



The Electronic Library

Mapping multilingual lexical semantics for knowledge organization systems

Shu-jiun Chen Hsueh-hua Chen

Article information:

To cite this document:

Shu-jiun Chen Hsueh-hua Chen, (2012), "Mapping multilingual lexical semantics for knowledge organization systems", The Electronic Library, Vol. 30 Iss 2 pp. 278 - 294

Permanent link to this document:

<http://dx.doi.org/10.1108/02640471211221386>

Downloaded on: 27 December 2014, At: 06:15 (PT)

References: this document contains references to 17 other documents.

To copy this document: permissions@emeraldinsight.com

The fulltext of this document has been downloaded 581 times since 2012*

Access to this document was granted through an Emerald subscription provided by 312137 []

For Authors

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com

Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

*Related content and download information correct at time of download.



EL
30,2

Mapping multilingual lexical semantics for knowledge organization systems

Shu-jiun Chen

*Research Center for Information Technology Innovation,
Academia Sinica, Taiwan, and*

Hsueh-hua Chen

*Department of Library and Information Science,
National Taiwan University, Taiwan*

278

Received December 2010
Revised March 2011
Accepted April 2011

Abstract

Purpose – The study aims to investigate, through mapping analysis, the operation of knowledge organization systems (KOS) in different languages (English and Chinese), the types of term equivalence and the degree of similarity between different conceptual structures, and issues related thereto.

Design/methodology/approach – Terms are selected from the Art & Architecture Thesaurus developed by Getty Research Institute in the USA (source language) and the National Palace Museum in Taiwan participating in Taiwan e-Learning and Digital Archives Program (target language). In respect of data analysis, mapping analysis and content analysis are both adopted.

Findings – Among the six types of term equivalence, “exact equivalence” appears most frequently. The degree of similarity between different conceptual structures can be divided into four types: similar structure, transferring part of the structure of source language into the system of target language by modeling; similar structure, but the structure of source language needs to be expanded or revised; dissimilar structure, the term in target language can be partially mapped to the term in source language; lack of structure, no matches between target language and source language.

Practical implications – The next step is to develop a more comprehensive conceptual structure that can incorporate multicultural perspectives. If the “partial equivalence” terms are further identified as important concepts, the problems of non-exact equivalence terms and dissimilar conceptual structures can be solved by addressing the needs of English users in searching for Chinese art collections.

Originality/value – The study serves as a pilot study applicable to the development of multilingual KOS in the domain of Chinese fine arts, building the first demonstration model for the interoperability of KOS in Chinese and English.

Keywords Arts, Classification, Digital libraries, Indexing, Languages

Paper type Research paper

1. Introduction

Knowledge organization systems (KOS) is a general term referring to various semantic tools that present human knowledge structure in an organized manner, including dictionaries, classification schemes, thesauri, semantic networks, ontology, and other information retrieval languages and indexing languages. In the environment of digital



libraries, KOS can provide users with the search results, further broadening or narrowing down the scope to encompass the gathering of identical concepts and the links to relevant concepts. At present, the contents of most digital library systems are based on a main language or different KOS, and when digital library services become available on the internet, the access needs of different languages and users are also brought into attention. To satisfy the user's need to conduct a cross-digital library search in their local languages and in the same ways depends on the degree of interoperability between these different KOS. Vocabulary mapping is one of the important approaches to solving the problem of interoperability between various KOS. The understanding of the vocabulary mapping types and structure similarities between different KOS can assist in assuring the quality of vocabulary mapping between different systems, and thus improve their interoperability. For instance, the research on vocabulary mapping types allows the digital libraries to rank the results of various equivalence relationships between vocabularies and present the user-preferred vocabularies, to provide users with the details of different precision criteria in relation to the search results, and to help identify the regularity of mapping between specific vocabularies and systems in order to facilitate the development of large-scale, automated vocabulary mapping. The research on structure similarities between different KOS can explore the process of interoperability and the types of issues related to conceptual structure, and further establish feasible principles, guidelines and solutions. In particular, mapping is the key to the multilingual interoperability of KOS, whose significance lies mainly in dealing with the identification and connection of identical concepts between different languages, and thus facilitates cross-lingual retrieval. Focusing on the domain of Chinese fine arts, the study has two main objectives: First is to gain an insight into the types of term equivalence between KOS of different languages (English and Chinese) through mapping. Second is to comprehend the types of matches between different conceptual structures of KOS in the West and East, and the issues related to the operative process of mapping.

2. Related work

Up to the present, relevant research has been mostly focused on the interoperability between different KOS in the same languages, domain-specific to general KOS in different languages, or KOS in the domain of science. Among them, research on the interoperability of different KOS in the same language is the majority. For example, the mapping between Dewey Decimal Classification (DDC) and Library of Congress Subject Headings (LCSH) conducted by Online Computer Library Center (OCLC), and the interoperability of geographic names dictionaries and thesauri service protocols developed by the Alexandria Digital Library (ADL) headquartered at the University of California at Santa Barbara, are both based on English (Zeng and Chan, 2004). This kind of interoperability research on KOS in the same language is more likely to match a great number of terms in an automated or semi-automated way.

