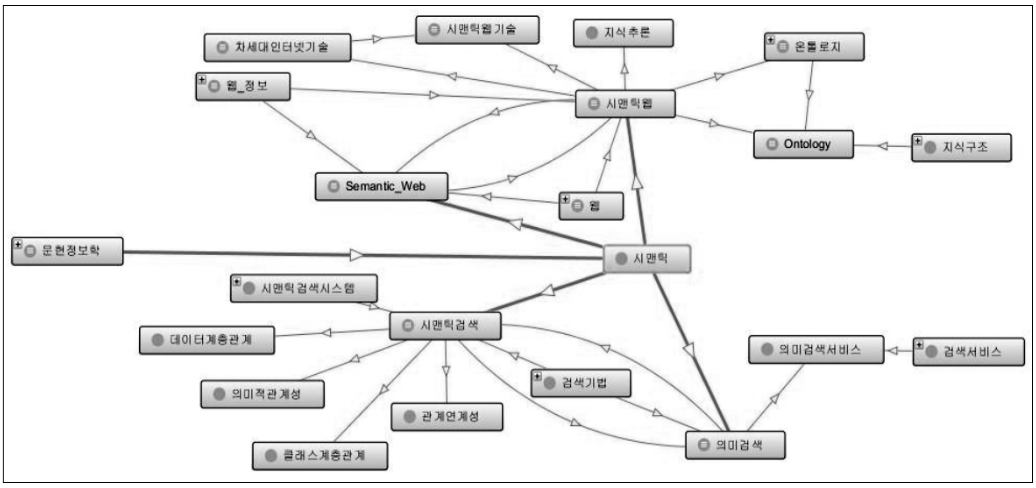
Object Property	Domain	Co-domain	Constraint	Meaning
hasNumber	Journal	No.	exactly 1	Journal Class has only one No. class as its constituent
hasVolum	Journal	Vol.	exactly 1	Journal Class has only one Vol. class as its constituent
hasPage	Journal	Page	exactly 1	Journal Class has only one Page class as its constituent
hasJournalName	Journal	JournalName	exactly 1	Journal Class has only one JournalName class as its constituent
hasJournalName En	JournalName	JournalName En	exactly 1	JournalName Class has only one JournalNameEn class as its constituent
hasJournalName Kor	JournalName	JournalName Kor	exactly 1	JournalName Class has only one JournalNameKor class as its constituent
hasTitle	Article ID	Title	exactly 1	Article ID Class has only one Title class as its constituent
hasTitle En	Title	Title En	exactly 1	Title Class has only one TitleEn class as its constituent
hasTitle Kor	Title	Title Kor	exactly 1	Title Class has only one TitleKor class as its constituent
hasAuthor	Article ID	Author	some	Article ID Class has Author class as its constituent
hasName	Author	Name	some	Author Class has Name class is its constituent
hasAuthorName En	Name	AuthorName En	some	Name Class has AuthorNameEn class as its constituent
hasAuthorName Kor	Name	AuthorName Kor	some	Name Class has AuthorNameKor class as its constituent
hasEmail	Author	Email	exactly 1	Author Class has only one Email class as its constituent
hasAffiliation	Author	Affiliation	some	Author Class has Affiliation class as its constituent
hasAbstract	Article ID	Abstract	exactly 1	Article ID Class has only one Abstract class as its constituent
hasAbstract En	Abstract	Abstract En	exactly 1	Abstract Class has only one Abstract En class as its constituent
hasAbstract Kor	Abstract	Abstract Kor	exactly 1	Abstract Class has only one Abstract Kor class as its constituent
hasPublication	Article ID	Publication	exactly 1	Article ID Class has only one Publication class as its constituent
hasPublisher	Publication	Publisher	exactly 1	Publication Class has only one Publisher class as its constituent
hasYear	Publication	Year	exactly 1	Publication Class has only one Year class as its constituent
hasInfluence	Domain	Domain	some	Domain Class has Domain class as its constituent
AnalyzedBy	Domain	Domain	some	Domain Class has Domain class as its constituent
AppliedTo	Domain	Domain	some	Domain Class has Domain class as its constituent
ComponentOf	Domain	Domain	some	Domain Class has Domain class as its constituent
HeadedBy	Domain	Domain	some	Domain Class has Domain class as its constituent
InfluencedBy	Domain	Domain	some	Domain Class has Domain class as its constituent
MeasuredBy	Domain	Domain	some	Domain Class has Domain class as its constituent
OperatedBy	Domain	Domain	some	Domain Class has Domain class as its constituent
Usingby	Domain	Domain	some	Domain Class has Domain class as its constituent
RelatedTo	Domain	Domain	some	Domain Class has Domain class as its constituent
ResultOf	Domain	Domain	some	Domain Class has Domain class as its constituent
hasComponent	Domain	Domain	some	Domain Class has Domain class as its constituent
hasExample	Domain	Domain	some	Domain Class has Domain class as its constituent
hasFunction	Domain	Domain	some	Domain Class has Domain class as its constituent
hasMember	Domain	Domain	some	Domain Class has Domain class as its constituent
hasResult	Domain	Domain	some	Domain Class has Domain class as its constituent
hasUsed	Domain	Domain	some	Domain Class has Domain class as its constituent

