ONTOLOGY BASED KNOWLEDGE RETRIEVING IN A WEB COLLABORATION ENVIRONMENT FOR CONSTRUCTION INDUSTRY

Cynthia Changxin Wang, PhD / Senior lecturer, cynthia.wang@unsw.edu.au
Faculty of the Built Environment, University of New South Wales (UNSW), Sydney, Australia

ABSTRACT

As the amount of information and knowledge that we deal with in construction projects are huge, computerized collaboration and management systems have been seen as effective tools for construction project participants. While a vast amount of information and knowledge can be stored in these systems, how to retrieve knowledge when needed is a challenge. Traditional keyword search usually results in high returns but low precision, as context and terminology difference are not considered. This research implements construction domain ontology into a web collaboration environment. Domain ontology provides a common understanding of a domain (a particular area) in which people and the application system communicates with each other. The ontology is composed of a network of concepts, which are clearly defined and interlinked based on their context. Knowledge items published in the web are annotated according to the ontology, and enable the semantic inference to locate a particular knowledge items during the retrieval process. In this paper, some knowledge items (knowledge stories) are published as blog entries in the web collaboration systems, and a comparison between traditional keyword search and ontology based retrieval is reported. The ontology based knowledge retrieving gives much more accurate returns, and therefore can facilitate the web-based knowledge sharing practice more efficiently in the construction industry.

Keywords: Ontology, Knowledge management, Knowledge retrieving, Construction industry, Web-based collaboration.

1. INTRODUCTION

Accounting for 6.4% of the Australia's GDP and employing nearly one million people (Year Book Australia, 2008), construction is by any measure a vast industry. As complex as it is large, construction projects are information and knowledge intensive. As the industry is project-based, and the teams in construction projects are often disbanded at the end of the project, the information and knowledge generated in the project is difficult to be tracked after the project is completed. The reason is that projects do not have any organizational memory, as they are temporary in nature (Fong, 2005). Compared with non-project organizations that are supported both by the organizational structure and knowledge-absorbing routines, where knowledge becomes reutilized and socialized into the organization, project-based organizations generally do not have support mechanisms that enable knowledge transfer to occur.

Thus, proposing a KM mechanism for project-based organizations, in order to facilitate innovation and avoid re-invent-the-wheel becomes a very important issue. In such a mechanism, knowledge generated in project lifetime can be collected, modeled, stored, and diffused across project boundaries and up to the organizational level.

In recent years, web-based collaboration and management systems have been seen as effective tools for managing project information and facilitating knowledge transfer. Project participants can use website to get real-time information, review the change request, view updated plans and track the progress of the project. Thousands of leading construction and engineering companies rely on online project information system to manage their projects. The most well-known commercial project information management systems include Aconex, Project Centre, Prolog Manager, Autodesk Buzzsaw, etc.

These systems are sufficient for managing project documents and information, but are lack of functions to support knowledge sharing between project participants. One important issue is how to retrieve knowledge items effectively. A web-based collaboration environment keeps a vast number of records of knowledge items (e.g. best practices, lessons learned and project stories). Traditional keyword search usually results in high recalls but very low precision. When dozens, or even hundreds of search results are returned, the user still needs to manually manipulate and interpret the information to extract the needed knowledge.

Another difficulty in obtaining accurate search results is caused by the different terminologies people use. As project participants can be from different culture background, work in different cities and countries, the terminologies may be quite different. In this research, construction domain ontology, which provides a common understanding of the industry by defining a network of concepts, is implemented into a prototype web collaboration environment. The system is utilized to experiment high-level knowledge management functions, and to compare the ontology-based knowledge retrieval with traditional keyword search function.

2. KNOWLEDGE SHARING FUNCTIONS IN A WEB COLLABORATION ENVIRONMENT

The prototype system used in this research is built on the base of a collaborative system – LiveNet. LiveNet was initially developed in the Faculty of IT at The University of Technology Sydney (Hawryszkiewycz, 2000). LiveNet provides facilities to support group work, and these include document management facilities (e.g. central repository, key-word search engine, etc.), various communication channels (e.g. discussion forum, chat room, email, internal messages, etc.), and work-plan tools (e.g. calendar, workflow, meeting schedule, etc.). In this research, LiveNet is customized to provide a web collaboration environment for construction organizations.

