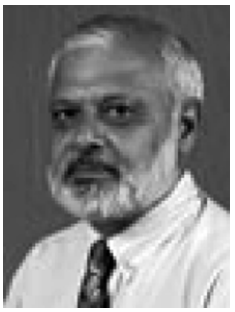
Support from Bibliographic Tools to Build an Organizational Taxonomy for Navigation: Use of a General Classification Scheme and Domain Thesauri

Zhonghong Wang*, Abdus Sattar Chaudhry**, and Christopher Khoo ***

* 625 Meadowlark Street, Troy, IL 62294 <wang_zhonghong@yahoo.com>
** Department of Library and Information Science, College of Social Science, Kuwait University, Kuwait 13060 <abdusattar.chaudhry@ku.edu.kw>
*** Wee Kim Wee School of Communication & Information, Nanyang Technological University, 31 Nanyang Link, Singapore 637718 <assgkhoo@ntu.edu.sg>

Dr. Wang obtained her PhD degree from the Nanyang Technological University, Singapore, in 2010. Before joining NTU, she had been a librarian in university libraries in China for more than 10 years. She currently is working in a public library in the United States. Her research interests are in the area of taxonomies, ontologies, metadata, and social tagging.



Dr. Chaudhry obtained his PhD from the University of Illinois at Urbana-Champaign in 1985. He was Head of Division of Information Studies, Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore, from 2003 to 2008. Before joining NTU, Dr. Chaudhry had held teaching and professional positions in USA, Saudi Arabia, Pakistan, and Malaysia. He is currently in the College of Social Sciences, Kuwait University. His current research focuses on information organization and knowledge management.

Dr. Khoo obtained his PhD at Syracuse University and his MSc in Library & Information Science at the University of Illinois, Urbana-Champaign. He is the head of the Division of Information Studies at Nanyang Technological University, Singapore, where he teaches courses in knowledge organization, information architecture, data mining and Web-based information systems. He has also worked for several years as a science reference librarian, cataloger and online information searcher at the National University of Singapore Libraries. His main research interests are in text mining (information extraction and text summarization), medical decision support system, knowledge organization, and human categorization behavior.

Wang, Zhonghong, Chaudhry, Abdus Sattar, and Khoo, Christopher. **Support from Bibliographic Tools to Build an Organizational Taxonomy for Navigation: Use of a General Classification Scheme and Domain Thesauri.** *Knowledge Organization, 37(4)*, 256-269. 25 references.

ABSTRACT: A study was conducted to investigate the capability of a general classification scheme and domain thesauri to support the construction of an organizational taxonomy to be used for navigation, and to develop steps and guidelines for constructing the hierarchical structure and categories. The study was conducted in the context of a graduate department in information studies in Singapore that offers Master's and PhD programs in information studies, information systems, and knowledge management. An organizational taxonomy, called Information Studies Taxonomy, was built for learning, teaching and research tasks of the department using the *Dewey Decimal Classification* and three domain thesauri (ASIS&T, LISA, and ERIC). The support and difficulties of using the general classifi-

cation scheme and domain thesauri were identified in the taxonomy development process. Steps and guidelines for constructing the hierarchical structure and categories were developed based on problems encountered in using the sources.

1.0 Introduction

Taxonomies are increasingly being used to organize content within organizations and to support navigation of web sites or digital repositories. Several writers have advocated using a top-down approach and classification schemes and thesauri as sources for building organizational taxonomies (Iyer 1995; Aitchison et al. 2000; Conway and Sligar 2002; Cisco and Jackson 2005). This would allow the taxonomies to leverage on the strengths and principles underlying existing classification schemes and thesauri (McGregor 2005; Saeed and Chaudhry 2002), and enable the taxonomies to be developed with less effort than starting from scratch (Wyllie 2005). At the same time, it has been pointed out that organizational taxonomies are different from classification schemes and thesauri in scope, components and roles (Wang et al. 2006). The coverage of organizational taxonomies depends more on the activities of the organizations and interests of the stakeholders. The hierarchical structure used for navigation is expected to be more flexible and simpler; and the categories of taxonomies must be intuitive to intended users. The construction of an effective organizational taxonomy that supports navigation needs to incorporate the organizational context and take into consideration its navigational role while using components of classification schemes and thesauri.

Several taxonomy projects (McGregor 2005; Saeed and Chaudhry 2002; Bertolucci 2003) have used classification schemes and thesauri to build taxonomies. These projects demonstrated that bibliographic tools have the potential of providing the knowledge context and terms of categories (Saeed and Chaudhry 2002). Taxonomies built based on them would share the consistency of the classification schemes and controlled vocabularies (McGregor 2005). But these projects did not incorporate the organizational context, the activities of the organizations, and interests of the stakeholders, in the taxonomy development process. The organizational context was missing in the medical taxonomy development process that used MeSH (McGregor 2005). The pilot study in the computer science domain of using *Dewey Decimal Classification* and IEEE Web Thesaurus (Saeed and Chaudhry 2002) did not define the application scope of the taxonomy. The Snoopy taxonomy built using *DDC* (Bertolucci 2003) was composed of 50 categories that indicated the narrow scope of the project. In the SeSDL educational taxonomy, the "subjects" facet was based on the ten main classes of *DDC*, and the other facets used the British Education Thesaurus as one source of categories. However, no details of the development process have been reported. The prototype of the taxonomy was accessible on the Internet. In other words, an empirical study of building an organizational taxonomy by using classification schemes and thesauri is still lacking.

We conducted an empirical study of building an organizational taxonomy using a general classification scheme and domain thesauri keeping in view the previous taxonomy projects. The objectives of the study are: 1) to review the capability of a general classification scheme and domain thesauri in supporting an organizational taxonomy that is used for navigation; and, 2) to develop steps and guidelines for constructing the hierarchical structure and categories. We hope that the report of advantages and problems we encountered in using the general classification scheme and domain thesauri will provide a lesson for using sources of bibliographic tools. We also hope that the steps and guidelines we developed will be helpful for other organizations to build taxonomies.

