How to Create Knowledge Organization Systems in Urban Historical Buildings — Take Wuhan University as an Example

Fei Yu and Meng Zhan

***Abstract*—This paper described to create an effective knowledge organization system based on "urban historical buildings," including analytic knowledge features of urban historical buildings, knowledge organization system being used to describe the urban historical buildings, definition of the relationship groups, searing for the relation factors, etc. Such a system can be used to index and classify the research direction of urban historical buildings and served for users on the internet to important application.**

***Index Terms*—Chinese library classification (CLC), knowledge organization system (KOS), knowledge organization system of urban historical building (UHB-KOS), Urban historical building (UHB).**

I. Knowledge Features of UHB

Historical buildings (HB), which gathered historical and cultural value, scientific research value, and artistic esthetic value in the development of human civilization, are immovable cultural relics [1]. Accordingly, UHB, which were built in the process of urban evolution, can be seen as the symbols of cities and some important heritages. In a very long period of time, such buildings were protected and damaged alternately, some of them were destroyed in the war and others were disappeared in the rebuilding of the cities. The concretes of destroyed UHB, at present, were vanished, but left a few of words, pictures, and photos merely.

Because UHB can reflect the development of human society well, they refer to multiple knowledge topics and contain various disciplines. Except some features in construction that related to the buildings themselves, for the construction and evolution of the buildings, it also related to the historical process of human development and their accomplishments. It, for the buildings in early time of Wuhan University, was designed by Americans, on the one hand, integrated some Chinese ancient elements into modern materials and technology to make the architectural style looks unusual. On the other hand, it embodies very obvious historical features, having been experienced the establishment of the university in 1928, the destruction of the university in the Second World War and the development of the university in the new century. Besides that, it also has a great deal of historical backgrounds and research achievements of Wuhan University in some literatures, materials, pictures, and private collections. Such corresponding knowledge of different kinds of disciplines can be organized together to reflect a real building and the connation and extension of its knowledge completely, meanwhile, they can also express the versatility of historical buildings. Such collections of different kinds of knowledge can be called the features of knowledge in the field of UHB.

It is still, now, in the exploratory stage to reveal the features of knowledge in the field of UHB. But, for the protection of UHB, every government in the word has drew up some legal systems to protect them, put money to renovate them and used modern technology to rebuild them in virtual environments. Our government always pays much attention to the protection of historical buildings, it has made some corresponding policies and rules to protect them, and then established enormous protection projects, e.g., the No 18 historical buildings on the edge of Huangpu River in Shanghai [2], based on such polices and rules. For another example, the buildings of early time in Wuhan University, which protected and renovated by the governments of all levels in China during a long time, have been arranged into the list of national key protected building by the Chinese Sate Council. But, instead of managing the corresponding historical materials normatively, some of the protection works, for such protection projects, are only to clean and renovate the buildings themselves. In addition, it neither set up the relation description between buildings and their historical materials nor to establish the information system that can be browsed on line to show the connotations and extensions of the knowledge in the field of historical buildings that record human cultures. Therefore, it is a very significant research work to analyze the features of knowledge in the field of UHB to create the UHB-KOS.

II. Effects of Using KOS to Describe UHB

KOS, which consists of classification methods, thesaurus, semantic networks, information retrieval languages, and indexing languages, is a semantic tool that can show and organize human knowledge structures organically. Besides that, it is also a sign system that can define and express the real word [3]-[9]. KOS connects user's needs with literature resources, reveals the knowledge structures and semantic relations in a single discipline, also, it can, among different disciplines, map and integrate various knowledge structures and inner semantic relations between different disciplines by using its interoperation to create the KOS, furthermore, to offer users an operable knowledge retrieval environment.

In the field of UHB, it can integrate the original materials

Manuscript received April 10, 2015; revised May 30, 2015.

The authors are with Wuhan University/Library, Wuhan, China (e-mail: yufei@lib.whu.edu.cn, mzhan@lib.whu.edu.cn).

DOI: 10.7763/IJIMT.2015.V6.605

of UHB and some related materials (e.g., historical background materials, architectural and technical materials, protection projects materials, and humans' achievements) into one collection, use the way of construction of KOS to design a knowledge relation model that related to UHB, apply the way of semantic interoperation to build a UHB-KOS.

