BIMTag: Semantic Annotation of Web BIM Product Resources Based on IFC Ontology

Ge Gao[1,2], Yu-Shen Liu[1], Meng Wang[1], Xiao-Guang Han[1,2]
[1]BIM Research Group, School of Software, Tsinghua University, Beijing 100084, China
[2]Department of Computer Science and Technology, Tsinghua University, Beijing, China
gg07@mails.tsinghua.edu.cn, liuyushen@tsinghua.edu.cn, wm0409@gmail.com

Abstract. With the rapid popularization of BIM technologies, BIM resources such as building product libraries are growing rapidly on the web. However, these numerous resources are usually from heterogeneous systems or various manufacturers, possibly with ambiguous expressions and unclear relationships for product descriptions, which do not effectively support information retrieval. Inspired by the Semantic Web techniques, this paper presents a new semantic annotation method for web BIM product resources based on IFC ontology. First, a domain-specific ontology is built to represent the knowledge in AEC for semantic annotation, which mainly utilizes the concepts and relationships defined in the IFC schema. Then with reference to the ontology, an automatic concept annotation method is proposed to annotate web BIM resources. A prototype system, named BIMTag, is developed and applied to a search engine for demonstrating the utility and effectiveness of our method.

1. Introduction

With the rapid popularization of Building Information Modeling (BIM) technologies in AEC (Architecture, Engineering and Construction) industries, BIM resources such as building product libraries are growing rapidly on the web. For instance, the well-known Autodesk Seek[1] is an online system, which provides a large repository of BIM products and allows users to search. Currently it has more than 65,000 building products from nearly 1,000 manufacturers. BIMobject[2] is another widely visited website containing over 450,000 BIM objects. Other web libraries (e.g. National BIM Library, SmartBIM) also contain a large number of BIM products. With the vast volumes of information of BIM resources becoming available, the problem of "information overload" is being increasingly recognized in AEC.

The typical libraries of web BIM resources (e.g. Autodesk Seek) consist of both structured BIM models and associated unstructured documents (including product specifications and descriptions). The structured BIM models normally are in their native file format dependent of various BIM software vendors (e.g. Autodesk Revit, Graphisoft ArchiCAD) or in industry-neutral file format (e.g. IFC/ifcXML). The instructed product specifications are some documents of textual descriptions for BIM models, which include product functions, dimensions, materials, performance, attainability, manufacturers, etc. These product documents are independent to file format of BIM models. In particular, much of knowledge is embedded in the textual product documents generated during design and construction phases. However, these numerous existing documents are usually from heterogeneous systems or various manufacturers, possibly with ambiguous expressions and unclear relationships for product specifications and descriptions, which do not provide effective support for information retrieval, interoperability and other applications.

Semantic annotation is about attaching additional information (e.g. names, attributes, comments, descriptions) to a document or a selected part in text (Kiryakov, 2004). It could

[1] Autodesk Seek. http://seek.autodesk.com
[2] BIMobject. http://bimobject.com/

help to reduce the ambiguity and unclearness of the natural language through expressing the notions and their relationships in a more formal language. A lot of researchers and developments have contributed in semantic annotation (Bueno, 2010; Rajput, 2011; Fiorelli, 2013; Mota, 2013), which lower the barrier of linking shared data with the web resources in various areas. However, the lack of domain-specific formal knowledge is obstructing the utilization of semantic annotation in BIM-related area. Therefore, the crucial problem is how to build the BIM-oriented formal knowledge and how to use the formal knowledge to annotate the web content of textual product specification documents in the semantic level.

Inspired by Semantic Web, this paper presents a novel semantic annotation method for web content of BIM product resources based on IFC (Industry Foundation Classes) ontology (BuildingSMART, 2014). In our method, the standardized concepts, properties and relations of building products within the IFC schema are utilized to mark precise information (e.g. names, types and attributes) within a specific document using the RDF (Resource Description Framework). Compared to the conventionally manual annotation/tagging approaches, which are time consuming and subjective, our semantic annotation method is automatic and more objective. It enriches unstructured specifications and descriptions of building products with a context that is further linked to the knowledge of BIM-specific domain.

BIM product resources of web pages are usually written in a way for human beings to read, but not formally described as the standardized concepts within the IFC standard. This results in that original product description documents are far from ideal in retrieval application. Many studies have shown that fundamental characteristics of human verbal behaviour result in the retrieval difficulties. Because of the tremendous variety in the vocabulary can be used to describe the same meaning or concept (synonymy), people will often use different words from information providers. As a result, the relevant concepts in the IFC schema will therefore be missed. On the other hand, since a single word often has more than one meaning (polysemy), the irrelevant concepts in the IFC schema will often be retrieved. For example, the terms (“coating”, “finishing” and “covering”) express a meaning of “a decorative texture or appearance of a surface” in AEC area, but only the term “covering” appears in IFC standard. Consequently, when given a query “covering”, some web pages of BIM product resources including “coating” or “finishing” but without “covering” will not be matched.