2.0 Research Approach

The empirical study was conducted in the context of an academic organization (a graduate school) in the information studies domain, the Division of Information Studies, School of Communication and Information, Nanyang Technological University, Singapore. The Division has 15 full-time faculty members and nearly 500 students, and offers three Master's programs by coursework: MSc in Information Studies, Information Systems, and Knowledge Management, and Master's and PhD programs by research. The students in the MSc coursework programs focus on courses and project reports in the Critical Inquiry course that involves small group research projects. The students in the research programs focus on research projects and theses. The Division has four main research groups: Information and Knowledge Management, Knowledge Organization and Discovery, Information Retrieval and Digital Libraries, and User and Usability Studies. We se-

lected the Division as an academic organization because it has explicit goals and divisions of people that are compatible with the two essential features of "goal-oriented" and "coordinated human activities toward the common goal" in slightly different definitions of organizations (Barnard 1938; Schein 1970; McAuley et al. 2007), and a considerable scale in library and information science education.

Dewey Decimal Classification (*DDC*) and three domain thesauri, ASIS&T, LISA, and ERIC thesauri, were chosen as sources. *DDC* was selected because its structure makes it easy to navigate and it has been used in previous projects related to navigation (Saeed and Chaudhry 2002; Vizine-Goetz 2002). The two thesauri (ASIS&T and LISA) in the library and information science area, and the ERIC education thesaurus, were selected based on their relevance to the subject coverage of the taxonomy.

An organizational taxonomy, called *Information Studies Taxonomy,* was built for the Division. We designed three phases to develop the organizational taxonomy keeping in view guidelines suggested in the literature (Roberts-Witt 2000; Conway and Sligar 2002; Choksy 2006, Lambe 2007; Sharma et al. 2008). The first phase is taxonomy needs identification. We examined the goals, major tasks of the Division, and the major stakeholders across the tasks. Interviews with 17 stakeholders were conducted to investigate the stakeholders' tasks, created knowledge assets, and problems encountered in locating information resources for performing tasks. The use of existing knowledge organization systems in the Division intranet was also examined. The second phase is taxonomy design. We determined the taxonomy objectives, roles, target users, organization scheme (facets), subject coverage, and target content based on problems that the taxonomy aimed to address, the activities of the Division, and the needs of stakeholders. The last phase is the taxonomy construction.

We constructed the taxonomy with a focus on the subject facet, keeping in view the research objectives. The hierarchical structure and categories of the subject facet was manually constructed via a combination of top-down and bottom-up approach and with a focus on the top-down. The construction of the hierarchical structure started from the top-level, the main categories, high-level categories (level 2, 3), and low-level categories (level 4, 5). In addition to *DDC* and the domain thesauri, sources related to the tasks of the stakeholders such as course materials and staff publications, sources from the community of library and information science schools (LIS) (e.g., course descriptions on websites of LIS schools), sources from relevant professional associations (e.g., IFLA Guidelines for Professional Library/Information Educational Programs 2000), and relevant domain taxonomies (Hawkins et al. 2003; Mentzas 1994; Doke and Barrier 1994; Cheung et al. 2005) in information systems and knowledge management were also selected. Keeping in view suggestions in the literature (Cheuk 2002; Hunter 2005; Pack 2002; Batley 2005; Raschen 2006; Dickson 2008), we made an effort to incorporate the stakeholders' interests and perspectives in the construction of the hierarchical structure and categories by employing relevant sources. We employed those additional sources when we found that *DDC* and the domain thesauri were not adequate. As recommended by Wyllie (2005), Lambe (2007), Dickson (2008), Singhal and Nath (2008), in the last construction step, we delivered the taxonomy draft to 11 stakeholders for review.

3.0 Information Studies Taxonomy

Figure 1 illustrates the objectives, roles, and intended users of the taxonomy. The taxonomy was expected to support the learning/teaching and research tasks in the Division. It focuses on two groups of users: graduate students and instructors. The taxonomy will support navigation and knowledge discovery of a digital repository that is relevant to the students and instructors' learning/teaching and research tasks, such as course materials, research project reports, dissertations, and so on.

The taxonomy draft comprises 7 facets and about 540 categories. The subject facet comprises about 440 categories ranging from 2 to 5 levels. A full listing of the taxonomy prototype with references and sources of labels has been reported in a previous paper (Wang et al. 2008). A brief display of the 7 facets and the 12 main categories (top-level) of the subject facet is shown in Fig. 2.

4.0 Support of a General Classification Scheme and Domain Thesauri

4.1 A General Classification Scheme (DDC)

We had assumed that most of the main categories (top-level of the subject facet) could be selected or adapted from *DDC*. But it is not often the case that the subject coverage of an organizational taxonomy falls in the main classes or sub-classes of a general classification scheme, and the main categories repre-

