

The Future of General Classification

Jens-Erik Mai

SUMMARY. Discusses problems related to accessing multiple collections using a single retrieval language. Surveys the concepts of interoperability and switching language. Finds that mapping between more indexing languages always will be an approximation. Surveys the issues related to general classification and contrasts that to special classifications. Argues for the use of general classifications to provide access to collections nationally and internationally. [Article copies available for a fee from The Haworth Document Delivery Service: 1-800-HAWORTH. E-mail address: <docdelivery@haworthpress.com> Website: <<http://www.HaworthPress.com>> © 2003 by The Haworth Press, Inc. All rights reserved.]

KEYWORDS. Interoperability, switching languages, general classification, special classification

Jens-Erik Mai, PhD, is Assistant Professor, Information School, University of Washington, Mary Gates Hall, Suite 370, Box 352840, Seattle, WA 98195-2840 USA (E-mail: jemai@u.washington.edu).

[Haworth co-indexing entry note]: "The Future of General Classification." Mai, Jens-Erik. Co-published simultaneously in *Cataloging & Classification Quarterly* (The Haworth Information Press, an imprint of The Haworth Press, Inc.) Vol. 37, No. 1/2, 2003, pp. 3-12; and: *Knowledge Organization and Classification in International Information Retrieval* (ed. Nancy J. Williamson, and Clare Beghtol) The Haworth Information Press, an imprint of The Haworth Press, Inc., 2003, pp. 3-12. Single or multiple copies of this article are available for a fee from The Haworth Document Delivery Service [1-800-HAWORTH, 9:00 a.m. - 5:00 p.m. (EST). E-mail address: docdelivery@haworthpress.com].

<http://www.haworthpress.com/store/product.asp?sku=J104>
© 2003 by The Haworth Press, Inc. All rights reserved.
Digital Object Identifier: 10.1300/J104v37n01_02

INTRODUCTION

The exchange of bibliographic records is common and the interest in interoperability is growing. The latter is the notion that users from one interface and using one search language will be able to search multiple databases and collections at the same time. Exchange of bibliographic records and interoperability is of interest, not only because of economic advantages, but also because of the possibilities of increasing the exchange of ideas across domains and the possibilities of more exhaustive search results.

For the purpose of analyzing the issues related to the development of such future information systems it might be sensible to separate the issues related to descriptive cataloging from the issues related to subject representation. This paper will treat the issues related to subject representation and will furthermore focus on the use of general classification schemes for accessing documents across domains and collections.

The core idea explored here is that if documents in different collections were organized and represented with a common classification system, then access to the material would be enhanced since the documents on the same subject matter would be classified under the same entry across all collections. This idea is far from new. In fact, this is the core of Dewey's "invention" of the decimal classification system by which he envisioned that all knowledge could be classified. Francis Miksa¹ explains:

There can be little doubt that Dewey fully intended to make an extensive, hierarchically arranged scheme, which covered all knowledge, a scheme that could be used in libraries to organize books according to the knowledge the books contained.

The real benefit of Dewey's invention was that the same books at different libraries would have the same call numbers and that it would be easier to navigate different libraries since they would be organized in the same way. This ensured that a person searching for books about a particular subject could be sure that the books would be classified under the same entries in any library.

If those libraries were connected in a network a user could, in fact, retrieve the bibliographic records for books on a particular topic even though only a fraction of the books were included in a particular library's collection and the rest were scattered in a number of libraries around the country—or even around the world.

This is not a novel idea. Dorothy May Norris² explains how English Franciscan monks in the late 13th century created a shared catalog, *Registrum Librorum Angliae*, of collections in 183 English monastery libraries. The monks traveled from monastery to monastery and used the shared catalog to keep informed

about what works were available at the different libraries that they planned to visit.

The traditional suggestion to solve the problem of interoperability has been to create a switching language to translate back and forth between different classification systems. Although this has generally been considered a very difficult task, the idea is gaining new attention, especially since the Internet has provided opportunities to search in more information retrieval systems simultaneously.

