Towards a Conceptual Model of Knowledge Organization Systems:

1. The Functions of Knowledge Organization Systems

Jonathan Rochkind

LIS 535

12 December 2005

Modern library science included at it's beginnings a comparative discussion of function in classification. Cutter's influential 1876 essay on "Library Catalogs" compares the functions served by a classified or systematic arrangement of subjects or topics in a catalog directory to those served by an alphabetic arrangement. This discussion became ossified in a general idea of two types of knowledge organization systems in the library: the classification, and the alphabetic index. In fact, each of these two types always collected together decisions structural, semantic, and representational features as well as ideas about function which do not need to determine each other, but instead represented tendencies arising in a particular environment.

The changes to the environment in which knowledge organization operates have made the model based on these two tendencies of less value. The browsing related functions of 'classification' take on a new prominence in the online environment, used in different ways. New types of knowledge organization have arisen to meet new functions in new ways and new environments. Serving different functions in different contexts may require different features and structures of a knowledge organization system. Yet the library and information science literature has not updated it's conceptual model of classification and knowledge organization. To give us a coherent framework for thinking

about knowledge organization and how to apply and improve it in various contexts, a new systematic conceptual model of knowledge organization systems is required, relating the possible functions knowledge organization systems may play for the user to the nature of particular knowledge organization systems.

This paper will begin by analyzing the received understanding of the distinction between classification and alphabetic index, in particular from Cutter's "Library Catalogues" essay. Next, I will synthesize an updated model of the functions of knowledge organization systems from the available literature. Finally, I will discuss the implications of this model of functions on a general conceptual model of knowledge organization systems, and the role of such a general model.

Received Understanding: Classification vs. the Alphabetic Index

Much of what discussion exists on functions of classification make the distinction between 'classification' and 'alphabetic (subject) index'. This common distinction represents two tendencies of actual knowledge organization systems which had become common in libraries at the midpoint of the 20th century. The first, 'classification' systems, as exemplified by the Dewey Decimal Classification or the Library of Congress Classification, are hierarchical systems with alphanumeric notation designed to allow use as a locating and a shelf order device, meant to place items on the shelf in proximity to other 'similar' items, and using discipline as a primary characteristic of division. Alphabetic subject indexes, exemplified by the Library of Congress Subject Headings, are simple lists of topical subject headings, with no arrangement among or relationship between terms beyond the alphabetical list, used for finding materials on a given topic.

An example of this distinction is provided by Vickery (1971), where in an article with very similar goals to those of this paper, he begins his discussion by writing:

> Initially, two functions [of retrieval languages] were recognized: to identify, and hence to select, items on a specific topic (the task of the alphabetic index); to group in proximity items on similar subjects (the task of classification).

Vickery calls the first function "specific reference" and the second "general survey." By "subjects" he appears to mean disciplinary-based categories (as a later example, he uses "Parasitism"), as opposed to "topic" by which he appears to mean something closer to 'aboutness'. We can immedialtely identify some conceptual ambiguities within this idealized binary model of knowledge organization systems and their functions. First, the idea that 'classification' systems are only for grouping and arrangement, and consequently for 'browsing'-like activity, and never for identification and selection. Of course, anyone that has spent some time observing reference in a library knows that librarians will respond to topical 'specific reference' questions with directions to a particular classification range. In fact, even Vickery forgets this point, later without comment changing his definition of the function of 'general survey' to "provide for the selection of a group of records lying within a certain subject field."

There is also an assumption that users are only interested in browsing or 'surveying' based on disciplinary characteristics, and are only interested in selecting a subset of documents for a 'specific' well-defined need on the basis of pure 'topic'. There is the assumption that the characteristics of distinction in a 'classification' are different ("subject" or disciplinary) than they are in an alphabetic index ("topic"). In fact both traditional classifications and alphabetic indexes contain some overlap in their characteristics of interest, both containing 'aboutness' information, as Chan (1990) writes

"Like subject headings, classification also serves as a device for representing the subject content of library materials." Also, both traditional classifications and subject indexes contain information and allow retrieval on characteristics that are neither disciplinary nor 'aboutness', including as noted by Bland (1990), various 'formal and purposive' aspects including form, genre, method of approach (e.g. fictional, criticism, etc.), and 'intended audience' (e.g. children, undergraduates, popular adult treatment, experts, etc.). Cutter (1876, p. 535) writes that some "alphabetically" organized catalogs recommended including genre headings as far back as 1856.