Even in the case of KOS in different languages, at present the research is mostly focused on the interoperability of KOS between general domains or between general domains and specific domains. For example, the collaboration between WordNet and EuroWordNet developed by Princeton University and Sinica BOW developed by Academia Sinica; the project of Multilingual Access to Subjects (MACS), in which Swiss National Library links between the German Subject Heading Authority File

(SWD)/Rules for the Subject Catalogue (RSWK), the French RAMEAU (Unified Encyclopedic and Alphabetical Subject Authority List) and the English LCSH, in order to provide users multilingual access to the bibliographic databases of libraries in different countries; and the Academic Subject Gateway Service Europe (Renardus) coordinated by National Library of The Netherlands, including members from Denmark, Finland, Germany, The Netherlands, Sweden and the UK, which aims to build the cross-browsing function based on DDC and integrate scattered and heterogeneous subject gateways in Europe (Vizine-Goetz *et al.*, 2004; Zeng and Chan, 2004; Vossen, 1998; Huang *et al.*, 2004).

The research on the interoperability of KOS in different languages under the specific domains is mostly related to mathematics or life science. For example, the Unified Medical Language System (UMLS) Metathesaurus, developed by National Library of Medicine, integrates thesauri, subject headings, classification systems, coding systems and controlled vocabularies in various languages from the domain of medicine, health, biology and other related disciplines; Food and Agriculture Organization (FAO) of the United Nations cooperates with the Scientific Documents Information Center, Chinese Academy of Agricultural Sciences to develop the multilingual agriculture thesaurus, which is to map between terms in AGROVOC and bilingual Chinese agricultural thesaurus; CAMed, developed by Columbia University and Kent State University, is an integrated English and French thesauri management and cross-thesauri retrieval system in the domain of complementary and alternative medicine (Zeng and Chan, 2004; Liang and Sini, 2006).

On the contrary, there is less research on the interoperability of KOS in different languages under the domain of humanities, art, or cultural assets. Up to the present time, most of the relevant research has been concentrated on the interoperability between English and European languages. For example, the interlingua developed by the European Heritage Network (HEREIN); the mapping between French terms from “Le thesaurus de l’architecture” developed by Merimee and English terms from AAT developed by the Getty Research Institute (GRI) and from English Heritage Thesaurus published by the National Monuments Record; the collaborative project of Spanish AAT between GRI and Centro de Documentación de Bienes Patrimoniales (Zeng and Chan, 2004; Baca, 2009).

Related research points out that the more two KOS differ in language and culture, the greater heterogeneity they will have in the conceptual structures (Liang and Sini, 2006). Besides, compared with the multilingual KOS focused on scientific domains such as medicine or agriculture, the terms and conceptual structures of the art and culture domain seem to be more closely connected with local cultures. However, as mentioned previously, current research is mostly focused on the interoperability between different KOS in the same languages, domain-specific and domain-general KOS in different languages, or KOS in the domain of science. In other words, in the research on the interoperability of KOS, a thorough and systematic examination on the semantic interoperability of multilingual terms with high heterogeneity, especially Chinese and English terms in the domain of art and humanities, is still lacking, and this is what the study will endeavor to fully explore. Besides, according to the previous research, the mapping between terms from KOS in different languages is faced with many similar problems, for example: when doing the Dutch AAT, the Netherlands Institute for Art History (RKD) encounters a lot of issues that are similar to those of

Chinese AAT, such as the equivalence mapping problems of multilingual terms due to different cultural factors, and the one-to-many relationship between target languages and source languages (Zelfde and Schaik, 2009; Chen, 2009).

3. Methodology

By means of content analysis, the study conducts mapping analysis on the terms from different KOS in different languages in the domain of Chinese art. More details with regard to the research subject and data analysis will be explained in the following.

3.1 *Research subject*

The study takes the terms from the Art & Architecture Thesaurus (hereafter AAT) developed by The Getty Research Institute in US, and the controlled vocabularies from the National Palace Museum (hereafter NPM) in Taiwan participating in Digital Archives Sub-Project of Antiquities in the National Palace Museum under Taiwan e-Learning and Digital Archives Program (hereafter TELDAP), as the principal research subjects. The study has two major data types and collection methods, including mapping analysis logs from the participants (including the researcher), and mapping analysis workshops of the researcher and participants, which will be further explained in the following.

3.1.1 Sample source. AAT terms are chiefly used for describing collection items related to art, architecture, decorative art, material culture and archives documents. The hierarchical thesaurus structure mainly consists of seven facets, including 34,000 concepts and descriptors (Harpring, 2010). As of January 25, 2009, Digital Archives Sub-Project of Antiquities in the National Palace Museum has digitized and uploaded to the TELDAP Union Catalog more than 19,000 collection images and metadata for public online access, establishing 1,500 controlled vocabularies in total for the digital collections.

Due to the expectation to explore as much of all the facets in AAT, the types of equivalence between AAT and TELDAP terms, and the types of homogeneity or heterogeneity between conceptual structures as possible, the study first fully examines the controlled vocabularies in the Metadata Specification of National Palace Museum, categorizing them according to the classification schemes of AAT facets, and finally selects representative heterogeneous samples. Specifically speaking, this study starts from selecting 60 Chinese controlled vocabularies in the Metadata Specification of Antiquities at National Palace Museum (version 1.2), and maps them with the English terms in AAT. These 60 controlled vocabularies (as shown in Figure 1) have included preliminarily six out of seven facets in AAT, which are Physical Attributes, Styles and Periods, Agents, Activities, Materials, and Objects. Owing to the features and limits of controlled vocabularies for NPM antiquities, the Associated Concepts Facet in AAT is not included in the study.

