The Boundaries of Classification

Jens-Erik Mai
Faculty of Information
University of Toronto
je.mai@utoronto.ca

Abstract: This paper discusses and analyzes the conceptual basis for classification work in the 21st century; it provides an account of classification that lays out the boundaries within which classification operate. The methodological and practical effects of the boundaries are discussed. The main point of the paper is to demonstrate that classifications are bound by particular contexts and conceptual frameworks.

Subject tags: classification; knowledge organization; epistemology

Introduction

Bibliographic classification has been haunted by the idea of a universal classification that would cover all topics, all material, for everyone ever since Melvin Dewey had his Eureka! moment that Sunday morning during a long sermon (Wiegand, 1996; 1998).

Dewey's idea was simply, clear and practical. Instead of having each library assign call numbers based on a book's physical location in the library – libraries should assign numbers that refer to books' topics in the universe of knowledge. This would mean that a book would have the same number no matter which library the book was placed in. This move would have significant advantages, incl. that users could more easily find books their were interested in, libraries could more easily share material and the whole enterprise could be industrialized.

While the idea made--and makes--sense from a practical point of view, it is haunted by a flawed conceptual basis. The notion that there is such a thing as the universe of knowledge has been challenged many times in modern philosophy and most recent work in classification rejects the notion. Most scholars accept that there are multiple universes of knowledge, that people come to the world from different perspectives, view the world differently, and understand the world differently. Further, most scholars today accept that things and topics cannot be classified independently from their function, interpretation, or relation to other things and topics – it is generally accepted that things and topics do not have innate essences. In this line of thought, things and topics make sense only in context, and they are always understood from a particular epistemic standpoint.

This paper will explore the implications that a pluralistic understanding has in classification work and practice and, as such, it will outline the boundaries for classification.

Different contexts, different meanings, different classifications

Miksa (1998) explores the notion of the 'universe of knowledge' especially within the context of early theorists contributing the foundation of bibliographic classification, and he found that,

> "As the universe of knowledge has become a legitimate object of investigation, classification theorists have discovered that it is a much more complex phenomenon than was previously supposed. In fact, the closer Richardson, Bliss, Sayers, Ranganathan, and those who have followed them got to the phenomenon, the more complex it appeared and the more difficult it was to describe. Much of this new sense of complexity was due to the new approach to subjects that arose from the

documentation movement and elsewhere, where subject complexity is a direct reflection of the way in which scientists, engineers, and specialists of all kinds search for information" (Miksa, 1998, p. 74).

While these early theorists might have expressed doubts about the notion of the universe of knowledge, their doubts seem to have had little impact on the subsequent development of classification theory and practice. More recently alternative understandings of the conceptual foundation of classification have been offered by a number of scholars, common for whom is the idea that the development of meaning, concepts and categories are bound temporally and spatially. Bowker & Star (1999, p. 10), for instance, state that, "classification is a spatial, temporal, or spatio-temporal segmentation of the world" making it clear that any classification is bound in time and space. Similarly, Hjørland (2009, p. 1529) concludes that we "should not consider concepts to be universal but to be linked to certain discourses and interests."

The 1986 ISO Standard for the development of thesauri states that,

"'Banks' would imply a broader term such as 'Financial institutions'; 'Computers' is *mentally associated* with 'Data processing'; and 'Amsterdam' implies *the* wider location 'Netherlands'. Any of these *mentally-associated* terms might serve as a user's approach to the subject index. These relations are document-independent, since they are *generally recognized* and could be established through reference to standard works, such as dictionaries and encyclopedias" [emphasis added] (ISO, 1986, p. 1).

The ISO Standard's call to mental constructs and relations that are generally recognized builds on the assumption that there are in fact such relations that are general and universal in nature and which can be assumed to be shared by everyone. While Netherlands certainly is one of the wider locations of Amsterdam, so is North Holland (the province in the Netherlands), and New York (the state in the US); it does seem a bit insular to claim "'Amsterdam' implies *the* wider location 'Netherlands'". The wider location of 'Amsterdam' seems to dependent on one's particular spatial outlook; it is easy to imagine that children in upstate New York would claim a different wider location to Amsterdam than children in North Holland. Likewise, to claim that, "'Computers' is mentally associated with 'Data processing'" is limited to a particular time in which this association was a norm, but it is also limited to a particular view of computers which was challenged even in the mid-1980s. 'Computers' will have different associations within different discourse communities and in different times, and while the relation to 'data processing' is one possible association there are many others that are equally possible and accepted. ISO's call to establish these relations using dictionaries and encyclopedias seems like reasonable advice, but the ISO does seem to assume that there exists *a* set of relations that are "generally recognized" and therefore universal in nature. The idea that the associated relations are mental in nature further implies a particular epistemic stance that nevertheless remains unvoiced in the document.