Cutter's Original Explication

Charles Cutter's seminal 1876 essay on "Library Catalogs" is very instructive in exploring the origin in practice and theory of this distinction between classified and alphabetic systems. American libraries originally provided subject access with a classed catalog, where entries were organized in categories arranged by their relation to each other (Chan 1990). But by 1876, many American public libraries were instead arranging their catalogs alphabetically, and Cutter provides an argument in favor of the alphabetical approach. He discusses the issue in terms of functions, providing what is still the basis of much of the discourse on function, clearly influencing Vickery's treatment.

Cutter admits there are advantages of the classified arrangement of a catalog. Such an arrangement is of use to "one who is pursuing any general course of study [who] finds brought together in one part of the catalog most of the books that he needs". Again, we point out that this is true only if the classified catalog is arranged by discipline ("course of study"). Cutter notes that was indeed the organizing principle of existing

classed catalogs, but we note that it is possible have a systematic (rather than alphabetic) arrangement of classes based on other characteristics as well. Cutter points out another value of the classed catalog is that it allows a kind of browsing or exploration, even when the user's inquiry is in fact 'topical', in that the user "sees not merely books on the particular topic for which he is interested, but in it's immediate neighborhood works on related topics, suggesting to him courses of an investigation which he might otherwise overlook." (p. 529). As such a strategy is often necessary even for simple topical searching, not just for surveys based on 'courses of study', the lack of related topic guidance function is a prime disadvantage of the alphabetic arrangement: If the library does not have a book on exactly the subject the inquirer wants, he might be served by a work covering a more general topic, or a more specific particular aspect of his topic, but "he will get no help whatsoever from any dictionary catalogue yet made" in finding these related classes, but "must trust to his own knowledge of the subject… to guide him." (p. 532-533). On the other hand, the main disadvantage of the classified catalog is that it requires some expertise to even know where to look, "a man may want a book on the badger without being much of a naturalist, but he could hardly find it in such a catalogue unless a naturalist should help him." Cutter decides that the classified catalog serves mainly those "who want to make a thorough study of some specific subject," or "who want to study fully some general class of subjects," but the alphabetic catalog serves "those who want something quickly." (p. 541). Because there are choices of how to arrange the classified catalog, none of which can serve everyone (p. 541), and presumably because he thinks the majority of inquiries are of the "want something quickly" variety, Cutter recommends the alphabetic arrangement.

Cutter does admit that an ideal solution would be to have both a classed and an alphabetic catalog, but in his 19th century context that was an unreasonable expense. Oddly, despite the considerable advantages he identified to the classed arrangement, he also concludes, without explanation, that such a double catalog would also "not perceptibly increase the practical value of the catalog" (p. 543) even were it feasible. Note well that Cutter is explicitly talking only about *arrangement* of the printed catalog, not about entirely separate systems. In our computer era, multiple arrangements can easily be provided for the same controlled vocabulary. The complete schedule of a traditional 'classification' (with classified arrangement) can easily be keyword searched, and if there are suitable term headings for each class in the classification, an alphabetical arrangement can easily be displayed as well. An alternate arrangement to a classified system could be provided in Cutter's time via an alphabetical index, which indeed the Dewey Decimal Classification provided, but Cutter mysteriously dismisses that possibility by writing vaguely that it was "not occurring or not pleasing those who were dissatisfied by classification" (p. 532).

Development and Decline of Classification as Opposed to Alphabetic Indexing

There is no reason that this discussion of two methods of arranging a printed listing for *display* had to result in two entirely separate *systems* of knowledge organization, based upon different characteristics of selection and including very different features. But for whatever practical historical reasons, it did. In fact, the traditional *classification* became used *only* for arranging books on the shelf, and *not* for serving any of those other finding functions Cutter identifies as its advantages, except

insofar as shelf browsing of open stacks can provide. Chan (1990) notes that "the classed catalog has all but disappeared from the American libraries over the last century," and that the Library of Congress classification was "designed and continues to be maintained solely as a shelf location and browsing device, nothing more." Dewey (1876) interestingly claimed that his classification system "was devised for cataloguing and indexing purposes," but was "found on trial to be very valuable for numbering and arranging books and pamphlets on the shelves." This second "accidental" use is indeed what the DDC had come to be maintained exclusively for, and despite the continued existence of the alphabetic 'relative' index to the classification, Markey (1986) found that the system would require significant changes to be ideal for subject access.

