

Ontology Implementation and Methodology Revisited Using Topic Maps based Medical Information Retrieval System

토픽맵 기반 의학 정보 검색 시스템 구축을 통한
온톨로지 구축 및 방법론 연구

Myongho Yi*

ABSTRACT

Emerging Web 2.0 services such as Twitter, Blogs, and Wikis alongside the poorly-structured and immeasurable growth of information requires an enhanced information organization approach. Ontology has received much attention over the last 10 years as an emerging approach for enhancing information organization. However, there is little penetration into current systems. The purpose of this study is to propose ontology implementation and methodology. To achieve the goal of this study, limitations of traditional information organization approaches are addressed and emerging information organization approaches are presented. Two ontology data models, RDF/OW and Topic Maps, are compared and then ontology development processes and methodology with topic maps based medical information retrieval system are addressed. The comparison of two data models allows users to choose the right model for ontology development.

초 록

트위터, 블로그, 위키 등과 같은 web 2.0 서비스는 구조화가 덜 되어 있고, 정보량 폭증을 감당하기 어렵다는 한계를 갖고 있는 기존의 정보조직 방법을 향상시킬 것을 요구하고 있다. 이 같은 정보조직 방법을 향상시킬 수 있는 방안의 하나로 지난 10년간 온톨로지가 연구자의 주목을 받았음에도 불구하고 현행 시스템에까지 이것이 적극 활용되고 있지는 않은 것으로 보인다. 이 연구는 온톨로지 구축 및 방법론을 제안함으로써 향후 온톨로지의 방향성을 제시하는 것을 목적으로 한다. 이를 위해 이 연구는 기존의 정보조직 방법론의 한계점을 살펴보고, 온톨로지 표현을 위한 데이터 모델을 서로 비교하고 분석하였다. 또한 토픽맵 기반 의학 정보시스템을 통해 온톨로지 구축 단계 및 방법론을 기술하였다.

Keywords: semantic web, ontology, topic maps, ontology methodology, ontology data model
시맨틱웹, 온톨로지, 토픽맵, 온톨로지 방법론, 온톨로지 데이터 모델

* Assistant Professor, School of Library and Information Studies, Texas Woman's University,
P.O. Box 425438 Denton, TX 76204 (myi@twu.edu)

■ 논문접수일자: 2010년 5월 12일 ■ 최초심사일자: 2010년 6월 4일 ■ 게재확정일자: 2010년 7월 18일
■ 정보관리학회지, 27(3): 35-51, 2010. [DOI:10.3743/KOSIM.2010.27.3.035]

1. Introduction

Organizing and searching for large amounts of poorly-structured information is a challenging task. The search results tend to be lengthy and irrelevant. To improve information organization and searching, three traditional approaches—terms lists, classification/categorization, and relationship groups—have been used (Zeng 2005). One of the emerging approaches is ontology. Ontology is one of the core standards in the Semantic Web. The Semantic Web emerged as a dynamic web for organizing and searching data on the current static web in 1998 (Berners-Lee 1998). The Semantic Web Architecture was released in 2000 and real-world applications were introduced. Cardoso (2008) shows how the Semantic Web can be used to develop course management. He introduces three components of Semantic Web to administer course management. These include the languages to represent knowledge (such as OWL), to query knowledge bases (SPARQL, RDQL) and to describe business rules (such as SWRL). Obama's administration (Peterson 2009) is set to utilize Semantic Web technologies to bring transparency to government. The recently launched recovery.gov website brought with it the promise that citizens would be able to view where the money was going and how it was going to be spent. However, there is little penetration into current web and information systems and the industry is still skeptical about the Semantic Web's potential. There are criticisms (Ian Horrocks, Bijan Parsia, Peter Patel-Schneider & Hendler 2005) of the Semantic Web. In order to understand causes

of little penetration of Semantic Web, this study reviews 10 years of research on one of the core concepts in Semantic Web: ontology. First, this study reviews traditional and emerging approaches to organize and search information. Second, this study presents and compares different ontology data models. Third, this study addresses ontology methodology as well as technical and non-technical aspects of ontology system development processes followed by a conclusion and future research directions.

2. Three Traditional Information Organization Approaches

The issue of organization and the searching of information is not a new subject. Bush (1945) addressed the issue of finding information about a half century ago. In order to enhance information representation, researchers have endeavored to find more efficient information organization methods. Their efforts can be summarized into three major categories of methods: term lists, classification/categorization, and relationship groups (Zeng 2005). Term lists, which contain lists of words, phrases, or definitions, give emphasis to lists of terms. Synonym rings, authority files or dictionaries are some examples. Classification and categorization sort information into groups of similar units. Taxonomies or faceted classification are some examples. In taxonomy, the broader/narrower relationship is used to build the structure. Using faceted classification, a developer can create a set of facets to provide multiple access points. Relationship groups, on the

