KSCE Journal of Civil Engineering (2013) 17(7):1654-1663

DOI 10.1007/s12205-013-1155-6

Information Technology

pISSN 1226-7988, eISSN 1976-3808
www.springer.com/12205

Ontology-based Construction Knowledge Retrieval System

Moonseo Park*, Kyung-won Lee, Hyun-soo Lee***, Pan Jiayi****, and Jungho Yu*****

Received March 25, 2012/Accepted January 23, 2013

Abstract

Managing knowledge effectively is critical to the competitive power of a company. Knowledge is used as an important resource in many industrial areas, and so it follows that there is a growing interest in knowledge management within the construction industry. Yet because of the unique characteristics of construction knowledge created during projects, there are limitations to its capture and reuse. The knowledge produced during construction projects is project-oriented, experiential, and context specific; due to these characteristics, the reuse of knowledge is difficult. In this research, research team focus on capturing and identifying the characteristics of construction knowledge, then propose a method for applying these characteristics to the development of an ontology-based construction knowledge retrieval system. Moreover, research team developed a system prototype that applies ontology in suggesting related search words during the search process and validated the effectiveness of the prototype in terms of precision and recall rate. By applying the prototype, the precision and recall rate was improved by approximately 10~30%.
Keywords: *knowledge retrieval, ontology, construction knowledge, information technology*

1. Introduction

Knowledge Management (KM) is a concept of systematically managing knowledge assets of an organization to improve performance, gain competitive advantage, transfer lessons learnt and develop collaborative practices. The construction industry is project-oriented, thus knowledge is generated primarily during the project process (In this paper, knowledge is limited to information or knowledge that are uploaded/stored in knowledge management systems). Therefore, effective knowledge management can enhance the performance of similar projects and prevent the recurrence of mistakes (e.g., project participants will not have to start from scratch each time). Furthermore, successful knowledge management can serve as the basis for innovation and overall improvement (Tan, 2007). Such benefits and the importance of knowledge management have been increasingly recognized in the construction industry, as well as in academia (Carrilo and Chinowsky, 2006; Kivrak *et al*., 2008).

Despite this growing awareness, there are limitations to knowledge management (Tan, 2007). Capturing and reusing project knowledge is particularly difficult because construction knowledge is project-oriented, context specific, and experiential. Indeed, due to the unique, temporary, non-routine, and non-repetitive nature of construction projects, it can be challenging to reuse project-generated knowledge with the current retrieval methods. Additionally, project knowledge is closely tied to the person who created it, and to the context in which it was created. Therefore, even though construction companies collect and store construction knowledge, insufficient knowledge retrieval makes it difficult for others to fully benefit from this valuable asset.

Although there has been significant effort to enhance knowledge management in the construction industry, there has been little associated research. This causes a gap between research and the actual application to practice. Therefore, a more focused study is required.

In this research, in order to enhance the reuse of both explicit and tacit knowledge generated during building construction projects, a more effective strategy for retrieval is proposed-a domain-specific knowledge retrieval system using ontology. An ontology provides a framework for representing, sharing, and managing domain knowledge through a system of concept hierarchies (taxonomies), associative relations (that link concepts across hierarchies), and axioms allowing semantic reasoning (El-Diraby *et al.*, 2005).

2. Knowledge Retrieval in Construction Projects

In construction projects, relevant knowledge is required to

*Member, Associate Professor, Dept. of Architecture, Seoul National University, Seoul 151-916, Korea (Corresponding Author, E-mail: mspark@snu.ac.kr)
**Research Engineer, Construction Strategy Research Institute, Hanmi Global Co., Ltd., Seoul 151-016, Korea (E-mail: pureblue71@ hotmail.com)
***Member, Professor, Dept. of Architecture, Seoul National University, Seoul 151-916, Korea (E-mail: hyunslee@snu.ac.kr)
****Assistant Professor, Dept. of Construction Management, Tsinghua University, China (E-mail: pjy@tsinghua.edu.cn)
*****Member, Assistant Professor, Dept. of Architecture, Kwangwoon University, Seoul 139-701, Korea (E-mail: myazure@kw.ac.kr)

support decision-making. Knowledge that was generated in similar projects can be useful in planning new ones. During a project, knowledge about construction methods, technologies, and failures can be used as indirect experience. In addition, knowledge is needed to solve problems like accidents, defects, and shortcomings.

Despite many requirements for knowledge, current retrieval fails to fully meet the user's needs. Retrieving the right knowledge is difficult because current search models (e.g., Vector Space Model, Extended-Boolean Model) require the words in a search query to exactly match those used in the knowledge. These problems are further broken down in the following paragraphs.

First, from the user's perspective, there is initial difficulty in creating a search query that includes the user's knowledge requirements. Such difficulty is identified in research about user behavior while searching for information (Holscher & Strube, 2000; Spink *et al.*, 2001). When the user needs knowledge about an unfamiliar area-applying a new construction method or participating in a new type of project-very little information can be included in the traditional search. It has been demonstrated that when searching outside of one's area of expertise, people are less certain of where to start, use less precise language, and have more difficulty evaluating search results (Holscher & Strube, 2000). Hence, retrieving knowledge is difficult when a user has no prior knowledge of a task.

Second, even if the user is sure of their search needs, the search query itself is limited, and so cannot fully express the users' needs. Typically, a search query can only express fragments of information by listing a few words; this influences a user to formulate search queries such that they are short and simply structured (Spink *et al.*, 2001). Thus because users have project-oriented and context-specific knowledge needs, many other factors must be considered to retrieve the right knowledge.

Third, in addition to the problems of formulating a search query, a limitation has been identified in the matching process, as the domain-specific content and context are not reflected in the knowledge surrogates. As mentioned above, only the exact matching of words results in knowledge retrieval. Along with this, the frequency of terms (TF: Term Frequency) and the frequency of knowledge including these terms (DF: Document Frequency) are used to statistically evaluate a word's importance. This measure (Tf-idf weight (Manning *et al*., 2008) is used in ranking knowledge retrieved, and is offered in terms of relevance to the search query. However, unlike searching in general areas, in a specific domain like construction, there are important concepts that are domain-specific and that play roles in presenting knowledge's main ideas. Thus, adopting the existing measure in construction projects can result in retrieving knowledge that is irrelevant but with a high frequency of a search word, rather than retrieving knowledge that is useful but with a low frequency of the search word.