Under the guidance of IFC standard, our annotation aims to discover and extract the formal concepts within a document to represent its semantics. For the above example, our method will annotate all similar terms (i.e. “coating” and “finishing”) appearing in the document with the same concept “covering” in IFC. Furthermore, a concept may have various hyponyms (subclass), for example, “covering” can be divided into many subclasses (“sleeving”, “roofing”, “ceiling”, “insulation”, “membrane”, “cladding”, “wrapping” and “flooring”), according to the type enumeration in IFC schema. These subclasses also need to be annotated with their superclass (i.e. “covering”) to enhance the annotation. We put forward a combination of IFC schema and WordNet (Miller, 1995) to make concept annotation widely for web BIM resources. Our annotation provides a mechanism linking the documents and queries with the standardized concepts, relationships and properties, which can also be used to many information retrieval applications, such as matching the similar documents, extracting the topics of the documents, et al. Our main contributions can be summarized as follows.

- The domain-specific ontology is built to represent the knowledge in AEC for semantic annotation, which mainly utilizes concepts and relationship defined in the IFC schema.
- With reference to the built ontology, an automatic concept annotation method is proposed to annotate web pages of BIM product resources.
- The presented annotation method is applied to a search engine in order to demonstrate

the utility and effectiveness.

2. Developing the IFC Ontology for Semantic Annotation

2.1 IFC and Ontology

In information science, *ontologies* are defined as "formal, explicit specifications of shared conceptualizations". Ontologies can be roughly divided into two categories: *general ontologies* and *domain ontologies*. The interest of general ontology is the whole world, while domain ontology focuses on specification of particular domain conceptualization. Although some general ontologies, e.g. WordNet (Miller, 1995), contain a large number of general concepts, they are not designed for specific domain, which may lead to inaccurate description of concepts in AEC. In contrast, domain ontology is a representation of semantics in a particular domain, which often consists of a hierarchical description of important concepts precisely defined on the domain, along with description of properties of each concept (Lukasiewicz, 2008). Domain ontology is considered as a key element to enhance domain-specific semantic annotation (Gruber, 1993; Rezgui, 2006; Li, 2008; Lin, 2009).

The IFC specification developed by the BuildingSMART plays a crucial role to facilitate interoperability. The IFC schema is written in EXPRESS, which is the same data definition language used in STEP or CIS/2. The IFC data exchange file (with suffix ".ifc") is an ASCII file format used to exchange building information between different applications. Today, the IFC standard has been widely supported by the market-oriented BIM software vendors. As the most widely used taxonomy and specification in BIM applications, the underlying IFC specification is therefore our preferred candidate semantic resource, which provides a sharable skeleton on which the BIM-oriented annotation ontology can be built.

Recently, several IFC-based ontologies have been studied for particular application needs. For instance, Pauwels (2011) applied an IFC ontology to semantic rule checking. Beetz (2009) converted the IFC schema into the OWL format. Zhang and Issa (Zhang & Issa, 2013) used an IFC ontology to extract partial model from a complete IFC model. Several studies used IFC ontologies for querying spatial information within a BIM model (Nepal, 2012; Borrmann, 2009). However, very little attention is paid to develop an IFC-based ontology for annotation of web BIM product resources. To develop such ontology, our work consists of two steps: (1) mapping EXPESS language to OWL and (2) trimming the ontology for concept annotation.

2.2 Step 1: Mapping EXPESS Language to OWL

Our BIM-oriented ontology is based on the latest version IFC4 specification, which is expected to reuse the rich semantic contents of newly developed IFC specification. The IFC4 enhances the capability in several areas of building elements, building service elements and structural elements, and also accompanying basic definitions. Currently, it contains 766 entities, 206 groups of enumeration types, 408 groups of property sets, 1691 individual properties, and lots of defined types, select types, quantity sets, functions and rules.

Domain ontologies have a variety of representative forms, among which OWL (Ontology Web Language) is selected as the modeling language in our study. With syntax extensions from the RDF – a light weight representation of data and knowledge, OWL has been proposed by W3C as the ontology language of the Semantic Web (OWL, 2012). In our approach, the EXPRESS schema of IFC is translated into OWL ontology by mapping different language structure between them. This step is similar to Beetz (2009) but with different purpose.