3.1.2 Data types and collecting methods. Among the two main data sources, mapping analysis logs of the participants (including the researchers) are the main data collecting method employed to understand the status of equivalence between TELDAP terms and AAT terms. In order to strengthen the quality and accuracy of mapping results, not only the researchers conducted mapping analysis on each term, but also TELDAP members were chosen to participate in the process. Through the mapping analysis workshops, the types of term equivalence, the degree of matches between conceptual

TELDAP terms (the contexts of terms are shown within the parentheses)	Initial mapping to AAT facets
蝴蝶形 (< 昆蟲式 < 動物 < 全器形制) Butterfly shape (< insects < animal < main shape)	PHYSICAL ATTRIBUTES
仰韶文化 (< 考古學文化) Yangshao Culture (< archaeological period)	STYLES AND PERIODS
后妃 (< 宮廷人物 < 人物 < 紋飾) empress (< monarch < people < decoration)	AGENTS
繡花 (< 技法 < 紋飾) embroidery (< technique < decoration)	ACTIVITIES
青銅 (< 金屬 < 礦物 < 質材) bronze (< metal < mineral < material)	MATERIALS
隸書 (< 書體 < 形式 < 款識) Clerical script/li shu (< script < style < inscription)	OBJECTS

Figure 1.
Samples of controlled vocabularies within Digital Archives Sub-Project of Antiquities in TELDAP

structures and other issues were further discussed by the researchers and participants. The procedures and characteristics of data collection are explained in the following.

The researchers recruited the participants from TELDAP, a total of 12 people including the researchers themselves. Because the mapping contents and the skills required are mainly focused on the domain of art, history and museum studies, the recruitment is based on the participants' academic backgrounds and work contents at TELDAP in relation to these disciplines. Besides, in order to reinforce the effectiveness of mapping, the domain experts from GRI and NPM are asked to further examine the terms with discrepant mapping results via e-mail communication, so that the validity of the research findings can be assured. To address the difference of opinions regarding the mapping results among the researchers, participants and domain experts, the study also establishes the principles of mapping as the guidelines for future large-scale equivalence mapping.

To begin with, the researchers held workshops to train all the participants before mapping, and the training includes: AAT hierarchies, TELDAP controlled vocabularies, TELDAP Union Catalog, and methods and examples of TELDAP-AAT term equivalence mapping. Furthermore, the mapping analysis is conducted three times in a gradual fashion, each time lasting five work days. At the first time only five terms were mapped, one of which was jointly mapped by all participants, but in the meanwhile the participants were asked not to discuss with each other, just recording their own mapping results, processes and comments. Another four terms were mapped respectively by five to seven participants concurrently. The

researchers also mapped each term by themselves. The previous method is adopted to add to the mapping results of a term more analyzers' perspectives and verification. The participants were required to write a mapping analysis log for each term, the content of which includes: duration of mapping, number of entries found in Union Catalog and their categories, the fields in Union Catalog where the term appears, English translation of the term and references, the term's corresponding facet, hierarchy and equivalent in AAT, and the analysis process.

After the first mapping analysis, the researchers and participants reached a consensus about equivalence mapping, and through observing and examining the mapping results, they modified their strategies for the second and third mapping analysis. In order for the participants to observe a more comprehensive context during mapping analysis, the participants were assigned terms under a certain context of NPM antiquities in TELDAP. Besides, in the mapping analysis log, the participants had to record the characteristics of the mapped AAT term that are distinguished from other terms at the same hierarchy, which is supposed to make the participants more concentrated on the mapping analysis of the specific context. The researchers also introduce to the participants the five types of term equivalence from ISO 5964-1985, so that they can follow a set of common standards when mapping Chinese terms (TELDAP) with English terms (AAT).

3.2 Data analysis

This part of the study includes identifying the types of term equivalence between different languages, recognizing the similarity of conceptual structures between different KOS, and discovering different concept types for the term.

In respect of data analysis, mapping analysis is adopted in the study. Mapping analysis, in the context of thesauri, refers to the process of recognizing the equivalence of terms, concepts and hierarchical relationships (Doerr, 2001). The method is one of the core approaches to establishing interoperability between multiple KOS, particularly in the integration of thesauri. The study can be divided into three parts in terms of range; the first part aims to identify the types of term equivalence between TELDAP and AAT, which includes verifying the English translation of the TELDAP term, and finding its possible equivalent in AAT. The second part aims to determine the type of term equivalence, observing and determining the type of conceptual structure similarity between the two terms in TELDAP and AAT. The third part aims to identify every controlled vocabulary in TELDAP, finding out what other concepts in the Union Catalog, in addition to those originally chosen as samples, the term can represent, and whether these concepts can also be mapped to other terms in AAT. Based on the principle that one term represents one concept, the main task of this part is to observe from different perspectives whether the term can represent different concepts simultaneously in the Union Catalog, and to induce different concept types of the term, finally returning to the first part to determine the types of equivalence between these concepts and AAT terms.

3.2.1 Identify the types of term equivalence between different languages. The study's method of identifying the types of equivalence is mainly based on mapping analysis, and through the researcher's testing and the participants' experiences and feedback, it is modified repeatedly and gradually grows into shape. The data analysis details of the main steps in the procedure will be explained in the following.

(1) *Verify the English translation of the TELDAP term.* Authoritative references are the best source and the foremost choice for the English translations of TELDAP terms. For one thing, the research subject is art-related terminology, for another, the controlled vocabularies of the TELDAP research sample “NPM antiquities” are mostly specialized terms. Therefore, the list of references includes general references (e.g. *Encyclopedia Britannica/Britannica Online Traditional Chinese Edition, Webster’s Online Dictionary with Multilingual Thesaurus Translation*) and art-related specialized dictionaries (e.g. *Dictionary of Art Terms and Techniques, Chinese-English/English-Chinese Dictionary of Culture & Archaeology*). Under certain circumstances, the participants should consult different references persistently to identify possible English translations of the terms and bibliographical information (see Table I).