Lastly, because multidisciplinary participants work on construction projects, and because much knowledge is experiential, different words can be used for the same concept. The current search model makes it is difficult to identify the same concept using different vocabulary. In other words, it is not possible to find knowledge relevant to the user's needs if the words used in knowledge are different from what s/he entered as a search query.

3. Ontology and Knowledge Retrieval

Among many different definitions of ontology, the one most prevalent and frequently cited is Thomas R. Gruber's: "an explicit specification of a shared conceptualization" (Gruber, 1993). Generally, ontology consists of concepts (or classes), properties, relationships, constraints, axioms, and instances. A 'concept (or class)' is a unit that represents an existing thing, whether physical ('truck', 'concrete') or abstract ('time,' 'process'). A 'property' can be either the character/attribute inherent in a concept (internal) or the relationship showing how concepts are connected (external). A 'constraint' is a rule or regulation about the 'property' and 'relationship' of a concept. An 'axiom' is a constraint that is always true and is the basis for reasoning. Finally, the 'instance' (also called the 'individual') is the most specific entity of a concept (or class). Fig. 1 shows the composition and example of ontology in the construction industry. In the example, 'Concrete' is defined by the ontology as a type of 'Material' that has properties of 'Strength' and 'Slump'. Also, 'R.C. Slab' is defined by the ontology as an 'Element' that consists of only 'Steel' and 'Concrete'. The definition of concepts can be described in more detail depending on the purpose of the ontology. For example, 'R.C. Slab' can be defined in more detail as an 'Element' that is placed on top of a 'Column' or 'Beam'.

Ontologies are used to facilitate knowledge sharing and reuse (Noy and Hafner, 1997). With ontology, it is possible for the users to reflect their intentions in the retrieval process, which

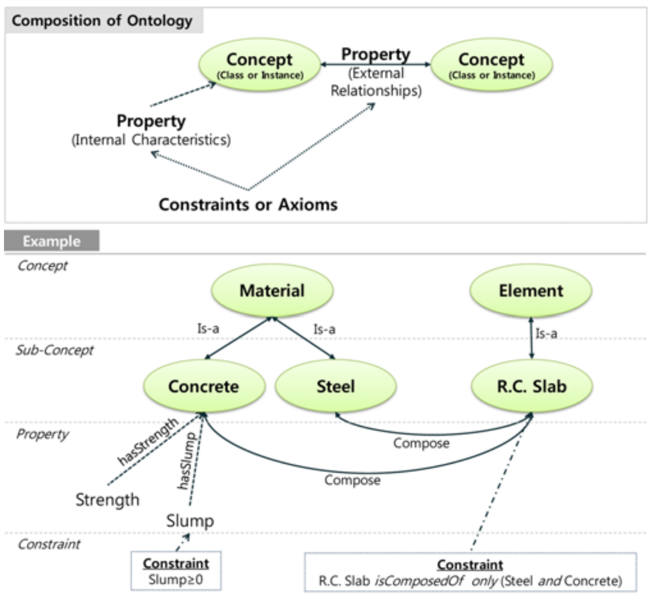

Fig. 1. Composition of Ontology and an Example (Lee, 2010)

Table 1. Definition of Precision and Recall Rate

Rate Type	Definition
Precision Rate	| A B | / | B |
Recall Rate	| A B | / | A |

| A | : total number of existing relevant documents
| B | : total number of documents retrieved
| A B | : number of relevant documents retrieved by a search

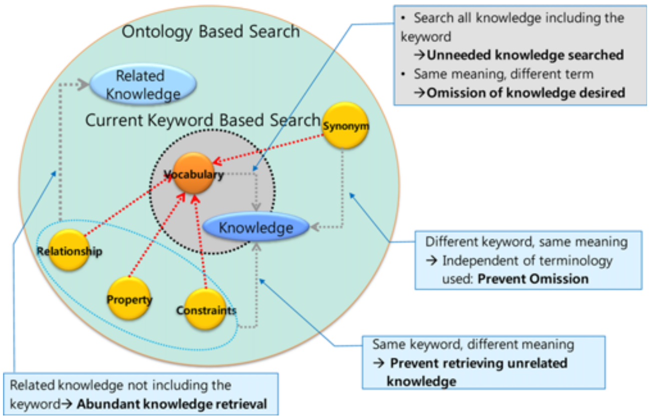

Fig. 2. Enhancement of Knowledge Retrieval by Application of Ontology (Lee, 2010)

increases the quality of their search, in terms of precision and recall rate. Precision rate shows how many knowledge needed by the user are retrieved in the total search results. Recall rate shows how many knowledge from the total knowledge repository are retrieved without omission. Table 1 summarizes the definition of the two measures.

Figure 2 summarizes the effect of ontology when applied to knowledge retrieval. First, precision rate can be increased by preventing the retrieval of irrelevant knowledge when the same term is used differently. Second, the precision rate can be increased by retrieving relevant knowledge using vocabulary that is different, but with similar meanings. Moreover, by establishing synonym relationships between concepts (terms), omission of relevant knowledge can be reduced and the recall rate can be increased.

The benefits of ontology are noticed and it is applied in information systems of different areas like electronic commerce area, medical field, legal field and digital contents area. Firstly, in the electronic commerce area, ontology is used to standardize product information (ex. ebXML, RosettaNet, UNSPSC). Secondly, in the medical field, it is used to enhance the compatibility of words used in medical treatments (ex. SNOMED-CT, UMLS). Thirdly, in the legal field, it is used to standardize legal words and make the retrieval of legal information more efficient. Lastly, in the digital contents area, it is used to express the metadata of the contents and make the retrieval of image and audio files.