3.3 Ontology Construction Results

Figure 1 and Figure 2 show the ontology structure established for this thesis. Figure 1 shows the ontology structure for the bibliography information of the journal articles, and Figure 2 shows the concept ontology (domain ontology). The bibliographic ontology describes information and relations of the journal articles, also showing property and property relations for the journal articles to enable search. The concept ontology referenced the Korean Decimal Classification (KDC), thesauruses of LIS and related fields and previously developed ontologies from similar science fields to find out the relations among concepts.



〈Figure 1〉 Ontology for Bibliography



〈Figure 2〉 Example of Concept Ontology

4. Ontology's Performance Evaluation

4.1 Data Structure

Experimental data is composed of 189 cases of journal data published in the *Journal of the Korean Society for Information Management* for a three-year span. This summary information shows bibliography information for individual articles including the title of the journal article (Table 2).

Based on this, an ontology in XML format was produced, and the relation between the XML tag and the OWL article ontology was induced based on the relation table seen in Table 1. Also, each individual class defines the property that expresses the relation with other classes and also describes the limits of instance that comprise the relevant class. This is called class description, and Figure 3 is an example of this.

〈Table 2〉 Journal Date Structure Example

논문ID	824138	
한글제목	인터넷 지식거래소와 저작권에 관한 연구	
영문제목	A Study on Internet Knowledge Markets and Copyright Issues in Korea	
저자	노영희 Younghee Noh	
출판사항	출판사	정보관리학회
	출판년	2007. 3. 1
저널정보	권	제 24권
	호	제 1호
	페이지	121～145
주제키워드	지식거래소; 저작권법; 학술지; 논문; knowledge market; copyright law; thesis article; academic journal	
초록	본 연구는 현재 상업적인 지식거래소를 통해서 유통되고 있는 다양한 유형의 지식콘텐츠들의 저작권 문제를 저작권법에 비추어 분석하고 있다. 지식거래소에 정보를 제공하는 주체는 수많은 개인, 원문DB제공업체, 국가 및 공공기관, 출판사 등 매우 다양하다. 그러나 이러한 정보제공주체와 지식거래소간에 이루어지는 저작물의 상업적인 거래에서 정작 원저작자인 개인 저자들은 빠져있다는 점을 주목할 필요가 있다. 원칙적으로 모든 저작물의 저작권은 원저작자에게 있으며 원저작자가 저작권 이양 동의서를 통하여 2차적 저작물 생성권 등을 포함한 모든 권리를 양도하지 않는 이상 저작물의 디지털화, 원문DB화 및 지식거래소를 통한 유통 등은 저작권에 위배된다는 결론을 내릴 수 있다. This study aims to identify copyright issues regarding the knowledge content currently circulated through knowledge exchange markets in the Republic of Korea. The content providers of knowledge exchange markets comprise government ; public institutions, full-text database companies, publishers and individuals. It is worth noting that commercial trade of copyrighted content or material among academic journals, database companies and knowledge exchange markets essentially exclude individual authors who are the actual copyright holders. In principle, the original author owns the copyright whether it has an explicit notice or not. Unless the author/owner officially agrees to transfer the copyright including the right for so-called "derivative works", content-making based or derived from the copyrighted material, digitalization of the copyrighted work as well as its registration on full-text database and circulation through knowledge markets are illegal.	
URL	http://dbpia.co.kr/view/ar_view.asp?arid=824138	

