Research on Construction Method of Semantic Dictionary in Architecture

Zhenbo Bi1*, Huiqin Wang2, Ying Lu2 and Hua Yang3

1 Information Engineering, Zhejiang Ocean University, Zhoushan 316022, China
2 Information and Control Engineering School, Xi’an University of Architecture and Technology, Xi’an 710055, China
3 Engineering department, Zhejiang Ocean University Donghai Science and Technology College, Zhoushan 316000, China
Email: bzb136@sina.com

Abstract. Creating a semantic dictionary is an important means of solving semantic problems in the BIM core technology system, and it is one of the cores of the application of BIM technology in the architectural field. At present, the research on IFD-like libraries in China is still in its infancy. How to construct a semantic dictionary suitable for the needs of the domestic civil engineering field is one of the key issues in domestic BIM technology research. In response to this problem, this paper proposes a method of constructing an IFD Library based on the ontology of the architectural domain. Firstly, it discusses the basic problems of IFD Library-like, then analyzes the root causes of semantic ambiguity in the application of BIM technology and the methods to solve semantic ambiguity, and finally explores and verifies the method of how to construct IFD-like Library based on the ontology of the architectural domain.

Keywords. Building Information Modeling, IFD library, ontology in the building domain, semantic dictionary in the building domain.

1. Introduction
Currently, the BIM method makes the seamless communication and sharing of information in the life cycle of construction projects a reality. At present, the core of BIM technology includes three types of technology: semantic technology [1], storage technology and process technology. Semantic technology is mainly realized through the standardization of terminology, such as the IFD Library (ISO 12006-3) standard. Storage technology is mainly realized through the formulation of a neutral data format, such as IFC (ISO 16739) standard. The process (to solve the timing and information exchange problems of application logic) is mainly realized through standardized process technology, such as IDM (ISO TC59 SC13) standards. The three are connected to each other to form a triangle. The side length represents the extent of their respective functions. The area of the triangle represents the application value of BIM technology. When the three are fully matched to the greatest extent, they form an equilateral triangle. At this time, the area of the triangle is the largest. The value of BIM technology can best be used.

Through literature research, in the core technology system of BIM, the development of semantic issues is relatively slow compared to the other two aspects. Compared with domestic and foreign research, the development abroad is relatively fast. For example, in the formulation of the IFD standard, countries such as Canada, the Netherlands, Norway and Portugal have participated in the formulation of the standard to varying degrees [2]. In addition to these countries, the United Kingdom,
Germany, the United States, and Japan have also made great achievements [2-4]. They have built a termbase, namely a semantic dictionary, for their own national languages and habits. It is still difficult to see such a standard term database in China. The key point is that the termbases that these countries have built are not suitable for China. China needs to build its own IFD Library-like semantic dictionary based on actual conditions, taking into account regions, languages, and habits. In view of the important status of semantic dictionaries in the BIM family and the reality of China's vigorous promotion of BIM technology, according to the domestic situation, this article will discuss the construction of Chinese IFD Library-like semantic dictionaries.

Currently, there are few documents on how to construct the IFD Library. Reference [4] introduces the IFD project JCCs in Japan, which is developed on the basis of ISO 12006-2 and 12006-3; based on Semantic Web technology, reference [5] studies an IFD library case of a specific building component (skylight), and the IFD library can be shared through web services. In the above literature, ISO 12006-2 and 12006-3 are essentially standard terminology classification systems in a specific field. Based on the system, it is helpful to extract specific concepts and relationships in the field. Similar classification systems include uniclass for building and construction practice in the United Kingdom, CSI and masterformat in the United States [4], and “digital definition of building objects” in China. The foundation of semantic web technology is ontology technology. Ontology is a concept framework for specific domain. It gives a set of vocabulary identification and a set of concepts. These words are terms. In essence, they are conceptual models with different levels of description. Therefore, we believe that the construction of China's own IFD Library-like semantic dictionary based on ontology technology, especially based on ontology in the field of architecture, is an important research direction and has potential research value.