This paper reviews the applicability of subject interoperability and evaluates the purpose of special and general classification schemes, and suggests that the purpose of general classification schemes is to facilitate interoperability among different information retrieval systems and that the purpose of special classification schemes, on the other hand, is to give access to the material at a greater level of exhaustivity and specificity.

INTEROPERABILITY

There is a sense in the information science community that, with the advent of the World Wide Web (WWW), the previous models for information use have become obsolete. William Moen³ states that “The convergence of computer and communications technologies of the late 20th century has profoundly affected information creation, distribution, access, and use,” and this has caused a change in how people access information; he continues: “Few would deny that the Internet, primarily through the World Wide Web, has increased users’ expectations for access to information.”⁴ Reflecting on personal experience, it is correct that many of us have changed our information behavior lately. In researching for this paper, for instance, I frequently searched different catalogs and databases for relevant literature right here at my desk and I printed the articles on the printer right next to me, and there is no doubt that I tend to expect that this is possible today. In that sense, one could say that the access to the material has changed significantly with the advent of these technologies. However, the meaning of the term access is crucial here. The way I have used access above has been to talk about how I gained access to the physical documents via a network connection and printed on a local printer. Nevertheless, the real question for the information profession is how these technologies have changed the way we provide access to the documents in terms of search facilities. That is, how has the advent of network technologies changed the way we build and use search mechanisms? In this sense, access refers to the knowledge organization and information retrieval tools that are used.

The creation of the WWW and networks has brought forth the idea of *interoperability*. This concept is used and referred to in many places in the lit-

erature and in current practice, but the exact definition of the term is still discussed. Arms et al.⁵ offer this definition:

The goal of interoperability is to build coherent services for users, from components that are technically different and managed by different organizations. This requires agreements to cooperate at three levels: technical, content, and organizational.

Of these three levels, the content level is of special interest here. This level covers “data and metadata, and includes semantic agreements on the interpretation of the information.”⁶ The descriptive representation of documents has received much attention among interoperability researchers, whereas subject representation is neglected, overlooked, or regarded as unproblematic.

Traditionally, there have been two avenues to ensure some degree of subject interoperability across different collections, as indicated by Lancaster⁷ (citing Neelameghan):

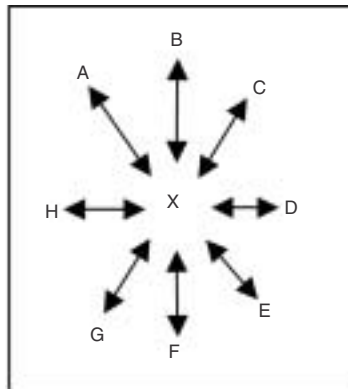
1. To use an intermediate language or switching language through, or by, which one moves from one information system to another.
2. To use the same or very nearly the same information storage and retrieval language in all the information systems.

There has been much research into the problems related to switching languages and compatibility of different indexing language systems. Dahlberg⁸ has produced an annotated bibliography that covers 450 entities and 40 years of work in the area. Switching language is the idea that a user could use the same indexing language to search in information systems that are indexed with different indexing languages. Lancaster explains, “A user could sit at a terminal in center A and enter terms from A’s vocabulary. The terms would be converted to the codes of X and switched to the appropriate terms in, say, D’s vocabulary.”⁹

The basic idea of a switching language is that the user would only have to learn one retrieval system and one controlled vocabulary and that the user would be able to use that one system to search a range of different systems simultaneously. For this to be possible, each of the indexing languages would have to be mapped against the common switching language (denoted as X in Figure 1). It would be rather simple to add another indexing language to the list of indexing languages that can be switched between them, and there is no limit to the number of indexing languages that could be attached to the switching language.

It must, however, be expected that “a reasonable level of commonality in subject matter . . . exist among centers; otherwise there would be little point in the activity.”¹⁰ It would be very difficult to switch between a thesaurus on

FIGURE 1. A Switching Language



physics and a thesaurus on sociology, whereas it would be easier to switch between a number of thesauri on either of the subjects fields.