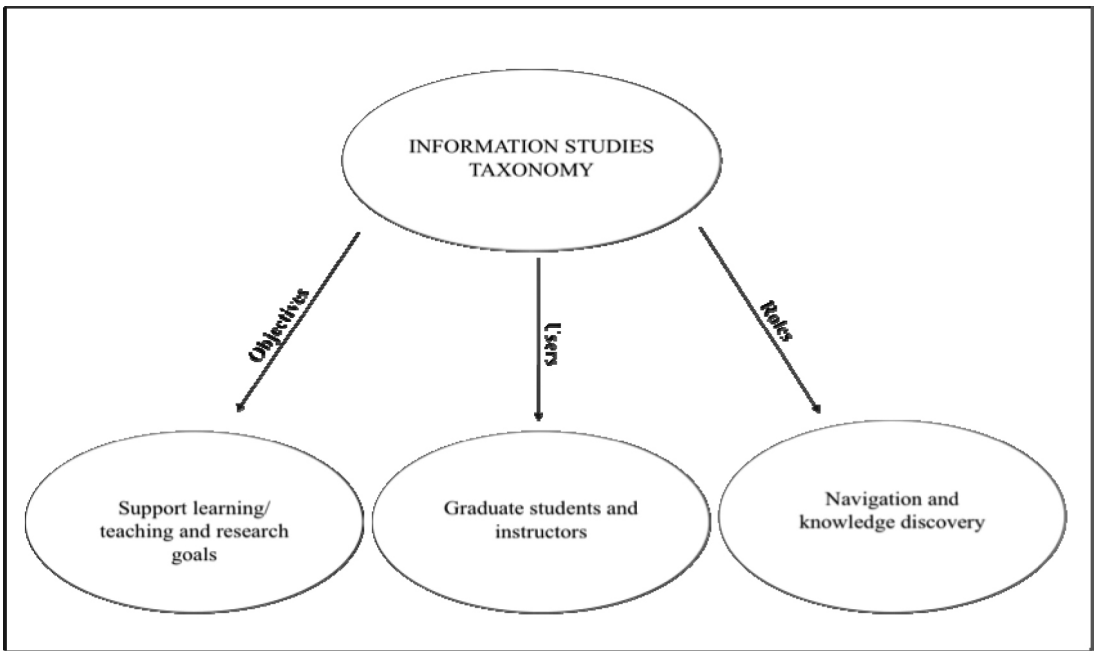


Figure 1. Objectives, roles, and intended users of the Information Studies Taxonomy

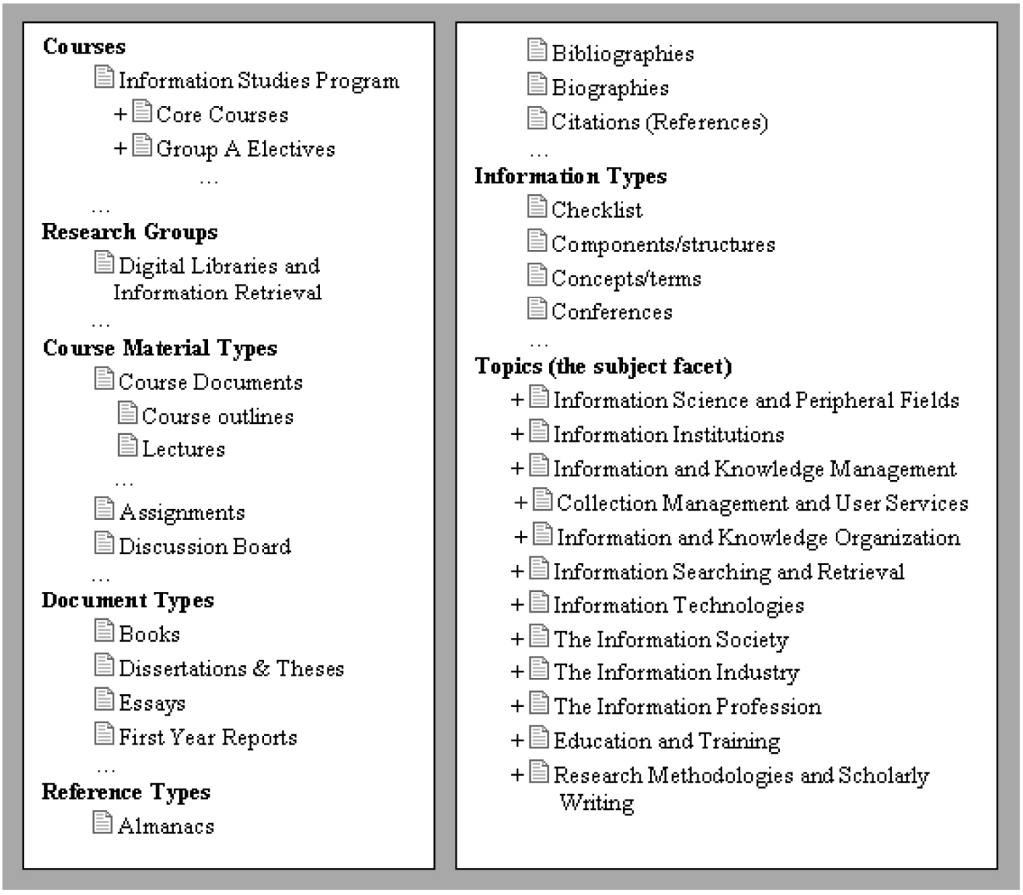


Figure 2. Overview of the Information Studies Taxonomy

senting the subject coverage can be selected. The subject coverage of this taxonomy that was determined based on the programs, research groups in the Division, and the tasks of the stakeholders, was not in line with the ten main classes and sub-classes in the library science and computer science schedule of *DDC*. The 12 main categories of the subject facet previously listed were not selected from *DDC*.

We also found that not all high-level categories (level 2, 3) within the main categories could be selected from *DDC*. The high-level categories within 5 of the 12 main categories (less than 50%) were selected from *DDC*. Table 1 lists the 5 main categories and relevant *DDC* classes. The 5 main categories were fairly similar to relevant sub-classes of *DDC*. For example, Information Institutions is similar to 026-027 Library and Information Sciences—Specific Kinds of Institutions; The Information Industry to 338.1-338.4 Economics—Production—Specific Kinds of Industries; and high-level categories within the two main categories of Collection Management and User Services, and Information and Knowledge Organization, were selected from the 025 Operations of Libraries, Archives, and Information Centers.