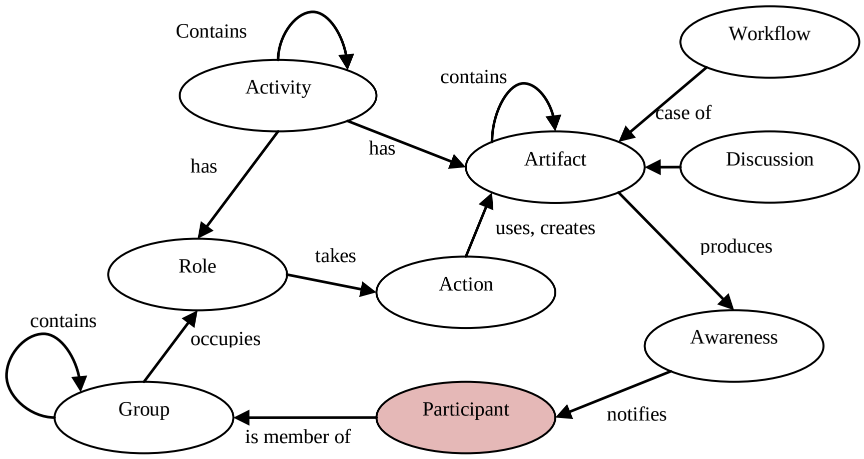

Figure 1: The concept model of the web collaboration environment.

The conceptual model of the web-based collaboration environment LiveNet is illustrated in Figure 1, where a LiveNet user, referred as a participant in the figure, “is a member of” one or more groups. An activity (such as a project) “has” any number of roles, and “has” any number of artifacts. An artifact can be documents, meeting schedules, discussion forums, project stories or workflow, etc. Artifacts could be “used or created” by an action, and awareness may be “produced” which reflects the cooperation status, and it can “notify” the participant via a number of communication means.

Livenet provides many collaboration functions, including forming group communities, discussion forums, finding experts, *etc*. and they have been presented in previous publications (Wang, 2010). In this paper, the knowledge management and retrieval functions will be presented in details. Figure 2 shows some artifacts in an ongoing project in LiveNet. The project has some participants taking different roles, and they have access to various artifacts (e.g., document folders, etc.). One of the artifacts is “unsolved problems”, where problems encountered in the project can be discussed here. Once the problem is solved, it is moved to “Solved problems”. The facilities in the system could encourage people to contribute to solving the problem, and also keep a record for future references.

Storytelling, no matter of a success or a failure, is a good culture to encourage employees to shout about their experience so that others can learn from it. LiveNet provides a hierarchical activity structure to record and organize all the project related elements. After a project completes, all the activities involved in the project were reviewed, categorized and stored. This information can be accessed by geographically dispersed project team members.