This division between *classification* maintained as a device for shelf location and (useful, helpful) arrangement, while the structurally simpler list of 'alphabetic' subject terms is used for subject access, is not, therefore, a necessary consequence of the natures of these two systems, but essentially an accident of history where each system came to occupy a certain ecological niche in actual practice, in the context of a certain technological and organizational environment. E.J.Coates (1978) reports that British libraries used a classified catalog into the 20th century, but nonetheless, discussion function of classification and it's relation to structure, relied on this distinction even in the thoughts of Vickery (1971), British himself.

This grouping of knowledge organization functions and structures between *classification* and *alphabetic index* began to break down when the technological environment that produced it changed with the introduction of computer retrieval. Alphabetic languages like LCSH began to acquire some of the structural features of

traditional classification, like extensive hierarchical relationships. Some proposed that certain features of traditional classification become not only more feasible, but especially useful, in an online environment for retrieval, and for online (as opposed to physical) browsing. Martha West (1984) for example suggests that "There are a number of reasons why this separation of function between indexing and classification may no longer be valid." Additionally, new varieties of knowledge organization systems, like thesauri, taxonomies, and ontologies were brought into existence, in large part to serve specific functions (Gilchrist 2003). The traditional distinction between alphabetic indexing and classification, never very systematic even at its origin in that it lumps together a variety of choices on both characteristics of division and structural features which need not determine each other, now serves especially poorly as a framework for understanding the different possible functions and structures of controlled vocabularies in the 21st century environment.

Consequently, I suggest that a new model of the functions of knowledge organization systems is required, initially treating all types of knowledge organization (classification, alphabetic index, taxonomy, thesaurus, etc.) collectively. The Classification Research Group (1955) noted that "some writers consider these techniques to be alternative ways of carrying out the same process, differing only in scope and effectiveness. Others consider them to be mutually exclusive, each suitable for only one aspect of information retrieval." In treating all these types together, I am not coming down on the first side, but instead suggesting that the functions filled by each are overlapping and not yet clearly defined. No doubt different species and genuses of knowledge organization systems accomplish different functions with differing degrees of

success, but we must first identify those functions before we can distinguish a typology of knowledge organization in a systematic way.

Discussions of Function in the Recent Literature

Systematic analyses of function are sparse in the recent literature. Vickery's 1971 article already discussed treats this topic more centrally than any other CRG paper found. While it was considered important enough to provide as a discussion paper for a symposium in 1997 revisiting the work of the CRG, I could find no citations to the paper in the literature. Another CRG paper (Mills 1969) provides a list of 'inadequacies' or 'issues' in classification systems, many of which are actually expressed in terms of function. The CRG generally treated what here is called 'knowledge organization systems' as a whole, with Vickery referring to the overall category as "retrieval languages" (later "information languages" in Coates 1988), and Mills meaning the same thing by "General classification schemes".

In the 1980s, a sub-genre of LIS literature emerged to reevaluate the use of traditional 'classification' in the emerging online environment. This task often required re-addressing the functions and roles of classification, although sometimes only peripherally and almost accidentally. Significant examples include Svenonius (1983, 1986) and Chan (1986, 1990) who both address the topic of "uses of classification in an online environment",, and additionally restrict their attention to traditional 'classification' specifically, rather than a more abstract treatment of the functions of knowledge organization in general. Nonetheless, since identifying uses of classification in the new online environment requires one to at least implicitly challenge the received wisdom

about the traditional role of 'classification' vs. 'alphabetic index,' this body of LIS literature provides the main identifiable contemporary treatment of functions of knowledge organization. In particular, among other concerns, both Svenonius and Chan explore the possibilities of using existing 'classification' systems for functions more traditionally thought to be filled solely by the 'alphabetic index'. An additional notable example in this area is the extensive study by Markey (1986) of uses of the DDC system specifically in an online environment, where functions DDC can serve online are identified and a possible implementation is tested empirically.

The most useful recent existing treatment of the functions of classification I found is in a 1999 'grey literature' report titled simply 'Final Report', for which the party responsible is bureaucratically identified as the Subcommittee on Metadata and Classification of the Subject Analysis Committee of the Cataloging and Classification Committee of the Associations for Library Collections and Technical Services, a division of the American Library Association. This goal of this committee, whose report will be identified in this paper as ALCTS (1999), was to "study the application of classificatory frameworks… as metadata for digital resources," specifically web pages. They found that in order to "develop criteria for evaluating the application of classification to organize Web resources," they first had to identify the functions of classification. To this end, they provide a list of functions, apparently devised from whole cloth by the committee itself, which serves as an excellent general outline of roles and functions of classification. Their list of functions was intended to be general, not initially restricted to those functions applicable in the online environment (in fact, they decide that one of their initial functions is 'outmoded' in the context of Web resources as cataloged objects), and

their perspective in 1999, well into the emergence of the digital networked environment, allows them to look at the functions of classification removed from the restrictions of the traditional 'classification' vs. 'alphabetic index' distinction.