Despite the benefits of using ontology for knowledge retrieval, retrieval can result in decreased generality. In other words, because the retrieval process follows rules based on ontology, any important concepts excluded from the ontology will not be captured or considered. Nevertheless, compared to search portals that deal with information from all areas, knowledge retrieval systems in specific areas can focus on the important concepts. Knowledge in a specific area, such as in construction projects, is formed around important concepts of the domain. In most cases, these main concepts represent the contents of knowledge. In addition, typical search portals have to fulfill the knowledge needs of various people, whereas in specific domains the requirements of knowledge are not as diverse. In these respects, it is useful to define a common vocabulary. Yet, to supplement the constraints, research team need to consider updating and maintaining an ontology or way of combining different methods.

4. Construction Ontology Framework for Knowledge Retrieval

No single ontology is able to fully cover a domain, nor can it satisfy the needs and preferences of every user (Gruber, 1993; Guarino and Welty, 2000. Therefore, it is important to set an explicit objective and scope for ontology's application. The purpose of ontology in this research is to enhance knowledge retrieval. Construction knowledge is primarily generated during the construction process and the main users of this knowledge are site managers and engineers. Therefore, this research focuses on knowledge retrieval in the construction phase, focusing on the engineering perspective. The knowledge targeted for retrieval is thus generated in the construction phase at the project level.

Though construction projects are non-repetitive in nature, as long as relationships between existing solutions and new problems can be established, knowledge and skills are transferrable (Fong, 2005). As ontology is the specification of a common vocabulary in which shared knowledge can be represented, it links knowledge from previous projects to the requirements of new ones. In the knowledge retrieval process, the link can be found between search queries and knowledge surrogates. In other words, ontology can be applied to the input of search words, as well as to formulating knowledge surrogates, in order to represent knowledge.

Research team have developed the proposed ontology by consulting Integrated Construction Information Classification (Korean Ministry of Land, Transport and Maritime Affairs (MLTM), 2006), Omniclass (OCCS Development Committee, 2006), e-COGNOS ontology (E-COGNOS project, 2005), Standard Construction Specification (MLTM, 2006), and construction textbooks. Research team also considered websites and blogs when determining categories of knowledge.

4.1 Main Concepts and Relationships of Construction Ontology

Construction ontology is developed by extracting the main concepts providing a link between a user's search requirements (query) and the existing knowledge. Construction projects are unique, non-routine, and temporary. Still, common factors can be found in the basic production processes. These processes

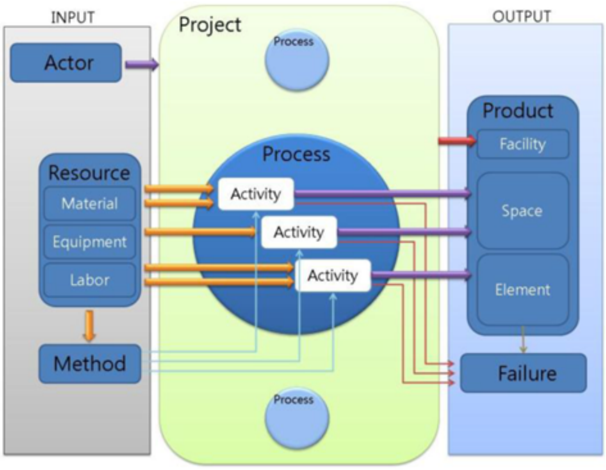

Fig. 3. Formalized Diagram of Main Concepts of Construction Ontology (Lee, 2010)

basically consist of steps of inputting resource and producing a result or product through certain activities. Fig. 3 illustrates the production process in a conceptual way. The seven main concepts of construction ontology (Project, Actor, Process, Resource, Product, Method, and Failure) are shown in Fig. 3.

As shown in the figure, the production process proceeds within a certain project. The resource and actors (e.g., managers, engineers, etc.) are inputted into processes, and the processes generate products out of the inputted resource. Also, while appropriate methods are utilized in the production process, failure can occur.

The five concepts (Project, Actor, Process, Resource, and Product) that consists part of the construction ontology are commonly used concepts in classifications (Gruninger *et al.*, 1997). Moreover, these five concepts are essential in models like Information/Integration for Construction (ICON) (Aouad *et al.*, 1994), Building Process Model (BPM) (Luiten, 1994), General Construction Object Model (GenCOM) (Froese, 1996), and Generic Process Modeling Method (GEPM) (Karhu, 2003), which conceptualize inputted resource (input) and proceeded activity and product (output). The concepts of ‘Method’ and ‘Failure’ are also important in representing the content of construction knowledge. They were derived by investigating the knowledge management systems of construction companies, blogs, and websites. It was found that out of 1200 knowledge in the knowledge management system of construction company D, 163 were about construction methods, and 165 were about failure or defects during construction projects. The proportion of knowledge about ‘Method’ and ‘Failure’ to the entire knowledge pool are about 10% each, and so cannot be ignored.

The seven main concepts selected comprise the top-level of the construction ontology. The relationship between each concept is shown in Fig. 4. These relationships are derived by considering both the conceptualization of the production process and the usage in the knowledge retrieval process.

As stated earlier, most construction industry knowledge is

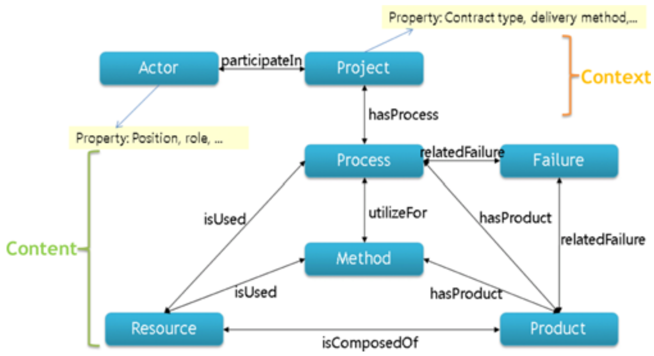

Fig. 4. Top-level of Construction Ontology (Lee, 2010)

generated in projects delivering a custom-built facility (Tan, 2007). To assist in the reuse of construction knowledge, the context of that knowledge should provide what project or process it was created for or used in. Showing what kind of project generated the knowledge allows for the judgment of whether the knowledge is appropriate and relevant for use elsewhere. In other words, context allows a project participant to judge whether the knowledge was generated or used in a similar way. Further, a project participant who knows how and when knowledge was generated will be able to retrieve it at the appropriate stage. Work environment and conditions are important to context, but including all these features will make retrieval too specific, and consequently, will make finding appropriate knowledge difficult.