2. IFD Library
IFD Library is called International Framework for Dictionaries (IFD) Library, which is one or more dictionary libraries [3, 6-8]. IFD Library is a standardized description of terms (or concepts) and attribute sets in construction engineering under the framework of international standards. Chinese means “IFD library”. The main content includes semantic norms, context norms, uniform identifiers, etc. It describes the semantics of architectural concepts in different situations.

Due to the diversity and ambiguity of natural languages, it is difficult to ensure that information providers and information requesters from different countries, regions, cultural backgrounds and languages have a completely consistent understanding of the same concept in the same construction project. For example, different language backgrounds have different understandings of the concept “door”. If the Norwegian “dor” is translated into English “door”, both have the meaning of “door”. But “dor” in Norwegian means “door frame”, which should correspond to “door set” in English, and “door” in English means “door leaf”, which corresponds to “dorblad” in Norwegian. In this way, the concept of “door” expressed in different languages has semantic inconsistencies, which will cause serious misunderstandings and errors in the project. In the Chinese language, “angle steel” is called “angle iron”, etc. There are many similar situations.

In view of the above situation, it is necessary to build a mechanism so that each concept in the project has a logo connected to it. The current IFD assigns a specific scope (global, country, region, industry, or construction project) to each concept or term with a globally unique identifier GUID (Global Unique Identifier), such as “3vHRQ8oT0Hsm00051Mm008” as the GUID of the concept “door”. Like a person’s ID number, each concept in the project has a unique identifier connected to it. As long as the Guid of the exchange information is provided, it doesn't matter which language and idiom the project members use to describe the information in any background [3], but the objects referred to are consistent.

3. The Root Causes of Semantic Ambiguity in Bim Application
In the construction project based on IFC, the information description and association of building entity's attribute set often conflicts and is not easy to identify. The fundamental reason is that IFC’s
attribute set uses the string of name attribute as identification [9]. String as a semantic information expression, although it has the advantages of easy to understand and remember. But there is uncertainty in describing a concept. The main reasons are as follows:

- A concept can be expressed in English, other languages and even Chinese.
- Even if the same language is used, there will be customary expression with local characteristics.
- There are also full names, abbreviations and proverbs.

When the computer processes BIM model information, it is necessary to identify the attribute set by the name of the attribute set. However, due to the uncertainty of the string type attribute set name, the existence of the attribute set may occur and the computer can not correctly identify the semantics, which will affect the automatic processing of the computer.

4. Solutions to Semantic Ambiguity

In IFC-based BIM applications, terms and attribute sets can be identified by GUID and stored in a specific global server for shared access by project participants. GUID, as a unique identifier for terms and attribute sets, establishes a bridge between different forms of string expressions of the same concept. For example, in the attribute description of “window” in figure 1 [10], different information sources have their own concept of “window”, and each source refers to a specific range of attribute sets of “window”. In fact, each information source only involves part of the information in the “window”. Some attributes can be shared among different information sources, while some are not. Here, IFD Dictionary records the attributes of all information sources about “window”, thus forming a concept of “window” which contains all possible attributes of “window”. At the same time, it records the initial information source of each “window” attribute. Thus, a complete dictionary of “window” is formed. Finally, each attribute is assigned a unique identification code in a specific range. When recognizing the semantic information of the concept of “window”, the string description is ignored, and the consistent understanding of the information expression can be ensured through the identification of the guid.

![Figure 1. “Window” attribute description from different source [10].](image-url)
In summary, building an IFD Library-like semantic dictionary is an effective method to resolve semantic ambiguity in BIM applications. Its basic tasks include the following aspects:

- Concept and concept attribute acquisition.
- Calibration of the relationship between the concept and its attributes.
- Attribute coding.
- Establish a database and store the IFD dictionary.

Among these aspects, the core is the first two aspects, this article will focus on the research. Among them, how to construct the IFD Library-like data dictionary, that is, which method and approach to construct, is the core of the research on the IFD Library-like data dictionary.