Bryant (2000), on the other hand, presents a pluralistic account of classification and demonstrates that classifications are part of temporal and spatial interests and discourses. She takes as the "starting point" for her exploration the "complexities of the natural world" and argues that "as a result of this complexity, numerous scientifically significant patterns of similarity or regularity exists according to which natural entities can be divided into kinds or classes" (Bryant, 2000, p. 16). Bryant goes on to develop an *explanation-based view* of classification, which requires that one acknowledge, "that in order to have a meaningful concept of anything at all, we must first understand the role that that particular kind of object plays in the world" (Bryant, 2000, p. 60), and to "grasp a concept, we require not only definitional, but also encyclopedic knowledge" (Bryant, 2000, p. 59), which "presupposes extralinguistic knowledge" and is "uncoordinated by nature" (Eco, 1999, p. 226). Broadfield gives as an example of this kind of explanation based classification work, that "if opinion is divided whether a feature of certain organisms is a backbone or not, this

reflects uncertainly not as to what they are, but as to how this particular feature of theirs is to be thought of in biology" (Broadfield, 1946, p. 20). Similarly, Eco in his discussion of the classification of the platypus observes that when: "there is a clash about whether something is milk or mucus, it is necessary to treat the perceptual experience in cultural terms too and to decide which criteria and chemical classifications allow us to distinguish milk from mucus" (Eco, 1999, p. 253). In such situations, the classification of the objects needs to build and extend from the shared discourses and interests and cannot rely on inspection and analysis of the objects themselves alone.

Bryant's work challenges classificationists to provide that extralinguistic knowledge which is required to understand particular classifications and the choices made in the design of the classification. It requires that one is explicitly aware and responsible for the particular view that the classification promotes and can base the classification within a particular discourse and set of interests. This pluralistic, temporal understanding of concepts and classifications creates two challenges and consequences within which bibliographic classification operates:

1. There is no one universe of knowledge in which books (and other information things) can be placed. Different people and groups of people have different categorical structures of the world – and these categorical structures are often incompatible. There is no overreaching, correct superstructure that all people recognize and operate within.
2. Books (and other information things) don't have topics independently of particular people, activities and perspectives. Any thing means something only within a particular spatial and temporal context; it does not have a topic independently and objectively.

An approach to classification theory and practice that assumes this complexity of phenomena bounds classification in time and space to particular interests and discourses. This requires a commitment to humans and their activities and classification activities would be grounded in local meanings and discourses. The two challenges and consequences have an effect in three different, but interrelated, areas:

I. <u>Conceptual</u>: Classification work and workers need to recognize and appreciate a pluralistic understanding of the world, which holds that not only do different people understand the world differently, but no one group holds a better or more correct understanding of the world, things and topics.
II. <u>Methodological</u>: The methods employed in classification need to be developed to account for a pluralistic understanding.
    a. For the design and construction of classificatory structures: The methods employed need to explicitly start with an understanding of the activities and interpretations that people engage in. The classificatory structures should be designed to emulate the categorical structure that is already in play in the epistemic community.
    b. For naming of things and topics: The methods employed to determine the meaning of things and topics for placement in classificatory structures need to tied to how the things and topics gain meanings in contexts through activities and interactions.
III. <u>Practical</u>: What is built and named in one context cannot be used in another context – unless the contexts share meaning constructing milieu. Sharing of records and structures across epistemic communities will create problems, misunderstanding, and flawed systems.

The methodological and practical commitments follow from the conceptual commitment. The pluralistic commitment has been best summed up by Hjørland (1997, p. 111), who argues humans interacting with information, "should be seen as individuals in concrete situations in social organizations and domains of knowledge." He continues by giving an example of what that means in practice,

    "A stone on a field contain different information for different people (or from one situation to another). It is not possible for information systems to map *all* the stone's

possible information for every individual. Nor is any *one* mapping the 'true' mapping" (Hjørland, 1997, p. 111).