The content of knowledge is another aspect to consider for its reuse. For a construction manager, the important concepts needed to represent the content of construction knowledge for reuse can be derived by investigating different parts of project management: cost, schedule, quality, safety, procurement, human resource, etc. Although representing each managerial part with ontology is beyond the scope of this research, general concepts in the project such as resource, product, and construction process can partially represent the knowledge. In cost and procurement managements, what resources are used is important in cost estimating. In schedule, quality and safety managements, the type of activity or process is progressed-in which safety accidents happened and the deficiency was discovered-can be of interest to a project participant. For a construction engineer, how the product was produced, and what kind of resources were used in recent projects, can be of interest.

Along with explicit knowledge, a lot of knowledge and skills based on experience reside in project participants’ heads and are not shared after the completion of a project. Showing what kind of project the knowledge creator belonged to and knowledge the person created can be a help to the knowledge seeker because these information can be referenced to find the right person to get help.

4.2 Composition of Construction Ontology

Ontology should be composed with concepts (terms) that reflect the knowledge seeker’s needs and that are commonly used. As mentioned previously, terms proposed are adopted from existing classifications, specifications, and textbooks. Because the proposed retrieval targets knowledge written in Korean, research team composed the ontology mainly using Integrated Construction Information Classification (MLTM, 2006), construction textbooks, and Standard Construction Specification (MLTM, 2006), as well as English sources such as Omniclass (OCCS Development Committee, 2006).

The sub-concepts that compose 'Resource', 'Product,' 'Method', and 'Failure' were collected employing a 'Process'-centered approach. In other words, the subconcepts were extracted from classifications and texts, based on the 'Process'(ex. 'earthwork', 'finishing work', etc.) it is related to. An overview of the seven main concepts is as follows:

- Project: This class is the most important for representing the context of knowledge. The properties of this class are 'Contract Type', 'Delivery Type', and 'Delivery Method'. Examples of 'Contract Type' include fixed price and reimbursable cost. Examples of 'Delivery Type' are design-build and turnkey. Lastly, examples of 'Delivery Method' include sequential delivery and fast track. The deliverables (facility/building type) of a certain project serve as an important reference when searching for relevant knowledge. Deliverables are represented by the relationship between 'Project' and 'Product'.
- Actor: This class defines major participants of a construction project. It includes both 'Personnel' and 'Organization'. 'Personnel' is a subclass, and includes individuals participating in a construction project. 'Organization', another subclass, includes construction companies, contractors, the government, and so on. This class has position and role as its property. To effectively share tacit knowledge among people, a human network that includes each person's expertise is needed. The instances of this class can be used as the basis for forming this human network.
- Process: This class consists of administrative and engineering processes. In this research, the ontology focuses on the field construction process of a building construction project. The 23 construction processes, such as earthwork and concrete work, compose this class. The processes come from the standard classification of construction works, as established by the Korean MLTM (Ministry of Land, Transport and Maritime Affairs).
- Resource: This class includes 'Labor', 'Material' 'Equipment' as its subclasses. The subclass 'Labor' includes concepts and instances that also exist in the class 'Actor'. For example, a 'Mechanical Engineer' can be an 'Actor' who participates in a project, and can also be a 'Resource' inputted into a process. The cognition of 'Mechanical Engineer' is different depending on what class the interpretation will be based on. Hence, the different interpretations of a concept are distinguished by their label as different classes. This is one of the roles of ontology-to represent the semantics of a concept.
- Product: The subclasses here are adopted from the e-COGNOS ontology, which includes most of IFC concepts, a considerable number of concepts from BS6100, and benchmarks leading e-procurement standards (such as the Common Procurement Vocabulary). The subclasses here are 'Basic Products' (doors, wall, etc.), 'Construction Complex/Facility' (highway, factory, etc.), 'Material' (natural-soil, manufactured-concrete, etc.), 'Construction Aids' (scaffolds, cofferdams, etc.) and 'Management Products' (safety report, budget, etc.). In this research, the focus is on the 'Construction Complex' subclass, so that research team are able to represent the final deliverable of a project.
- Method: This class represents techniques and mechanisms used to accomplish different 'Processes'. For example, 'Earth Anchor Method' is used in 'Earth work Process'. Unlike other classes, the instances of this class are difficult to develop, as new techniques with new names appear frequently. Nevertheless, the name of every 'Method' is denoted by a word like '~ method' or '~ technique', making a capture possible, even when a technique is new.
- Failure: The concept of 'Failure' expresses the defects or accidents that occur during or after a construction activity. Incidents are comprised of accidents involving humans ('Fall' and 'Electrocution'), structures or equipment ('Collapse' and 'Breakdown'), and defects ('Boiling' and 'Bleeding'). Providing related knowledge about failure by using relationships of the construction ontology, a project participant can foresee failure and be prepared. Also, if knowledge about failures is retrieved after a failure occurrence, related concepts can provide the engineer with an idea about the failure's causes, and can help the engineer make a counterplan.

5. Ontology-based Construction Knowledge Retrieval System Prototype

The construction ontology discussed in the previous section acts as a basis for the knowledge retrieval system. Fig. 5 summarizes this application, which research team will now detail. First, the ontology is employed to support the user's input process of knowledge needs, expanding the query with concepts and words related to the inputted search word. Then, ontology is used to represent the knowledge; words matching concepts in the ontology are extracted as representative keywords. The extracted keywords help identify main factors of construction knowledge and make for specific retrieval. Next, ontology is used in the ranking of knowledge. Using the relationships defined in the ontology, knowledge is weighted among the extracted keywords, and presented in the search results based on weight. Finally, any keywords extracted based on the ontology are used in showing the basic information of knowledge to the user, aiding in the selection of appropriate knowledge.