Main Categories (relevant)	*DDC*
Information Institutions	026, 027
Collection Management and User Services	025.2, 025.5, 025.6, 028
Information and Knowledge Organization	025.3, 025.4
Information Technologies	004, 005, 006, 384.1-384.6
The Information Industry	338.47001-338.47999
Main Categories / hierarchies (non-relevant)	**Reasons**
Information and Knowledge Management / Archives management	No related classes
Information Searching and Retrieval / Information storage and retrieval systems	Not compatible with the users' perspectives (025.04, 025.06)
The Information Profession	No detailed structure (020.9)
Education and Training	No detailed structure (020.07)

Table 1. Relevant main categories and examples of non-relevant main categories

High-level categories within other 7 main categories were not from the *DDC* in three situations. Table 1 lists examples of the main categories and the reasons why. First, some areas did not fall in the main classes of *DDC*. For example, no specific classes were related to the areas of archives management, document management, knowledge management, and scholarly writing. Second, *DDC* represented some areas in one class but could not provide detailed structures, for example, library and information science education (020.7), information professionals (020.9), and research methodologies (001.4). These two situations are probably typical because organizational taxonomies are different from general classification schemes in the nature of the subject coverage, and taxonomies used for navigation would require more detailed categories for tagging resources other than book collection. Another possible situation is that the structures provided by the classification schemes might not be compatible with the perspectives of the intended users, as experienced by Bertolucci (2003). For example, the structure of the 025.04 class provided by *DDC* was not adopted because it organized types of information retrieval systems by subjects and persons. Such a structure would not fit needs of the students and instructors for learning or teaching and research in the area. Our findings suggested a general classification scheme would not be sufficient to construct the hierarchical structure.

Within the 5 relevant main categories, 15 out of 29 (52%) categories at level 2 and 44 out of 106 (41.5%) categories at level 3 were identified from *DDC*. Table 2 lists examples of the categories. These categories were identified from classes or relative index terms of the *DDC*. For example, within the main category of Information Institutions, the three categories at level 2, Archive, Libraries, and Information Centers, were identified from the 026 and 027 classes. Also, within the main category of Collection Management and User Services, the three categories at level 2 and some at level 3 were identified from the 025, 028 classes, and Relative Index terms.

In addition to the high-level categories, we observed that general classification schemes might provide support for the low-level categories, especially those based on the "genus/species" division. Within the 5 relevant main categories, lower categories at level 4 and 5 (14 out of 31 hierarchies) were identified from the *DDC*: 9 of the 14 hierarchies are based on the "genus/species" division. For example, the lower categories within the hierarchy of Classification Schemes, Controlled Vocabularies, Special Material Cataloging, which are based on the "genus/species" division, were identified from the 025.3 and 025.4 classes.

Main Categories	Hierarchies and Categories	DDC
Information Institutions	**Archives**	026 Libraries, archives, information centers devoted to specific subject and disciplines 027 General libraries, archives, information centers
	Libraries	026, 027
	/ Academic libraries	027.7 College and university libraries Relative Index term: academic libraries
	/ Government libraries	027.5 Government libraries
	/ National libraries	027.4 Public libraries
	/ School libraries	027.8 School libraries
	/ Special libraries	027.1-027.3, 027.6 Relative Index term: special libraries
	/ Subject libraries	026
	Information centers	026, 027
Collection Management and User Services	**Collection management**	Relative Index term: collection management – library science
	/ Collection development	025.21 Collection development
	/ Material acquisitions	025.23 Acquisition through purchase 025.26 Acquisition through exchange, gift, deposit
	/ Special collection development	025.28 Acquisition of and collection development for materials in Special forms
	/ Archival collection development	025.2814 Archival materials – library acquisitions
	/ Audiovisual collection development	025.2873 Motion pictures – library acquisitions
	/ Digital collection development	025.284 Electronic resources – library acquisitions
	/ Serial collection development	025.2832 Serials – library acquisitions
	/ Collaborative collection development	Relative Index term: cooperative collection development
	/ Collection maintenance	025.8 Maintenance and preservation of collections
	/ Collection preservation	025.8
	User services	025.5 Services for users
	/ Circulation services	025.6 Circulation services
	/ Reader advice services	025.54 Reader advisory services to individuals and groups
	/ Reference services	025.52 Reference and information services
	/ Reference sources	Relative Index term: reference works
	/ User instruction	025.56 Orientation and bibliographic instruction for users
	Reading and use of information media	028 Reading and use of other information media
	/ Reading interests and habits	028.55 Reading interests and habits of young people

Table 2. Examples of categories identified from *DDC*

However, we found that the *DDC* might not fully support the high-level categories within the 5 relevant main categories; while it would play a major role. In this case, about half of the high-level categories within the 5 main categories were not from *DDC*. These categories were not from the *DDC* in three situations. The first situation was that it was difficult to make use of the discipline-based main classes, to represent categories within the main category of Information Technologies from the 000 schedule that focused on computer science. Similarly, categories within Archive and Information Centers were selected from the ASIS&T and LISA thesaurus because the 020 schedule focused more on librarianship. The second situation was that it was difficult to fully make use of classes that allows number-building, for example, to identify categories within the main category of The Information Industry from the 338.47001-338.47999 class. Similarly, lower categories in the hierarchy of Computer Applications in Information and Knowledge Organization could not be from *DDC* because *DDC* used only one class (025.30285) to represent this area. The above two situations could be typical because organizational taxonomies and general classification schemes are different in the nature of the subject coverage and applications. The third situation was that *DDC* did not cover some new concepts, such as, metadata and social tagging. The high-level category of Resource Description within the main category of Information and Knowledge Organization was thereby added from the Division's course materials to accommodate the new concepts. Similarly, categories such as Ontologies, Cataloging Outsourcing, Semantic Networks, and Mobile Communications were added from the thesauri. Also, categories could be added for the purpose of adding the taxonomy features. For example, the category of Collection Measurement was added within the hierarchy of Collection Development to make the hierarchy comprehensive and consistent with other hierarchies. The above findings, as previously pointed out, indi-

Components of the Subject Facet		**Support of DDC**
Main categories		None support
Hierarchies	High-level categories	Part support (five of 12 main categories)
		Part support (About half of the first and second-level categories within the five relevant five main categories)
	Low-level categories	Part support (categories at the third and forth level of 14 out of 31 hierarchies within the five relevant main categories)
	Divisions	None support
Categories	Category terms	Part support (using Relative Index)
	Scope of categories	Support (using hierarchical structure)
	Labels of categories	Support (using Relative Index)
	UF references	Support (using Relative Index)
Components of the Subject Facet		**Support of the Thesauri**
Main categories		None support
Hierarchies	High-level categories	Part support (using the hierarchical index)
	Low-level categories	Most support (using term relationships)
	Divisions	None support
Categories	Category terms	Most support
	Scope of categories	Support (using term relationships)
	Labels of categories	Support
	UF references	Support (using term relationships)

Table 3. Support of *DDC* and the domain thesauri for the Information Studies Taxonomy

cated that more than a general classification scheme had to be employed to construct the hierarchical structure of organizational taxonomy.