other hand, emphasize relationships among terms. Thesauri, concept maps, or ontologies are some examples of the relationship group approach. Equivalence, hierarchical, and associative relationships are used in thesauri to add relationships among resources (ANSI/NISO Z39.19-2005 2005). The American National Standards Institute (ANSI) / National Information Standards Organization (NISO) defines equivalence relationships as “the relationships between or among terms in a controlled vocabulary that leads to one or more terms that are to be used instead of the term from which the cross-reference is made” (ANSI/NISO Z39.19-2005, 2005, p.17). Equivalence relationships include synonyms, lexical variants, quasi-synonyms, and factored or unfactored forms of compound terms. Hierarchical relationships show levels of superordination and subordination. Hierarchical relationships use broader term (BT) and narrower term (NT) relationships to build connections between concepts. A hierarchical relationship is a basic feature of a thesaurus and an important factor in the enhancement of recall and precision performance (Aitchison, Gilchrist & Bawden 2000). The third thesaural relationship, the associative relationship, is found among terms that are related conceptually, but neither hierarchically nor equivalently (Aitchison, et al. 2000). Wordnet (Miller 1995), thesaurus development (Jun 2006; Tseng 2002), query expansion (Efthimiadis 1996; Greenberg, 2001; Khan & Khor 2004), and thesaurus-based search (Shiri & Revie 2006) are some approaches that use equivalent, hierarchical, and associative relationships for organizing and retrieving information. The associative relationship has received

great attention from researchers. Perreault (1965) published a classified list of 120 relationships, including proposals from various classification specialists, such as Ranganathan. In the medical field, the Unified Medical Language System (UMLS) uses 49 non-hierarchical relationships. Yet, while many studies indicate the usefulness of associative relationships, the creation of associative relationships is not an easy task (Bean 1998). In other words, associative relationships can be more powerful than other relationship types because they can describe more complex relationships, but they can also be more difficult to define than equivalence or hierarchical relationships due to this complexity. In addition, assigning associative relationships to information units can be risky (Aitchison, et al. 2000) since there are a limited number of agreed upon rules to define these relationships (Schmitz-Esser 1999). Many thesauri lack rules for defining and creating associatively-related inter-term links; thus, constructing related terms causes inconsistency, idiosyncratic application, and incomplete pathways through information systems (Molholt 1996). Jonassen and Grabowski (1993) defined concept mapping as “visually describ[ing] the relationship between ideas in a knowledge domain” (p. 443). Concept mapping has been used to facilitate cognitive learning (Rice, Ryan & Samson 1998; Roth & Roychoudhury 1993; Williams 1998) since the publication of *Learning How to Learn* (Novak & Gowin 1984). Finally, Gruber (1993) defined ontology as “the specification of one’s conceptualization of a knowledge domain.” In recent years, several studies have attempted to explore mixed traditional approaches, such as faceted taxonomies

(Tzitzikas 2006, 2007), dynamic taxonomies (Demo & Angius 2007; Rolletschek 2007; Sacco 2007) and faceted search (Freeman 2006; La Barre 2007; Ross & Janevski 2005; Smith, et al. 2006). One interesting study is the Flamenco faceted search system (Hearst, n.d). The system shows key facets such as country, affiliation, prize, and year to find information about Nobel Prizes. Yet, one of the limitations of any faceted approach is that much information is not structured by these particular facets. Additionally, even though one can add facets, it is extremely costly to implement and maintain such a system because domain experts must generate facets manually (Hearst 2006; Otwell 2002).

3. Emerging Information Organization Approaches

There are two types of associative relationships: associative relationships between terms belonging to the same hierarchy (symmetric) and those between terms belonging to different hierarchies (asymmetric) (ANSI/NISO Z39.19-2005, 2005). The ANSI/NISO (ANSI/NISO Z39.19-2005, 2005) presents a few associative relationships among terms belonging to different hierarchies. For example, a cause and effect relationship is an associative relationship between terms belonging to different hierarchies. However, these 13 typical situations are not adequate to show the range of associative relationships among terms. The necessity of studies on associative relationships between terms belonging to different hierarchies has

recently motivated the study of ontology.

3.1 Ontology

Guarino(1998) defined ontologies as “constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words” (p. 4). Ontology is the core framework of the Semantic Web. Ontologies are an emerging approach of making the concepts within a domain, as well as concepts across diverse domains, and the relationships between these concepts machine understandable (Krishnamurthy 2006; Müller, Kenny & Sternberg 2004). Because ontologies can encode semantic relationships between data elements, rather than just basic information about the elements themselves, ontologies are more capable of describing complex, less-structured data than are typical databases (Krishnamurthy 2006). Ohlms (2002) claims that ontology is the key to the next generation of information retrieval systems and knowledge management applications because ontology provides rich semantic relationships between information elements, making valuable knowledge accessible. Enhancing the semantic relationships of information through ontology is a promising alternative approach for information organization, and this improved information organization is vital for effective and efficient searching. Research has demonstrated that ontology improves recall and search time (Guo, Pan & Heflin 2004; H. Kim 2005). Ontology also has been used to build associations on web pages (Naing, Lim & Chiang 2006; Ras

& Dardzinska 2004). Some domain-specific ontologies are also built to represent preferred terms, related terms, and synonyms: for example, the gene ontology (Leroy & Chen 2005) or the learning ontology (Qin & Hernandez 2006). Research shows that an ontology-based search engine improves retrieval performance (H. Kim, Rieh, Ahn & Chang 2004; Müller, et al. 2004).