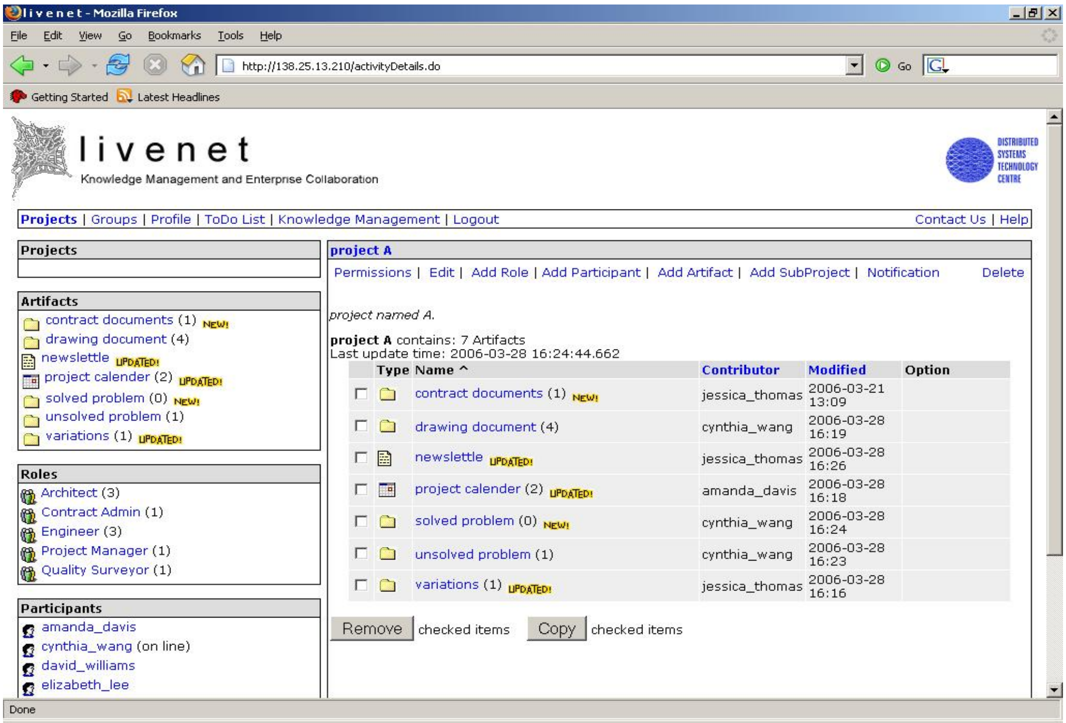

Figure 2: An ongoing project in *LiveNet*

A blogging site has also been implemented within the LiveNet environment. LiveNet users can publish their knowledge and experience in the blogging site and share with others. Blogging in this research has been treated as an organizational wide knowledge sharing mechanism. Blogs are easy to use and can provide rich communication context, and help the blog writers to gain social status. As mentioned before, teams in construction projects are often disbanded at the end of a construction project, a project-related, organizational wide blog can take an important role to collect and disseminate knowledge. The blogs created in every individual project can be shared by all the members in the organization.

Blogs can be published by all the project participants, therefore thousands of blog entries may be created in a web collaboration system. To facilitate retrieving the blog entries, the construction domain ontology is implemented in the LiveNet system. The domain ontology and how it can be used to retrieve blog entries effectively is explained in the next section.

3. ONTOLOGY AND SEMANTIC WEB TECHNOLOGY

Ontology is a term that comes from philosophy, which has become a key technology for the semantic web. Ontology is designed in order to generalize and abstract knowledge in a specific domain, which interview human understanding of symbols with their machine processability. Ontology is composed of a network of concepts, which are clearly defined (creating a common vocabulary), interlinked based on their context (using a set of relationships), and bound to certain behaviors (through a set of rules). Domain ontology provides a common understanding of a domain (a particular area) in which people and the application system communicates with each other.

Ontologies are developed for different areas. Some researchers have set up various ontologies for construction industry (Lima et al. 2003, El-Diraby et al. 2005). Among them, e-Cognos (El-Diraby et al. 2005) is the most comprehensive and mature one, and it was developed by an IST (Information Society Technologies) founded project in Europe. e-Cognos is also compatible with IFC (Industry Foundation Class) and several other classification systems (such as BS6100 and UniClass). In this research, e-Cognos is adopted as it provides adequate details for the proposed semantic blogging system and also satisfies the compatibility requirement for the industry's software integration plan.

About 15,000 concepts are identified and classified in e-Cognos ontology document. In e-Cognos, construction knowledge is encapsulated in several overlapping systems, and these relationships are illustrated in Figure 3. In construction activities, "Actors" use some "Resources" to produce a certain kind of "Products" by following some "Processes" that form a "Project" according to some boundary conditions. There are a set of "Technical Topics" which provides information to support the project processes run smoothly within the work environment.

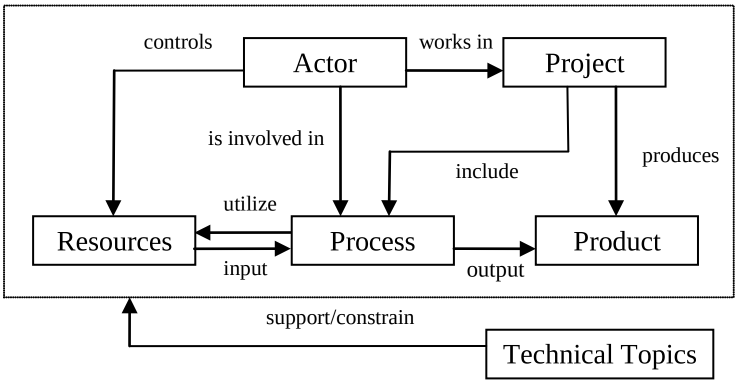

Figure 3: Higher level concepts and their relationships in e-Cognos ontology

To facilitate efficient retrieval of knowledge items from the web-based collaboration system, knowledge items (in this paper, we mainly use knowledge stories published as blog entries) are annotated according to the ontology when they are recorded, and semantic inference helps to locate them during the retrieval process. The annotation process and semantic inference are further explained below.