With modern information technologies the work involved in mapping between the indexing languages and the switching language is reasonable. However, the main problem related to switching languages goes beyond technology and labor and is really a problem that is buried in language and meaning. The main problem can be summed up to that fact that there very seldom exists an indexing term Q in an indexing language A that has the exact same meaning as an indexing term Q' in indexing language B , even for indexing languages that cover the same subject field. In other words, the mapping will always be an approximation.

There are three main reasons for this:

1. Pre-coordination. The levels and kinds of pre-coordination vary from indexing language to indexing language, so a concept might not be expressed using the exact same coordination of terms in two indexing languages.
2. Hierarchical structure. The meaning of any indexing term in an indexing language is determined by the term's place in the hierarchical structure, so even if the same term exists in two indexing languages the meaning of the terms might not be the same.
3. Absence of concepts. The level of specificity might vary from indexing language to indexing language, so concepts that are expressed in one indexing language are not expressed in the other indexing language.

In an international environment, where more natural languages are involved, the absence of concepts takes another twist. It is only seldom that there exist

one-to-one relationships between words in different natural languages. Birger Hjørland,¹¹ following Hjelmslev, gives the example in Figure 2 of the words “tree,” “wood,” “woods,” and “forest,” which shows that these words do not have one-to-one counterparts across six different natural languages.

The consequence for this is that switching between indexing languages in different natural languages will fail because the meanings of the words are determined from the use of the words in particular contexts. In other words, the German word “wald” can be translated into either “woods” or “forest” depending on the context. It is obvious that a single tree is not a forest, but when is a woods large enough be a forest? A person translating German into English will have to make a decision about when to use the English word “woods” and when to use the word “forest”; a decision that is not based on the actual German words but on the use and context.

The three aforementioned reasons suggest that it would only be possible to switch between systems that are fairly close in subject coverage and classificatory structure. It might therefore be worthwhile to consider the second avenue to ensuring interoperability between different collections, namely that of using the same indexing language for multiple collections. This takes us back to Dewey’s idea about the classification of the universe of knowledge that I refer to in the introduction.

GENERAL AND SPECIAL CLASSIFICATION

The basic idea behind general classification is that a single classification covers the entire universe of knowledge, that is, all subject fields. There are today a number of such systems in use across the world. Some of these are common and used in many libraries, some are widely known but not used in many (or any) libraries, and some are constructed specifically for one particular library. There exist, furthermore, a number of special classification systems that only cover a part of the universe of knowledge, in the sense that these systems focus on specific subject fields.

FIGURE 2. Cultural Relativity in Word Meanings

English	German	Danish	French	Italian	Spanish
	Baum	Træ	Arbre	Albero	Arbol
Wood	Holz		Bois	Legno	Lena Madera
Woods	Wald	Skov		Basco	Basque
Forest			Foret	Foresta	Selva

The distinction between general and special classification is generally taken to be that, “A general classification covers the whole of being; a special classification some parts of it.”¹² However, this distinction is not the most helpful because any “special” classification scheme will, in reality, provide for a broad range in subject matter. As Derek Langridge explains: “There is no such thing as a special library in this sense: all special libraries contain a range of subjects in addition to their central interest.”¹³ It might be more helpful to think of any classification scheme as a classification scheme that has a special purpose.

In spite of the fact that there exist a number of general classification systems, the nature and status of general classification systems have been discussed in the literature and it has often been concluded that the construction and use of such general systems is unattainable. Brian Vickery¹⁴ lists five reasons why existing general classification systems are unsatisfactory:

1. Most of them do not give adequate detail for accurate specification of the highly complex subjects in papers and reports that documentation must handle today.
2. Despite the comprehensiveness and variety of certain general schemes, they do not fully cater to the special viewpoints of each particular library and information center.
3. Even if they are varied in viewpoint, they do not sufficiently provide for the flexible combination of terms which highly specific subject headings demand.
4. Even if flexible, they achieve such flexibility only by unnecessarily lengthy or complicated notational means.
5. They fail to give optimum helpfulness in filing order.