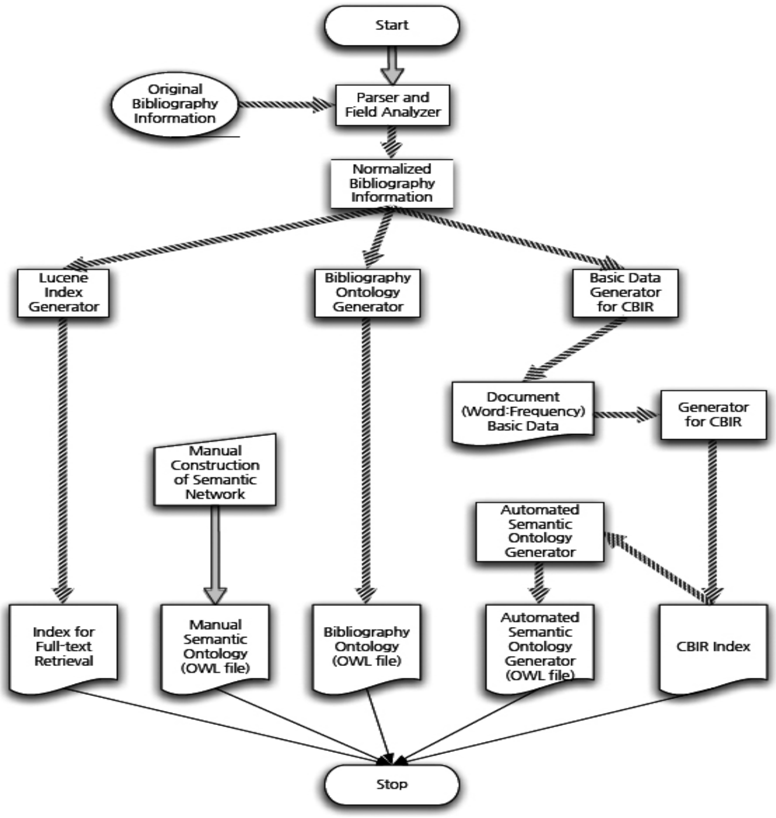
```xml
<?xml version="1.0"?>
    <!--
    ///////////////////////////////////////////////////////////////////////////////////////
    //
    // Individuals
    //
    ///////////////////////////////////////////////////////////////////////////////////////
     -->

                                ~ 중략 ~

    <!-- http://www.semanticweb.org/ontologies/2011/2/Ontology1300787436640.owl#824138 -->

    <owl:NamedIndividual rdf:about="&Ontology1300787436640;824138">
        <rdf:type rdf:resource="&Ontology1300787436640;ArticleID"/>
        <hasNumber rdf:resource="&Ontology1300787436640;1"/>
        <hasContents rdf:resource="&Ontology1300787436640;1._서론;"/>
        <hasPage rdf:resource="&Ontology1300787436640;121~145"/>
        <hasContents rdf:resource="&Ontology1300787436640;2._지식거래소의 지식거래 유형;"/>
        <hasYear rdf:resource="&Ontology1300787436640;2007"/>
        <hasVolume rdf:resource="&Ontology1300787436640;24"/>
        <hasContents rdf:resource="&Ontology1300787436640;3._지식거래유통과 관련된 법률;"/>
        <hasContents rdf:resource="&Ontology1300787436640;4._거래유형에 따른 저작권문제;"/>
        <hasContents rdf:resource="&Ontology1300787436640;5._결론 및 제언;"/>
        <hasArticle rdf:resource="&Ontology1300787436640;A_Study_on_Internet_Knowledge_Markets_and_Copyright_Issues_in_Korea"/>
        <hasJournal rdf:resource="&Ontology1300787436640;Journal_of_the_Korean_Society_for_Information_Management"/>
        <hasAbstract rdf:resource="&Ontology1300787436640;This_study_aims_to identify copyright issues regarding the knowledge content currently circulated
through  knowledge exchange markets in the Republic of Korea. The content providers of knowledge exchange markets comprise government &amp; public
institutions, full-text database companies, publishers and individuals. It is worth noting that commercial trade of copyrighted content or material among
academic journals,  database companies and knowledge exchange markets essentially exclude individual authors who are the actual copyright holders. In