From this literature I have assembled an initial synthesized list of general functions of knowledge organization systems.

A Synthesized Simple Taxonomy of Functions of Knowledge Organization Systems

1. Class Retrieval

In a broad sense, all knowledge organization systems are arguably intended for retrieval, ultimately for pointing to the user toward documents of interest more efficiently than if he were to browse through a random ordering of documents. (Mills 1969). This statement has sometimes stopped the inquiry into functions of knowledge organization systems in it's tracks before it has begun. However, as we will see, there are many distinct avenues toward this ultimate goal (and possibly other goals) that are worthwhile to identify. What we identify here as the function of *Class Retrieval* indicates a specific method of retrieval, involving collocating items belonging to specific and intentionally chosen classes, or an intentionally chosen post-coordinated algebraic combination of specific classes. The user identifies classes or terms from the controlled vocabulary that will contain documents meeting certain criteria or having certain characteristics of interest, and then retrieves the selection of documents belonging to those classes. A Classification Research Group Memorandum (1955) proposes a model for information acquisition which assumes this type of retrieval function exclusively:

1. Identifying the exact subject of a search.

2. Locating the subject in a guide which refers the searcher to one or more documents.
3. Locating the documents.
4. Locating the required information in the documents.

This very idealized view of retrieval is just one function of knowledge organization. Note that we include under this heading this type of class retrieval, without distinction of *which* characteristics of interest the user may be interested in retrieving on, or the system may support. So we include both Vickery's (1955) "provide for selection of a set of records likely to be relevant to a particular topic," and "provide for the selection of a group of records lying within a certain subject field." The key to this function is the selection of records formally collocated by the knowledge organization system. Svenonius (1983) refers to this role of classification simply as to "improve recall and precision."

It's worth noting that a given knowledge organization system may provide for the user choosing a class according to multiple characteristics, at multiple levels of generality. The introductory material to the BC2 classification (Mills and Broughton 1977, pg 30) use the example of a document on 'industrial relations in the motor vehicle industry of Great Britain during the war,' specifying that it "should be retrievable in response to any combination of it's constituent element classes" (e.g. "Industrial relations", "motor vehicle industry", etc.), and also at different hierarchical levels ("Manufacturing industry" as more general than "Motor vehicle industry", or "Motor accessories industry" more specific). Different systems may make different characteristics of selection available, as well as more or fewer characteristics of selection. Chan (1990) mentions that one of the benefits of the Library of Congress Classification

used for retrieval is that it's organization makes it easier to search at a more general level, and also makes it easier to organize a retrieval set based on disciplinary perspective. Markey (1986) also notes that DDC groups items together according to different characteristics of interest than LCSH.

2. Browsing

The ALCTS (1999) report specifies a "browsing" function as "examining adjacent resources within similar class, or across adjacent classes. One of the uses of "traditional classification" in an online environment that Svenonius (1983) identifies is "to provide a structure for meaningful browsing. " Exactly what "browsing" consists of is often not defined in these discussions of classificatory functions, but it is some sort of exploratory or investigatory, probably iterative, interaction with a corpus, to be contrasted with the more specifically directed aims of Class Retrieval. It is to some extent what has been thought of as the domain of traditional 'classification' as opposed to alphabetic subject languages. Note that browsing can happen among physical documents arranged on a shelf, or online, or in a printed catalog of some kind—with different possibilities and requirements made of the system. The ALCTS (1999) report writes that in order for browsing to be supported "displays of resources need to be listed in class number order", but of course online this is just one way to support browsing, and other multi-dimensional methods are also possible. An example of another kind of browsing possible in an interactive online environment is given by Marchionini (2003). The expanded possibilities of various forms of browsing in an online retrieval system, rather than

among a physical arrangement, are a driving force in the re-evaluation of the function of knowledge organization systems.