4. Representing Ontologies

4.1 Major Standards for Representing Ontologies

Knowledge Interchange Format (KIF), CycL, Topic maps, and RDF/OWL are major standards for representing ontologies. KIF is a logic language much used for representing ontologies and it can represent concepts and inference rules. The Suggested Upper Merged Ontology (SUMO) uses a form of KIF known as SUO. The CycL is a proprietary Lisp-like ontology language used for Cyc. The Topic maps is a logical inference that supported through tolog. The OWL is an ontology language, used for describing your ontologies formally. With the support from the World Wide Web Consortium, Resource Description Framework (RDF) and the Web Ontology Language (OWL) have been used to represent ontology-based information systems. Even though there are four standards representing ontologies, most Semantic Web Architecture-related researchers have focused on RDF/OWL. All of the studies mentioned above used Resource

Description Framework (RDF) to represent ontology.

There have been many debates about whether Topic Maps or RDF is a better model. While RDF is optimal for making inferences about information, Topic Maps is better for finding information. RDF is suitable for the physical sciences or biomedical domains where terms are less ambiguous. Pepper (2002) defines Topic Maps as an International Organization for Standardization (ISO) standard used to describe knowledge structures and associate them with information resources. Topic Maps is usually implemented using XML Topic Maps (XTM). As opposed to RDF, Topic Maps is more centralized because all information is contained in the map rather than associated with the resources (Garshol 2002).

Topic Maps brings together the best of traditional classification schema, taxonomies and thesauri, with the features of a back-of-the-book index, and adds to these the advanced capabilities of Web technologies, such as hyperlinking (Pepper 2002). The topics, occurrences, and associations that comprise Topic Maps allows it to describe ontologies. Each data element in Topic Maps is called a topic. Any term in a thesaurus can be seen as a topic in Topic Maps. Each topic can be given multiple names, and it is not necessary to distinguish between topics with the same names. This means that the topic Seoul City can be entered as “Capital of Korea”, “서울”, or the “The Seoul”, thus solving the synonym problem. This is parallel to an equivalence relationship in a thesaurus (UF and USE) (Garshol 2004; Pepper 2002). In the Topic Maps standard, names can also be given a scope, or a set of topics within

which the name is appropriate (Garshol 2004, p. 385). In addition, creators can assign types to topics. For example, the topic “Seoul” can be assigned a type “place.” Dividing topics into types in this way can be a powerful tool to help users filter information and eliminate erroneous search results, another feature of back-of-the-book indexes duplicated and enhanced in Topic Maps. Occurrences show where information about a topic can be found (similar to an index). Occurrences can also have types, such as twitter, facebook, user-created content (UCC), podcasts, wikis, music videos, blogs, tutorials, etc. Scope can be applied to an occurrence, allowing users to limit results, for example, to only those appropriate for adults or for children six and under (Garshol 2004). Perhaps the most powerful aspect of Topic Maps is the ability to create associations between topics. Using Topic Maps, almost any association can be described: both associative relationships between terms belonging to the same hierarchy and those between terms belonging to different hierarchies. Associations can explain relationships like metadata “describes” data. Another benefit of using Topic Maps instead of thesauri is that, while thesauri are often limited in the number of associative relationships they can represent due to the use of controlled vocabularies, ontologies use open vocabularies. This means that ontology creators can define terms as they go along, adding and changing terms to better fit the dataset, and they do not have to follow any guidelines as to what terms can or cannot be used. Because of this, Topic Maps based ontology is capable of describing almost any possible

relationship between terms (Garshol 2004, p. 386). Additionally, not only can Topic Maps help users navigate through vast amounts of information by representing more complex relationships between terms than thesauri, but according to Garshol, thanks to their rich descriptive capabilities, they can also serve as information resources themselves (Garshol 2004, p. 388). For example, users could learn from searching a Topic Maps that the topic “봄 여름 가을 겨울” is a type of “band” comprised of the members “전태관”, and “김종진.” A few studies have been conducted to examine various aspects of Topic Maps based ontology. Ramalho, Librelotto, and Henriques (2006) used Topic Maps to enable a conceptual navigation among heterogeneous information systems. Gulbrandsen (2006) discussed strengths and weaknesses of using Object Role Modeling (ORM) and Unified Modeling Language (UML) Class Diagrams for conceptual modeling of Topic Maps. Leuenberger, Grossmann, Stettler, and Herget (2006) utilized Topic Maps to develop an information system for a digital collection of different types of visual resources. Kim, Shin, and Kim (2007) addressed how to merge Topic Maps. Yi (2006) implemented Topic Maps based ontology for information technology jobs. Vassallo created a Topic Maps based system where users can browse different aspects of their cultural heritage. However, even though a few Topic Maps studies have been conducted, there are few studies to date comparing user performance on a Topic Maps based ontology information system and a thesaurus-based information system.