principle, the original author owns the copyright whether it has an explicit notice or not. Unless the author/owner officially agrees to transfer the
copyright including the right for so-called derivative works, content-making based or derived from the copyrighted material, digitalization of the
copyrighted work as well as its  registration on full-text database and circulation through knowledge markets are illegal"/>
        <hasAuthorEn rdf:resource="&Ontology1300787436640;Younghee_Noh"/>
        <hasKeywords rdf:resource="&Ontology1300787436640;academic_journal"/>
        <hasKeywords rdf:resource="&Ontology1300787436640;copyright_law"/>
        <hasKeywords rdf:resource="&Ontology1300787436640;knowledge_market"/>
        <hasKeywords rdf:resource="&Ontology1300787436640;thesis_article"/>
        <hasAuthorKor rdf:resource="&Ontology1300787436640;노영희;"/>
        <hasNumber rdf:resource="&Ontology1300787436640;1;"/>
        <hasKeywords rdf:resource="&Ontology1300787436640;지식거래소;"/>
        <hasAbstract rdf:resource="&Ontology1300787436640;본 연구는 현재 상업적인 지식거래소를 통해서 유통되고 있는 다양한 유형의 지식콘텐츠들의 저작권 문제
를 저작권법에 비추어 분석하고 있다. 지식거래소에 정보를 제공하는 주체는 수많은 개인, 원문DB제공업체, 국가 및 공공기관, 출판사 등 매우 다양하다. 그러나 이러한
정보제공주체와 지식 거래소간에 이루어지는 저작물의 상업적인 거래에서 정작 원저작자인 개인 저자들은 빠져있다는 점을 주목할 필요가 있다. 원칙적으로 모든 저작물
의 저작권은 원저작자에게 있으며 원저작자가 저작권 이양 동의서를 통하여 2차적 저작물 생성권 등을 포함한 모든 권리를 양도하지 않는 이상 저작물의 디지털화, 원문
DB화 및 지식거래소를 통한 유통 등은 저작권에 위배된다는 결론을 내릴 수 있다.;"/>
        <hasArticle rdf:resource="&Ontology1300787436640;인터넷 지식거래소와 저작권에 관한 연구;"/>
        <hasKeywords rdf:resource="&Ontology1300787436640;저작권법;"/>
        <hasJournal rdf:resource="&Ontology1300787436640;정보관리학회지;"/>
        <hasKeywords rdf:resource="&Ontology1300787436640;학술지;"/>
        <hasContents rdf:resource="&Ontology1300787436640;참고문헌;"/>
        <hasKeywords rdf:resource="&Ontology1300787436640;논문;"/>
        <hasPublicasher rdf:resource="&Ontology1300787436640;한국정보관리학회;"/>
    </owl:NamedIndividual>
```