The conversion of EXPRESS simple types (e.g. String, Integer, Real, Binary, Boolean, Logical) is direct, as they have equivalents in OWL (XSD types). For example, String type is mapped into `xsd:string` type. EXPRESS entities are used to define the concepts from the real world, which have properties characterizing them. We directly map entities as well as their subtypes and supertypes to OWL classes and inheritance. The OWL data property is used to represent EXPRESS simple attributes, as well as the OWL object property represents named attributes. The OWL cardinality is used according to the EXPRESS attribute's optional flag. Some EXPRESS inherited attributes from a supertype entity can be renamed or retyped according to the user's needs. These redefined attributes are mapped using OWL classes-specialization. Regarding EXPRESS inverse attributes, which are pointers to the relating entity, the OWL inverse property is available for the same purpose. The constructed data type in EXPRESS enumeration data type is a concept common to many other languages, and defines a set of names to be used in a specific domain. Enumeration and Select types are mapped through the use of OWL clauses `owl:oneOf`.

2.3 Step 2: Trimming the Ontology for Concept Annotation

Although the IFC specification provides a unified standard for construction engineering data exchange, most IFC-based applications still stay in the level of data exchange, not from the perspective of annotation and retrieval. One of the limitations is that the scope defined in the IFC specification is too broad for information retrieval purpose, which will greatly reduce the accuracy of annotation and search due to a large number of irrelevant terms included in the original IFC specification. Therefore, these terms in IFC specification that are relevant to annotation and retrieval should be kept in the ontology, while the ones that are irrelevant to annotation and retrieval should be removed. In this paper, only the concepts (e.g. walls, beams or doors) inherited from the entity *IfcObject* are kept, which are the generalization of concrete thing or process. Consequently, the built ontology is composed of six aspects, which are the subclass of "objects": "Actor, Control, Group, Process, Product, and Resource".

Most of previous studies (Beetz, 2009; Pauwels, 2011) focused on mapping between the IFC schema and the Semantic Web, while ignoring the measure of semantic similarities between BIM domain concepts. Their ontologies, which are directly converted from the original IFC schema, are in graph-like structures, as a result, in which there will be many cyclic references among entities and attributes. In contrast, a tree-like structure (Yan, 2011) is a more ideal structure often used for representing ontologies in information retrieval. Since the similarity measure is the foundation of document ranking in ontology-based retrieval, we typically trim the IFC ontology as a tree-like structure with convenience of further retrieval applications.

The IFC schema is designed for computer instead of human readability, so it has a special formalized naming convention. If the ontology is directly translated from the IFC schema, the terms and concepts cannot directly use the plain English words. In fact, the data items in most existing IFC ontologies are still following the direct naming convention (Beetz, 2009; Pauwels, 2011). For example, the data item names for types, entities, rules and functions start with the prefix "Ifc" and continue with the English words in Camel Case naming convention (no underscore, first letter in word in upper case); the attribute names within an entity follow the Camel Case naming convention with no prefix; the property set definitions that are a part of this standard start with the prefix "Pset", etc. In practice, these concepts or names will be frequently used as query words in the retrieval process. Therefore, they should be refined by the process of segmenting the names of entity, removing the prefix and eliminating redundancy, to construct ontology linguistically suitable for retrieval applications. This process is carefully and semi-automatically handled in IFC ontology in Step 2.

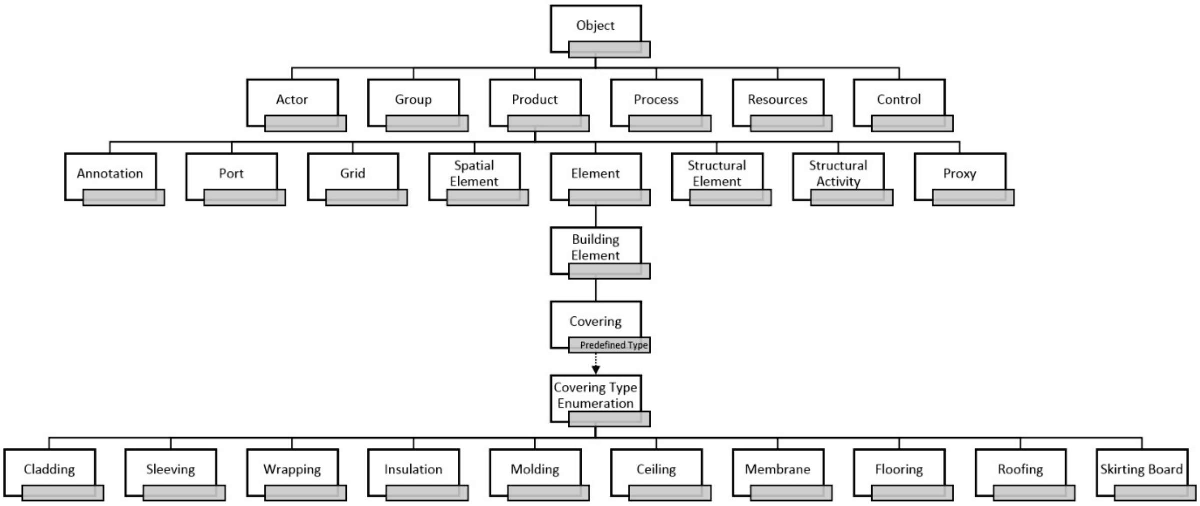

Figure 1: Part of the IFC ontology developed for semantic annotation.