5. Solutions to Semantic Ambiguity

5.1. The Relationship Between Architecture Domain Ontology and Ifd Library

IFD is essentially a terminology library or ontology standard. To build an IFD Library, the basic concepts or terms of the domain should be extracted first, and then the attributes of these concepts should be defined [11-12]. The ontology of the architectural domain is a clear formal specification of the conceptual model of the shared architectural domain, which contains four meanings: conceptual model, clarity, formalization and sharing. Conceptual model is a model obtained by abstracting some concepts related to phenomena in the field of architecture, and is independent of the specific environment; clarity means that the types of concepts and the constraints on the use of these concepts are clearly defined; formalization refers to the precise mathematical description that can be understood by computers in non natural languages; sharing refers to a collection of recognized concepts in the architectural field, and is aimed at sharing by groups in the field rather than exclusive use by individuals. As a strictly defined conceptual model, the architectural domain ontology is generally used to describe the following aspects [11]:

- Individual: basic or basic object;
- Class: collection or object type;
- Attributes: the characteristics and parameters that the object may have or share.
- Relationship: the way objects can be related to each other;
- Event: Change of attribute or relationship.

In the above aspects, the individual may be a door in a specific construction project or a product in a product database. IFD does not intend to hold such an individual. Therefore, except for individuals, IFD contains almost everything. In IFD, a concept is not only described by a name and defined attribute set expressed in multiple languages, but also identifies the association between a concept and another concept, as shown in figure 2 [10]. It contains many concepts and sub-concepts (represented by rectangular boxes), such as “door”, “inner door”, “out door”, “sliding door”... In addition, there are various relationships, such as “a type of”, “Consist of”, “part of”, “can be”... They indicate the relationship between two adjacent concepts in the same line, for example, “inner door” and “door” are “is a type of” relationship, “door “And “door leaf” are “consist of” relations, and there are other relations such as “is part of” and “relates to”. Comparing IFD and architectural domain ontology, it is found that the two have many similarities:

- They are all about the collection of concepts, they are all about the clear and standardized description of the concepts, and the concepts have attributes;
- They are all about the collection of relations between concepts, which are embodied in a certain relationship between specific concepts, such as IS-A, Part-of, SubClass-of and Consist-of.
Therefore, in essence, both are the ontology of some special form. But from the perspective of the BIM core technology system, IFD is a special realization based on the architecture domain ontology, and it is a certain form of standardization, as shown in figure 3. The bottom two layers are the IFD Library and the architectural domain ontology. The connection between the two indicates the source. The source may be extracted, filtered, and standardized in some form to form the IFD Library. Therefore, building IFD Library Based on architecture domain ontology is an effective construction method. It can not only make use of all kinds of existing architectural domain ontologies, but also be a standardized construction method which is not easy to cause ambiguity and avoid repeated labor.

Figure 3. The relationship between IFD Library and building domain ontology.
5.2. Construction of IFD Library-like Semantic Dictionary Based on Architectural Domain Ontology

According to the previous article, the use of IFD Library is related to the BIM model through its concepts, and the scope of use includes most entities in IFC, or independently exchangeable entities and resource entities that comply with IFC specifications. Therefore, the method of constructing the Chinese IFD Library based on the architectural domain ontology is as follows:

- Collection of existing architectural domain ontologies in the field of Architecture

At present, the application of ontological technology in the field of architecture has been increasing [13-15]. The construction of domain ontology is a relatively time-consuming project. In order to save time and speed up the efficiency of constructing Chinese IFD Library, we can use the top-level architectural domain ontology that has been built in the industry to build IFD Library on its basis. Therefore, it is necessary to gather various domain ontologies that have been built in the industry, especially domain ontologies with common characteristics.