3. Relationship Navigation

ALCTS (1999) and Chan (1990) both specifically address a kind of hierarchical navigation. The ALCTS report specifies "hierarchical movement", and Chan talks of how "and one can move up or down a set of records that are displayed in call number order, to broaden or narrow a search." Of course, using call number order is just one way of accomplishing this function, and hiearchical relationships of broadening and narrowing are also just one kind of relationship that can be followed. Mills (1969) notes that "a given class cannot fail to be related to many other classes in addition to those in which it is juxtaposed" in a single linear order as in traditional classification, and that the shelf order therefore "must be supplemented by indexes (catalogues)" to "utilize every facility possible for indicating the multiple connections between classes." While relationship navigation may be traditionally considered a form of browsing, also recall Cutter's (1876) argument that relationship navigation serves as recourse when one's initial class selection does not provide satisfaction. See also *Negotiation* below.

4. Identification

The Identification function is served by listing assigned class or term information on a record so that the user knows more about the nature of the document indicated. Traditionally, subject tracings on library cards served this function, perhaps

unintentionally. ALCTS (1999) describes one way to serve this function as "when a searcher sees a class notation in a listing and uses this information to identify the subject content of the website being classified." Of course, an actual natural language description or heading (as in subject tracings) may likely serve the user better than an obscure notation.

5. Locating

This is the traditional function of traditional classification as a 'shelf location device' (Chan 1990, p. 9): A means to identify exactly where to find a given known document. The ALCTS (1999) report identifies this function, but considers it an "outmoded" function of classification when web pages are the items classified, because URLs are instead used as locating devices. Chan (1990) suggests that one could look up a known item in a catalog by class number, although it's not entirely clear why one would want to. This function could be correlated to the third step in the CRG (1957) model of information acquisition identified above: "Locating the documents."

6. Ordering.

Knowledge organization systems can be used to provide one or more useful arranged sequences or orders to a set of documents. Vickery (1971) describes one aspect of this function as "to provide for the ordering of a group of records into a meaningful sequence." This may mean the ordering of a retrieval set online, or of a catalog, and we also include here the traditional physical shelf ordering function of classification, as identified by Mills (1969), who also notes that while shelf ordering requires a single

arrangement, multiple helpful arrangements are always possible and can be provided by the system. What Vickery identifies as a separate function we also include under this heading: "to provide for the sequencing of a set of selected records according to probable relevance to a particular topic." The ALCTS (1999) report does not draw this out as a separate function, but instead uses classificatory ordering as the operationalization of support for browsing. We note that an ordering is just one way to support browsing, and that ordering may additionally be used for other purposes.

7. Surveying

Knowledge organization systems can be used to allow the user to get a general overview of what exists in a corpus. The ACLTS (1999) report calls this function "profiling", described as "giving a portrait of a collection of… resources." Marchionini and Brunk (2003) identify this as a benefit of their Relationship Browser interface, "enabling understanding the scope and extent of the corpus through active exploration of different 'slices' defined by different attribute-value juxtapositions." While Marchionini and Brunk use a classificatory structure that resembles (but may not be) a facetted classification to drive their interface, it is clear that even traditional non-facetted classification has functioned to help users gain a sense of scope and extent of a corpus. Something like this function is identified by Cutter (1876) as being one of the advantages of the classified arrangement: "So that he can take a general survey of the ground before he chooses his route". This indicates a somewhat ambiguous relationship here between the *Surveying* function and the *Browsing* function often thought of as the central role of the classified (as opposed to alphabetic) system of knowledge organization. And recall

also that Vickery's version of the "general survey", we assigned to the *Class Retrieval* function.

8. Dealing with Large Result Set

Classification can be used to help users deal with large result sets returned by interactive online systems. Markey (1986) found that users were unlikely to view retrieval sets exceeding 200 items (p. xxxvii), and that users "expressed the need for an online catalog capability to limit search results," and that classification could provide functionality in that area (p 373). Some research suggests that providing search results in a classified or categorized display allows users to analyze search results more quickly (Vizine-Goetz and Thompson, 2003). The use of knowledge organization and classificatory-like structures (including computer generated collocations) has become of wide interest in the age of online retrieval. The ALCTS (1999) report referred to this function as "Limiting/Partitioning." They describe partitioning being when a retrieval set is returned in a classified arrangement with labeled sub-headings, and limiting being a function of further restricting a result set by class. We note that partitioning is effectively the *profiling* function carried out on a result set, while limiting is related to the *class retrieval* function.