(2) *Find the possible equivalent term in AAT.* Mapping analysis can achieve the goal of interoperability by linking equivalent terms from different glossaries (McCulloch *et al.*, 2005). The study adopts “concept-based” rather than “term-based” mapping analysis. In other words, when doing equivalence mapping for each term, what should be considered above all is the concept represented. As for “concept” and “concept-based” mapping analysis, they are described as follows. The meaning of “concept” is to use terms to represent abstract ideas and physical attributes (e.g. shapes, patterns, colors, styles or periods, activities, performers, materials, objects, and visual and verbal communication forms discussed in art, literature, architecture and material culture). In this case study, a “concept” represents the terminology related to the “general concepts” information that is necessary for cataloguing the domain of art. But proper names such as people, organizations, geographic places, named subjects or events are not included in AAT. The so-called “concept-based” mapping analysis has two steps in the study: one is to determine the context of the represented concept according to the source of the TELDAP term; the other is to find the corresponding facet, hierarchy and equivalent term in the conceptual structure of AAT (see Table II).

Table I.
The process of verification for TELDAP term “peacock blue”

TELDAP term	English translation	References	TELDAP term source
孔雀藍	Peacock blue	Ralph Mayer, <i>Dictionary of Art Terms and Techniques</i> (2002) [Chinese version]	Digital Archives Sub-Project of Antiquities in the National Palace Museum – Metadata Specification version 1.2), p. 20

Table II.
The process of mapping for TELDAP term “embroidery”

TELDAP term	Context representing concept	Equivalent concept in AAT
繡花	<i>technique</i> > embroidery	Term: embroidering Hierarchy: Processes and Techniques Facet: Activities Facet Note: embroidering (needleworking (process), < needleworking and needleworking techniques > . . . Processes and Techniques)

(3) *Determine the type of term equivalence.* When linking terms between different KOS by means of mapping analysis, the issue we are faced with is identifying the type of term equivalence. For example, according to Chaplan and others, there are 19 types of term equivalence between Laborline Thesaurus and Library of Congress Subject Headings. Based on the findings of Chaplan and others (2008) further conducted mapping analysis between several KOS, such as AAT, Library of Congress Subject Headings, MeSH, UNESCO, and the Dewey Decimal Classification, and they found that the types of term equivalence can be reduced down to 9. The previous equivalence mapping is conducted between different systems in the same language, while the study, in comparison, is dealing with the equivalence mapping between different systems in different languages (see Table III).

3.2.2 *Recognize the similarity of conceptual structures between different KOS.* The issues of conceptual structures involve the hierarchical and associative relationships between different terms, the core aspect of which is whether the conceptual structure of each language has to be the same, or the existence of dissimilarity can be allowed (IFLA, 2005). The study tries to first judge whether two different KOS have similar conceptual structures. If they differ in structure, what follows is to discover their different characteristics. The way of analyzing the homogeneity and heterogeneity of TELDAP and AAT conceptual structures is to observe and compare the two systems in accordance with the context represented by the term itself through the aforementioned process of term equivalence mapping. For example, the context of “gold” in the study sample is “Materials” (gold < metal < mineral < materials) rather than the relevant context produced by the “color” concept. Furthermore, related terms such as “silver,” “copper” and “bronze” in the same TELDAP context are also observed.

3.2.3 *Discover different concept types for the term.* In the design and development process of controlled vocabularies, different terms representing the same concepts will be connected by synonym sets, that is, term groupings based on their senses in terms of linguistics or the so-called “synset.” For the same terms representing different concepts, it is necessary to make clarification and interpretation, which is to make sense distinction in terms of linguistics (IFLA, 2005; Huang *et al.*, 2008). During the research process, it is found that many terms in the TELDAP collections can represent different concepts, so we have to distinguish between different concept types for these terms and make clear markings. According to Huang *et al.* (2008), this situation

TELDAP term	Equivalent concept in AAT	Type of equivalence	Note
繡花	embroidering	Exact equivalence	

Table III.
The type of equivalence
for TELDAP term
“embroidery”

TELDAP term	TELDAP – Union Catalog
Clam shell (material) 蚌殼 (質材)	Describe the material of an object
Other concepts	Describe the shape of an object
	Describe the object itself (1) the content of the pattern
	Describe the object itself (2) and the entity

Table IV.
Different concept types of
the same term

corresponds to one of the five basic principles of sense determination – “one sense one entry.” If one word form has more than one sense, every pair of “word form – sense” is regarded as an individual entry. Through distinguishing between different concept types of the same terms, the study also increases more and more new concept terms in the mapping analysis (see Table IV).

4. Results

4.1 *Types of term equivalence: frequency*

“Exact equivalence” (35, 58 per cent) is the type that appears most frequently in the study, followed by “partial equivalence (species-genus relationship)” (18, 30 per cent), “non-equivalence” (6, 10 per cent) and “inexact equivalence” (1, 2 per cent).

4.2 *Valid types of term equivalence*

According to the study, there are six types of term equivalence between TELDAP and AAT (see Figure 2).