4.2 Three Different Perspectives on Topic Maps and RDF

The goal of Topic Maps and RDF is similar, and some efforts to make these two data models interoperable have also been conducted. The World Wide Web Consortium (W3C) and ISO have set up a task force to make these two standards interoperable. The Semantic Web Best Practices and Deployment (SWBPD) Working Group support the RDF/Topic Map Interoperability Task Force (RDFTM) to help users who want to combine data from W3C RDF and ISO topic maps. However, there are some differences when it comes to choosing the right data models. The topics, associations, and occurrences that comprise Topic Maps allow users to describe ontologies. Each data element in a Topic Map is called a topic. Any term in a thesaurus can be seen as a topic in Topic Maps. Each topic can be given multiple names, and it is not necessary to distinguish between topics with the same names. This means that the topic Tim Berners-Lee can be entered as “Tim BL”, “TBL”, or the “Inventor of World Wide Web”, resulting in the same information, thus solving the synonym problem (Pepper 2002a, 2002b). Associations express relationships between topics, e.g. “Tim Berners-Lee” made “<http://www.w3.org/RDF>.” Associations are inherently multidirectional. The statement Tim Berners-Lee made <http://www.w3.org/RDF> automatically implies the statement <http://www.w3.org/RDF> was made by Tim Berners-Lee. While RDF does not have a mechanism to cope with this confusion, Topic Maps provide a subject identifier and subject indicator to

resolve this confusion. Users cannot rely on names because of synonym, homonym, and multiple language problems. To resolve these issues, users need to use identifiers that are clear both to humans and machines. A subject identifier is an URI used by a machine to identify a subject and a subject indicator is information used by humans to identify a subject. The topic “apple” can be identified by a machine using <http://psi.fruit.org/#apple>. A subject indicator about “apple” can be used for a human to identify it. Both subject identifier and indicator refer to the same subject in the real world. Topic Maps provide rich representations of a topic by using three different kinds of topic characteristics: topics, associations, and occurrences. RDF has only one way to make assertions about things: triple (subject, predicate, object), and triplet notation is not expressive enough (Schaffert 2001). One of the main differences between Topic Maps and RDF is the structure of the representation (Garshol 2002). Garshol asserts that RDF relates one thing to another, while Topic Maps can relate any number of things. In Topic Maps, users can discern between the relationships that are represented which makes it easier to build complex relationships.

5. Ontology Methodology

In the computer and library science field, the highly cited ontology definition is adopted from Gruber where an “ontology is a formal, explicit specification of a shared conceptualization.” Shared reflects that ontology has commonly used terms among people.

Fonseca (2007) proposed a differentiation between ontologies of information systems and ontologies for information systems. Fonseca (2007) defines ontologies of information systems as support the creation of modeling tools. He also defines ontologies for information systems as support the creation of ontology-driven information systems. Development of ontology systems involves several steps and Figure 1 shows those processes.

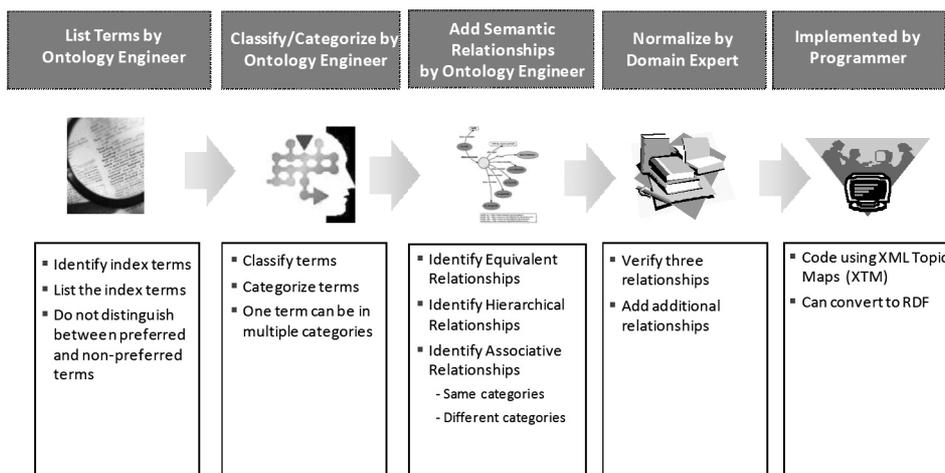
5.1 List Terms

The first step is to identify index terms. Term lists, which contain lists of words, phrases, or definitions, give emphasis to lists of terms. If there are synonym rings, authority files or dictionaries available, we can use these resources to identify terms. Unlike the taxonomy, ontology does not require distinguishing between preferred terms or non-preferred terms. One of the issues in social tagging is non-pre-

ferred terms usage. These issues easily can be resolved with ontology. In fact, the more non-preferred terms there are the better to express a concept. In other words, when a user looks for H1N1, he or she can put in different terms. Furnas, Landauer, Gomez, and Dumais (1987) show that the probability of two people favoring the same term in every case is less than 0.20. They argue that many alternative access terms are necessary for users to determine relevant information. Users do not use lengthy expressions to search for information. Jansen, Spink, and Saracevic (2000) show that an average of 2.21 terms are used in each search.

5.2 Classification/Categorization

Using classification and categorization approaches the index terms need to be sorted into groups of similar units. If there are existing taxonomies or faceted classification available, we can use those resources. The



<Figure 1> Development Process of Ontology System

broader/narrower relationship can be easily built automatically using automated processes.

Some of the classifications and categorizations are as follows;

- H1N1, Swine Flu, 돼지독감
- Flu, 감기, 독감
- Virus, 바이러스
- Symptom, 증상
- Treatment, 치료
- Medicine, 약

Not only are popular terms clustered but also different languages are associated with classified terms.