9. Keyword Match Enhancement

When a user executes a simple keyword search, English words from classification schedules or thesauri attached to bibliographic records will increase the recall of the result set. Markey (1986) calls this type of a search a "Direct search" of subject-enhanced

bibliographic records, and notes that combining different vocabulary different knowledge organization systems can enhance recall yet further. This function is least well treated in the literature, but is included here because it seems that the ad hoc collocation of documents that results from such a search is distinctly different from the *class retrieval* function, likely requiring different features of the knowledge organization system to support. It is also a function likely to be of special interest in the contemporary environment where users are believed to prefer a keyword search of this type to formal *class retrieval*. Vickery (1971) perhaps touches upon this function when he specifies a function to provide for "the selection of bibliographic records with unique characteristics; for example any record containing a specified string of juxtaposed words, such as 'enzyme activity in mammalian cells.'"

10. Negotiation

One of the functions of knowledge organization that Vickery (1971) identifies is "to give aid to the searcher in his choice of search terms," the traditional domain of the thesaurus. The CRG (1955) memo suggests that "an information retrieval system should be designed… to help even the ignorant user pass from the vague formulation of a subject in his mind to its precise formulation in the system," associated with the first and second stages in their model of information acquisition. Svenonius (1983) writes that "perspective hierarchies [contained in a classification system] can be used to contextualize the meaning of vague search terms, enabling the computer to simulate in part the negotiation of a search request carried on by a reference librarian." While Cutter

(1871) instead identifies a classified arrangment in a single printed catalog as an obstacle to this this function, in that one must be an expert in order to find one's place.

Functions Not Included in the Present Model

Some functions from the literature have been purposefully left out of this model for now. Vickery (1971) and Svenonius (1983) both write of a function of classification as a mapping/switching language. Vickery mentions as a function to "provide for the automatic conversion of index entries from one form to another," of which I'm confused about it's purpose. Svenonius (1983) also specifics "representation and retrieval of non-bibliographic information" as a use of classification, which seems to me just a different context for the same functions identified for bibliographic information, albeit a context which may prioritize functions differently. Svenonius also mentions "automatic retrieval" as a use, which seems to me instead to be just a different method of attempting to fulfill the same functions.

Flaws in the Systematic Nature of this Model

The functions identified here are not all of the same kind. Some of the functions are user tasks (e.g. class retrieval, browsing) , while others are system tasks (ordering). Some of the tasks identified may actually be instrumental to or components of other tasks: Clearly, neither *ordering* nor *hierarchical movement* are ends in themselves, but instead are means to accomplishing other ends, perhaps some of the other functions identified. *Dealing with large results sets* is a situation that users may find themselves in, which they may use any number of functions to respond to. *Negotiation* may include

other functions as components (some authors, e.g. Chan (1986) p 185, seem to suggest that it's by *browsing* and following connections between related terms or classes that a knowledge organization system provides a negotiation function), but may be an instrumental step to *Class Retrieval* or some other function. In fact, as online searches are interactive and iterative processes, identifying appropriate classes for *Class Retrieval* may often involve some aspects of *Browsing*, where it is supported. The CRG model of information acquisition would suggest that most of the functions identified here are ultimately instrumental to *Class retrieval*, an argument we do not address here. The relationship between other functions identified is not clear: *Surveying* and *Browsing* overlap in ill-defined ways, and what we are here not entirely precise by what we mean by *Browsing* in the first place. Traditional classification as shelf-arrangement relies on *Ordering* to also allow certain *Relationship Navigation* on the shelf.

A model including ten separate functions is already more than might be considered ideal, yet there are probably other functions not properly included here yet. The taxonomy of functions offered here is just a beginning, but does represent a synthesis of what attention has been given to this topic. This outline can serve as a starting point for a more systematic and logically consistent treatment. That complex relationships requiring further investigation can be found between the functions identified here suggests the starting point provided here can be fruitful. Connecting this model of functions to a model of user behavior in using knowledge organization systems might prove a profitable way to rationalize these functions further.

Towards a Conceptual Model of the Knowledge Organization System

In writing an encomium of Brian Vickery's contributions to knowledge organization, E.J. Coates (1988) makes a gentle critique:

> "In choosing not to take the risk [of adding his own value judgements] Vickery has to some extent forgone the opportunity to present the field of classification itself as an ordered pattern of concepts. In bringing many of these patterns into the daylight, but in a still partially disordered context, he has ensured that classification remains highly significant in the information professions, while being still, perhaps unnecessarily, hard to assimilate."

This situation remains true today. This lack of coherent organization of our knowledge of classification and knowledge organization, leading to difficulty in assimilation, presents a barrier to the effective application of knowledge organization to the novel environments of expanded possibilities we find ourselves faced with.