The general criticism here is that the general classification system fails to represent the documents at a level of specificity that is required or desired by the users. This is related to the notion of “the force of discrimination” (as David Blair¹⁵ has recently named it), which dictates that the intellectual content of a document should be represented and described in as much detail as possible for any given information retrieval system. The idea is that if documents in a particular information retrieval system are described and represented as fully and specifically as possible, then the users of the system would be able to query the system for exactly those documents that are relevant to the user at that particular time.

It is a classic recommendation in the classification and indexing literature that documents should be represented as specifically as possible; it dates back to Cutter’s recommendation of the *specific entry*.¹⁶ Jack Mills¹⁷ stated very clearly that “a classification should provide maximum detail in specification, both in the enumeration of isolates within categories and in facilities for combing these to

indicate compounds.” All leading textbooks and guidelines on indexing and classification follow this recommendation and advocate that the intellectual content of documents should be represented as specifically as possible.

However, the specificity of any particular document is relative to the context and use of that document. In other words, for any document it would only be possible to identify the specificity of that document relative to a particular user group, time and place, i.e., the purpose of the classification system will determine the interpretation and representation of the documents for that system.

A general classification will not have a single purpose, since the classification is not serving a particular domain or particular interests, and will therefore, in effect, not serve any purpose. As Broadfield¹⁸ says, “a general purpose classification is likely to display no purpose at all.” He continues by saying that classification for general purposes is impossible:

While it is possible to classify the parts of medical science as specific ways in which the general medical purpose is advanced, and to classify the parts of mathematics, economics, or anthology on the same principle, to arrange these four sciences so that they all contribute to one purpose is impossible without knowing to what purpose they do contribute.¹⁹

A few pages later, Broadfield²⁰ calls on the advocates of general classification schemes to set forth the purposes that the general classifications serve,

On those, therefore, who wish to find a theory of general classification the onus is placed of showing that there is a general purpose, that it takes various forms related as the species to the genus, and that “characteristics” essential to the working out of such a purpose throughout the system can be found.

Although special classifications only treat parts of the universe of knowledge, that treatment is relative to a specific understanding of the whole universe of knowledge. A general classification, on the other hand, assumes no special understanding or treatment of the universe of knowledge and de facto claims a neutral and objective representation of the universe of knowledge. It has been shown in numerous papers that any classification is, in fact, biased and it is generally accepted that classifications cannot be neutral and objective.²¹

The distinction between special and general classifications is that special classifications serve a specific purpose and will explicitly incorporate that specific purpose in their structure and in their representation of documents using the systems. The systems and the representation of documents will aim at being as specific as possible since this will enhance and ease the retrieval of documents. General classifications, on the other hand, have as their purposes to structure and represent potentially all subjects, and the problem in using these

systems is that any document could potentially be represented in a number of ways using the system depending on context and potential use.

CONCLUSIONS, DISCUSSION AND IMPLICATIONS

There is no doubt that the future of libraries includes providing access to collections that are remote—physically, culturally, and subject wise. Also, there is no doubt that there is a pressure to facilitate access to these remote collections using modern information technology. The real question for the classification and indexing community is how these demands will and should be met.

This paper has reviewed some of the core problems related to two possible solutions to the demands, namely, that of creating switching languages and providing access to documents using a range of indexing languages and providing a mechanism to search across a number of collections simultaneously, and that of using a general classification scheme to provide access to the whole universe of knowledge via one single classification. I have shown that both of these solutions are theoretically problematic and in practice unattainable.

Subject representation is tied to the purposes, cultural, and contextual circumstances in which the representation is produced. This creates problems when the subject representation is exported from one country to another, especially if the subject representation is exhaustive and specific.