One would approach the development and use of classifications conceptually differently, if one assumed that there are multiple equal correct classifications, than if one assumed that there is *a* universe of knowledge. In other words, when taking a pluralistic conceptual approach, one would be concerned with creating a map of something for certain people, for particular purposes, it would not be the goal to create a map of *how things really are*. In this sense, classification is an epistemological exercise not an ontological exercise. Alternatively, a commitment to a universal universe of knowledge will, naturally, generate a different set of goals and expectations – most likely one in which it is possible to create the one, best system for all material and where the naming can be done neutrally and objectively. In this tradition, one would strive towards a classification of how things really are, for an *unbiased* system – and assumably, one could make calls to *generally accepted or recognized* understanding of concepts and relations. A pluralistic approach, on the other hand, acknowledges that classifications are inherently biased and that one should instead ask how systems could be designed "to incorporate bias in a responsible manner" (Feinberg, 2007). A responsible manner would be one that is based on explicit discourses and encyclopedic knowledge of concepts and their relations, in this sense we should,

> "consider as encyclopedic knowledge only those items that the Community has in some way registered publicly. But there is no doubt that, with regard to the events and objects of this world … that there is always new facts to learn, and therefore those who find the encyclopedic format hard to handle are not wrong" (Eco, 1999, p. 226).

While the notion of the 'universe of knowledge' has a long history in bibliographic classification work and is well established as a framework for understanding classification practice, a pluralistic understanding offers a framework that better handles the diversity of meanings and understandings that most classifications encounter.

Approaches to classification
Classifications are developed from within particular conceptual frameworks for specific contexts, and while the conceptual framework provides the general approach taken, the specific context further bounds the practice and the methods employed.

It may at this point be useful to divide these contexts of knowledge organization[1] practice into three sets of knowledge organization problems (KOPs):
1. Big KOP. This is the organization and representation of large quantities of information for unrecognizable many people; people with varied interests, beliefs, positions, knowledge, expertise, etc. The Web is the prototypical example of such a KOP, large academic and many public libraries are also Big KOPs.
2. Medium KOP. These are information collections for particular, stated, clear, objective, and specific purposes to be used by people with particular, similar interests, beliefs, positions, knowledge, expertise, etc., which can be known, understood, and articulated by those in charge of the collection or service. A company's intranet, a web portal, a store, and some special libraries are typical examples.
3. Small KOP. This is individuals' information management challenges and collections. These KOPs are particular to an individual's (or a few individuals') personal information collection and will typically be managed by the same individual(s). The information could be emails,

---

[1] I use the term 'knowledge organization' to denote the broader sphere of issues that dealt with the representation and organization of information and other artifacts for retrieval. I use the term 'classification' more narrowly to denote the acts of establishing the sense of various entry points and their relationships.

documents, files, photos, etc., which will be collected, searched, and used by individuals for their own usages.[2]

Classification work within library and information science has historically focused on Big KOP with a focus on general, universal systems that organize the entire universe of knowledge. There has been some interest in Medium KOPs, especially in CRG's work on the development of special classification and Hjørland's work on the domain-centered approach. However, the distinct natures of the difference between Big KOPs and Medium KOPs have seldom been used as an approach to understand various practical and methodological approaches to classification.

We can now combine the two general conceptual approaches (universal universe of knowledge vs. pluralism) to classification, with two kinds of KOPs (Big and Medium) to compare and contrast the various approaches to classification. I suggest doing this along four lines of inquiry:

1. Design of classificatory structures: The general approach that will be taken in the design and development of the classificatory structures or the system as a whole. The classificatory structure can be more or less complex or complete.
2. Naming of things and topics: The assumption and approach taken in representing the material for future retrieval and/or discourse. This could be done by professionals or as a collaborative process among users.
3. The success factors: The factors that used as measures against which it is determined whether the system is a success; these are also the factors against which the system's effectiveness is evaluated.
4. Overall goal of the classification system: The overall objective of the system that outlines the general values that the system strives towards.