4.3 *Analysis of similarity between conceptual structures*

Relevant studies point out that the more apart two KOS are in respect of language and culture, the more heterogeneity they have in conceptual structures (Liang and Sini, 2006). According to the study’s preliminary findings, the degree of similarity between different conceptual structures can be divided into four types:

- (1) Similar structure, transferring part of the structure of source language into the system of target language by means of modeling.
- (2) Similar structure, but the structure of source language needs to be expanded or modified.
- (3) Dissimilar structure, the term in target language can be partially mapped to the term in source language.
- (4) Lack of structure, no matches between target language and source language.

More details will be presented in the following.

4.3.1 *Similar structure transferring part of the structure of source language into the system of target language by means of modeling.* The study uses “similar structure” to denote that the target KOS (NPM) has the same hierarchical logic as the source KOS (AAT), which is not conflicting but rather compatible with the contextual structure of NPM. Through the study, it is found that AAT conceptual structure has more or broader hierarchies than NPM does, so part of the AAT conceptual structure can be transferred to NPM by means of modeling.

4.3.2 *Similar structure but the structure of source language needs to be expanded or modified.* TELDAP terms often represent concepts unique to Chinese culture, so it is necessary to further expand the current conceptual structure of AAT to depth, or create a whole new set of concepts at the broader hierarchies, such as Chinese culture-specific concepts of “kiln system,” “archaeological period,” “main shape” “script” and “decoration.”

4.3.3 *Dissimilar structure – the term in target language can be partially mapped to the term in source language.* “Dissimilar structure” in the study indicates that TELDAP and AAT differ in classification logic of conceptual structures, but the TELDAP term

Equivalence type	Meaning	Example	
		TELDAP term	AAT term
1. Exact Equivalence	TELDAP term and AAT term represent exactly the same concept.	仰韶文化(< 考古學文化) Yangshao Culture(< archaeological period)	Yangshao (< Styles and Periods)
2. Exact Equivalence (cross ref.)		掐絲(< 技法) filigree(< technique)	filigree enameling (< Processes and Techniques)
3. Inexact Equivalence	TELDAP term and AAT term represent generally the same group of objects, phenomena or concepts, but with minor differences.	刻花(< 技法 < 紋飾) engrave (< technique < decoration)	bright cutting (< engraving (action) ... < Processes and Techniques)
4. Partial Equivalence (species-genus relationship)	TELDAP term can be mapped to a broader term in AAT with the similar meaning.	乳丁紋(< 點狀紋 < 幾何 < 紋飾) Nipple nail pattern/ Ruding pattern(< dots < geometric < decoration)	dots (< Design Elements)
5. Non-Equivalence (cultural uniqueness)	TELDAP term is not equivalent to any AAT term, for it represents a cultural-specific concept totally new to AAT.	隸書(< 書體 < 款識) Clerical script/li shu (< Chinese scripts < inscriptions)	x
6. Non-Equivalence (beyond the scope)		彌勒佛(< 佛道人物 < 人物 < 紋飾) Maitreya Buddha(< Buddhist figure < people < decoration)	x

Figure 2.
Six types of term
equivalence

can be exactly or partially mapped to an AAT term, such as the conceptual structure of “glaze color.”

4.3.4 *Lack of structure - no matches between target language and source language.* “Lack of structure” indicates that the conceptual structure of the TELDAP term is not similar to AAT, and related terms also cannot be mapped to any AAT term, such as the conceptual structure of “second shape.”

5. Discussion

5.1 *Types of term equivalence*

According to McCulloch and Macgregor (2008), the study that maps AAT, LCSH, MeSH and UNESCO with DDC indicates the ranking is “partial equivalence (species-genus relationship)”, “concept equivalence” and “exact equivalence” in order of frequency. From the previous results we can find out, “exact equivalence” and “partial equivalence (species-genus relationship)” are the most frequent types among the term equivalence relationships between different KOS in the same languages, different KOS in different languages, domain-specific and domain-general KOS, or between domain-specific KOS themselves.

There are two types of partial equivalence: genus-species relationship (mapped to a narrower term) and species-genus relationship (mapped to a broader term) In the preliminary findings, it is found out that all the 18 terms interpreted as partial equivalence are in the species-genus relationship, demonstrating that TELDAP terms are more in-depth, many of which can only be mapped to broader terms in AAT. For example, in the aspect of geometric motifs, there is “dots” under “geometric motifs” in AAT, but this is the bottom of the hierarchy; while in TELDAP, there is “dots” under “geometric pattern,” and below “dots” are “nipple nail pattern”, “grain pattern,” “plum blossom dots,” etc. These terms need to be further examined to see if they represent important concepts. If they do, then it is necessary to add new concepts for these terms under “dots” in AAT.

5.2 *Transferring part of the structure of source language into the system of target language by means of modeling*

Take the conceptual structure of “people” for example, the two sample terms “empress” and “Maitreya Buddha” have hierarchical structures as follows:

"people > monarch > empress"

"people > Buddhistfigure > MaitreyaBuddha"

When mapping with AAT conceptual structure, we can see that the two systems have similar structures, but AAT’s structural design contains a bigger context. Therefore, there are more guide terms or terms existing between “people” and “monarch > empress” in the hierarchy, such as < people by occupation >, < people in government and administration > and “rulers.” So if part of the AAT conceptual structure is going to be transferred to TELDAP by means of modeling, more related terms should be added to the hierarchy of TELDAP (terms highlighted in gray), as shown in Figure 3.