2.4 Overview of Developed IFC Ontology

After the above two steps are done, the resulting ontology is constructed. It is organized in a tree-like data structure, in which each node represents a concept and each arc represents a relationship. A portion of our IFC ontology is shown in Figure 1. In this figure, the concept, e.g. *covering*, is inherited from the root node *object*, and it has a property *PredefinedType* to map to an enumeration *covering type enumeration.*

In this paper, we only consider building information for web BIM product resource retrieval purpose, which is mainly relative to object class, so the resulting IFC ontology is a sub-tree of IFC schema rooted with *Object*. An *Object* is the generalization of any semantically treated thing or process. In IFC4 specification, *Object* is the supertype of *Actor*, *Control*, *Group*, *Process*, *Product*, and *Resource*, which consist of the first level of IFC ontology (see Table 1).

Table 1: The outline of concepts in the resulting IFC ontology.

Taxonomies	Number of concepts	Example of concepts	Original IFC entity
Actor	2	occupant	*IfcActor*
Control	11	project order	*IfcControl*
Group	12	building system	*IfcGroup*
Process	4	task	*IfcProcess*
Product	326	covering	*IfcProduct*
Resource	8	crew resource	*IfcResource*

3. The Process of Semantic Annotation

This section introduces the annotation process (Figure 2), which consists of four parts: pre-processing, concept identification, reasoning and linkage, with regard to several data modules: web resources, IFC ontology, WordNet and IFC schema website.

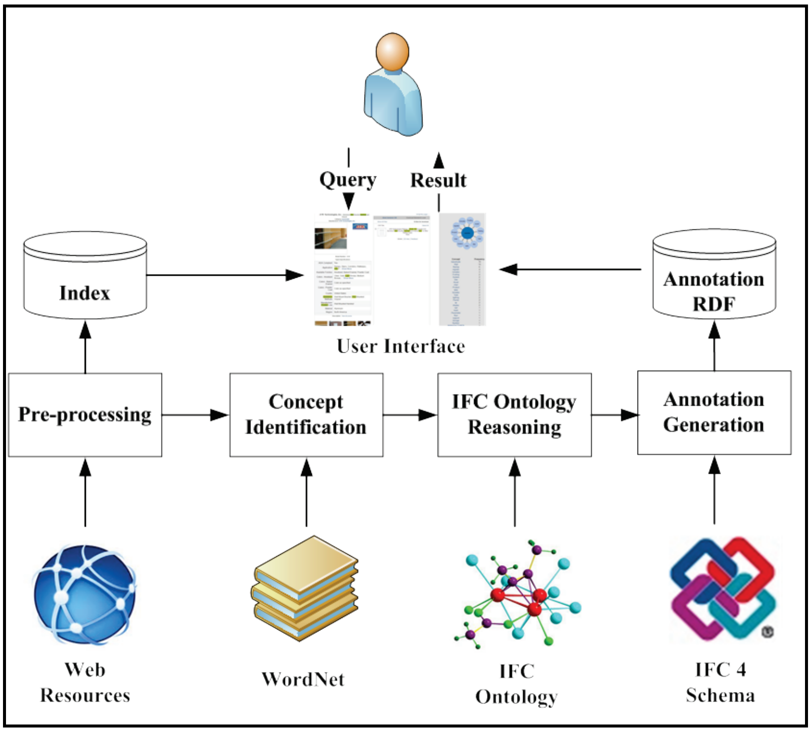

Figure 2: The process of semantic annotation.

3.1 Pre-processing

Before doing annotation, document collection must be prepared in the pre-processing. In our system, a crawler is used to obtain the data set (i.e. the document collection of BIM product resources on web pages) for annotation, where Hetrix Web crawler is selected to collect the BIM-related documents from the web. After the documents are crawled from the web, they should be properly processed on a syntactic level. Its main goal is to build a structured representation of web BIM resources. This process can be divided into six steps:

1) Split the tokens at whitespace. This is the step making paragraph and sentences into the terms which can be processed by the following procedure.
2) Divide the text at non-letter characters and lowercase, where all the characters should be stored in lowercase for convenience.
3) Remove the special words called stop words, which are assumed to carry very little standalone meaning, such as “a”, “are”, “by”, “for”, “on”, etc.
4) Tokenize the terms based on sophisticated grammar, which aims to recognize the special structures like email addresses, acronyms and alphanumeric.
5) Stem and transform the terms to reduce the word form (e.g. “doors” to its elemental root “door”).
6) Storage the terms and their associated frequency and positions in the documents.