Of course, this paper usefully takes one of Vickery's attempts to add coherence to our knowledge of classification as one of it's central texts. But Vickery's (1971) approach, as well as some of the others we have synthesized here, have a limited vision of the possible functions of classification, sticking too closely to a simple model of information retrieval emphasizing our *class retrieval* function, without the context we now have of living in an environment of a networked computing environment powerful than Vickery could have imagined. A model of the functions of knowledge organization systems is necessary to allows us to investigate how to best apply our traditional systems and knowledge of systems of knowledge organization to the new online environment. The ALCTS (1999) committee discovered this, and in formulating their own model provided us with a useful starting point. A model of the functions of knowledge organization systems directs us to the questions: What functions did traditional knowledge organizations serve? How have the importance of these functions changed in

the present environment, and what new functions present themselves? How are functions best achieved in the new environment?

Lee and Olson (2005) in a recent article undertake a thoughtful and overdue exploration Yahoo! Web navigation understood as a form of classification, and a method of access as opposed to keyword searching. With a model of the functions of knowledge organization systems, we can put their research questions and findings in a coherent context. What functions does a Yahoo! style navigation serve (browsing, surveying, relationship navigation, class retrieval), compared to keyword searching? Lee and Olson suggest that:

> Two types of research are suggested…: continuing studies, both quantitative and qualitative, of users' performance and preferences in hierarchical navigation; and, the development and testing of experimental systems and interfaces that offer easier and more effective hierarchical navigation than is currently available.

Indeed, both of these types of research are desperately needed, not just for hierarchical navigation, but for many features of knowledge organization systems in general. But a third type of research is needed as well: Theoretical elaboration and modeling of knowledge organization systems, in order to provide a framework for creation and testing of knowledge organization systems. Theory and models provide us with a way to organize our empirical observations into a coherent whole and provide us with direction for future development and study. Our current understanding of knowledge organization systems has become fragmented and without an overall picture. A conceptual model of function gives us a way to compare traditional knowledge organization with 'new' types of (or competition with) knowledge organization: What functions are served by full-text retrieval or 'folksonomy', and how effectively, compared with other forms of knowledge organization? There is a role for empirical investigation in these questions, but there is

also a need for a conceptual framework to give us the questions to ask and to organize our findings.

This initial attempt at listing of functions can serve as one component of a general conceptual model of knowledge organization. Another important part is a model of the structural and representational features of a knowledge organization system. But structural features, after Vickery, we mean those features that can largely be considered in the context of a knowledge organization system as a formal language in itself, without discussing the relationship of the system to the actual world. Such as relationships (of various types), lead-in vocabularly, facetted structure, etc. Vickery (1971) provides an initial treatment of this topic as well. We need such a model to understand which features support which functions in which ways, and which features may work against other functions. In what ways are features built into our legacy knowledge organization systems optimized to support some functions at the expense of others that may be important in an online environment? How can different systems be used in tandem to support complimentary functions? For example, are there different optimum structures for a 'lead-in vocabulary' to support *negotiation* than to support *keyword search enhancement*? Does a lead-in vocabulary have a role in *identification*? This model also encourages us to distinguish between *representational* features of systems, such as the fact that DDC and LCSH choose different characteristics expressed in different ways to hilight, and *structural* features, which are also different in these two systems. The reconsideration of the dichotomy between traditional *classification* and *alphabetic indexing* shows us that any combination of characteristics of division and structural features is possible, and may serve different functions in different ways.

Much of the existing discussion of the relationship between features and functions in the literature discussed here focuses excessively on *notation*. In the context of a model of functions, we can understand notation as a clever and efficient way of fulfilling several functions at once: providing a *locating* function, within a single *ordering* for physical shelves, that attempts to provide a useful *browsing* function, and where expressive, allows for a certain kind of hierarchical *relationship navigation*. But notation is just one way of fulfilling these functions, albeit a very efficient one in the physical environment. In the 80s, some made the mistake to think that, for instance, an expressive notation was the only way to allow relationship navigation (e.g., Chan 1990). But of course, so long as the system can present the user with a network of relationships, relationship navigation can be supported, and notation is not the only way to do this. This fact is obvious to us now when considering Yahoo! as a classification, for instance. The ALCTS (1999) report likewise focuses excessively on notation, oddly assuming that entering a class number is the only way support the *retrieval* function, and displaying a class number the only way of supporting an *identification* function. Lee and Olson's (2005) investigation of Yahoo! as classification, a system with no notation at all, shows the shortcomings of a notation-bound approach. Even when dealing with 'traditional' legacy classification systems, as the ALCTS (1999) report was, there are various ways to support these functions. Even if the materials of the corpus are physical books, a notation may only be necessary to provide *locating* functions. The conceptual model of function shows us that notation is just one way of carrying out certain functions, and suggests we consider it's strengths and weaknesses in a given context.