Such a UHB-KOS is an interdisciplinary system, which refers to the knowledge structures and inner relations in Library Science, Architecture and History. It can connect the knowledge concepts with the semantic relations in such different kinds of disciplines by the theme of historical building. Such a KOS, can provide some complete knowledge frameworks and inner relations for the projects that related to the protection of historical buildings in our country. Meanwhile, it can also provide the metadata description system, technical support and a retrieval system based on knowledge operation for the resource system of historical building, network retrieval system and the users respectively.

III. Study on the Creatment of UHB-KOS

The KOS, which involves all ways of information organization and knowledge management, combines user's needs with literature resources. The structure of KOS, at present, can be divided into three types:

First type is Term Lists, (e.g., authority files, authority dictionaries, glossaries, and gazetteers).

Second type is the Classification and Categorization, (e.g., headings, classification schemes, category list, and list of classification system).

Third one is the Relationship Groups, (e.g., thesaurus, semantic networks, ontology, and concept maps) [10]- [12].

The above structure is a general way to express the structure of KOS. But the UHB-KOS, which structure must as the same as the one in a general KOS, is an interdisciplinary KOS and should to be designed according to some corresponding features of knowledge in the field of UHB.

A. Ways of Organization of the Term Lists in UHB-KOS

The so called term lists, is the minimal collection of knowledge units or knowledge concept words in the structure of KOS. Only if such words' meaning must as the same as the real word they stand for, can, in some sense, make resource organizers and users trust the system. In a single discipline, it can refer to the existing authority files and professional dictionaries to choose and make up the corresponding term lists easily. Conversely, it may be more complicated to design the term lists in different kinds of disciplines in UHB-KOS than the single one because of the interdisciplinary research.

In our research, on the one hand, knowledge features of the UHB can be divided into five topic areas, namely, geographical topic area (e.g., evaluation urban geographical terms), historical topic area (e.g., background materials, historical persons, and some corresponding information), architectural topic area (e.g., design, structures, materials, styles, and environments), heritage protection topic area (e.g., evaluate of the historical heritages and some corresponding protective policies), and cultural topic area (e.g., documents, pictures, achievements, and specialists) by faceted classification in UHB. On the other hand, to extract some corresponding words from the existing authority files, professional dictionaries, professional thesaurus, and classification tables to make up the typical words in such five term lists respectively.

Because the knowledge of UHB is a very small branch in every topic area, it cannot establish the term lists based on the existing authority files directly, but to extract some knowledge concept words related to UHB accurately from different kinds of professional dictionaries to mapping from the words in such five term lists to the topic area of UHB accurately, moreover, to make the term lists in UHB-KOS as the subclasses of the authority files in different kinds of fields. In order to make term lists well, we design the entity resources and some corresponding documents as data resources of extraction, and then apply the way of two step's operation of post-control thesaurus, just as the following description:

First, make the buildings in early time of Wuhan University as a research object, to investigate, collect, and manage some entity data that related to such five topic areas in this object widely, (e.g., historical background materials, architectural design drawings, drawings of repairing, materials of architecture styles and features, geographical position of the historical buildings, some corresponding laws and documents of historical heritage protection, and the papers, pictures, and persons that related to such historical buildings).

Second, to extract some corresponding notional words from the existing authoritative files and professional dictionaries in different kinds of disciplines to form five self-built term lists that related to the buildings of early time in Wuhan University based on the core topical terms in such materials. Obviously, such a way of the formation of knowledge features words, which originate from a single real historical building, is a post-control retrieval method to make term lists. As time goes by, it will find some materials that related to the UHB constantly, so the function of post-control retrieval in such term lists calls for establishment of some maintenance systems with expansibility of term lists in UHB-KOS.

Now, take the establishment of the term lists in geographical topic area as an example:

First, collects some cultural and historical materials that related to the buildings in early time of Wuhan University, (e.g., geographical name information of the location of the existing historical buildings, historical books, articles, and pictures that describe the buildings of early time in Wuhan University).