The ontology-applied retrieval process is developed into a system model, as presented in Fig. 6. The system model consists of an indexing module, searching module, construction ontology, and knowledge repository. First, when new knowledge is

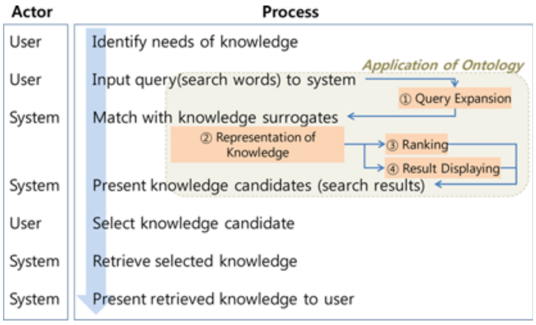

Fig. 5. Application of Ontology in the Retrieval Process

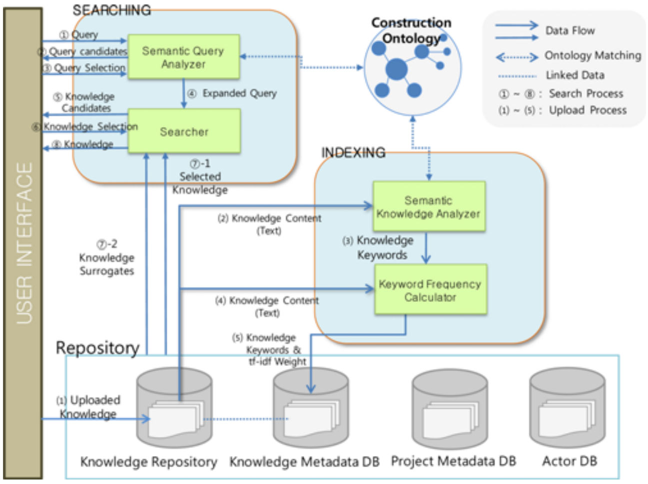

Fig. 6. Ontology-based Knowledge Retrieval System Model (Lee, 2010)

uploaded, it is stored in the knowledge repository ((1)). Additionally, the indexing module extracts keywords and features of the knowledge ((2), (3)). The semantic knowledge analyzer scans knowledge content and extracts words from the construction ontology ((2)). These words are stored as metadata in the repository. The keyword frequency calculator then counts the frequency of extracted keywords in knowledge content, and derives a weight for each keyword ((3)). Using the relationships in the ontology, a modified weight is derived. In the searching module, the semantic query analyzer searches a corresponding concept in the construction ontology that matches the search word, and retrieves the matching and related concepts (words) from the ontology (①). The retrieved and related words are presented to the user for selection of supplemented search words (②). After the user selects existing and related search words (③), the searcher matches these words with knowledge surrogates (the weighted keywords explained above) (④). The weights are summed up, and knowledge with the highest total weight is presented first in the search results (⑤). The repository consists of the knowledge repository, knowledge metadata database (DB), projects metadata DB, and actor DB. The knowledge repository stores knowledge, and the knowledge metadata DB stores basic information of knowledge. Projects metadata DB stores information about the construction projects from which the knowledge were generated and used as well as the creators (that are included in the actor DB) and users of the knowledge. The projects metadata DB is required for retrieving knowledge based on context and also required for finding the right person to get help from.

5.1 Search Word Expansion

In the proposed system, ontology is used to support the input of a user's knowledge needs. This is done by the semantic query analyzer, which provides the user with related and expanded search words. Research on search behaviors (Holscher *et al.*, 2000; Spink, 2001; iProspect, 2006), shows that users have difficulty forming sufficient search words. Moreover, search words are usually insufficient because different people use different vocabulary, and it is hard to include all of one's needs in just a few words. Therefore, considering that search words are the only way to input one's knowledge needs into a retrieval system, presenting the user with a collection of relative terms solves such limitations. Further, presenting the user with a collection of terms often leads to more precise queries, as users may encounter additional applicable terms that they otherwise would not have thought of (Davies *et al.*, 2003).

Figure 7 is a flow chart showing how the related and expanded search words are presented by the system. First, the user inputs a search word. The system then searches the construction ontology to find a concept (term/word) that is the same as the search word. When a matching concept is found, matching terms are also found (as they have matching concepts). The result of this process is presented to the user as related concepts, or in other

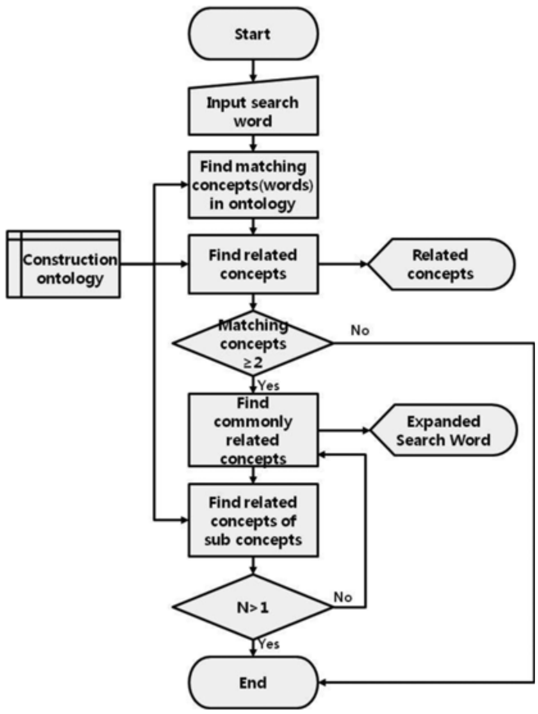

Fig. 7. Algorithm of Related Search Word & Expanded Search Word (Lee, 2010)

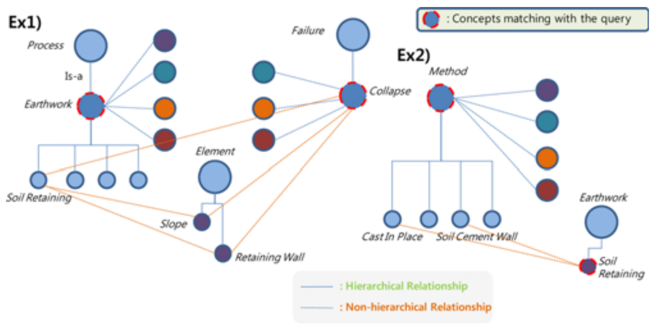

Fig. 8. Example of Expanded Search Word (Lee, 2010)

words, as related search words.