The development of a conceptual model of knowledge organization systems is critical to coherently guide us in the application of knowledge organization to novel environments. This paper has attempted to lay the groundwork for such a conceptual model by beginning a systematic discussion of the various functions of knowledge organization.

References

ALCTS CCS SAC Metadata and Classification (1999). SAC Subcommittee on Metadata and Classification, Final Report, ALA Annual Conference 1999. http://archive.ala.org/alcts/organization/ccs/sac/metaclassfinal.pdf

Bland, Robert N. (1990). Toward the Catalog as a Tutorial Guide to the Literature. *Cataloging & Classification Quarterly* 11 (1), pp 71-82. (Special issue representing the proceedings of the Annette Lewis Phinazee Symposium).

Chan, Lois Mai (1986). Library of Congress Classification as an Online Retrieval Tool: Potentials and Limitations. *Information Technology and Libraries*, September 1986. pp 181-192.

Chan, Lois Mai (1990). The Library of Congress Classification System in an Online Environment. *Cataloging & Classification Quarterly* 11 (1), pp 7-25. (Special issue representing the proceedings of the Annette Lewis Phinazee Symposium).

Classification Research Group (1955). The Need for a Facetted Classification as the Basis of All Methods of Information Retrieval. Reprinted in *From classification to "knowledge organization" : Dorking revisited, or, "Past is prelude"* 1997, Alan Gilchrist (ed.), The Hague: International Federation for Information and Documentation (FID).

Coates, E.J. (1978). Classification in Information Retrieval: The Twenty Years Following Dorking. *Journal of Documentation* 34(4) p.288-299. Reprinted in *From classification to "knowledge organization" : Dorking revisited, or, "Past is prelude"* 1997, Alan Gilchrist (ed.), The Hague: International Federation for Information and Documentation (FID).

Coates, E.J. (1988). The Role of Classification in Information Retrieval Action and Thought in the Contribution of Brian Vickery. *Journal of Documentation*. 44(1). 216-225.

Cutter, C.A. (1876). Library Catalogues. In *Public Libraries in the United States of America: Their History, Condition, and Management, Special Report, Part I.* Department of the Interior, Bureau of Education. Washington DC: Government Printing Office. 526-622

Dewey, Melvil. (1876). Catalogues and Cataloguing: A Decimal Classification and Subject Index. In *Public Libraries in the United States of America: Their History, Condition, and Management, Special Report, Part I.* Department of the Interior, Bureau of Education. Washington DC: Government Printing Office. 623-643.

Gilchrist, Alan (2003). Thesuari, taxonomies and ontologies – an etymological note. *Journal of Documentation*. 59 (1). pp. 7-17.

Lee, Hur-Li and Hope Olson. Hierarchical Navigation: An exploration of Yahoo! Directories. *Knowledge Organization* 32 (1). pp. 10-24.

Marchionini, Gary and Ben Brunk. Toward a General Relational Browser: A GUI for Information Architects. *Journal of Digital Information*, 4 (1), Apr 2003.

Mills, J. (1969) Inadequacies of Existing General Classification Schemes. In *Classification and Information Control: Papers representing the work of the Classification Research Group during 1960-1968*. London: The Library Association.

Svenonius, Elaine (1983). Use of Classification in Online Retrieval. *Library Resources & Technical Services*, January/March 1983, 76-80.

Svenonius, Elaine (1989). An Ideal Classification for an On-Line Catalog. In *Classification Theory in the Computer Age: Conversations Across the Disciplines*. Conference Proceedings, Nov 18-19 1988. Albany: NY: Rockefeller College Press.

Vickery, B.C. (1971). Structure and Function in Retrieval Languages. *Journal of Documentation* 27 (2), p. 69-82. Reprinted in *From classification to "knowledge organization" : Dorking revisited, or, "Past is prelude"* 1997, Alan Gilchrist (ed.), The Hague: International Federation for Information and Documentation (FID).

Vizine-Goetz, Diane and Roger Thompson (2003). Towards DDC-Classified Displays of NetFirst Search Results. In *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Sattellite Meeting held in Dublin, OH, 14-16 August 2001… ,* I.C. McIlwaine (ed.), Munchen: K.G. Saur.

West, Martha W (1984). Classification as a Retrieval Tool. *National Online Meeting*.