Second, extracts some geographical nouns from the existing materials, (e.g., Hubei, Wuhan, Wuchang, Luojia Hill (罗家山), Luojia Hill（珞珈山）, Meiyuan, and Yingyuan). After that, it can retrieve the geographical nouns mention above and their explanatory notes in the published geographical dictionaries. Some normative geographical nouns can be got in geographical dictionaries, but some others, e.g., "Meiyuan" and "Yingyuan," cannot be found. These unemployed nouns, however, having some significant

relations corresponding historical persons and events, must be added in to the term lists of geographical nouns in UHB-KOS one-by-one by means of post-control retrieval according to the canonical formats in geographical dictionaries. Of course, such added nouns and their explanatory notes must come from the factual account in some corresponding historical materials.

From the above, it can not only make every notional word of the five term lists in the field of UHB, connect with the popular semantic environments, but also, in some sense, retain the regularity and natural language environments, by ascertaining the ways of the establishment of the term lists in core knowledge concept words based on extract notional words and their means from corresponding documents.

The ways of the establishment of the other four term lists in UHB-KOS, are similar with the ones in geographical nouns mentioned above.

B. Classification Ways of UHB-KOS

The term lists, which established by the ways mentioned above, do not connect the words with each other. The classification or categorization in UHB-KOS is to establish some semantic relations between these words. UHB-KOS can improve the establishment of word relations between different term lists by chose the existing dictionaries and classification.

There are five term lists in UHB-KOS. Geographical term lists, among them, establish the classification relations of the words based on geographical dictionaries directly; others, accordingly, establish the classification relations of the words based on CLC.

Words relations in geographical dictionaries, present as a tree-based form and with a single structure. When establish them, it can use the words in the term lists in UHB-KOS to compare with the ones in dictionaries. For the words contained in dictionaries, introduce their relations directly. But for the ones cannot be found in dictionaries, add them into correct positions in dictionaries according to their logical relations. The words relations, established by such a way, maintain some linear relations of the original dictionaries and become a subclass of them, moreover, with extensibility.

It is more complicated to manage the relations of words classification in the term lists of buildings, history, heritage protection, and cultural materials established by CLC. Because, CLC is not a table with a tree-based structure, but a knowledge classification scheme with scalability, consisting of category names, explanatory notes, tables of imitative classification, and tables of compound classification, it should make some rules to analyze the knowledge points in CLC to let them have a tree-based structure before classify UHB-KOS by the means of CLC. The specific methods are as the follows:

For compound categories, conceptualize them singly, e.g., a compound category “the Revolution of 1911 and the Provinces Uprising” can be got two independent words “the Revolution of 1911” and “the Provinces Uprising” after being analyzed singly. Besides that, these two words have some parallel relations in the classification relations, namely, sharing the same category number.

For neutral words in categories, pack them by concept limited according to upper and subordinate knowledge features; for the abbreviations, manage them normatively according to knowledge features, e.g., after managing the abbreviation “principle” to “design principle,” it can get two neutral words “design theory” and “design principle,” furthermore, it can also get two independent notional words “theory of architectural design” and “principle of architectural design” by limited its upper category.

Due to the limited space of this paper, the following analytical ways will not be illustrated by examples.

For explanatory notes, to extract categories with independent knowledge concepts to form new category names; for descriptive explanatory notes, to form several category words with the capacity of independent recognition through the analytical ways of independent notional words extraction and concept packaging according to the semantic relations between explanatory and category names. After that, such categories inherit the number of the original ones.

For imitative classification, manage them with imitative combination directly according to corresponding regulations in order to make the recessive imitative categories to be explicit and get some new names and numbers, moreover, extend the categories.

For compound classification, it consists of special subdivision and general compound classification. The general compound classification table can be seen as a special imitative list which can be used in all of categories and has the same regulations as the imitative categories. Such results can also extend the categories.

Such analytical methods, discussed above, use the scalability of CLC to turn implicit categories into explicit ones, make many discipline knowledge points to become more detail and become independent notional words.

In UHB-KOS, it doesn’t analyze the whole classifications, but to analyze the classifications that related to UHB-KOS, and then, creates some classifications relations of historical, heritages protection, and cultural materials in the UHB-KOS according to their structures.

Through analyzing the linear relations of the words in CLC, it can draw the conclusion that each word has a corresponding classification number, which can be considered as an element to be used in the interoperation between two heterogeneous systems. The words classification and the broadly classified in the term lists of buildings, history, heritages protection, and cultural materials in UHB-KOS, use the analyzed CLC. Besides that, relations of architectural words, which established based on the dictionaries of building, can be used in CLC by the way of semantic correlation to introduce the numbers of CLC into their term lists. As a result, every word in four of five term lists in UHB-KOS has its corresponding classification number in CLC. Of course, not all of the words in UHB-KOS are exists in CLC, but for the words that are not contained in CLC, it can assemble them the corresponding numbers according to words semantic relations and the extension principles of classification.