4. ANNOTATION PROCESS

"Annotation" is one of the important facilities to support efficient knowledge management. Blog entries are organized chronologically by default, but they can also be grouped by topics. Categorizing is a method of organizing blog entries by assigning each entry to a predetermined topic. Another form of annotation is Tagging. Tags are collections of keywords that can be assigned to blog entries, and they can be used to describe and categorize the entry.

Ontology based annotation uses a top-down approach. It provides hierarchical and domain specific vocabulary which describes the elements of the domain and their relationships. In LiveNet, blog entries are annotated with machine-understandable meta-data according to the ontology. These meta-data make the blog contents to be found easily and precisely. According to Cayzer (2004), because the machine "semantically" understands the blog entries by using conceptual inference, semantic blogging can provide richer processing features.

Annotation can be done automatically, semi automatically and manually. For the first two annotation methods, once a blog entry is created, it will firstly be analyzed by text analysis tool. The text analysis tool extracts key sentences from the document based on the frequency of words and phrases occur in the document. The extracted keywords and the relationships between the keywords are compared with the concepts defined in the ontology; the concepts with highest similarity will be used for annotation. In the current research, we focus more on the retrieval process, therefore manual annotation is adopted to simplify the process. Figure 4 shows how to choose a concept from the ontology tree and append it to a blog entry's annotation list. Multiple tagging is supported in this system; this means one blog entry can be tagged by more than one ontology concepts.

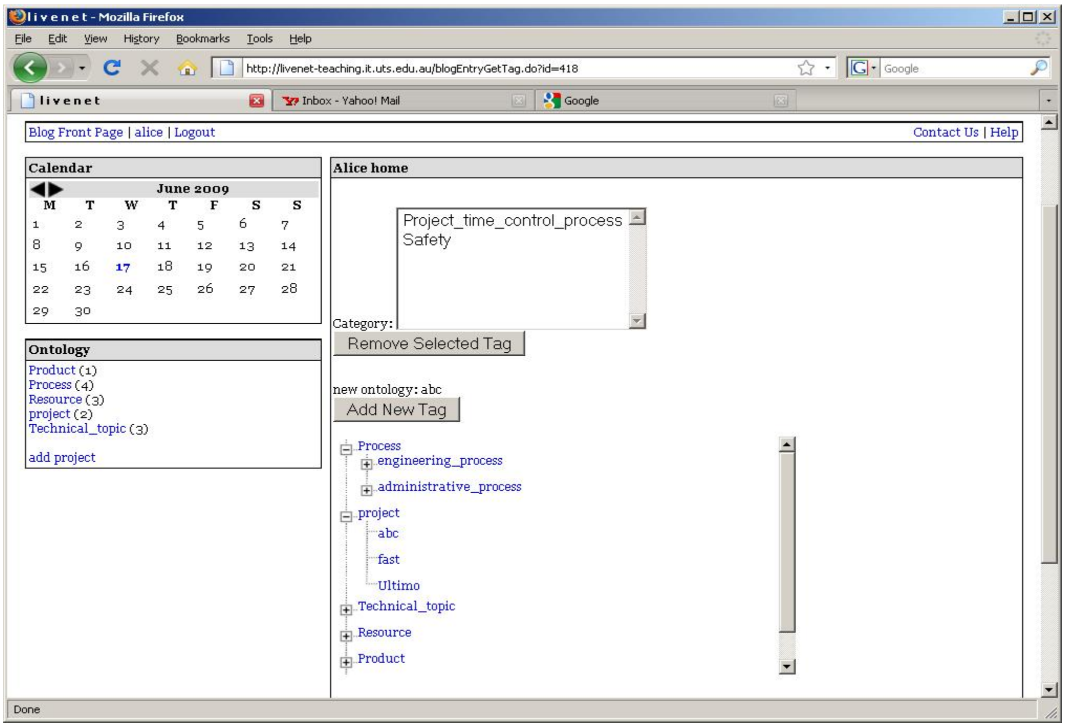

Figure 4: Manual annotation processes

5. SEMANTIC INFERENCE

After the annotation process, LiveNet users can navigate the system by using the machine-understandable meta-data to find blog entries effectively. The dynamically constructed inference can help users to browse related items and follow labeled annotations to find useful information and knowledge items.

The semantic inference engine can also use the ontology to help the users building up semantic queries and set up semantic subscription. Blog users can subscribe to a blog category such as "how to use a particular building material in a special construction process." The query and subscription processes are based on blog aggregation.