- Construction of ontology not available in architectural field based on Ontology Technology

For the ontology that does not exist, it can be constructed by the popular “skeleton method”, “evaluation method”, “bernardas”, “metontology method” and “sensus method”, which can not only meet the construction needs of ifdlibrary, but also build other applications based on the ontology. In the domain ontology construction, it can be obtained by sorting out the existing term definitions, through the terminology standards and various specifications in construction engineering documents, such as ISO 12006-3 “organization of construction engineering information” Part 3: object oriented information framework, civil and architectural terminology standards, digital definition of building objects, architectural design standards, etc. In addition, we can also get the common related objects, attribute sets and classes through the engineering documents and forms of construction enterprises and construct the standardized domain ontology through the specific extraction technology, ontology representation language and reasoning tools.

- Extract concepts and attributes based on the ontology of each architectural domain

Each domain ontology represents a conceptual model of its domain, so at this stage, concepts and attributes should be extracted one by one for the domain ontology. The extraction method is as follows:

1. Open the ontology file (RDF/OWL format).
2. Starting from the top level of the domain ontology, through the selected tree search algorithm, read the concept $C_i$ and its attribute $P_{ij}$ layer by layer (representing the $j$-th attribute of $C_i$).
3. For the searched concept $C'_i$, search for the concept $C'_j$ that has a specific relationship $R_m$ (Kind-of, Part-of, Attribute-of, SubClass-of, Memo-of, etc.). For the searched concepts that satisfy these relationships, a dictionary description of $C'_i$ is established as shown in figure 2, and the description satisfies the rule shown in equation (1):

$$C'_i \rightarrow R_m C'_j (i \neq j, i \in N, j \in N) \quad (1)$$

4. Encode the related concept $C'_j$ of $C'_i$ that satisfies the formula (1) (using 22-bit encoding, such as “3vHRQ8oT0Hsm00051Mm008”), the encoding method adopts a random number generation algorithm combining letters and numbers.
5. Use the same method to complete the creation of dictionaries for all concepts in the selected ontology file that meet the filtering conditions in multiple cycles;
6. Read the next ontology file and construct the dictionary in the order of step 1 to step 5 until all domain ontology is completed.

For the above methods, in order to make the built-like IFD Library easier to understand and use, this article believes that a certain user participation mechanism can be added in the construction process (described in the following example part).

5.3. Application Examples

At present, there have been many studies and applications of BIM in the field of ancient building protection [16-17]. The construction of the IFC-based method in the ancient building information
model is generally carried out by combining the expansion of the entity definition and the expansion based on the attribute set. The loadable family method provided by Revit is the most flexible and effective way to construct ancient architectural families. Based on previous research work, this article illustrates how to apply the above methods and verify their effectiveness. Here is an example of the ontology (part) of the ancient architecture domain (as shown in figure 4). According to the above construction method, the mapping relationship shown in table 1 can be obtained. For the relatives-to relationship, user participation is sometimes required to supplement and improve (some ontology may lack this relationship, and users can add it through the user participation interface program).

![Figure 4. Ancient building domain ontology (part).](image-url)
Table 1. The corresponds between the specific relationship in domain ontology and the relationship in IFD Library.

| Relationship name | Effect                                                                 | Examples                                                                                                                                 |
|-------------------|------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
| Kind-of           | For similar attributes, they belong to the same class, corresponding to the “a type of” relationship in figure 6 | For example, Kind-of (palace architecture, Ming and Qing palaces), that is, Ming and Qing palaces are a type of palace architecture, and others are similar. |
| Part-of           | Describe the relationship between the whole and the parts, corresponding to the “consistent of” relationship in figure 6 | For example, part of (Ming and Qing palace, roof), that is, the roof is an integral part of Ming and Qing palaces, and other similar. |
| Attribute-of      | Describe the relationship between classes and attributes, partly corresponding to the “Relates-to” relationship in figure 6 | For example, the attribute of (palace building, the highest level), that is, the highest level building among the architectural styles, the others are similar. |
| SubClass-of       | Describes the relationship between classes and subclasses, corresponding to the “can be” relationship in figure 6 | For example, subclass of (veranda roof, single eaves veranda roof), that is, single eaves veranda roof is a subclass of veranda roof, and the others are the same. |
| Memo-of           | Describe the “notes” and further information, corresponding to the “Relates-to” relationship in figure 6 | For example, related to (single eaves veranda roof, important building), that is, the single eaves veranda roof building is an important architectural form of architecture, other similar. |