If there are more than two terms in the search word that match the terms in the construction ontology, the commonly related concepts (terms) of the matching concepts are searched for. For example, if the user inputs 'Earthwork Collapse Accident' as the search word, 'Earthwork' and 'Collapse' are found in the construction ontology. Several related concepts of these two terms exist, but out of the related concepts, 'Soil Retaining' is the only term that is relevant to both 'Earthwork' and 'Collapse' (see Fig. 8). The commonly related concepts, which are called "expanded search words" in this paper, are presented to the user as support in the retrieval process. Subsequently, the sub-concepts are identified among the previously-found related concepts, and the process of finding commonly related concepts is repeated with the sub-concepts. For example, in Fig. 8, the related concept of 'Earthwork' is 'Soil Retaining', and the related concepts of 'Collapse' are 'Retaining Wall' and 'Slope'. 'Soil Retaining' is identified as a sub-concept of 'Earthwork', so the related concepts of 'Soil Retaining' are searched for. The concepts of 'Retaining Wall' and 'Slope' are found to be commonly related concepts of 'Soil Retaining' and 'Collapse'. As result of this search word expansion, the system provides 'Soil Retaining', 'Retaining Wall', and 'Slope' for additional search using the boolean operator 'AND(&)' when the search word is 'Earthwork Collapse Accident'.

5.2 Weighting and Ranking

Ranking knowledge is important because search engine users tend to look at only the first few pages of search results. A research by iProspect (2006) shows that 62% of search engine users click on a search result within the first page of results, and by the time three pages have gone by, 90% of users have click on a result.

The ontology applied in the weighting and ranking of knowledge is used to modify the term frequency-inverse document frequency (tf-idf) weight, as summarized in the equations below. The tf-idf weight, often used in information retrieval and text mining, is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. A word's importance increases proportionally to the number of times it appears in the document, but is offset by the frequency of the word in the corpus. Variations of the tf-idf weighting scheme are often used by search engines as a central tool in scoring and ranking a document's relevance given a user query. One of the simplest ranking functions is computed by summing the tf-idf for each query term; many more sophisticated ranking functions are variants of this simple model (Manning *et al.*, 2008).

◆ tf-idf Weight(w) (Manning *et al.*, 2008): $(tf\text{-}idf)_{i,j} = tf \times idf_i$

$$tf_{ij} = \frac{n_{i,j}}{\sum_k n_{k,j}} : \frac{\text{number of occurrences of the considered term } (t_i) \text{ in document } d_j}{\text{sum of number of occurrences of all terms in document } d_j}$$

$$idf_i = \log\frac{|D|}{|\{d : t_i \in d\}|} : \tag{1}$$

$$\left(\begin{array}{l} |D| \text{:total number of documents in the corpus} \\ |\{d : t_i \in d\}| \text{:number of document where the term} (t_i) \text{appears} \end{array}\right)$$

The weighting is processed by the indexing module, which consists of a semantic knowledge analyzer and a keyword frequency calculator. The semantic knowledge analyzer scans uploaded knowledge to find and extract terms that are included in the construction ontology. The extracted terms are saved in the knowledge metadata as keywords. Subsequently, the keyword frequency calculator calculates the tf-idf weight of each keyword. As shown in Fig. 9, the tf-idf weights are summed up based on the relationship defined in the ontology. For example, if concepts 'A', 'a', 'b', and 'c' are found in knowledge, and 'a', 'b', and 'c' are related words of 'A', the final weight of 'A' is the sum of the weights of 'a', 'b', and 'c'. In addition, rules are made to prevent infinite summation when relationships form a loop. As mentioned previously, the classes of 'Method' and 'Failure' are important in representing the content of knowledge. Also, as shown in following sections, the sub-concepts of 'Process', 'Material', 'Method,' and 'Element' are frequently used terms in knowledge title. Therefore, the adjusted weight is only given to the terms of 'Process', 'Method', 'Element', and 'Failure'. The terms of 'Resource' are neglected because their document frequency is high and they fail to represent the main content of knowledge. Such terms are used only to give enough additional weight on the four classes that the adjusted weight is applied.

5.3 Search Results Display

Effectively displaying search results is essential for an efficient selection process. Typically, the title and abstract of each search

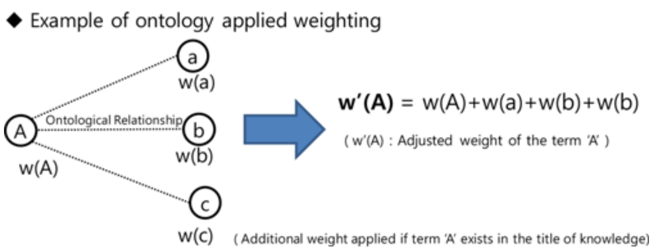

Fig. 9. Ontology Applied Weighting (Lee, 2010)

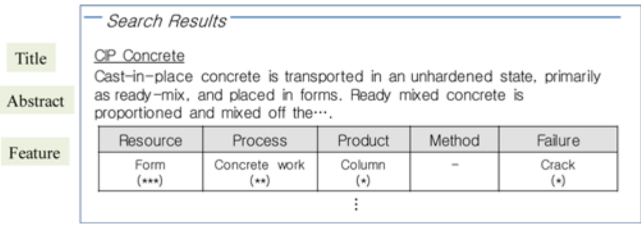

Fig. 10. Example of Search Results Display with Knowledge Features (Lee, 2010)

result is presented to the user to judge whether the knowledge has the desired contents. If the first selection fails to satisfy the user's needs, the selection process is repeated. Therefore, to minimize repetition and make the selection process efficient, construction ontology is used to show more information about the content of knowledge, focusing on the main concepts of construction domain. Fig. 10 is an example of displaying a search result using the knowledge features extracted based on construction ontology. Keywords summarizing the content of knowledge are extracted based on the ontology, and then the importance of each keywords is marked (e.g., '*' in Fig. 10), depending on the relative frequency of keywords. This helps the user to capture the main ideas of knowledge, thus supporting the selection process.