5.3 Relationships among Topics and Occurrences

The relationships among resources can be automatically built from a structured document such as a database. The relations among topics are shown in Table 1. All the examples shown in Table 1 are an example of different hierarchies between terms. There are two types of associative relationships: associative relationships between terms belonging to the

same hierarchy (symmetric) and those between terms belonging to different hierarchies (asymmetric) (ANSI/NISO Z39.19-2005, 2005).

Occurrences are the actual resources that are linked to topics or associations. The occurrences can be typed as well. Occurrences show where information about a topic can be found (similar to an index). Occurrences can also have types, such as twitter, facebook, podcasts, wikis, videos, blogs, tutorials, etc. Scope can be applied to an occurrence, allowing users to limit results, for example, to only those appropriate for adults or for children six and under (Garshol 2004). Some examples of occurrences are the following;

- Twitter
- Websites
- Video
- Blog
- Wikis
- News
- Flyers
- Brochures
- Podcasts

<Table 1> Relationships among Topics

Instances/Topic Types	Associations	Instances/Topic Types
H1N1/Disease	Treats/Treated by	Tami Flu/Medicine
H1N1/Disease	Has symptom/is Sign of	Fever/Symptom
H1N1/Disease	Protects/Protected by	Flu/Vaccines
H1N1/Disease	Spread/Spread by	coughing by people with influenza/Spread
H1N1/Disease	Caused by/Causes	H1N1/Virus
Tami Flu/Medicine	Makes/Is made by	Pfizer/Manufacturer
Tami Flu/Medicine	Has side effect/Example of Side Effect	Fever/Side Effect

5.4 Ontology modeling

The ontology modeling process involves building relationships or associations among resources. All resources must be semantically connected. H1N1 ontology modeling is displayed in Figure 2 below.

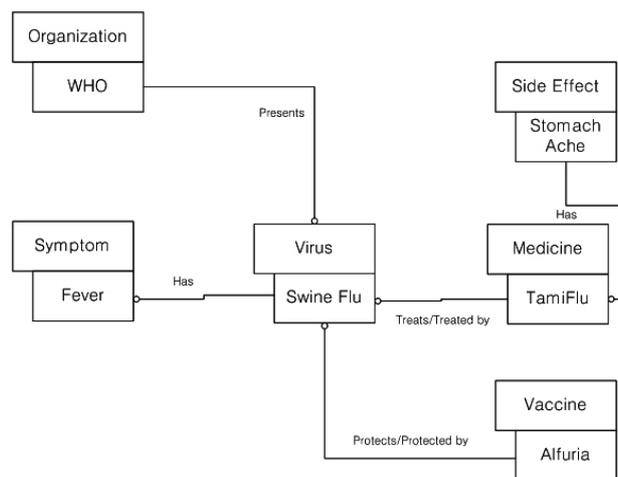
Figure 2 shows relationships among terms, and these relationships provide seamless connections among terms. More specifically, when a user searches for the virus of a specific disease such as Swine Flu, relationships in Figure 2 allow users to navigate different viruses and related symptoms and medicine. The useful association for this search query will be “treats/treated by” between virus and medicine. Another example for H1N1 ontology is as follows: When a user wants to find out the vaccine that protects against Swine Flu. A user can find relationships between the virus and the vaccine very easily by browsing relationships between virus and vaccine.

6. Ontology Implementation

After collecting topics (terms), modeling the topics, associations among topics and classifying occurrences, implementation follows. The H1N1 ontology is written in Topic Maps XML (XTM) are shown in Figure 3.

Relationship between id12 (vaccines) and id 36 (manufacture) are shown as “Make vaccines.” (See Figure 3)

As shown in Figure 4, this system has three features that help users navigate and find information effectively and efficiently. First, this system shows the types of information. The users with limited medical knowledge know what Swine Flu is by looking at the type of information, disease. All the resources are semantically connected. Therefore, they can browse the same type of information to find out any possible semantic relationships. Second, this system displays the same information in different



<Figure 2> H1N1 Ontology

```

<?xmlversion="1.0"encoding="utf-8"standalone="yes"?>
<topicMap xmlns="http://www.topicmaps.org/xtm/1.0"xmlns:xlink="http://www.w3.org/1999/xlink"
id="reified-id356">
  <topic id="id522">
    <instanceOf>
      <topicRef xlink:href="#id11"></topicRef>
    </instanceOf>
    <baseName>
      <baseNameString>Makes vaccines</baseNameString>
    </baseName>
  </topic>
  <topic id="id673">
    <instanceOf>
      <topicRef xlink:href="#id12"></topicRef>
    </instanceOf>
    <instanceOf>
      <topicRef xlink:href="#id36"></topicRef>
    </instanceOf>
    <subjectIdentity>

```

<Figure 3> XTM code for the Topic Maps based Medical Information Retrieval System

Swine Flu Type(s): Disease

Untyped Names (1)

- **Swine Flu**
 - H1N1 - Scope: *Other Name*
 - 猪流感 - Scope: *Chinese*
 - 신종플루 - Scope: *Korean*

Associations (18)

- **Caused by**
 - H1N1, H1N2, H3N1, H3N2, and H2N3
- **Has Symptoms**
 - Body Aches
 - Chills
 - Cough
 - Diarrhea
 - Fatigue
 - Fever
 - Runny Nose
 - Sore Throat
 - Stuffy Nose
 - Vomiting
- **Protected by**
 - Afluria
 - Flumist
 - Fluvirin
 - Fluzone