3.2 Concept Identification

After the document collection is finished, the concepts can be identified according to the IFC ontology. The tokens of documents generated in pre-processing are extracted and looked up in the IFC ontology. If a term is matched with a concept in IFC ontology, the concept is identified, which will be reasoned and linked to the URL in IFC website (see Section 3.3). However, the IFC specification only contains the limited terms in AEC, often leading to missing many words which are not appearing in IFC ontology. Therefore, the IFC ontology should be combined with some general ontologies for a wider spread. In this work, we combine the synonyms of general ontology WordNet into an extended IFC ontology. When the terms in documents include the terminologies that are not appearing in IFC specification, for instance that a collected document includes a term "shaft" that is not a standard entity defined in IFC4, this extension will be able to translate this term into the corresponding standard IFC term "beam" and then a concept will be also identified in this way.

Since the above-mentioned annotation is a word-level annotation method that is of the finest granularity, our approach bypasses the need for predefined external templates, which is used in the sentence-level or paragraph-level annotation. A list of candidate concepts can be produced according to the noun phrases in the document, and then be used for constructing the mapping. One limitation of this word-level simple mapping is that contextual information is not utilized, which may result in incorrect mapping between terms of documents and concepts of IFC ontology. In WordNet, a term may have various meanings, but not every meaning should be annotated with its corresponding concept in IFC ontology. For example, the term "covering" in WordNet has several synonyms (including "cover", "screening", "masking", "coating" and "application"), of which "application" should not be annotated with the IFC concept "covering" in BIM-specific domain.

In this section, we propose a context-based method to reduce the incorrect mapping, which is inspired by the classical TF-IDF method. Our idea is based on an assumption that if a term in a document represents a concept in IFC ontology, the concept and its related concepts in IFC also often occur in the context of the term. In particular, we present a context-based similarity measure to judge whether a term in a document should be mapped as its candidate concept in IFC. The similarity measure is defined as:

$$Sim(t,c) = \sum_{c_i \in O(c)} TF(c_i) IDF(c_i) / |O(c)| \qquad (1)$$

$$TF(c_i) = \frac{n_i}{\sum_k n_k} \quad \text{and} \quad IDF(c_i) = \log \frac{|D|}{1 + |\{j : t \in d_j\}|}$$

Where: t is a term in the target document, c is a candidate IFC concept for term t, $O(c)$ is the related concepts of c in IFC ontology (e.g. subclass, superclass, etc.), c_i is one of concepts in $O(c)$, n_i is the term frequency of c_i in the document, $|D|$ is total number of documents in the corpus, $|\{j : t \in d_j\}|$ is the number of documents where the term t appears. In Eq. (1), $TF(.)$ rewards the concepts co-occurring frequently with query term t, and $IDF(.)$ penalizes the concepts occurring frequently in the collection. $Sim(t,c)$ is finally compared with a predefined threshold to decide if the term t has strong enough relationship with its candidate concept c.

3.3 IFC Ontology Reasoning

For each concept identified, we can reason about its related classes, relationships and properties in IFC ontology. The useful results can be added to annotation to improve the retrieval and navigation of documents. There are various strategies for reasoning, for example that different relationship and property can be extracted from the superclass and/or subclass within a certain level. According to our experimental results, we use the direct superclass and *PredefinedType* property (to find the related enumeration types). Under this condition, the superclass and predefined type can be annotated with an identified concept. For example, in IFC ontology the entity *IfcFlowController* is the superclass of *IfcDamper*, then the "flow controller" is added to the indexed term of "damper". As a result, when given a user query "flow controller", the resources that have a description about "damper" will be matched.

Our system will find the relative concepts to the query concept and compute their weights according to their relationships. In this way, the related IFC concepts (superclass, properties, etc.) are added to the annotation/index. For example, the class *IfcCoveing* has a property *PredefinedProperty* linked to the *IfcCoveringTypeEnum*, which has the enumeration items *membrane*, *cladding*, *insulation*, et al. The term "covering" will also be annotated to the indexed term "membrane" based on our ontology reasoning. In practice, when we execute the retrieval, the reasoned terms are assigned to the lower weights according to their relationships with the identified concept in the search engine.

3.4 Annotation Generation

We follow the above-mentioned procedure to generate semantic annotation for the document collection, where the RDF format is used to record all relations between documents, concepts, terms, and their relations. RDF does not contain the contents of documents, but it records which terms are included in the document, and which concepts are generated inside the document. This information can be used in the navigation and retrieval on the documents of web pages. This function is implemented by a toolkit Apache Jena, which is semantic framework that provides the system with many tools to treat ontologies and OWL language.