〈Figure 3〉 Example of Class Description for Bibliography Ontology

4.2 System Structure

The system structure for building the ontology for LIS is shown in Figure 4. The first step is to extract normalized bibliography information from the original through a parser and field analyzer to produce the ontology. The concept ontology was manually established, and the bibliography ontology was automatically constructed to produce the OWL concept ontology and the OWL bibliography ontology, respectively. To evaluate the performance, a Lucene automatic indexer was used to compile index files for full-text retrieval.

To evaluate the performance of the ontology constructed through this study, searches were made targeting the ontology and index file, each created by different methodologies. Lucene was chosen to search the general inverted index file. The Lucene search engine (2011) is an open-source search engine which enables index compilation and full-text searches by utilizing 100% Java. The Lucene search engine is the result of Apache's Jakarta project (Lucene 2011) and is developed in many different sources such as C, .net or Ruby, even though it started as Java. Lucene is one of the very famous search engines and is widely adopted in various known service sites such as Twitter.



〈Figure 4〉 Concept Ontology and Inverted Index File Constructing Process

Jena was used to evaluate the ontology. Jena is a Java framework designed to develop network-based framework applications and provides RDF, RDFS, OWL and SPARQL programming environments, including a rule-based inference engine. Jena's framework includes RDF API (Carroll et al. 2004; Kevin et al. 2003). Also, SPARQL includes the Semantic Web data access protocol and RDF query language. SPARQL is a SQL-like language type, preceded by RDQL. SPARQL has a Terse RDF triple language-based (Turtle-based) graph pattern grammar structure.

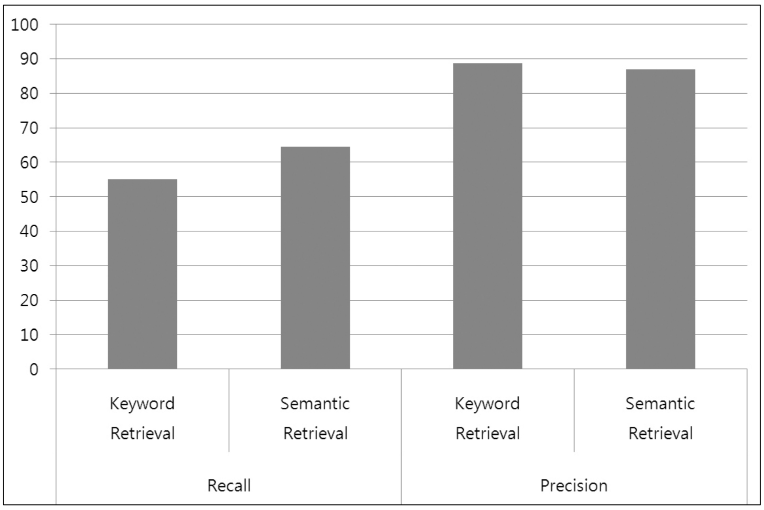
4.3 Performance Comparison Evaluation Result

This study compared the performance of the knowledge base of an ontology using the Jena search engine with the performance of an inverted index file using the Lucene search engine; then, the results were compared as shown in Table 3 and Figure 5.

As for the precision rate, keyword searches showed higher performance than semantic searches by 12.8%, meaning that there was no significant difference in precision rate between the two engines. As for the

〈Table 3〉 Performance Evaluation Comparison Table for Keyword Search and Semantic Search Results

Query	Recall		Precision	
	Keyword Retrieval	Semantic Retrieval	Keyword Retrieval	Semantic Retrieval
1	100	100	75	100
2	66.67	100	50	50
3	50	100	100	100
4	60	60	100	100
5	42.86	42.86	100	100
6	50	50	100	100
7	50	50	50	100
8	20	20	100	2.27
9	100	100	100	100
10	33.33	66.67	100	40
11	100	100	100	100
12	100	100	100	100
13	100	100	100	100
14	25	25	100	100
15	8.33	8.33	100	100
16	20	20	100	100
17	66.67	66.67	100	66.67
18	33.33	50	100	100
19	0	50	0	100
20	73.33	80	100	80
Average	54.98	64.48	88.75	86.95



〈Figure 5〉 Performance Evaluation Comparison Diagram for Keyword Search and Semantic Search Results recall rate, the semantic searches showed higher performance then keyword searches by 9.5% ; this was because semantic searches expanded the concepts, retrieving relevant subject words as query words.