A general classification scheme may not support the division criterion that selects categories at the same level for organizational taxonomies. General facets such as subjects, geography, and persons used in *DDC* were not appropriate for the taxonomy with a narrow coverage. Table 3 summarizes the support of components of *DDC* for the organizational taxonomy.

4.2 Domain Thesauri

As we expected, most categories (concepts and terms) of the subject facet were from the three domain thesauri. But we found that they were not sufficient for reflecting the organization, interests of the stakeholders, and features for navigation. About 16.6%, 71 out of 427 categories, were not from the thesauri. These categories can be grouped into new concepts, compound terms, and terms particularly related to the taxonomy domain and the organization. Table 4 lists examples of these categories. New concepts, such as media resource centers, collaborative tagging, mobile information retrieval systems, and digital watermarking, reflected the interests of the stakeholders. Among the new concepts, terms in the area of information management and knowledge management were not from the thesauri because they treated these areas from a broad perspective, such as information science or knowledge, and these concepts were used in the Division more from the organizational communication perspective. Compound terms, such as archival collection development and audiovisual material cataloging, were necessary for the hierarchical structure to support navigation. Other terms, such as knowledge management professionals, knowledge management education, library and information science schools, and information system development methodologies, reflected the organization as well as the taxonomy domain. The above findings suggest that more than the domain thesauri had to be used to collect category terms and concepts for the organizational taxonomy.

We observed that the hierarchy of terms in a thesaurus has the potential for supporting high-level categories. For example, the five categories at level 2 within the main category of The Information Society were identified from the Hierarchical Index of ASIS&T thesaurus. We found that the term relationships of thesauri were helpful for identifying low-level categories. Most relevant low-level categories were identified from the term relationships of the thesauri. A small number of low-level categories were not from them in two situations related to the scope of the thesauri and division's interior for creating narrow

terms. For example, narrow terms provided by the ERIC thesaurus in the area of quantitative research methodologies were not adopted because they were not in line with the "genus/species" division used in the hierarchy. As concluded by Saeed and Chaudhry (2002), we found that term relationships of thesauri were also helpful for identifying the scope of terms. For example, as previously mentioned, the terms in the area of information and knowledge management provided by ASIS&T and LISA thesauri were not adopted because they were not appropriate for the organizational context. The scope of these terms was identified by their term relationships. Table 4 lists the support of components of the domain thesauri for the organizational taxonomy.

New concepts	– Media resource centers – Collaborative tagging – Mobile information retrieval systems – Digital watermarking – Mobile commerce – Information development – Information audit – Information distribution – Information sharing – Information utilization – Knowledge audit – Knowledge development – Knowledge capture – Knowledge sharing – Knowledge utilization
Compound terms	– Archival collection development – Electronic collection development – Audiovisual collection development – Audiovisual material cataloging – Cartographic material cataloging – Digital resource cataloging
Terms related to the organization	– Knowledge management professionals – Information science & systems education – Knowledge management education – Library and Information Science Schools – Information system development methodologies – Oral presentation skills

Table 4. Examples of categories from sources other than the three thesauri

5.0 Difficulties Encountered

The major difficulties we encountered involved partial support of *DDC* and domain thesauri, manipulation of multiple sources, and incorporation of stakeholders' interests and perspectives in the construction of the hierarchical structure and categories. The main categories and hierarchical structures within some main categories had to be constructed without the help of the *DDC*. Concepts and terms of categories had to be collected from sources more than the domain thesauri.

The selected sources (classification schemes and thesauri) may represent the same concepts from different contexts and using different terms. For example, for the concept of classification schemes, the *DDC* used it in the librarianship context; the ASIS&T thesaurus represented it by taxonomies. And for the concept of automatic thesaurus generation, the ASIS&T thesaurus chose automatic taxonomy generation; and the LISA thesaurus combined the two terms of automatic construction and construction of thesauri. Also, vocabularies from the 020 schedule of *DDC* focused more on librarianship, and the terms of the ERIC thesaurus focused on general education.

The sources might provide different structures/term relationships for the same terms/concepts. For example, for the concept of ontologies, *DDC* reflected it using the 006.33 class of knowledge-based systems; the ASIS&T treated it as a narrow term of controlled vocabularies; and the LISA thesaurus related it to semantic web, controlled vocabularies, and thesauri. Also, the term relationships in a thesaurus may not be rigorously applied. For example, the LISA thesaurus put the three terms of archival description, archives law, and national archives that are not at the same conceptual level as narrower terms of archives.

6.0 Steps and Guidelines for Constructing the Hierarchical Structure and Categories

6.1 Steps

We developed the steps for addressing the difficulties we encountered in using the general classification scheme and domain thesauri. We used multiple sources in addition to the *DDC* and domain thesauri. In particular, sources from the organization were used to reflect the stakeholders' interests and perspectives such as in collecting the concepts and terms of categories, and determining the mapping of low-level categories with high-level categories. Sources related to

the professional association (IFLA) and the community (library and information science schools) were employed to reflect the taxonomy domain and for determining the main categories. We designed the steps keeping in mind the need to manipulate multiple sources and incorporate the stakeholders' interests and perspectives. The steps involve constructing facets, main categories, category concepts and terms, hierarchies, and labels of categories.