"Aggregation" is a very useful knowledge management feature. By generating RSS (Really Simple Syndication) feeds (a machine-readable file which summarizes the web contents), blog users can subscribe to particular blog entries. Blog aggregator can automatically check a list of RSS feeds regularly on behalf of a user and supply relevant blog entries (blog articles) to the user. This is an advanced knowledge management facility, as the knowledge can be actively "pushed" to the users rather than passively wait for users to find and access.

The inference engine provides two important functions in these modules: classifying concepts and identifying semantic links:

- Classify concepts: The inference engine is able to compute the relationships between concepts (in IT terms, the relationships between sub-classes and super-classes). This also helps to define the "properties" of a certain concepts. For example, "Clay" is a kind of "Natural_material", i.e., "Clay" is a subclass of "Natural_material". Therefore, "Clay" has all the properties that "Natural_material" has (in IT terms, this is called "property inheritance"). Therefore if a query is to find all the blog entries on "Natural_material", then the blog entries on "Clay" will be included.

- Identify semantic links: All concepts in the ontology are related to each other via certain relationships; for example, a "process" utilizes "resources" in a "project" to produce a "product" (see Figure 3). As a result, blog entries, after being annotated by these concepts, are also linked to each other semantically. Semantic inference can find and track these links and process complex queries. For example, semantic inference can identify the blog entries on using materials M, which is supplied by supplier S, used in process P and used by contractor C in project T.

The inference engine serves as the "search engine" in the ontology-based knowledge retrieving process, and the following example demonstrates how ontology-based retrieval is more effective than the traditional keyword search.

6. EXAMPLE OF KNOWLEDGE RETRIEVAL

To illustrate the functions of retrieving knowledge items effectively from *LiveNet* system, an example is given here. Some "knowledge events" from previous researchers are used in this example. A few mock blog entries (see below: Mock blog 1 to 5) are created based on the audio diary recorded by some UK researchers (Boyd & Xiao, 2006). Details of Mock Blog 1 is given below, Blog 2-5 are not included here due to limited space. Mock Blog 6 is created based on some trade technique handbooks.

Mock Blog 1: Building sand did not arrive as planned

T, a site manager on a small site, had ordered sand for one day but the sand did not arrive. This was planned for making mortar to lay bricks that day. His problem-solving dilemma was whether to lay off the bricklayers for the day and upset them as they do not get paid, keep the brick-layers on but lose the cost of the work time, or buy some very expensive sand from the retail source which loses time as well cost. The solution taken was a compromise in that he talked to the bricklayers to engage them with his problem and lay them off for only half a day (loss of 4 hours' money).

(Ontology Tag: Human_resource_management, Labour_force)

Mock Blog 2: Client continue to use the service yard during construction

Mock Blog 3: Shower channels were stolen

Mock Blog 4: Groundwork subcontractor has to be replaced

Mock Blog 5: Client's new QS did not like face to face communication

Mock Blog 6: An easy step-by-step guide on laying bricks

Move the stringline for the outer face of the wall on the hurdles, then lay the bricks out dry for the first course along the foundation to work out their correct placement. Now lift them to one side and set this course accurately to the stringline by laying a mortar bed. Using a trowel. To lay a bed, chop off a *section of mortar with the trowel that's about the same size as it. Separate it from the bulk of the mortar with a clean slicing action and shape it into a curve, as near to a sausage shape as possible. Sweep the trowel underneath to lift it from the mortar board in one go. Slide the trowel sharply backwards to lay the mortar in a sausage shape, and use the tip in a stepping action down the middle to spread it out. Lay enough for at least two bricks at a time. Lay the first brick in place and tap down with the trowel handle until the top outer corner of the brick rests against the stringline. The mortar should make a 10mm joint. Buttering. For the next brick, butter one end to create the vertical join before you lay it. Hold the brick in one hand and scrape on mortar with your trowel to one 'perp' end. Place the buttered end against the first brick to form a vertical joint 10mm thick and firm down with the trowel handle, again so the top outer edge for the bricks rests against the stringline. Repeat this for the next four or five bricks, still following the stringline. Test with your spirit level that the bricks are truly horizontal. Carefully scrape off excess mortar with your trowel. When the first course is laid, you can dispense with the hurdle line and use bricklayer's pins and line, gauge rod and spirit level instead.* (Ontology Tag: Productivity)

Using traditional keyword searching, "brick" is entered in the search function, and both mock blog 1 and blog 6 are returned. Figure 5 is the screen shot of the search results.