5. Discussion and Conclusion

This study constructed an ontology for scholarly journal articles and evaluated its performance to see whether it could be expanded and adopted to large scale scholarly journals. The performance of the triple structured ontology was compared with that of inverted index file system, which is now most widely used. The ontology target for this study was 189 cases of scholarly journal articles published in the *Journal of the Korean Society for Information Management*, extracted from the DBpia of Nuri Media. The database included a bibliography of the articles, contents, abstracts and author's information in a meta-data format and also included original document information. This study constructed meta-data with all the meta-data except the original document information. To enhance convenience in building the ontology, Chinese characters were converted to Korean ones.

The first step was to extract normalized bibliography information from the original through a parser and field analyzer to produce the ontology. The concept ontology was manually established, and the bibliography ontology was automatically constructed to produce an OWL concept ontology and an OWL bibliography ontology, respectively. To evaluate the performance, a Lucene automatic indexer was used to compile index files for full-text retrieval. This study compared the performance of the knowledge base of the ontology, using the Jena search engine with the performance of an inverted index file using the Lucene search engine.

For the overall analysis of the performance evaluation comparison, the Lucene search engine excelled in precision but showed a low recall rate because it doesn't extend semantics. Therefore, it is appropriate for massive scale databases that require fast search results. On the other hand, the Jena search engine showed a high recall rate because semantics could be extended based on triple relations; however, it was a bit time consuming when preparing the framework to extend concepts from OWL.

This study manually constructed an ontology, and the limit of such manual implementation is difficulty in efficiently constructing knowledge information with resources from various industries in a semantic network format. Therefore, future research should focus on how to build an efficient semantic network (ontology) to enable related searches of the semantic network base.

References

Carroll, J. J., L. Dickinson, C. Dollin, D. Reynolds, A. Seaborne, A., and K. Wilkinosn. 2004. "Jena: Implementing the semantic web recommendations." *Proceedings of the 13th International World Wide Web Conference*, New York. 74-83.

Cho, Hyun-Yang and Young-Joon Nam. 2004. "A study on the interchangeability between a thesaurus and an ontology." *Journal of the Korean Society for Information Management*, 21(4): 27-47.

Chung, Hee-Joon, Myong-Hwan Yoo, Kang-Chang Lee, Jae-Hong Min, and In-Seong Chung. 2002. "Design of ontology system for the semantic web." *Korea Information Processing Society*, 9(2): 23-25.

Cranefield, S., Stefan Haustein, and Martin Purvis. 2001. "UML-Based ontology modelling for software agents." *The Information Science Discussion Paper Series*.

Cui, Juntao, Jiamao Liu, Yujin Wu, and Ning Gu. 2004. "An ontology modeling method in semantic composition of web services." *Proceedings of the IEEE International Conference on E-Commerce Technology for Dynamic E-Business* (CEC-East'04).

Dong, Z. and Q. Dong. 1999. "HowNet." [cited 2009. 12.30]. <http://www.keenage.com/zhiwang/e_zhiwang.html>.

Đuric, D., Gaševic, D., and Devedžic, V. 2005. "Ontology modeling and MDA." *Journal of Object Technology*, 4(1): 109-128.

Han, Sang-Kil. 1999. "A study on the basic term relationship of thesaurus." *Korea Library and Information Science Society*, 30(4): 107-136.

Han, Sung-kook and Hyun-sil Lee. 2007. "Architectural reference model for semantic library." *Journal of the Korean Society for Information Management*, 24(1): 75-101.