6.1.1 Selecting Facets

The facets were not selected from the *DDC* or the domain thesauri. They were determined based on the taxonomy application context. The specific steps are as follows:

1. Select facets from the application context of the taxonomy, such as tasks and roles of the stakeholders, and types of target content.
2. Create labels for facets.

6.1.2 Determining the Main Categories (Level 1) of the Subject Facet

We designed the main categories as a separate step because they are at the top-level, represent the subject coverage of the taxonomy, and general classification schemes or domain thesauri are not expected to be useful in specifying the main categories. The specific steps are as follows:

1. Identify major areas and concepts of the stakeholders' interests by reviewing sources related to the stakeholders' tasks.
2. Select the main categories to cover concepts of interest and subject areas from industry/community sources (documents from professional associations such as IFLA Guidelines, course descriptions from library and information science school websites), and domain taxonomies (e.g. Information Science Taxonomy).
3. Create additional main categories to cover concepts of interest and subject areas not found in community/industry sources and domain taxonomies.
4. Select labels from the sources and construct labels for main categories not found in the sources.

6.1.3 Collecting Category Concepts and Terms

In the medical taxonomy project, McGregor (2005) used a term list representing the online journal content as a basis to select terms from the MeSH headings. We designed a term list representing the stakeholders' interests to select terms. The selection is in three steps as follows:

1. Create a list of concepts and terms related to the stakeholders' interests by selecting and consolidating terms from sources related to the stakeholders' tasks.
2. Select terms from the general classification scheme (*DDC*, class captions, and Relative Index terms), domain thesauri, and domain taxonomies based on the relevance to terms in the lists.
3. Consolidate concepts and terms from different sources.

6.1.4 Constructing the Hierarchies

Saeed and Chaudhry (2002) proposed three steps for constructing the hierarchical structure using classification schemes and domain thesauri: selecting hierarchies from the classification schemes, selecting terms from domain thesauri, and mapping the selected terms into the hierarchies. Based on their proposal, we designed four steps to construct the hierarchies. In addition to the three steps of high-level categories, low-level categories, and mapping, we inserted a step of division criteria to create neat categories at the same level. The specific steps are as follows:

– **High-level Categories (Level 2)**
  1. Based on relevance to the main categories, identify and reconsolidate the high-level categories (concepts and terms) from structures/term relationships of the general classification scheme (*DDC*), the Hierarchical Index of the thesaurus (ASIS&T), and relevant domain thesauri (mainly Information Science Taxonomy).
  2. Review whether the selected high-level categories cover the lower category terms. Create new high-level categories based on the lower category terms when the selected high-level categories cannot cover them.

– **Division Criteria (To Create Categories at the Same Level)**
  3. Determine the division criteria by identifying knowledge structures inherent in sources in the organization (e.g. course lectures).

– **Low-level Categories (Level 3)**
  4. Identify and reconsolidate the low-level categories from the multiple sources using the high-level categories as the starting points.
  5. Determine the low-level categories based on the chosen division criteria.

– **Mapping low-level categories and expansion of hierarchies (Level 4 or 5)**
  6. Map the low-level categories to the main categories and high-level categories by identifying knowledge structures inherent in sources in the organization (e.g. course lectures).
  7. Build cross references for categories that can be mapped into more than one perspective or cannot be mapped into the ideal "hierarchies".
  8. Expand the hierarchies by further identifying low-level categories and mapping them to the main categories and higher categories.
  9. Balance the levels of hierarchies within 5 levels.

6.1.5 Determining Labels of Categories

We created guidelines for selecting labels from various terms that were collected from multiple sources. To support user navigation, labels should fully reflect the concepts at hierarchical levels, be simple expressions, and be consistent. We reviewed whether the terms were appropriate for the organization, stakeholders, and at the target hierarchical levels. Similar guidelines can be found in the literature, but we had to address aspects of labels and terms from different sources. The specific steps are as follows:

1. Select labels from category terms according to the guidelines we created.
2. Determine the scope of the labels (terms) based on their uses in organization sources.
3. Modify the vocabularies (class captions and relative index terms) from the general classification scheme (*DDC*) into simpler expressions and in a style reflecting the taxonomy domain.
4. Modify the terms from the thesauri into expressions fully reflecting concepts at the target hierarchical levels.
5. Create labels for higher-level categories when labels cannot be found in the category terms.
6. Format the labels according to the thesaurus construction standard (ANSI/NISO Z39.19-2005 standard).
7. Organize the labels of categories at the same level alphabetically (e.g. "genus/species") or logically (e.g. "aspects" and "procedure") based on the division criteria used, group categories based on the same division criterion together when more than one division is employed at the same level.
8. Build UF (used for) references used for category terms that were not chosen as labels.

Table 5 lists examples of labels modified from *DDC* or the domain thesauri, and the reasons why.

6.2 Guidelines

We developed guidelines based on the difficulties we encountered in using the general classification scheme and domain thesauri, and experience we had in constructing the hierarchical structure and categories. The specific guidelines are as follows:

6.2.1 Selecting Facets

1. Labels of facets are not likely to be found in a general classification scheme or domain thesauri because they are usually very broad concepts or a combination of concepts. They usually have to be home-created.