C. The Definition of Relationship Groups

According to the ways above, it has established a great number of words and their relations in UHB-KOS. But, in

order to improve it, it must extend or related to the words and their relations to the objective things they stand for, namely, to establish some relationship groups to connect the notional words with the real objects. The main way to achieve it is to find the relation factors between the notional words and the real objects, and then, connects them together based on the relation factors according to some specific methods.

Searching for relation factors: real objects in UHB-KOS can be divided into architectural entities, architectural drawings, historical materials, network resources, and literature resources etc; it must search for the relation factors between real objects and the notional words in UHB-KOS before establish some relationship groups between them.

Every term list in UHB-KOS can have its normative notional words of the name. Except the geological term lists, others also have the explanatory notes and the classification numbers. Besides that, it can choose the notional words of the names and the classification numbers as a relation factor between the term lists and real objects because of their integrity and regularity. Therefore, features of the relation factors in UHB-KOS can be defined as the name, as well as the number, with normative and standardized description. Meanwhile, if every real object can have its notional words of the name and the number of CLC, it can solve the problem of how to choose relation factors for real objects and the factors may as the same as the ones in the five term lists in UHB-KOS. To achieve this, it must catalogue every real object.

Before cataloguing, some parts of real objects must be digitalized, it including real photos of buildings, architectural drawings, historical relics, historical materials, etc. The digital standards and the regulations of data indexing in UHB-KOS, certainly, abide by the ones in digital libraries.

Hence, it not only exist some heterogeneous system of the self-built resources, but also have some other resources on the internet that related to the topic area of UHB-KOS. In order to establish a complete knowledge relationship group, it must index and connect to all kinds of systems uniformity, which core point is every resource structure in each system must has a normative notional word and a corresponding number in CLC, namely, has a relation fact that can connect with other systems.

Research the ways of relationship: to research the ways of relationship in KOS is to research the interoperation between different kinds of knowledge system. At present, people have put forward some kinds of technology and methods that can realize interoperate, (e.g., derivative indexing, translation, serialization classification, mapping directly, the co-occurrence mapping, mediation dictionary, macro-vocabulary, etc) [13], [14]. It is a feasible way, according to the description of different kinds of resources and features of relation factors in UHB-KOS, to establish some relations between several kinds of resources based on using the method of direct mapping of the relation factors. Such a method including the mapping based on literal correlation and the one based semantic correlation.

When refers to the mapping based on literal correlation, it means, in the literal sense, the same correlation factor for both sides to establish the same meaning.

As to the mapping based on semantic correlation, it means, in the literal sense, the correlation factor has different form for both sides, but has the same knowledge to establish the same relation.

Such two mapping method can be reflected in the following table, as the examples of discipline classification (DC) and the CLC.

TABLE I: MAPPING OF DC AND CLC

DC	CLC	Mapping Relations
560.10 Architectural History	TU-09 Architectural History	Based on literal correlation
0831 Architecture	TU Architecture Science	Based on semantic correlation

In the practice, it can, for the mapping based on literal correlation, managed automatically by programming. But for the mapping based on semantic correlation, only by the way of human intervention, can maintain the accuracy of the established relations.

Through the ways above, it can create an interdisciplinary UHB-KOS, which can break through the limitation of original disciplines, merge several terms from different disciplines, integrate different types of historical materials, connecting several concepts and mapping to real objects respectively.

REFERENCES

[1] S. Zhang, *Introduction to Historical City Protection: An Integral Method Used in Cultural Heritage and Historical Environment Protection*, Shanghai: Shanghai Science and Technology Press, 2001, pp. 1-38.

[2] Y. Huang, "Analyze historical heritage protection and its application," *Shanxi Architecture*, vol. 34, no. 9, p. 60, March 2008.

[3] C. Xue, *Study in Create a Knowledge System and Its Application in Network Environment*, Nanjing: Dongnan University Press, 2009, pp.42-60.

[4] L. Zeng, "Knowledge organization based on the internet environments," *Modern Library and Information Technology*, no. 1, p. 2, Jan 2004.