Jeong, Do-Heon and Tae-Soo Kim. 2003. "A study on the thesaurus-based ontology system for the semantic web." *Journal of the Korean Society for Information Management*, 20(3): 155-176.

Kevin, W., C. Sayers, and H. Kuno. 2003. "Efficient RDF storage and retrieval in Jena2." *Proceedings of First International Workshop on Semantic Web and Databases*, 131-151.

Kim, Su-Kyoung and Kee-Hong Ahn. 2008. "Web ontology modeling based on description logic and SWRL." *Journal of the Korean Society for Information Management*, 25(1): 149-171.

Kim, Tae-Soo. 2001. "A study on the development of thesaurus using terminological definitions." *Journal of the Korean Society for Information Management*, 18(2): 231-254.

Kim, Yi-Ran. 2001. *(A) Study on the Creation of RDF Schema Using the Semantic of Ontology*. Master's Thesis, Yonsei University Graduate

School, Library and Information Science.

Kwack, Chul-Wan. 2005. "Human knowledge structure analysis for designing ontologies: Using library and information science field." *Korea Library and Information Science Society*, 36(1): 459-477.

Lee, Hye-won and So-Young Yoon. 2011. "A study on the model of history ontology: A focus on Korean modern historical person." *Journal of the Korean Biblia Society for Library and Information Science*, 22(1): 263-280.

Lee, Hye-won and Tae-soo Kim. 2007. "A study of the extended model of event-aware ABC ontology for music resources." *Journal of the Korean Society for Information Management*, 24(1): 273-300.

Lee, Hyun-Ja and Jun-ho Shim. 2005. "Ontological modeling of E-catalogs using description logic." *Journal of KIISE*, 32(2): 111-119.

Lee, Hyun-Sil and Too Young Lee. 2003. "A study of the design of ontology-based prescription knowledge management system of oriental medicine." *Journal of the Korean Society for Information Management*, 20(1): 341-371.

Lenat, D. B. 1995. "Cyc: A large-scale investment in knowledge infrastructure." *The Communications of the ACM*, 38(11): 33-38.

Lim, Soo-Yeon. 2005. "Relevance feedback based on medicine ontology for retrieval performance improvement." *Journal of the Korean Society for Information Management*, 22(2): 41-56.

Lucene. 2011. <http://lucene.apache.org/java/docs/index.html, http://lucene.egloos>.

Mahesh, K. 1996. *Ontology Development for Machine Translation: Ideology and Methodology, NMSU.* Computing Research Laboratory. Technical Report MCCS-96-292. New Mexico.

Miller, G., C. Beckwith, D. Fellbaum, K. J. Gross. 1990. "Five papers on WordNet." *International Journal of Lexicography*, 3(4).

Nam, Young-Joon. 2004. "A study on the extend guideline for the equivalence relationship in thesaurus." *Journal of the Korean Society for Information Management*, 2(2): 1-21.

Oh, Hyun-mok. 2005. *(A) Study on Growth Direction and Development Strategies of Standardization for Semantic Web*. Korea National Computerization Agency.

Posada, J., Carlos Toro, Stefan Wundrak, and André Stork. 2005. "Ontology supported semantic simplification of large data sets of industrial plant CAD models for design review visualization." *Lecture Notes in Computer Science*, 3683(2005): 184-190.

Protégé. 2011. <http://pretege.stanford.edu>.

Soergel, D. 2004. "The Arts and Architecture Thesaurus (AAT)." [cited 2010.01.30]. <http://www.dsoergel.com/cv/B47_long.pdf>.

Son, Dae-Hyung and Tae-Soo Kim. 1998. "A Study on thesaurus construction using facet classification." *Journal of the Korean Society for Information Managemen, Proceedings of the 5th Conference*, 8: 235-238.

WordNet. 2011. <http://www.w3.org/2001/sw/BestPractices/WNET/wordnet-sw-20040713.html>.