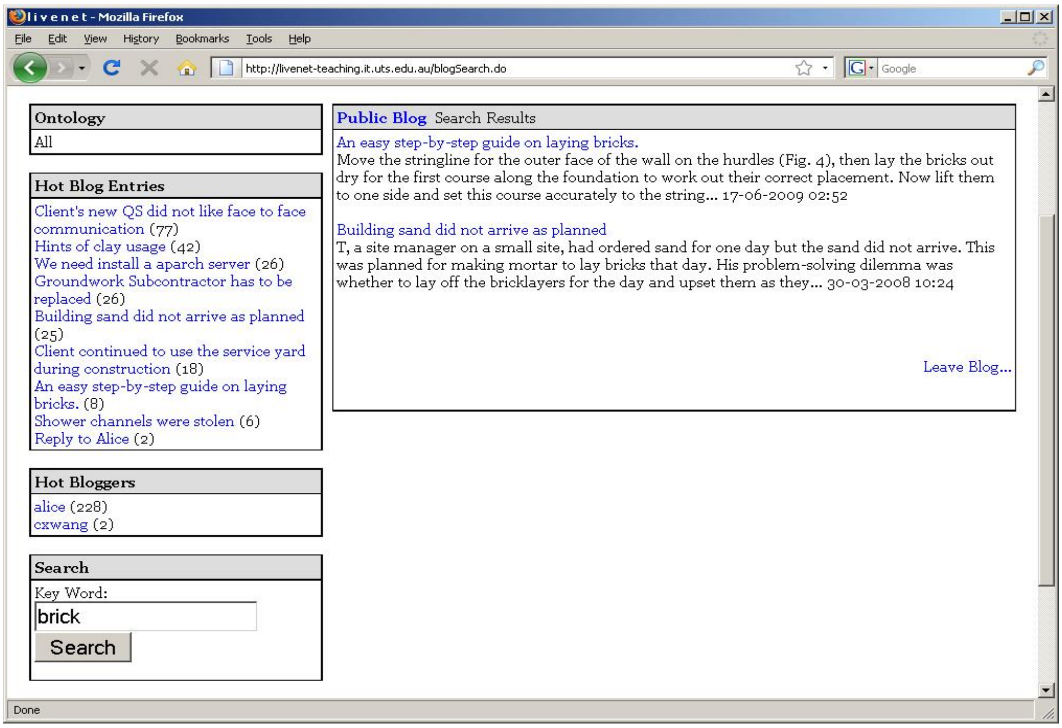

Figure 5: Search blog entries by keyword

In the e-CONGNOS ontology hierarchy, the following concepts and relationships are listed:

Process -> Administrative_process -> Human resource management ->…

Technical topic -> Productivity -> …

Therefore the Mock Blog 1 is under the ontology "Administrative Process", and the Mock Blog 6 is under "Technical topic" although both are related to "bricks". When using ontology enhanced search, e.g. "administrative process" is put in the ontology restrictions, the search returns only Mock Blog 1 as the semantic inference can identify that Mock Blog 6 does not satisfy the requirement. The screen copy of this ontology-based retrieval outcome is in Figure 6.

The above examples demonstrated that conventional keyword-based search may return irrelevant results. Ontology based search returns all the knowledge items in the category that the knowledge seeker is intend to find. Ontology-constrained keyword search narrows the search results and provides more accurate search outcome to the users. More examples are to be input to the system and they will be reported in later research publications.