[5] C. He and S. Li, "Agricultural ontology — The establishment of agricultural knowledge organization system," *Agricultural Library and Information Science*, vol. 16, no. 10, p. 41, Oct. 2004.

[6] J. Wang and L. Zhang, "Research status and its long-term trend of network knowledge organization system," *Journal of Chinese Library*, vol. 34, no. 173, pp. 65-69, Jan 2008.

[7] L. Sun, "Research of knowledge system and its long-term trend under network environment," *Library Theory and Practice*, no. 3, pp. 28-30, March 2010.

[8] J. Zhang and Q. Ying, "Study in the simple knowledge organization system (SKOS) and its application," *The Modern Intelligence*, no. 11, pp. 20-21, Nov 2008.

[9] L. Si, "Interoperability and its implementation among knowledge organization systems," *Modern Library and Information Technology*, no. 3, pp. 29-33, March 2007.

[10] L. Si and L. Zhou, "Research status and their long-term trend of English knowledge system at abroad in recent years," *Library Forum*, vol. 30, no. 158, pp. 220-226, p. 208, Oct 2010.

[11] X. Zeng, "Comparative study of ontology formalization of Chinese thesaurus and SKOS," *Journal of Intelligence*, vol. 36, no. 186, pp. 99-106, March 2010.

[12] D. He, X. Qiao, and L. Zhu, "Study in the trends of the word system both at home and abroad," *Digital Library Forum*, no. 6, pp. 44-48, June 2008.

[13] L. Si, "The interoperability and its realization of knowledge organization system," *Modern Library and Information Technology*, no. 3, pp. 29, March 2007.

[14] L. Si, L. Xu, and H. Chen, "The application of knowledge organization systems in Chinese digital libraries and their interfaces research," *Information Science*, vol. 25, no. 3, March 2007.

Fei Yu was born in Wuhan, Hubei, P.R. China in 1983. He earned the bachelor degree in computer science in Huazhong University of Science and Technology, Wuhan, Hubei, P.R.China in 2006; then, get the master degree in the history of science and technology in Wuhan University, Wuhan, Hubei, P.R.China in 2009; and now are majoring in doctor degree in architecture science in Wuhan University, Wuhan, Hubei, P.R.China.

From 2009 to now, as a librarian, he is working in Wuhan University and has published 5 papers, e.g., Fei Yu, "Knowledge management and application in reference working in university libraries-take Wuhan University Library as an example," in the academic annual meeting of Hubei Province Library Association, Enshi, Hubei, P.R.China, 2014, pp. 322-326; Fei Yu, "Research of subject knowledge revealing and specialized documents searching," in International Conference on E-Management and E-Education, Beijing, 2014, to be published (will be indexed by EI). His previous research interests are computer science and water history, but now he interested in information science, information management, and reference working.

Mr. Yu now is a membership of Chinese Library Association, as well as Hubei Province Library Association. He has also earned some awards for his published paper in the field of library, e.g., the first prize for his papers that has published in the proceedings of Hubei Province Library Association both in 2011 and 2014.

Meng Zhan was born in Changning, Hunan, P.R.China in 1954. She earned the bachelor degree in library science in Wuhan University, Wuhan, Hubei, P.R.China in 1997.

From 1985 to 2000, she was working in the library of Wuhan University of Hydraulic and Electrical Engineering, as a deputy director; From 2000-2009, worked in Wuhan University Library, also as a deputy director; From 2009-now, she is working in the library of Huazhong University of Science and Technology Wuchang Branch, as the director. Besides that, during Nov 2012 to April 2013, as a scholar, she visited to the Ease Asian Library in the University of Pittsburg in US. During these years, according to her working experience and practice, she has published several excellent papers, e.g., Meng Zhan, "The application of "the multiple thesaurus of classification" in retrieval system of Chinese bibliography database," Modern Library and Information Technology, No. 5, pp. 50-52, May 2005. Her previous and current research interests are digital library, information management, and Chinese Library Classification.

Mrs Zhan is a member of the communication and cooperation committee of Chinese Library Association; a deputy director of the Library Work Committee of Hubei. She has also got many awards for his working, e.g., the first prize in China for his working of the construction of CALIS characteristic database. Except that, her published papers are indexed by SCI, ISTP, and EI many times.