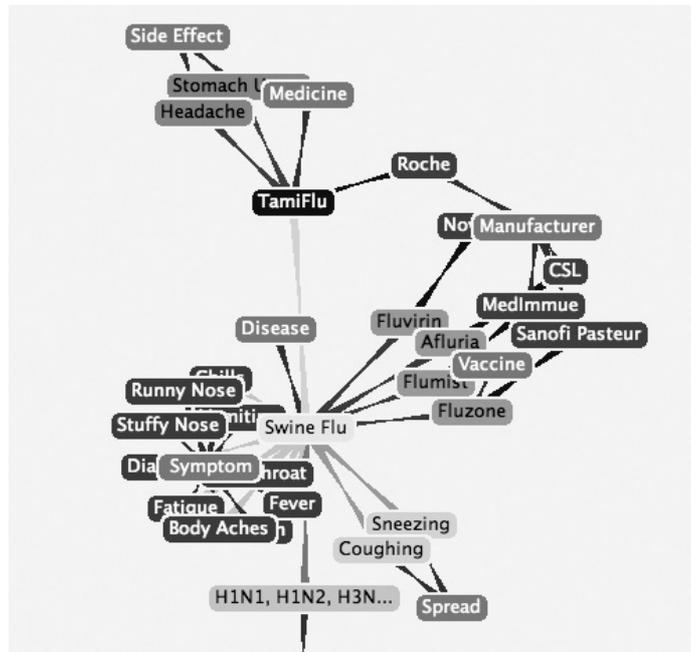
<Figure 4> Screen Shot of Ontology-based Medical Information Retrieval System

languages. For the purpose of this study, this system shows information in English, Korean and Chinese. This feature allows users to find and browse information using their native language. They can search with their language and the system will match the search results with other languages. For example, if a user searches “돼지독감”, the system returns “Swine Flu.” Third, this system shows relationships among resources. Even for global Internet users without prior medical knowledge or with a language barrier, the system shows explicit relationships among resources. For example, a user will learn how the Swine Flu is treated, the causes of Swine Flu, the symptoms of Swine Flu, and the vaccine for Swine Flu. The system shows that Swine Flu can be treated with four types of medicine, Alfluria, Flumist, Fluvirin,

and Fluzone. With these three features, international users with different languages, limited medical knowledge, and different cultures can effectively and efficiently search and browse medical resources. Figure 5 shows that the above information can be search via graphic mode. The graphic mode allows users to see the whole relationship among resources.

7. Conclusion and Future Research Directions

With the advent of various web services such as twitter, blog, wiki, etc as well as the huge amount of information created, the three traditional information organization approaches – term lists, classification/



<Figure 5> Graphic Representation of the H1N1 Ontology

categorization, and relationships groups—have limitations in representing ill-structured data. Therefore, information organization and information retrieval in the digital age is a challenging task. To address this issue, several studies have attempted to explore mixed traditional approaches, such as faceted taxonomies (Tzitzikas 2006, 2007), dynamic taxonomies (Demo & Angius 2007; Rolletschek 2007; Sacco 2007) and faceted search (Freeman 2006; La Barre 2007; Ross & Janevski 2005; Smith, et al. 2006). Researchers are also interested in ontology because many researchers claimed that ontology is a very promising alternative to traditional information organization approaches. Ohlms (2002) claims that ontology is the key to the next generation of information retrieval systems and knowledge management applications because ontology provides rich semantic relationships between information elements, making valuable knowledge accessible. Enhancing the semantic relationships of information through ontology is a promising alternative approach for information organization, and this improved information organization is vital for effective and efficient searching.

However, ontology has little penetration into the current web or systems. This study presents and com-

pares different ontology data models. Even though several major data models are available to represent ontologies, most researches have used RDF to represent ontologies. In this study, the other data model, Topic Maps, is compared with RDF. These two data models compliment rather than compete with each other; however, there are some differences. The comparison of two data models allows users to choose the right model for ontology development. This study also addresses ontology methodology and technical and non-technical aspects of ontology system development processes using Topic Maps based medical information retrieval system. The Topic Maps based medical information retrieval system provides many features that traditional information organization approaches can't support such as allowing semantic relationships among resources to be shown. Even for users without prior knowledge, the system can show explicit relationships among resources. For future study, many ontologies have been created but the research on integration or sharing of ontologies has received little attention. More research on ontology integration will be conducted to encourage and promote significant ontology penetration into our information retrieval systems.