Not only the relationships between documents and concepts are recorded, but also their external references valuable for end users are caught for showing the semantics of concepts in the documents. This annotation process assigns the concepts in the text linked to their semantic descriptions. In our system, the annotated terms are typically linked to the corresponding part of the IFC4 schema on the buildingSMART's official website. Via the use of external or background knowledge, those concepts can be coupled with the formal descriptions, which thus provide more semantics and connectivity to the web.

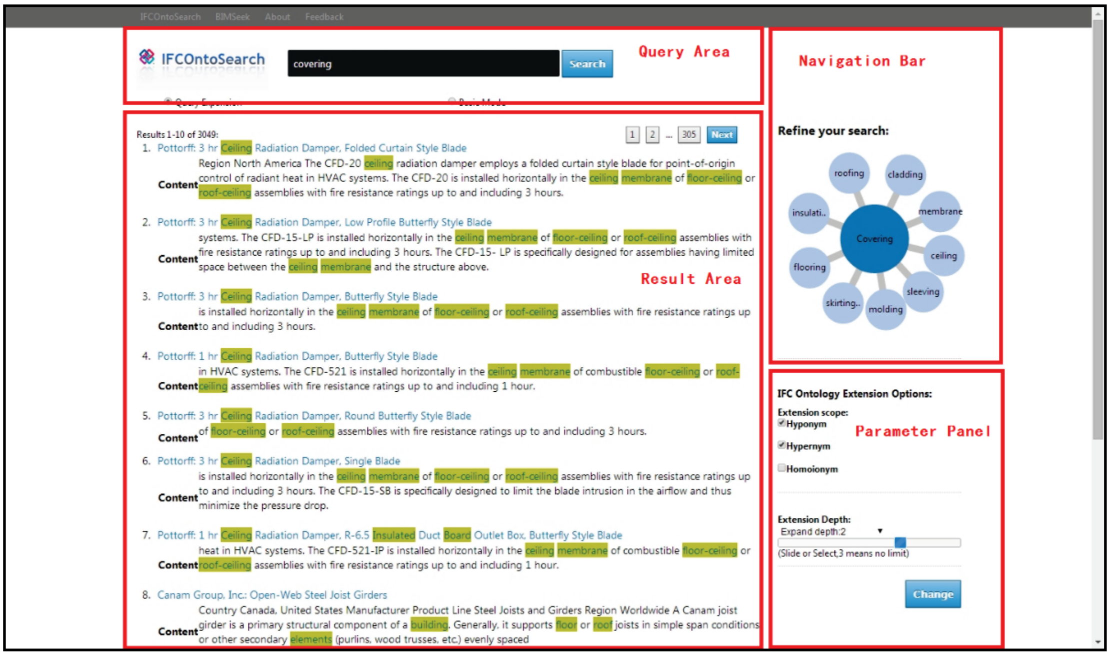

Figure 3: The query interface of search engine combing the presented semantic annotation.

3.5 System and Application

The proposed semantic annotation can enable various types of applications, such as highlighting, indexing and retrieval, categorization, smooth traversal between unstructured text and available relevant knowledge. The annotation is applicable for both web pages and local documents. We design and implement a prototype system according to the methodology and a web BIM product resources scenario outlined above. The prototype system of semantic annotation, named BIMtag, has two main interfaces: the *query mode* (Figure 3) and the *browse mode* (Figure 4). The query mode, named IFCOntoSearch (see Figure 3), implements the generic function for end users to search for BIM product resources, which includes a query area, a navigation bar, a parameter panel and the result area. When the user inputs a query to IFCOntoSearch, the system will identify the concept related to the query. Here the navigation panel on the top right displays the concepts, which are related to the query concept in IFC ontology to help the user to refine their query. Then the system executes a semantic search based on the annotated documents. Finally, the search results are shown in the left hand side of the result viewer.