5.4 System Prototype User Interface (UI)

The system prototype was developed as part of the 'Web-based Distributed Lean Construction Information System' research project, carried out by the Lean Construction Research Center (LRC2). The LRC2 Project, initiated in 2005 and terminated in 2010, was supported by a $7,784 million grant from the Construction Technology Innovation Program, funded by the Korean government's Ministry of Land, Transport, and Maritime Affairs. The overall aim of the LRC2 Project was to develop a web-based system that could be used to support tasks involved in design, supply, and construction phases. The system was also intended to effectively manage knowledge created in each phase by developing a blog-based information system. The system prototype proposed in this research is developed to improve upon the search of knowledge within those blogs.

The user interface (UI) of the system prototype is shown in Figs. 11 and 12. Fig. 11 shows the UI for knowledge search. A search word is inputted into the system, and through the process proposed in this research, related and expanded search words are presented to the user. The related search words can be selected and applied in additional searches using the Boolean operator 'OR(|)', and expanded search words with 'AND(&)'.

Figure 12 shows the UI for managing the ontology. The UI consists of five hierarchical trees (Resource, Process, Method, Failure, and Product), to which you can add or delete concepts. The bottom half of the UI shows information about the concept selected from the tree, specifically name and superclass. Synonyms and related terms can be added.

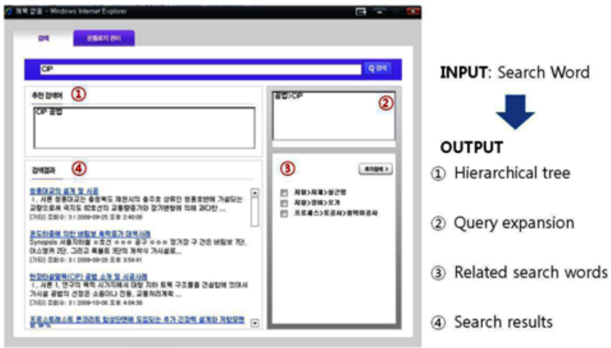

Fig. 11. User Interface: Search (Lee, 2010)

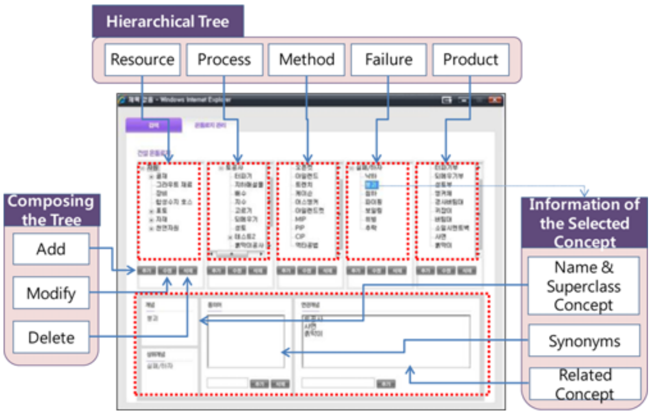

Fig. 12. User Interface: Ontology Maintenance (Lee, 2010)

The system prototype is currently limited to providing related search words to the user, and doesn't fully utilize ontology. Further development of the system is in progress.

6. Validation

In this section, the construction ontology and proposed knowledge retrieval system are validated using a case ontology on earthwork. First, the usability and comprehensiveness of ontology is validated to see if the concepts in the construction ontology appear frequently in construction knowledge. Then, the ontology-based knowledge retrieval system is evaluated by two measures, precision and recall rate. The evaluation of these two measures is completed by manually following the proposed search rules.

6.1 Comprehensiveness and Usability of Construction Ontology

The comprehensiveness and usability of ontology are validated by counting how many knowledge titles include the concepts in the case ontology. The topic of knowledge for validation is 'Earthwork', and the source of knowledge is from a construction company's knowledge management system, as well as websites and blogs. 647 knowledge titles are used in the validation. Table 2 shows validation elements and the results. 518 titles of knowledge, out of 647, included the terms in the case ontology-about 80%. Research team can thus infer that construction

Table 2. Result of Search Evaluation: Precision and Recall Rate

Search Word	Current		Ontology based		Related Search Words
	Precision	Recall	Precision	Recall	
Earthwork Collapse Accident	0.47	0.38	0.78	0.67	Retaining wall collapse, Slope collapse, Slurrywall collapse & etc.
Slurrywall method	0.93	0.32	0.83	0.73	SCW, Slurrywall, Bentonite, Guide wall, Clamshell & etc.
CIP method	0.76	0.84	0.84	0.84	Piling, Concrete pile, Earth auger, Mortar, Steel bar, Aggregates & etc.

ontology can represent the content of most construction knowledge. Especially in knowledge titles, terms included in certain classes (‘Material’, ‘Method’, ‘Process’, and ‘Element’) frequently appear. Titles of knowledge represent the main content, and thus serve an important role as surrogate in knowledge retrieval. Hence, as mentioned before, terms included in classes ‘Material’, ‘Method’, ‘Process’, and ‘Element’ should be covered when designing search algorithms. Through the validation, it can be seen that when the topic of knowledge is out of the ontology’s scope (procurement, delivery, and business, e.g.), knowledge titles don’t include terms from the case ontology. Also, if the knowledge title is too general, the construction ontology will not cover it.

6.2 Search Evaluation: Precision and Recall Rate

The measures used in evaluating search results are precision and recall rate, and the Boolean search model is applied in the evaluation. The Boolean model is a simple search model that retrieves knowledge by judging whether the search word is included (true), or not included (false) (Baeza-Yates *et al.*, 2001). Using the Boolean model, a search word is decided and the retrieval process is progressed first by not applying ontology, then by applying ontology. Finally, the precision and recall rate of each case is compared.

When applying the Boolean model, a total of 11,526 knowledge titles from the knowledge management system of company D, as well as blogs and websites, is searched. The results show that precision and recall rate generally increase after applying construction ontology. For example, as mentioned in Fig. 8, if the search word is ‘Earthwork Collapse Accident’(‘Earthwork’ is a sub-concept of ‘Process and ‘Collapse’ is a sub-concept of ‘Failure’ in the ontology) the system presents the user with ‘Retaining Wall’(sub-concept of ‘Element’) and ‘Slope’(sub-concept of ‘Element’) as expanded search words, improving the recall rate. Also, by neglecting knowledge with a title such as ‘Ways to Prevent Collapse of Forms’, the precision rate increases. Still, search words like ‘CIP (Cast-in-Place) Method’ (sub-concept of ‘Element’) which are specific and include only one concept of the ontology, see little change to precision and recall rates. However, by providing the user with related concepts, the users can refine the search goal and improve their chance of retrieving the desired knowledge.