The previous terms highlighted in gray are the original terms in TELDAP, and other unmarked < guide terms > and terms could be further added through

Agents Facet > **People** > people (agents) > <people by occupation> > <people in government and administration> > rulers (people) > **monarchs** > **empresses**

Agents Facet 代理者層面

.... **People** 人物(層級)

..... people (agents) 人物

..... <people by occupation> <依職務區分之人物>

..... <people in government and administration> <政府與行政機構之人物>

..... rulers (people) 統治者

..... **monarchs** 宮廷人物

..... **empresses** 后妃

Figure 3.

translation. Take the conceptual structure of “type” for another example, the “type” in NPM includes: “copperwork, metalwork,” “ceramics,” “jade and stone work,” “wood, bamboo and lacquer work,” and “enamel and glassware.” When mapped to AAT, we can see that “Object Genres (Hierarchy Name)” under Objects Facet has the same structure. Therefore, part of the AAT conceptual structure can be transferred to TELDAP by means of modeling. But in order to build up the equivalence relationship between TELDAP and AAT, some modifications have to be made in respect of terms and structure of the two systems:

5.2.1 *Modification of TELDAP.* Here TELDAP terms are often displayed in compound nouns, such as “copperwork, metalwork,” “jade and stone work,” “wood, bamboo and lacquer work,” and “enamel and glassware.” After mapping with AAT, we can see that the hierarchical relationship between “copperwork” and “metalwork” are actually “species-genus relationship.” In other words, “copperwork,” along with “silverwork” and “ironwork,” is one kind of “metalwork.” According to this, the study suggests that the compound concept “copperwork, metalwork” should be divided into two related concepts “copperwork” and “metalwork” in the hierarchy. It is the same with “jade and stone work,” which can be adjusted into the hierarchy of “jade” and “stone work” (species-genus relationship); with the same principle, “wood, bamboo and lacquer work” can be broken into three single concepts, “woodwork,” “bamboo work” and “lacquerware,” and “enamel and glassware” can be broken into “enamel” and “glassware.”

5.2.2 *Modification of AAT.* The hierarchy of AAT shown in the following indicates that part of the TELDAP terms related to “type” are not included yet, such as “bamboo work” and “enamel.” Another case is that even though there are corresponding terms under the “Object Genres” in AAT, those terms are marked as non-preferred parent here. For example, “jade” and “stone work” in TELDAP can be mapped to “stone (rock [N])” in AAT, but the term’s preferred parent belongs to Materials Facet rather than Objects Facet that is in line with the TELDAP context. Due to the fact that there are many collections of “jade” in TELDAP, which is also an important archive feature, it is suggested that “jade” be added under the hierarchy of Object Genres under Objects Facet in AAT. Besides, “enamel” and “bamboo work” should also be incorporated into

AAT. Through this, the hierarchy of Object Genres in AAT can include important object types of East Asia.

Mapping the hierarchy of “type > copperwork, metalwork” with AAT, the results are as follows (terms highlighted in gray are the original terms in TELDAP) (see Figure 4).

5.3 Similar structure – but the structure of source language needs to be expanded or modified

Take “script” for example, in the case study, TELDAP terms such as “clerical script,” “regular script” and “seal script” are important forms of Chinese characters, often used in Chinese paintings, too. But they are not included in AAT. However, there are 12 kinds of scripts under < Arabic scripts > in AAT, the classification of which is similar to the historical development process of Chinese scripts (e.g. oracle bone script, bronze inscriptions, seal script, clerical script, standard script, running script, and cursive script). So it is suggested to add a set of < Chinese scripts > with the similar hierarchy under “scripts (writing) – < scripts by form >” in AAT. As shown in Figure 5.

5.4 Dissimilar structure – the term in target language can be partially mapped to the term in source language

TELDAP and AAT differ in the classification of ceramics glaze, for example: TELDAP uses “glaze color” as a kind of classification for describing the form of ceramics glaze, while AAT uses “composition,” “form” and “technique” as the basis of classification. Therefore, terms representing the same concept (“青花” in

Objects Facet > Object Genres (Hierarchy Name) > <object genres (Guide Term)> > <object genres by material> > metalwork > copperwork

.... Objects Facet 物件層面
 Object Genres (Hierarchy Name) 物件類型(層級)
 <object genres (Guide Term)> <物件類型(先導詞)>
 <object genres by material> <依材料區分之物件類型>
 ceramics (objects) 陶瓷器(物件)
 glassware 玻璃器
 lacquerware 漆器
 metalware 金屬器皿
 metalwork 金屬器
 copperwork 銅器
 stone (rock) [N] 石材
 <stone by quality> <石材依質地區分>.....jade 玉
 woodwork 木器

Figure 4.