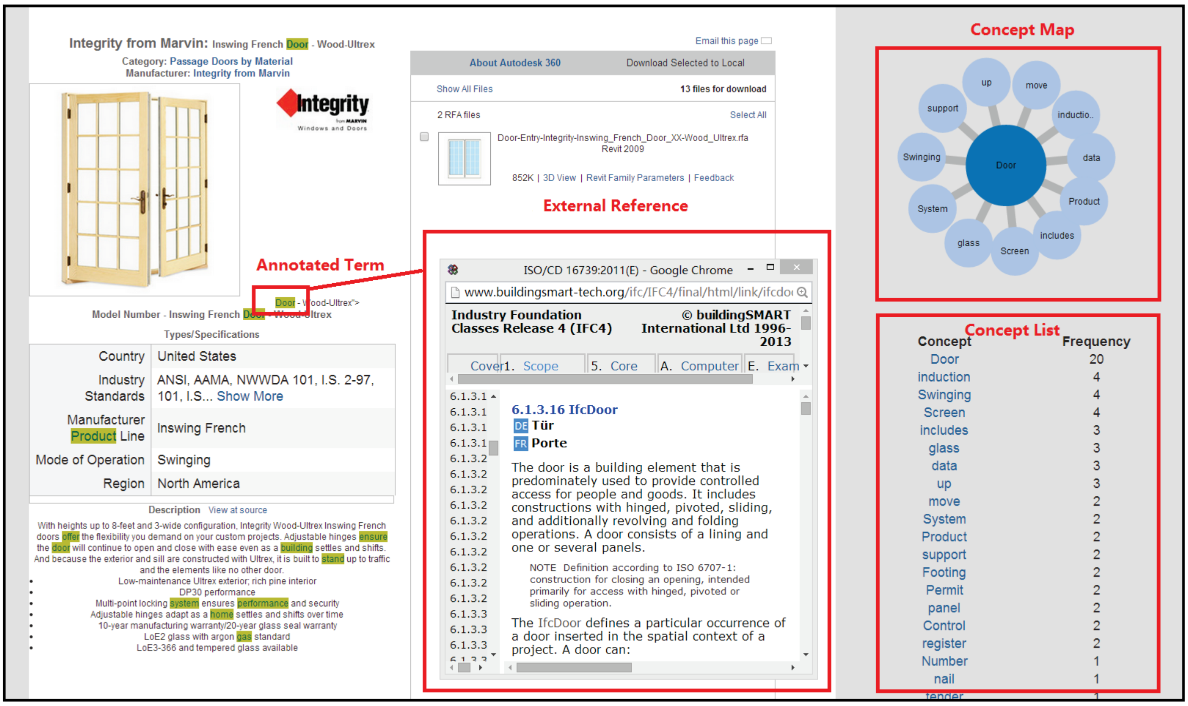

Figure 4: The browsing interface of our BIMtag system.

In the browse mode (Figure 4), when the user chooses to browse one of the resulting documents, the system will display the document in the left hand side of the interface, in which the annotated terms will be marked up with different color. If the user moves his cursor to a marked area, the corresponding annotated concept will be prompted out, where a popup window displays its external reference page of the IFC4 schema at buildingSMART website. The right side of the browser is divided into two sections. The top one displays the simplified "Concept Map" to show 11 frequently occurred concepts contained in this document. The list on the bottom ("Concept List") shows all the concepts that show up in this document, which provides information of how many times each concept occurs in this document. In addition, each concept can be clicked to navigate these documents containing it, so that the user can focus on one concept and retrieve its relevant documents among all document collection.

In order to evaluate the effectiveness of the proposed annotation method, we run an experiment of information retrieval application on a document collection. The document collection consists of a set of BIM resources pages crawled from Autodesk Seek website, along with a sequence of test queries. To measure the performance of the retrieval system, the set of relevant documents for each test query is extract from the classification system on the website. Currently, BIM resources in the collection are marked with OmniClass code, which is an implementation of ISO 12006-2. The OmniClass code is treated as “ground truth” for our retrieval evaluation. Hence the type or topic of the resource can be determined.

In the experiment, we compare the retrieval results of BIM product resources without using annotation and with using our semantic annotation. We use the standard evaluation procedure from information retrieval, namely *precision-recall* curves, for evaluating the retrieval results. The precision-recall curves describe the relationship between precision and recall for an information retrieval method. It is desirable to achieve both high precision and recall, but unfortunately this is rather difficult to achieve, especially for the text-based retrieval problem. Figure 5 shows the average precision-recall curves for the retrieval without using annotation and with using out semantic annotation. The results show that our semantic annotation method performs better retrieval than the traditional text-based matching method without using semantic annotation.

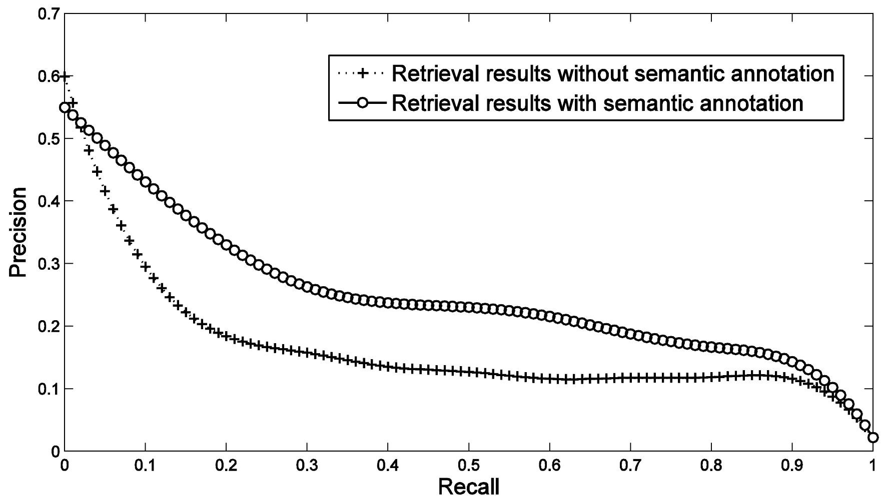

Figure 5: Comparison of the average precision-recall curves for the retrieval without using annotation and with using out semantic annotation.