References

- Aitchison, J., A. Gilchrist, and D. Bawden, 2000. *Thesaurus Construction and Use: A Practical Manual* (4th ed.). Chicago: Fitzroy Dearborn.
- ANSI/NISO Z39.19-2005. 2005. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Bethesda,

- MD: NISO Press.
- Bean, C. A. 1998(August 25-29). *The Semantics of Hierarchy: Explicit Parent-Child Relationships in MeSH Tree Structures*. Paper presented at the Proceedings of the 5th International ISKO Conference, Wurzburg, Germany.
- Berners-Lee, T. 1998. What the Semantic Web can represent. Retrieved March 5, 2004, from <<http://www.w3.org/DesignIssues/RDFnot.html>>.
- Bush, V. 1945. As We May Think. *The Atlantic Monthly*.
- Cardoso, J. 2008. Developing Course Management Systems Using the Semantic Web *Semantic Web and Beyond Computing for Human Experience* (6: 169-188): Springer.
- Demo, B. and A. Angius. 2007. *E-learning: Coupling Course Management Systems and Dynamic Taxonomies*. Paper presented at the FIND'07 Workshop on Dynamic Taxonomies and Faceted Search, Regensburg, Germany.
- Efthimiadis, E. N. 1996. Query Expansion. In M. E. Williams (Ed.), *Annual Review of Information Science and Technology (ARIST)* (31: 121-187). Medford: Learned Information, Inc.
- Fonseca, F. 2007. "The double role of ontologies in information science research." *Journal of the American Society for Information Science and Technology*, 58(6): 786-793.
- Freeman, R. T. 2006. "Topological tree clustering of web search results." *Intelligent Data Engineering and Automated Learning - Ideal 2006, Proceedings*, 4224: 789-797.
- Furnas, G. W., T. K. Landauer, L. M. Gomez, and S. T. Dumais. 1987. "The Vocabulary Problem in Human-System Communication." *Communications of the ACM*, 30(11): 964-971.
- Garshol, L. M. 2002. What Are Topic Maps. Retrieved February 12, 2006, from <<http://www.xml.com/pub/a/2002/09/11/topicmaps.html>>.
- Garshol, L. M. 2004. "Metadata? Thesauri? Taxonomies? Topic Maps! Making sense of it all." *Journal of Information Science*, 30(4): 378-391.
- Greenberg, J. 2001. "Optimal query expansion (QE) processing methods with semantically encoded structured thesauri terminology." *Journal of the American Society for Information Science and Technology*, 52(6): 487-498.
- Gruber, T. 1993. What is an Ontology? Retrieved December 12, 2004, from <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>.
- Guarino, N. 1998. Formal ontology and information systems. In N. Guarino (Ed.), *Formal ontology and information systems* (3-15). Amsterdam: IOS Press.
- Gulbrandsen, A. 2006. Conceptual Modeling of Topic Maps with ORM Versus UML. In L. Maicher & J. Park (Eds.), *Charting the Topic Maps Research and Applications Landscape* (3873: 93-106). Berlin/Heidelberg: Springer.
- Guo, Y., Z. Pan, and J. Heflin. 2004. *An Evaluation of Knowledge Base Systems for Large OWL Datasets*. Paper presented at the Third Interna-

- tional Semantic Web Conference, Hiroshima, Japan.
- Hearst, M. A. 2006. "Clustering versus faceted categories for information exploration." *Communications of the ACM*, 49(4): 59-61.
- Hearst, M. A. (n.d). Nobel Prize Winners. Retrieved February 12, 2008, from.
<<http://orange.sims.berkeley.edu/cgi-bin/flamenco.cgi/nobel/Flamenco>>.
- Ian Horrocks, Bijan Parsia, Patel-Schneider, Peter and J. Hendler. 2005. Semantic Web Architecture: Stack or Two Towers? In François Fages & S. Soliman (Eds.), *Principles and Practice of Semantic Web Reasoning* (37-41). Dagstuhl Castle: Springer.
- Jansen, B. J., A. Spink, and T. Saracevic. 2000. "Real life, real users, and real needs: a study and analysis of user queries on the web." *Information Processing & Management*, 36(2): 207-227.
- Jonassen, D. H. and B. L. Grabowski. 1993. *Handbook of Individual Differences: Learning and Instruction*. Hillsdale, NJ: Lawrence Erlbaum.
- Jun, W. 2006. "Automatic thesaurus development: Term extraction from title metadata." *Journal of the American Society for Information Science and Technology*, 57(7): 907-920.
- Khan, M. and S. Khor. 2004. "Enhanced Web document retrieval using automatic query expansion." *Journal of the American Society for Information Science and Technology*, 55(1): 29-40.
- Kim, H. 2005. "ONTOWEB: Implementing an ontology-based Web retrieval system." *Journal of the American Society for Information Science and Technology*, 56(11): 1167-1176.
- Kim, H., S. Rieh, T. Ahn, and W. Chang. 2004. *Implementing an Ontology-Based Knowledge Management System in the Korean Financial Firm Environment*. Paper presented at the ASIST AM 04, Providence, Rhode Island.
- Kim, J. M., H. Shin, and H. J. Kim. 2007. "Schema and constraints-based matching and merging of Topic Maps." *Information Processing & Management*, 43(4): 930-945.
- Krishnamurthy, M. 2006. "Library portals: Ontological representation of knowledge on the web." *Information Studies*, 12(2): 75-84.
- La Barre, K. 2007. "Faceted navigation and browsing features in new OPACs: Robust support for scholarly information seeking?" *Knowledge Organization*, 34: 78-90.
- Leroy, G., and Chen, H. 2005. "Genescene: An ontology-enhanced integration of linguistic and co-occurrence based relations in biomedical texts." *Journal of the American Society for Information Science and Technology*, 56(5): 457-468.
- Leuenberger, M., S. Grossmann, N. Stettler, and J. Herget. 2006. Topic Maps for Image Collections. In L. Maicher & J. Park (Eds.), *Charting the Topic Maps Research and Applications Landscape* (3873: 107-111). Berlin/Heidelberg: Springer.
- Miller, G. A. 1995. "WORDNET - A LEXICAL DATABASE FOR ENGLISH." *Communica-*