7. Conclusions

This paper focuses on the problems that a user of knowledge retrieval systems confronts when searching for knowledge in construction projects. Whether or not a user knows what to find, the user has difficulty forming the right search query, which is the first step to finding the right knowledge. This limits the usage of a knowledge management system, because the users can’t gain knowledge matching their needs.

To enhance the knowledge retrieval process, in this research, the concept of ontology is applied to the retrieval system. This addition assures that related search words are suggested to a user, and helps the users express their needs correctly to the machine (or computer). First, main concepts that can represent knowledge were extracted from existing models, classifications, and the knowledge management system of a construction company. These main concepts were then used to compose the construction ontology framework. Finally, the framework was applied to the retrieval system prototype, which suggested related search words to a user. This paper thus suggests a system model that includes further applications of the construction ontology to retrieval systems.

Applying the concept of ontology and making the knowledge retrieval system more construction-specific increases recall and precision rates. In other words, the ontology-based system leads the user to the knowledge of the user’s needs more precisely, without omitting useful knowledge. This can contribute to the success of knowledge management in construction projects, as reusing knowledge effectively is an essential aspect of knowledge management. Still, this paper is limited to applying the basic concept of ontology to retrieval systems. Further studies on fully applying ontology are required. Also, because no ontology can include every concept, the maintenance of an ontology considering the application should be further discussed.

Acknowledgements

This research was supported by a grant (05CIT-D05-01) from the Construction Technology Innovation Program, which is funded by the Korean government’s Ministry of Land, Transport, and Maritime Affairs.

References

Aouad, G., Brandon, P., Brown, F., Cooper, G., Ford, S., Kirkham, J., Oxman, R., Sarshar, M., and Young, B. (1994). *Integration of Construction Information (ICON)*, Final Report, University of Salford, UK.

Baeza-Yates and Ribeiro-Neto (1999). *Modern information retrieval*, ACM, pp. 25-27.

Carrillo, P. and Chinowsky, P. (2006). "Exploiting knowledge management: The engineering and construction perspective." *Journal of Management in Engineering*, ASCE, Vol. 22, No. 1, pp. 2-10.

Davies, J., Fensel, D., and Harmelen, F. V. (2003). *Towards the semantic web: Ontology-driven knowledge management*, Wiley, p. 157.

E-COGNOS project (Consistent knowledge management across projects and between enterpriSes in the construction domain) (2005). IST-2000-28671<http://www.e-cognos.org>.

El-Diraby, T. A. and Kashif, K. F. (2005). "Distributed ontology architecture for knowledge management in highway construction." *Journal of Construction Engineering and Management*, ASCE, Vol. 131, No. 5, pp. 591-603.

El-Diraby, T. A., Lima, C., and Feis, B. (2005). "Domain taxonomy for construction concepts: Toward a formal ontology for construction knowledge." *Journal of Computing in Civil Engineering*, ASCE, Vol. 19, No. 4, pp. 394-406.

European Union (2001). *Common Procurement Vocabulary (CPV)*, Regulation (EC) No 2195/2002 of the European Parliament and of the Council of 5 November 2002 on the Common Procurement Vocabulary (CPV).

Fong, P. S. W. (2005). "Aspects of learning and knowledge in construction projects." *Construction Research Congress 2005*, ASCE.

Froese, T. (1996). "Models of construction process information." *Journal of Computing in Civil Engineering*, ASCE, Vol. 10, No. 3, pp. 183-193.

Gruber, T. R (1993). "A Translation approach to portable ontology specifications." *An International Journal of Knowledge Acquisition for knowledge-Based Systems*, Vol. 5, No. 2, pp. 199-220.

Guarino, N. and Welty, C. (2000). "Towards a methodology for ontology-based model engineering." *Proc., IWME-2000-International Workshop on Model Engineering,* Sophia Antipolis and Cannes, France.

Holscher, C. and Strube, G. (2000): "Web search behavior of internet experts and newbies." *Computer Networks: The International Journal of Computer and Telecommunications Networking*, Vol. 33, Nos. 1-6, pp. 337-346.

iProspect (2006). *Search engine user behavior study*, iProspect <http://www.iprospect.com/>, p. 12.

Karhu, V. (2001). "A view based approach for construction process modeling." *Journal of Computer-Aided Civil and Infrastructure Engineering*, Vol. 18, No. 4, pp. 275-285.

Kivrak, S., Arslan, G., Dikmen, I., and Birgonul, M. T. (2008). "Capturing knowledge in construction projects: Knowledge platform for contractors." *Journal of Management in Engineering*, ASCE, Vol. 24, No. 2, pp. 87-95.

Korean Ministry of Land, Transport and Maritime Affairs (MLTM) (2006). *Standard construction specification*, Kimoondang.

Lee, K. (2010). *Ontology-based construction knowledge retrieval system*, MSc Thesis, Seoul National University, Seoul, Korea.

Luiten, G. (1994). *Computer aided design for construction in the building industry*, PhD Thesis, Delft University of Technology.

Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to information retrieval,* Cambridge University Press.

Noy, N. and Hafner, C. (1997). *The state of the art in ontology design: A survey and comparative review*, Al Magazine, Vol. 18, No. 3, pp. 53-74.

OCCS Development Committee (2006). Omniclass <http://www.omniclass.org>.

Spink, A., Wolfram, D., Jansen, B. J., and Saracevic, T. (2001). "Searching the web: The public and their queries." *Journal of the American Society for Information Science*, Vol. 53, No. 2, pp. 226-234.

Tan, H. C. (2007). "Development of a methodology for live capture and reuse of project knowledge in construction." *Journal of Management in Engineering,* ASCE, Vol. 23, No. 1, pp. 18-26.