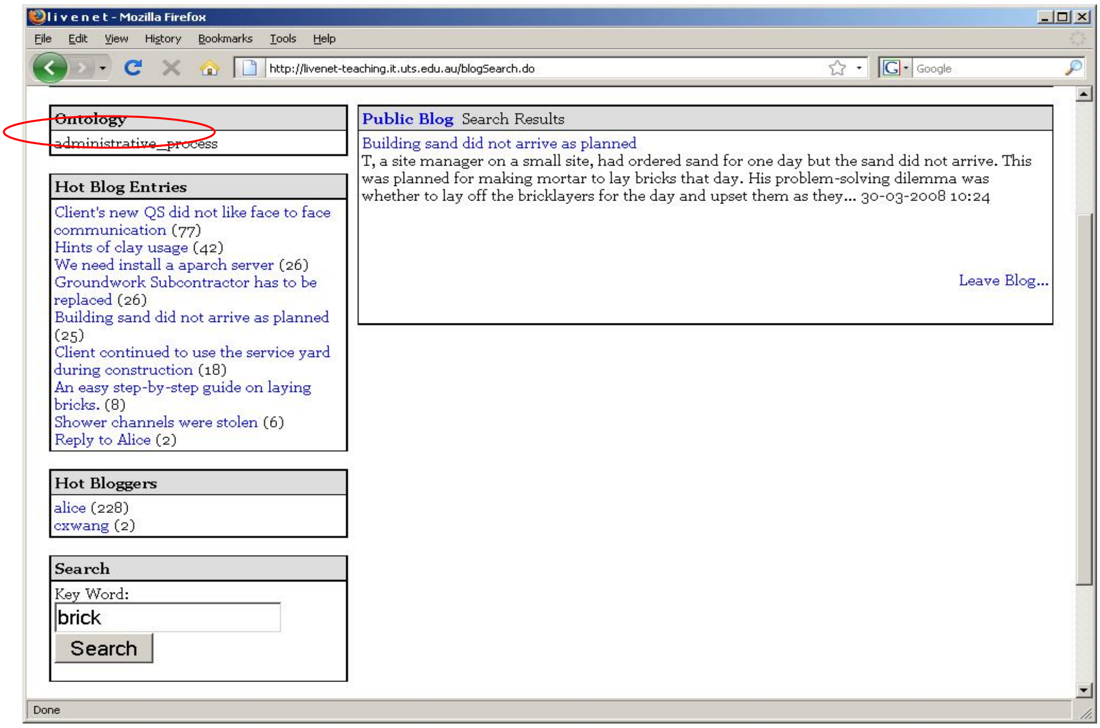

Figure 6: Search blog entries by keyword and ontology

7. CONCLUSIONS

This research shows advanced information technology, such as domain ontology and semantic web, can promote and facilitate knowledge sharing activities more efficiently in the construction industry. The contribution of this research is it demonstrates the possibility of using construction domain ontology in semantic blog sites to facilitate knowledge transfer. The proposed ontology-based system annotates knowledge items with machine-understandable meta-data, which can facilitate information categorization and retrieval. Semantic inference can help users to build up quires and set up subscription, and therefore, not only locate knowledge items quickly and accurately from the system, but also has the potential to "push" information and knowledge to the user automatically.

Following the spread of the usage of social network software in our everyday life, and the trend of publish personal option by using blog technology, it is reasonable to believe that a general collaborative software combined with dedicated designed construction domain ontology module will provide a lower cost but more effective approach for managing knowledge in construction industry. Hence it is clear that the next step of this research should be how to modular the proposed framework and make it ready to be linked with some commercial collaborative software, and encourage more company and professionals to use it.

ACKNOWLEDGMENTS

This research is supported by the CIOB Australasia special research grant 2008-2009.

REFERENCES

Australian Bureau of Statistics (2008), Year Book Australia 2008, 582-583.

Boyd, D., & Xiao H. (2006) “Turning problems into knowledge-events: Knowledge management for construction SMEs”, *Construction Information Quarterly*, Vol. 8, 2006.

Cayzer, S. (2004) “Semantic Blogging and decentralized Knowledge Management.” *Communications of the ACM*, Volume 47, Number 12, 47-52.

El-Diraby, T. A., Lima, C. & Feis, B. (2005). “Domain Taxonomy for Construction Concepts: Toward a Formal Ontology for Construction Knowledge.” *Journal of Computing in Civil Engineering* Vol.19, No.4, 394-406.

Fong, P. S. W. (2005) “Managing Knowledge in Project-based Professional Services Firms: An International Comparison.” *Management of Knowledge in Project Environments*. P. S. W. F. a. Z. I. Peter E. D. Love, Elsevier.

Hawryszkiewycz, I. T. (2000) “Evolving Knowledge Intensive Community Networks.” in *Proceedings of Ausweb 2000*, Cairns, June 2000.

Lima C., S. J. & Böhms M. (2003) The bcXML: supporting ecommerce and knowledge magagement in the construction industry. ITcon. Vol. 8, 293-308.

Wang**, C.** (2010), “An Advanced Knowledge Management and Collaboration System in Semantic Web Environment for Construction Organisations”, in *Proceedings of the 8th International Conference on Construction and Real Estate Management*, 1-3, Dec. 2010, Brisbane, Australia