		Big KOPs	Medium KOPs
Universe of knowledge	*Design*	Logical division. Facet analysis.	Logical division. Facet analysis.
	Naming	Neutral. Objective.	Neutral. Objective.
	Success factors	Correct representation. Consistency.	Correct representation. Consistency.
	Goal	Unbiased. Interoperability.	Unbiased. Interoperability.
Pluralism	*Design*	Facilitate collaboration among users.	Domain analysis.
	Naming	Democratic indexing. Social tagging.	Domain or user centered.
	Success factors	Involvement of users. Facilitate diverse interpretation.	Ability to analyze and understand domain and users' needs.
	Goal	Facilitate conversation and shared meaning making.	Understand and match users' needs and domain's structure.

Figure 1. Comparison of approaches to classification

The chart in figure 1 sketches out three general approaches to classification:

1. Authoritarian (upper section): In an attempt to adhere to the traditional values of transparency, consistency, interoperability, and stability this approach accepts the conceptual commitments made in the assumptions that follow the notion of the universal universe of knowledge.

---

[2] I will here focus on Big KOP and Medium KOP; while Small KOP is of interest and could inform Big and Medium KOP, it has developed its own area of research (personal information management) with it own unique agenda and literature.

2. Social tagging (lower left section): In the realization that it is difficult, if not impossible, to represent and classify huge masses of information for large groups of people in a meaningful way, social tagging offers a collaborative, democratic approach to this challenge.
3. Domain analysis (lower right section): In reaction to the need to represent and classify information for specific user groups, domain analysis offers an approach that attempts to capture users' discourse and interests.

This contextualization of various classification approaches provides a better insight into the assumptions and commitments of the various approaches. It further demonstrates that there are indeed viable alternatives to the dominating practice-based approach to classification and that this dominating approach is based on particular conceptual and epistemic assumptions and commitments.

Conclusion: Boundaries
Bibliographic classification has been grounded solidly in the notion of a universe of knowledge and has strived to develop systems that would potentially serve everyone, and organize all kinds of material. However, modern conceptual approaches to classification and pressures from social technology movements have forced a shift in the practice and theoretical foundation for classification.

A pluralistic understanding of classification runs up against mainstream practice on at least two fronts: First, when classification is bound by the activities, interpretations and understandings that people engage in within certain spatial and temporal settings, then classifications are bound to particular epistemic communities and make sense only within that context. It cannot be assumed that people from other epistemic communities unproblematically can interact with or work with classifications from other communities. Secondly, the placement of particular things and topics in a classificatory structure is bound to the understanding of those things and topics within the epistemic community and is only classifiable with structures that reflect the community. This sort of interoperability and sharing of records is a norm--and a goal--in many practices today, it is bound to a conceptual understanding that accepts the notion of *the* universal universe of knowledge.

As classification practice starts to embrace a pluralistic foundation, the goals for bibliographic classification shift and become focused on serving its users in their particular contexts and help them create meaning.

References

Bowker, G.C. & Star, S.L. (1999). *Sorting things out: Classification and its consequences.* Cambridge, MA; MIT Press.

Broadfield, A. (1946). *The philosophy of classification*. London: Grafton & Co.

Bryant, R. (2000). *Discovery and decision: Exploring the metaphysics and epistemology of scientific classification.* Cranbury, NJ: Associated University Presses.

Eco, U. (1999). *Kant and the platypus: Essays on language and cognition.* London: Secker & Warburg.

Feinberg, M. (2007). Hidden bias to responsible bias: An approach to information systems based on Haraway's situated knowledges. *Information Research*, *12*(4). [Available at http://InformationR.net/ir/12-4/colis/colis07.html]

Hjørland, B. (1997). *Information seeking and subject representation: An activity-theoretical approach to information science*. Westport, CT: Greenwood.

Hjørland, B. (2009). Concept theory. *Journal of the American Society for Information Science and Technology, 60*(8), 1519-1536.

ISO. (1986). *Documentation—guidelines for the establishment and development of monolingual thesauri. 2nd edition*. Geneva: International Organization for Standardization. ISO 2788-1986.

Miksa, F. (1998). *The DDC, the universe of knowledge, and the post-modern library*. Albany, NY: Forest Press.

Wiegand, W.A. (1996). *Irrepressible reformer: A biography of Melvil Dewey*. Chicago: American Library Association.

Wiegand, W.A. (1998). The 'Amherst method': The origins of the Dewey Decimal Classification scheme. *Libraries & Culture, 33*(2), 175-194.

* * *