Labels	Terms	Reasons
Subject classification schemes	(025.46) Classification of specific disciplines and subjects (DDC class)	To be simpler
Special material cataloging	Special materials – cataloging (DDC Relative Index term)	To be simpler
Information visualization	Electronic visualization (ASIS&T)	More often used in the organization
Software design	Program design (DDC Relative Index term)	To be compatible with the hierarchy of Software Engineering
Professional certification	Certification (ERIC)	To be compatible with the main category of The Information Profession
Collaborative cataloging	Cooperative cataloging (DDC Relative Index term)	To be consistent (collaborative)
Collection measurement	Collection assessment (ASIS&T, LISA)	To be consistent (measurement)

Table 5. Examples of labels modified from *DDC* and the thesauri

6.2.3 Constructing the Main Categories (level 1) of the subject facet

1. The main categories are not likely to be found in a general classification scheme or domain thesauri. However, it depends on whether the subject coverage of the taxonomy matches the main classes of the general classification scheme or its sub-classes.
2. When the subject coverage of the taxonomy does not match the main classes of the general classification scheme or its sub-classes, the main categories should be selected from sources in the organization, industry/community sources, and relevant domain taxonomies.
3. The size of the coverage and the width of the subject facet need to be considered to determine the number of main categories.

6.2.4 Selecting Category Concepts and Terms

1. Most of the category concepts and terms are likely to be found in domain thesauri. However, it depends on the availability of thesauri and the scope of the thesauri. Some category concepts and terms are likely to be found in the general classification scheme (class captions) and relevant domain taxonomies.
2. New concepts, compound concepts, and concepts representing organization tasks and activities (rather than academic subjects) may not be found in the domain thesauri.
3. The concepts and terms should be annotated with their sources.
4. See references provided by the domain thesauri should be kept with terms.

6.2.5 Constructing Hierarchies

– **Hierarchies: high-level categories (level 2)**
   1. If the main categories match the main class or sub-classes of the general classification scheme, about half the high-level categories (concepts and terms) are likely to be found in the general classification scheme.
   2. When the main categories are not related or relevant to the main classes or sub-classes of the general classification scheme, the high-level categories are likely to be found in the Hierarchical Index of a domain thesaurus and relevant domain taxonomies.
   3. The high-level categories (concepts and terms) can be selected from classes at different levels as well as related relative terms of the general classification scheme.
   4. The division criteria inherent in the selected high-level categories should be reviewed to ensure they are at the same conceptual level.
   5. The categories at level 2 should be based on one division criteria.
   6. When the high-level categories cannot be selected from the general classification scheme as well as domain thesauri and relevant domain taxonomies, they have to be selected from sources in the organization or home-created.
   7. The selected high-level categories should be annotated with their sources.

– **Hierarchies: division criteria (used to create categories at the same level)**
   8. The division criteria are not likely to be found in the general classification scheme. However, it depends on the size of the subject coverage of the taxonomy. For a taxonomy involving several subjects and with a narrow coverage, the division criteria have to be home-created.

– **Hierarchies: low-level categories (level 3)**
   9. Most low-level categories (concepts and terms) are likely to be found in the domain thesauri.
   10. More than one division criterion can be used to determine the low-level categories, depending on the number of categories and levels of hierarchies. These divisions are expected to be at the same conceptual level.

– **Hierarchies: mapping low-level categories and expansion of hierarchies (level 4 or 5)**
   11. The selection of categories at level 4 or 5 depends on the stakeholders' interests. Areas of high-level categories that are not the stakeholders' major interests are not expected to have many lower categories and levels.
   12. The number of levels can be shortened by using more than one division criterion at the same level.

– **Determining Labels of Categories**
   13. Prefer labels from sources in the organization
   14. Choose the same terms to represent the same concepts.

7.0 Conclusion

We conducted an empirical study of the extent that the *DDC* and three relevant domain thesauri can be used in developing an organizational taxonomy for an academic department in the information studies domain. We started with the assumption that a general classification scheme and several relevant domain thesauri would provide excellent support for the subject facet of an organizational taxonomy, particularly in an academic organization. Some modifications to the scope of the concepts, terms (labels), or concept/term relationships selected from these sources would of course be necessary to adjust them to the organizational context. But we found that the *DDC* and the domain thesauri were far from being sufficient for the organizational taxonomy. In particular, *DDC* could not provide support for the top-level categories of the taxonomy because the taxonomy is not discipline-based. Its subject coverage depends on the activities of the organization, and the tasks of the stakeholders. The *DDC* could also not provide complete support for the high-level categories (level 2 or 3 of the subject facet) for the same reasons and because of the focus on supporting navigation. The two selected domain thesauri in the area of library and information science also could not provide support for concepts and terms in the area of information and knowledge management because the organization treats them from a different perspective – the perspective of organizational communication.

Organizational taxonomies are different from general classification schemes and domain thesauri in their application scope and navigation roles. We used additional sources, for example, course materials and research publications of the organization to reflect the stakeholders' interests, and a domain taxonomy (Information Science Taxonomy) and those from the professional association and the websites of sibling organizations to help with the top-level categories. The steps we used to construct the hierarchical structure and categories took into consideration the necessity of manipulating multiple sources, the requirements for navigation, as well as requirements for a good taxonomy. The guidelines we developed were based on the issues encountered in the developing the taxonomy.

The findings of the study and the solutions implemented are limited to some extent to the context of our study. For example, due to the domain the study, we made use of partial schedules of *DDC*. Also, we used other knowledge organization systems such as domain taxonomies to complement the resources of *DDC* and domain thesauri. We did not cover ontologies. For example, the GEM ontology seems to be a good starting point for collecting terms and term relationships in the field of education. The findings about using *DDC* and the domain thesauri, as well as the steps and guidelines for constructing the hierarchical structure and categories, need to be validated in other types of organizations, knowledge domains, and using other knowledge organization systems.

The prototype of taxonomy was implemented in the University's e-learning platform using a taxonomy software, TLE-Equella version 3, to support navigation. The evaluation of the effectiveness of the taxonomy for supporting end-users' navigation has been conducted to identify the problems of the taxonomy construction steps. The evaluation methods we used and issues found in the evaluation results will be reported in a separate paper.

References

Aitchison, Jean et al. 2000. *Thesaurus construction and use: A practical manual.* (4th ed.). London: Aslib IMI.

Barnard, Chester. I. 1938. *The functions of the executive*. Cambridge: Harvard University Press.