However, the international exchange of bibliographic records and the interoperability across information retrieval systems will be possible if general classification systems are used in conjunction with special indexing systems. The general classification systems will then serve as the broad organization of knowledge and the sciences, and the special systems will serve as the domain oriented organization and presentation of documents.

The purpose of general classifications will then not be to represent the documents as specifically as possible but to represent the documents for the purpose of stating which domain and sub-domains that the documents potentially could be used in. Such systems will not be useable for retrieving documents on specific subjects but will be very helpful in navigating the international production of knowledge. The upshot of this suggestion is that the documents need to be represented for two quite different purposes, namely, for a potential worldwide audience and for the local users. The global users are served best if the material is represented using two different classification systems—a general and a special.

The future task for classification research is to explore the theoretical foundations and principles for the construction and use of general classification systems that serve a worldwide audience for the purpose of organizing knowledge and the sciences.

REFERENCES

1. Francis L. Miksa, *The DDC, the Universe of Knowledge, and the Post-Modern Library*. Albany, NY: Forest Press, 1998.
2. Dorothy May Norris, *A History of Cataloguing and Cataloguing Methods*. London: Grafton, 1939. pp. 30-34.
3. William E. Moen, "Interoperability for information access: Technical standards and policy considerations," *Journal of Academic Librarianship*, 26: 2 (2000): 129.
4. *Ibid.*
5. William Y. Arms, Diane Hillmann; Carl Lagoze, Dean Krafft, Marisa Richard, John Saylor, Carol Terrizzi, and Herbert Van de Sompel, "A Spectrum of interoperability: the site for science prototype for the NSDL," *D-Lib Magazine*, 8: 1 (2002). <http://www.dlib.org>.
6. *Ibid.*
7. F. W. Lancaster, *Vocabulary Control for Information Retrieval*, 2nd edition. Arlington, VA: Information Resources Press, 1986. p. 80.
8. Ingetraut Dahlberg, "The compatibility guidelines—a re-evaluation" in *Compatibility and Integration of Order Systems: Research Seminar Proceedings of the TIP/ISKO Meeting, Warsaw, 13-15 September 1995*. Warsaw: Wydawnictwo, 1996. pp. 32-45.
9. as 7, p. 189.
10. as 7, p. 190.
11. Birger Hjørland, "Principia informatica: foundational theory of information and principles of information services," in *Emerging Frameworks and Methods: Proceedings of the Fourth International Conference on Conceptions of Library and Information Science*. Greenwood Village, CO: Libraries Unlimited, 2002. p. 113.
12. W.C. Berwick Sayers, *An Introduction to Library Classification, Theoretical, Historical and Practical, with Readings, Exercises and Examination Papers*. London: Grafton, 1944. §118.
13. Derek W. Langridge, "Alternative starting points in classification," *Cataloging & Classification Quarterly* 19: 3/4 (1995): 7.
14. Brian C. Vickery, *Faceted Classification: A Guide to Construction and Use of Special Schemes*. London: Aslib, 1960. p. 7.
15. David C. Blair, "The Challenge of commercial document retrieval, Part I: Major issues, and a framework based on search exhaustivity, determinacy of representation and document collection size," *Information Processing and Management* 38 (2001): 273-291.
16. Charles A. Cutter, *Rules for a Printed Dictionary Catalog: US Bureau of Education, Special Report on Public Libraries*. Washington DC: US Government Printing Office, 1876.
17. Jack Mills, "Inadequacies of existing general classification schemes," in *Some Problems of a General Classification Scheme: Report of a Conference Held in London, June 1963*. London: Library Association, 1964. p. 33.
18. A. Broadfield, *The Philosophy of Classification*. London: Grafton, 1946. p. 87.
19. *Ibid.*
20. as 18, p. 89.
21. Hope Olson, "Mapping beyond Dewey's boundaries: Constructing classificatory space for marginalized knowledge domains," *Library Trends* 47: 2 (1998): 233-254.
22. Paul Miller, "Interoperability: What is it and why should I want it?" *Ariadne* 24 (2000), <http://www.ariadne.ac.uk/>.