4. Conclusions

This paper developed a domain-specific ontology based on IFC specification, which encodes the knowledge in AEC area for semantic annotation. With reference to the built ontology, an automatic concept annotation method, combining the IFC ontology and WordNet, is proposed for annotating the web content of BIM product resources. The proposed semantic annotation method can enable various types of applications, such as highlighting, indexing and retrieval, categorization, smooth traversal between unstructured documents and available relevant knowledge. In addition, a prototype system BIMTag is implemented and applied to a search engine for demonstrating the utility and effectiveness of our method. The preliminary experiment shows that our method improves the user’s experience and performs better than the traditional keywords based retrieving method. One of our future work is to develop an evaluation metric to precisely measure the performance of semantic annotation. Besides, since this is a word-level annotation, word sense disambiguation will also be implemented to decrease the number of incorrect annotations in the future. Finally, although IFC is the most widely accepted BIM standard, other possible AEC semantic resources rather than IFC can be experimented later.

Acknowledgements

The research is supported by the National Science Foundation of China (61272229, 61003095) and the National Technological Support Program for the 12th-Five-Year Plan of China (2012BAJ03B07). The second author is also supported by the Chinese 973 Program (2010CB328001) and Chinese 863 Program (2012AA040902).

References

Beetz, J., Leeuwenand, J., Vries, B. (2009). IfcOWL: A case of transforming EXPRESS schemas into ontologies. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 23(1), 89–191.

Borrmann, A., Rank, E. (2009). Specification and implementation of directional operators in a 3D spatial query language for building information models. Advanced Engineering Informatics, 23(1), 32–44.

Bueno, F., García-Serrano, A., Martínez-Fernandez, J.L. (2010). Enrichment of text documents using information retrieval techniques in a distributed environment. Expert Systems with Applications, 37(12), 8348–8358.

BuildingSMART (2014). Industry Foundation Classes (IFC). Available from: http://www.buildingsmart-tech.org/specifications/ifc-overview/ [Accessed 17/4 2014].

Fiorelli, M., Pazienza, M.T., Stellato, A. (2013). A comprehensive framework for semantic annotation of web content. 5th International Conference on Knowledge Engineering and Ontology Development, 245-252.

Gruber, T.R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, 5, 199–220.

Kiryakov, A., Popov, B., Terziev, I., et al. (2004). Semantic annotation, indexing, and retrieval. Journal of Web Semantics, 2, 49–79.

Li, Z., Raskin, V., Ramani, K. (2008). Developing engineering ontology for information retrieval. Journal of Computing and Information Science in Engineering, 8(1), 737–745.

Lin, K., Soibelman, L. (2009). Incorporating domain knowledge and information retrieval techniques to develop an Architectural/Engineering/Construction online product search engine. Journal of Computing in Civil Engineering, 23(4), 201–210.

Lukasiewicz, T., Straccia, U. (2008). Managing uncertainty and vagueness in description logics for the semantic web. Journal of Web Semantics, 6(4), 291–308.

Miller, G. (1995). WordNet: A lexical database for English. Communications of the ACM, 38(11), 39–41.

Mota, M.S., Medeiros, C.B. (2013). Introducing shadows: Flexible document representation and annotation on the Web. In: ICDE Workshops 2013, 13–18.

Nepal, M.P., Staub-French, S., Pottinger, R., Webster, A. (2012). Querying a building information model for construction-specific spatial information. Advanced Engineering Informatics, 26(4), 904–923.

Pauwels, P., Van Deursen, D., Verstraeten, R., De Roo, J., De Meyer, R., Van de Walle, R. (2011). A semantic rule checking environment for building performance checking. Automation in Construction, 20(5), 506–518.

Rajput, Q., Haider, S. (2011). BNOSA: A Bayesian network and ontology based semantic annotation framework. Journal of Web Semantics, 9(2), 99–112.

Rezgui, Y. (2006). Ontology-centered knowledge management using information retrieval techniques. Journal of Computing in Civil Engineering, 20(4), 261–270.

Yan, X., Lau, R.Y.K., Song, D., Li, X., Ma, J. (2011). Toward a semantic granularity model for domain-specific information retrieval. ACM Transactions on Information Systems, 29(3), article 15.

Zhang, L., Issa, R. (2013). Ontology based partial building information model extraction. Journal of Computing in Civil Engineering, 27(6), 576–584.