Batley, Susan. 2005. *Classification in theory and practice.* Oxford: Chandos Publishing.

Bertolucci, Katherine. 2003. Happiness is taxonomy: Four structures for Snoopy. *Information outlook* 7n3: 36-44.

Cheuk, W-Y B. 2002. Real-world taxonomy development: Creating a taxonomy that makes sense to your employees. *InsideKnowledge* 5n6. Available http://communication.sbs.ohio-state.edu/sensemaking/art/artabscheuk02kmtaxon.html (accessed July 20, 2007).

Choksy, Carol E. B. 2006. 8 steps to develop a taxonomy. *Information management journal* Nov/Dec. Available http://findarticles.com/p/articles/mi_qa3937/is_200611/ai_n16871474 (accessed December 4, 2006).

Cisco, Susan. L. and Jackson, Wanda. K. 2005. Creating order out of chaos with taxonomies {Electronic version}. *Information management journal* May/June: 45-50. Available from EBSCO Academic Search Premier database (accessed December 4, 2005).

Conway, Susan and Sligar, Char. 2002. Chapter 6. Building taxonomies. In Susan Conway and Char Sligar. *Unlocking knowledge assets*. Redmond, Wash.: Microsoft Press, pp. 105-24.

Dickson, Ian. 2008. *Taxonomy: Some guidelines for effective design of taxonomies*. Available http://drupal.org/node/81589 (accessed January 31, 2009).

Hunter, Anthony. 2000. *Taxonomies*. Available http://www.cs.ucl.ac.uk/staff/a.hunter/tradepress/tax.html. (accessed July 10, 2005)

Iyer, Hemalata. 1995. *Classificatory structures: Concepts, relations and representation*. Verlag, Frankfurt/Main: INDEKS.

Lambe, Patrick. 2007. O*rganizing knowledge: Taxonomies, knowledge and organizational effectiveness.* Oxford: Chandos Publishing.

McAuley, John et al. 2007. *Organization theory: challenges and perspectives.* Harlow, England; New York: Prentice Hall/Financial Times.

McGregor, Bruce. 2005. Constructing a concise medical taxonomy. *Journal of the Medical Library Association* 93: 121-23.

Pack, Thomas. 2002. *Taxonomy's role in content management*. Available http://www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=867 (accessed July 10, 2005).

Raschen, Bill. 2006. A resilient, evolving resource: How to create a taxonomy {Electronic version}. *Business information review* 22: 199-204. Available from Sage Publications database (accessed August 15, 2006).

Roberts-Witt, Sarah. L. 2000. *Practical taxonomies*. Available http://www.destinationKM.com/articles/default.asp?articleid=684 (accessed July 15, 2005).

Saeed, Hamid and Chaudhry, Abdus. S. 2002. Using Dewey Decimal Classification Scheme (*DDC*) for building taxonomies for knowledge organization. *Journal of documentation* 58: 575-84.

Schein, Edgar H. 1970. *Organizational psychology* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Sharma, Ravi. S. et al. 2008. Developing corporate taxonomies for knowledge auditability: A framework for good practices. *Knowledge organization* 35: 30-46.

Singhal, M. and Nath, S. S. 2008. Chapter 27. An approach toward taxonomies. In I V Malhan and Shivarama Rao K ed. *Perspectives on knowledge management*. Lanham, Maryland: Scarecrow, pp. 391-406.

Vizine-Goetz, Diane. 2002. Classification schemes for Internet resources revisited. *Journal of Internet cataloging* 5n4: 5-18.

Wang, Zhonghong et al. 2006. Potential and prospects of taxonomies for content organization. *Knowledge organization* 33: 160-69.

Wang, Zhonghong et al. 2008. Using classification schemes and thesauri to build an organizational taxonomy for organizing content and aiding navigation. *Journal of documentation* 64: 842-76.

Wyllie, Jan. 2005. *Taxonomies: Frameworks for corporate knowledge* (2nd ed.). London: Ark Group, in association with Inside Knowledge.

Appendix:
Sources of the Information Studies Taxonomy

Cheung, C. F. et al. 2005. A multi-facet taxonomy system with applications in unstructured knowledge management {Electronic version}. *Journal of knowledge management* 9n6: 76-91. Available from Emerald Full text database (accessed August 14, 2006).

Doke, E Reed. and Barrier, Tonya. 1994. An assessment of information systems taxonomies: Time to be re-evaluate? *Journal of information technology* 9: 149-57.

Education Resources Information Center. U.S. Department of Education. *ERIC Education Thesaurus*. Retrieved December, 2006 from the Education Resource Information Center database. Access provided by Nanyang Technological University.

Hawkins, D. T. et al. 2003. Information Science Abstract: Tracking the literature of information science. Part2: A new taxonomy for information science {Electronic version}. *Journal of the American Society for Information Science and Technology* 54: 771-81. Available from ProQuest ABI/INFORM Archive Complete database (accessed July 10, 2005).

IFLA. Education and Training Section. 2000. *Guidelines for Professional Library/Information Educational Programs – 2000*. Available http://www.ifla.org/VII/s23/bulletin/guidelines.htm (accessed November 3, 2006).

Mentzas, G. 1994. A functional taxonomy of computer-based information systems {Electronic version}. *International journal of information management* 14: 397-410. Available from ScienceDirect database (accessed July 12, 2005).

Proquest – CSA Social Science. *LISA library and information science thesaurus*. Retrieved December, 2006 from the Library and Information Science Abstracts database. Access provided by Library of Nanyang Technological University.

Redmond-Neal, Alice and Hlava, Marjorie. M. K. 2005. *Thesaurus of information science, technology,*

and librarianship, 3rd ed. Medford; New Jersey: Information Today.

OCLC. 1978. *Web Dewey*. Available http://connexion.oclc.org/ (accessed December, 2006). Access provided by WKW School of Communication & Information, Nanyang Technological University.