- tions of the *ACM*, 38(11): 39-41.
- Molholt, P. 1996. *Standardization of Inter-concept Links and their Usage*. Paper presented at the 4th International ISKO Conference, Washington, DC.
- Müller, H.-M., E. E. Kenny, and P. W. Sternberg. 2004. "Textpresso: An ontology-based information retrieval and extraction system for biological literature." *PLoS Biology*, 2(11): 1984-1998.
- Naing, M.-M., E.-P. Lim, and R. H. L. Chiang. 2006. "Extracting link chains of relationship instances from a Web site." *Journal of the American Society for Information Science and Technology*, 57(12): 1590-1605.
- Novak, J. D. and B. D. Gowin. 1984. *Learning How to Learn*. New York: Cambridge University Press.
- Ohlms, C. 2002. *The Business Potential of Ontology-based Knowledge Management*. New York: McKinsey & Company.
- Otwell, A. 2002. Comparison of Content Organization Methods Retrieved January 12, 2008, from <<http://www.heyotwell.com/work/ia/contentorgmethods.doc>>.
- Pepper, S. 2002. The TAO of Topic Maps. Retrieved January 12, 2006, from <<http://www.ontopia.net/topicmaps/materials/tao.html>>.
- Perreault, J. 1965. "Categories and relators: a new schema." *Revue Internationale De La Documentation*, 32(4): 136-144.
- Peterson, D. 2009. Obama's Groundbreaking use of the Semantic Web. Retrieved July 1, 2010, from <<http://www.sitepoint.com/blogs/2009/03/19/obama-groundbreaking-use-semantic-web/>>.
- Qin, J. and N. Hernandez. 2006. "Building interoperable vocabulary and structures for learning objects." *Journal of the American Society for Information Science and Technology*, 57(2): 280-292.
- Ramalho, J., G. Librelotto, and P. Henriques. 2006. Metamorphosis - A Topic Maps Based Environment to Handle Heterogeneous Information Resources. In L. Maicher & J. Park (Eds.), *Charting the Topic Maps Research and Applications Landscape* (3873: 14-25). Berlin/Heidelberg: Springer.
- Ras, Z. W. and A. Dardzinska. 2004. "Ontology-based distributed autonomous knowledge systems." *Information Systems*, 29(1): 47-58.
- Rice, D. C., J. M. Ryan, and S. M. Samson. 1998. "Using concept maps to assess student learning in the science classroom: Must different methods compete?" *Journal of Research in Science Teaching*, 35(10): 1103-1127.
- Rolletschek, G. 2007. *How Search can Profit from Dynamic Taxonomies*. Paper presented at the FIND'07 Workshop on Dynamic Taxonomies and Faceted Search, Regensburg, Germany.
- Ross, K. A. and A. Janevski. 2005. "Querying faceted databases." *Semantic Web and Databases*, 3372: 199-218.
- Roth, W. M. and A. Roychoudhury. 1993. "The Concept Map as a Tool for the Collaborative

- Construction of Knowledge - a Microanalysis of High-school Physics Students.” *Journal of Research in Science Teaching*, 30(5): 503-534.
- Sacco, G. M. 2007. *Research Results in Dynamic Taxonomy and Faceted Search Systems*. Paper presented at the FIND’07 Workshop on Dynamic Taxonomies and Faceted Search, Regensburg, Germany.
- Schmitz-Esser, W. 1999. “Thesaurus and beyond. An advanced formula for linguistic engineering and information retrieval.” *Knowledge Organization*, 26(1): 10-29.
- Shiri, A. and C. Revie. 2006. “Query expansion behavior within a thesaurus-enhanced search environment: A user-centered evaluation.” *Journal of the American Society for Information Science and Technology*, 57(4): 462-478.
- Smith, G., M. Czerwinski, B. Meyers, D. Robbins, G. Robertson, and D. S. Tan. 2006. “FacetMap: A scalable search and browse visualization.” *Ieee Transactions on Visualization and Computer Graphics*, 12(5): 797-804.
- Tseng, Y. 2002. “Automatic thesaurus generation for Chinese documents.” *Journal of the American Society for Information Science and Technology*, 53(13): 1130-1138.
- Tzitzikas, Y. 2006. “Revising faceted taxonomies and CTCA expressions.” *Advances in Artificial Intelligence, Proceedings*, 3955: 600-604.
- Tzitzikas, Y. 2007. “Evolution of faceted taxonomies and CTCA expressions.” *Knowledge and Information Systems*, 13: 337-365.
- Williams, C. G. 1998. “Using concept maps to assess conceptual knowledge of function.” *Journal for Research in Mathematics Education*, 29(4): 414-421.
- Yi, M. 2006. *Ontology as a Technology of Convergence: Developing a Classification of the ITWF Using Topic Maps*. Paper presented at the Annual Conference of the Association of Internet Researchers (AoIR), Brisbane, Australia.
- Zeng, M. L. 2005. “Using software to teach thesaurus development and indexing in graduate programs of LIS and IAKM.” *Bulletin of the American Society for Information Science and Technology*, 11-13.