Example : Chinese Scripts

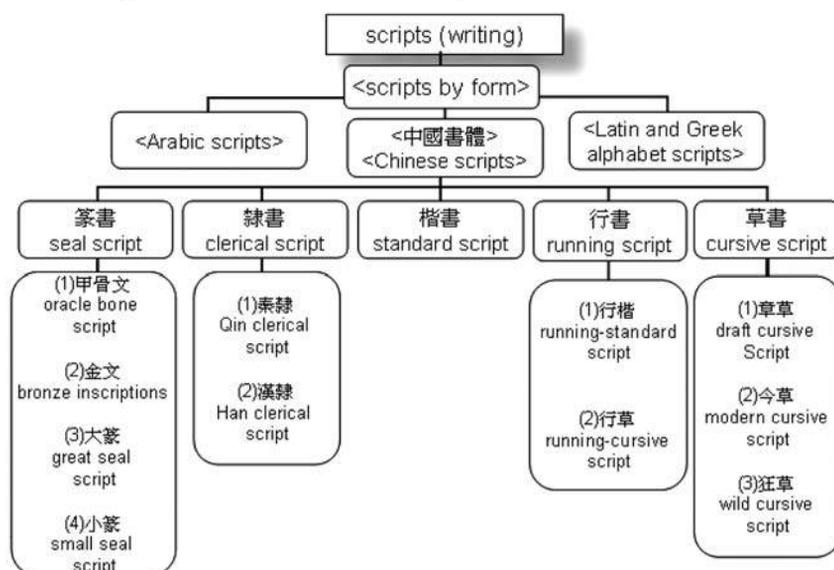


Figure 5.
The conceptual structure
of Chinese scripts

TELDAP and “blue-and-white (ceramic glaze)” in AAT) can appear in different conceptual structures. “青花” in TELDAP is categorized under “underglazing,” and “blue-and-white (ceramic glaze)” in AAT is put under < ceramic glaze by composition or origin >, as shown in Figure 6.

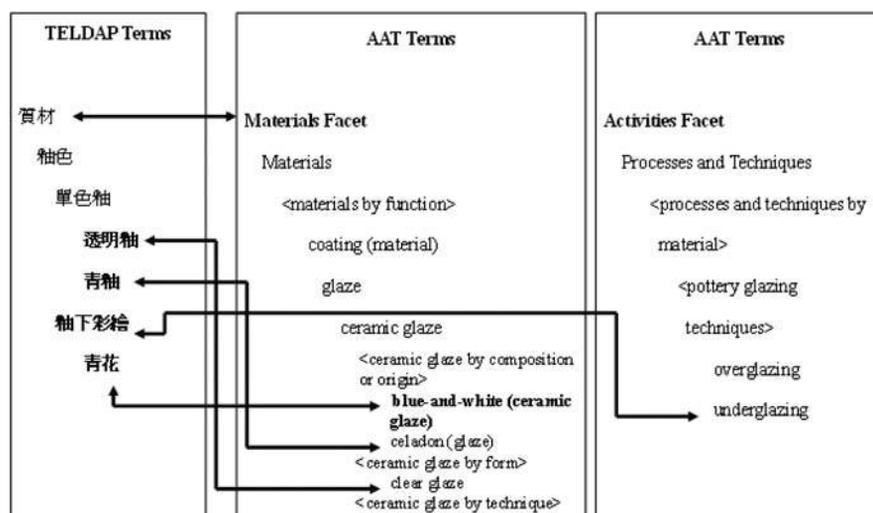


Figure 6.
The conceptual structure
of ceramics glaze in
TELDAP and AAT

6. Conclusions

In this paper we have presented an exploratory study that took the terms in AAT developed by The Getty Research Institute in US, and the controlled vocabularies from the NPM participating in Digital Archives Sub-Project of Antiquities in the National Palace Museum of TELDAP, as the principal research subjects, trying to conduct mapping analysis between Chinese and English terms, in order to gain an insight into the types of term equivalence and the types of matches between conceptual structures in the Eastern and Western art. The research results demonstrate that “exact equivalence” is the type of term equivalence that appears most frequently, second to which is “partial equivalence.” In the meantime, the study discovers six types of term equivalence to varying degrees, which are: exact equivalence, exact equivalence (cross ref.), inexact equivalence, partial equivalence (species-genus relationship), non-equivalence (culture uniqueness), and non-equivalence (beyond the scope). As for the analysis of similarity between conceptual structures of different systems, the degree of similarity can be divided into 4 types, including: similar structure, transferring part of the structure of source language into the system of target language by using the method of modeling; similar structure, but the structure of source language needs to be expanded or revised; dissimilar structure, the term in target language can be partially mapped to the term in source language; lack of structure, no matches between target language and source language.

According to the study, the terms in different languages and KOS under the same domain can be highly compatible (including the degree of term equivalence, and high similarity between conceptual structures), though there might be cultural-specific concepts and terms that are not mutually included yet. The next step is to further develop and design a conceptual structure that can incorporate more perspectives and viewpoints. As to the design and development of Chinese AAT, through the case study of TELDAP, preliminary findings indicate that the conceptual structure and terms of AAT can fulfill most of the needs of TELDAP, providing larger contextual structures and linking with in-depth terms in TELDAP. This has great significance for the development and interoperability of KOS, also saving the cost of building a whole new system. On the practical side, this research is quite beneficial for the collaboration between TELDAP and Getty’s AAT. It is suggested to establish bilingual mutual links between “exact equivalence” terms. With regard to “partial equivalence” terms, if they are further identified as important concepts, we can solve the problems of non-exact equivalence terms and dissimilar conceptual structures in the study by addressing the needs of English users in searching for Chinese art collections. Through the addition of and association with Asian perspectives, AAT can thus become more comprehensive and be enriched on a multicultural scale.

This study is based on the interoperability between two highly heterogeneous cultures. The application of the findings in the multilingual search in digital libraries can not only contribute to the information sharing among scholars and artists of Chinese and English language backgrounds, but also benefit the design of multilingual KOS in other cultures and the study of digital humanities, the most concrete example of which is that the Spanish, German and Dutch versions of AAT are also faced with many mapping problems caused by language and cultural differences (Nagel, 2010).