Figure 5. Similar IFD description of “Palace Architecture”. 
6. Conclusion
In summary, the advantage of constructing a Chinese IFD Library-like semantic dictionary based on the ontology of the architectural domain is very obvious. The semantic dictionary is an important semantic class library for information exchange throughout the life cycle of a BIM-based construction project. At present, it is difficult to see specific term definitions and attribute set definitions for the information requirements of the construction field in China, especially at the national level. Therefore, it is not only important but also urgent to seek a quick and reasonable method to build a Chinese IFD Library. The architectural domain ontology is a standardized, formal, and clearly stated semantic model, which includes concepts, attributes, and relationships, and has special similarities and connections with IFD Library. Therefore, it is more convenient to create an IFD Library-like semantic dictionary based on the construction of the ontology in the architectural field, avoiding the randomness and experience dependence of the semantic dictionary construction, and has the potential to use computer automation to build. This article believes that the method of constructing an IFD Library-like semantic dictionary based on the ontology of the architectural domain will be able to provide a methodological reference for how to build one's own IFD Library in China, which has important guiding significance.

Acknowledgments
This research was financially supported by Zhejiang cultural relics protection science and technology project (2017016), Project of Department of Education of Zhejiang Province (Y201840248), Xi'an University of Architecture and Technology Project (GX1704).

References
[1] Niknam M and Karshenas S 2017 A shared ontology approach to semantic representation of bim data Automation in Construction 80 22-36
[2] Li L 2012 The Preliminary Research On Building Collaborative Platform Based On Bim Technology Shanghai Jiaotong University
[3] Svetel I and Pejanovic M 2010 The role of the semantic web for knowledge management in the construction industry Slovenian Society Informatika 34(3) 331-336
[4] Terai T 2008 Development of the Construction Classification System in Japan (JCCS) Tsinghua Science & Technology 13(S1) 199-204
[5] Shayeganfar F, Mahdavi A, Suter G, et al. 2008 Implementation of an IFD library using semantic web technologies: A case study Proc., ECPPM 2008 eWork and eBusiness in Architecture, Engineering and Construction p 539-544
[6] He G P 2011 The three pillars of BIM Journal of Information Technology in Civil Engineering & Architecture 2(5112) 1547–1553
[7] Kim C, Kwon S, You S, et al. 2010 Using information delivery manual (IDM) for efficient exchange of building design information 27th International Symposium on Automation and Robotics in Construction-ISARC

[8] http://dev.ifd-library.org/index.php/Ifd:IFD_in_a_Nutshell#Introduction (2015-9-3)

[9] http://www.buildingsmart-tech.org/specifications/ifc-releases/ifc4-release (2015-9-3)

[10] IFD Library White Paper [R/OL] 2008 USA: IFD Library for BuildingSmart [2014-12-5] http://www.ifd-library.org/images/IFD_Library_White_Paper_2008-04-10_I_.pdf

[11] A A G M, B M R A and C M M 2020 Bim and semantic web-based maintenance information for existing buildings Automation in Construction 116

[12] http://dev.ifd-library.org/index.php/Ifd:IFD_in_a_Nutshell#Introduction

[13] Lee S K, Kim K R and Yu J H 2014 BIM and ontology-based approach for building cost estimation Automation in Construction 41 96-105

[14] Ma Z and Liu Z 2017 Ontology- and freeware-based platform for rapid development of BIM applications with reasoning support Lean & Computing in Construction Congress-joint Conference on Computing in Construction

[15] Jiang S H, Wang N and Jing W 2018 Combining bim and ontology to facilitate intelligent green building evaluation Journal of Computing in Civil Engineering 32(5)

[16] Sun W X 2013 Research on Parametric Information Model of Ancient Buildings of Ming and Qing based on BIM Xi’an University of Architecture and Technology

[17] Liu H 2014 Research on Building Ancient Architecture Information Model Based on Point Cloud Beijing University of Civil Engineering and Architecture