References

- Baca, M. (2009), "The Getty Vocabularies and issues in building multilingual thesauri", paper presented at TELDAP International Conference, Taipei.
- Chaplan, M.A. (1995), "Mapping Laborline thesaurus terms to Library of Congress subject headings: implications for vocabulary switching", *Library Quarterly*, Vol. 56 No. 1, pp. 39-61.
- Chen, S.J. (2009), "Chinese-language Art & Architecture Thesaurus: methods and issues", *CIDOC 2009, Santiago, 29 September*.
- Doerr, M. (2001), "Semantic problems of thesaurus mapping", *Journal of Digital Information*, Vol. 1 No. 8, available at: <http://jodi/ecs.soton.ac.uk/Articles/v01/i08/Doerr/>
- Harpring, P. (2010), *Art & Architecture Thesaurus: Editorial Guidelines*, J. Paul Getty Trust, Los Angeles, CA, available at: www.getty.edu/research/conducting_research/vocabularies/editorial_guidelines.html#aat
- Huang, C., Chang, R. and Lee, S. (2004), "Sinica BOW (Bilingual Ontological Wordnet): integration of bilingual WordNet and SUMO", *LREC 2004, Lisbon*, pp. 1553-6.
- Huang, C., Hsieh, S., Hong, J., Chen, Y., Su, I., Chen, Y. and Huang, S. (2008), "Chinese Wordnet: design, implementation, and application of an infrastructure for cross-lingual knowledge processing", *9th Chinese Lexical Semantics Workshop (CLSW2008), Singapore*, pp. 169-86.
- ISO (International Organization for Standardization) (1985), "Guidelines for the establishment and development of multilingual thesauri", ISO 5964, International Organization for Standardization (ISO), Geneva.
- Liang, A. and Sini, M. (2006), "Mapping AGROVOC and the Chinese agricultural thesaurus: definitions, tools, procedures", *New Review of Hypermedia and Multimedia*, Vol. 12 No. 1, pp. 51-62.
- McCulloch, E. and Macgregor, G. (2008), "Analysis of equivalence mapping for terminology services", *Journal of Information Science*, Vol. 34 No. 1, pp. 70-92.
- McCulloch, E., Shiri, A. and Nicholson, D. (2005), "Challenges and issues in terminology mapping: a digital library perspective", *Electronic Library*, Vol. 23 No. 6, pp. 671-7.
- Nagel, L. (2010), "The Spanish-LANGUAGE VERSION of the AAT: Tesouro de Arte & Arquitectura", paper presented at Multilingual Terminology Working Group Meeting, Los Angeles, CA, 23-26 August.
- Vizine-Goetz, D., Hickey, C., Houghton, A. and Thompson, R. (2004), "Vocabulary mapping for terminology services", *Journal of Digital Information*, Vol. 4 Nos 4, available at: <http://jodi.tamu.edu/Articles/v04/i04/Vizine-Goetz/>
- Vossen, P. (1998), *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*, Kluwer Academic Publishers, Dordrecht, available at: www.hum.uva.nl/~ewn/
- Zelfde, R. and Schaik, A. (2009), "AAT in Dutch language", *CIDOC 2009, Santiago, 29 September*, available at: www.koalawebhosting.com/cidoc/009.pdf
- Zeng, M.L. and Chan, L.M. (2004), "Trends and issues in establishing interoperability among knowledge organization systems", *American Society for Information Science and Technology (JASIST)*, Vol. 55 No. 5, pp. 377-95.

Further reading

- Riesthuis, J.A., Chan, L.M., Landry, P., Leth, P., McIlwaine, I., Kunz, M., McGarry, D., Naudi, M. and Zeng, M.L. (2005), "Guidelines for multilingual thesauri", *IFLA 2005, Oslo*, available at: www.ifla.org/VII/s29/pubs/Draft-multilingualthesauri.pdf

About the authors

Shu-jiun Chen received her MA degree in Library and Information Management from the Information Studies Department of University of Sheffield, UK in 1997, and is now a PhD candidate in the Department and Graduate Institute of Library and Information Science at the National Taiwan University. Ms Chen is the Assistant Research Engineer at the Research Center for Information Technology Innovation, Academia Sinica. She currently also serves as Program Manager of the Program Office of the Taiwan e-Learning and Digital Archives Program, and the Co-PI of the Metadata Architecture and Application Team, sponsored by the National Science Council, Taiwan. She is also the Chair of the MCN Taiwan (Museum Computer Network, Taiwan Chapter) Special Interest Group on Standards. Her research interests include digital libraries, metadata and knowledge organization systems. Recently, she initiated the Research Project of Chinese-English bilingual Art & Architecture Thesaurus with Getty Research Institute, USA. Shu-jiun Chen is the corresponding author and can be contacted at: sophy@gate.sinica.edu.tw

Hsueh-hua Chen received a Doctor of Education degree from the University of Georgia in 1986, a Master of Education degree from the same university in 1981, and a BA in Educational Media and Librarianship from National Taiwan University in 1977. Currently she is University Librarian at the National Taiwan University in Taipei, having previously been the University Press Director and Professor and Chairperson of the Department of Library and Information Science at the same university. Prior to that Hsueh-hua was Associate Professor and Director of the Department of Research and Development at the National Open University in Taipei.

This article has been cited by:

1. Krystyna K. Matusiak, Ling Meng, Ewa Barczyk, Beth Thomsett-Scott. 2015. Multilingual metadata for cultural heritage materials: the case of the Tse-Tsung Chow Collection of Chinese Scrolls and Fan Paintings. *The Electronic Library* 33:1. . [\[Abstract\]](#) [\[PDF\]](#)
2. Anne R. Diekema Design and Development Considerations for a Multilingual Digital Library 1222-1233. [\[CrossRef\]](#)
3. Anne R. Diekema Design and Development Considerations for a Multilingual Digital Library 1-12. [\[CrossRef\]](#)