Rick Szostak (University of Alberta, Edmonton)

Universal and domain-specific classifications from an interdisciplinary perspective

Abstract
A universal non-discipline-based classification is a complement to, rather than substitute for, domain-specific classifications. Cognitive work analysis suggests that especially interdisciplinary researchers but also specialized researchers would benefit from both types of classification. Both practical and theoretical considerations point to complementarity. The research efforts of scholars pursuing both types of classification can thus usefully reinforce each other.

The recent interchange of articles between Birger Hjørland and me in the *Journal of documentation* (64, n. 3, 2008) served to highlight two contrasting views of classification:
(1) a view which urges the development of a superior universal classification that would facilitate the access of especially interdisciplinary scholars to insights generated by all communities of scholars;
(2) a view that concepts are ambiguous and thus it is best to classify documents only within particular domains and in terms of concepts understood inductively as scholars within the community in question understand these.

Scholars of knowledge organization (KO) have long debated the relative merits of these two views. Yet these two positions might be seen as complements rather than substitutes. As an interdisciplinarian I am guided to pursue “both/and” analysis rather than “either/or” analysis: that is to seek to integrate the best of contrasting views rather than conclude that one is entirely misguided. Since these two views are both widely held within the KO literature (witness the many comments on the León Manifesto <www.iskoi.org/ilc/leon.htm> on the one hand, and the many articles by Hjørland and colleagues on the other), it is important that this possible complementarity be carefully elucidated so that we can as a KO community work together toward shared goals rather than be diverted into false controversies.

This paper will establish complementarity by outlining a set of key arguments regarding both the desirability and feasibility of pursuing the two types of classification in concert. The first section reviews the (compensating) strengths and weaknesses of the two types of classification for especially the interdisciplinary researcher. The second reviews a handful of theoretical and practical reasons to see the two as complements. The third suggests how empirical research on the two types of classification both can and should proceed in a complementary fashion.

1: Cognitive work analysis of specialized and integrative researchers
As Mai (2008) notes, the design of classifications should be grounded in an understanding of user needs. Yet the descriptive literature on user behavior provides little guidance. He thus suggests that we identify the constraints faced by particular user groups and then attempt to cope with these. In this section, we briefly apply Mai’s analysis to the situation of both interdisciplinary and specialized scholars. Integrative scholarship and specialized scholarship play a symbiotic role in the academy. “Studying
a thing in isolation and studying it in context are two halves of modern scholarship, and neither may be safely neglected. The structure of knowledge representation must continue to shift to reflect this balance” (Iyer 1995, 27). Specialized scholars should occasionally look beyond their domain for information. Integrative scholars need to move between domains but strive to understand the internal workings of particular domains.

Mai advises us to first look at the work environment. Universities are generally organized around disciplines. The vast majority of scholars thus interact mostly with scholars with a shared disciplinary perspective. Specialized scholars will benefit from a domain-specific classification that reflects these shared understandings. If they need to look beyond their domain, however, they will need to rely either on the kindness of strangers or on useful guidance from the field of information science. Interdisciplinary scholars will be hobbled if they must master several domain-specific classifications or seek out many disciplinary experts for advice.

What sort of work is done? University research, it is widely appreciated, is increasingly interdisciplinary. A growing body of scholars defines themselves as interdisciplinary. This must mean that they do not take any one domain-specific terminology for granted, but must consciously master multiple domains. Even specialized researchers are expected to have some familiarity with how their research fits within the broader scholarly enterprise. A point to stress here is that when scholars are doing highly specialized research, they do not really need information science: they follow a handful of journals religiously, attend certain conferences, and need little else. It is only when scholars lift their gaze beyond their own specialization that they need information science. Yet this is precisely when it fails them. Domain-specific classifications can help them look at other fields in their discipline, but something broader is necessary to guide them beyond that.

The greatest breakthroughs in scholarly research almost always come from drawing connections across previously distinct bodies of scholarly understanding (Root-Bernstein 1989). There is thus a huge body of “undiscovered public knowledge”: links between existing insights that have not yet been juxtaposed. The concept of undiscovered public knowledge has only rarely been highlighted in the KO literature (but see Beghtol 1995). Yet its implication is profound: if a guiding goal of KO is to facilitate the advance of human understanding (or more narrowly the success of scholarly research), it should seek to guide researchers to novel connections. Retrieval systems are universally judged only in terms of whether users find what they were looking for; they should also be evaluated in terms of whether they guide users to novel connections (Warner 2000, 36-37). As Davies (1989) has noted, the most important connections lie across quite distinct domains. If so, isolated domain-specific classifications will signally fail to point researchers toward such connections.

How is research organized? Most researchers still interact primarily with other scholars from their discipline. As noted above, they thus have little idea of how or where to look for relevant information from other disciplines. An increasing number of scholars operate within interdisciplinary teams. These almost universally face “translation” problems: scholars speak past each other because they use words in slightly different ways. Boundary work across disciplines is thus characterized by the use of “pidgins”: limited dialects that allow cross-disciplinary communication (Klein
1990). Since pidgins are limited in scope, it would be difficult to classify all works relevant to such a cross-disciplinary endeavor in terms of the shared dialect. Conversation across all disciplinary boundaries would be better served by the sort of meta-language embedded in a universal classification.

What sort of tasks, decisions, and search strategies characterize the work? With regard to the last of these, the important point to stress is that in the absence of reliable guidance from information science, scholars wishing to pursue questions across disciplinary boundaries are forced to rely on a host of time-consuming and ineffective search strategies: seeking out scholars in other fields, chasing citations, following “big names” in other fields, and so on (Palmer 1996). With respect to tasks and decisions, the key point would be that all scholars absorb sub-consciously the research methodology of their domain. Interdisciplinarity requires that they consciously re-evaluate their research strategies. Yet they can really only do so after they have encountered the research methodologies of other disciplines. One solution to this dilemma – that one needs to become interdisciplinary before one knows how to do so – lies in the development of an interdisciplinary research methodology (Repko 2008). Yet as Repko makes clear the interdisciplinary literature search is one of the most difficult components in interdisciplinary research precisely because existing classifications and subject search tools do not guide researchers readily to the information they need.

Finally, Mai guides us to interrogate the experience, expertise, and training of workers. The vast majority of (even interdisciplinary) scholars received their training exclusively within one discipline (with perhaps a token course or two outside). They are thus not taught how to access information from outside their discipline (and perhaps even sub-discipline). That is, scholars tend to know a lot about one thing, but have absolutely no training in how to connect that knowledge to related understandings in other fields. They rely on information science to facilitate this.

2: Theoretical and practical arguments for complementarity

2.1: Ambiguity. Language is clearly ambiguous. The degree of ambiguity lessens within groups that regularly interact (though it does not disappear). Yet it is equally clear that conversations across groups are possible (or international tourism would be impossible). Thus ambiguity differs only by degree between universal and domain-specific classifications, though that difference of degree is likely quite significant.

Psychologists (especially evolutionary psychologists) suspect that there are universals in how human beings think and perhaps in how they organize their conceptual maps. Neelameghan argued at length in the 1970s that it should be possible to develop subject representations grounded in these universals (Iyer 1995, 184). It could well be that universals are more powerful in certain domains of human activity (evolutionary psychologists would suggest that this is more likely when discussing activities that humans undertook during the millennia we operated as hunter-gatherers). If so, universal classifications may face greater difficulties in some realms than others. Yet it seems likely that any innate human classification system will be organized around phenomena rather than disciplines: “If […] the patron asks for a book on birds, he expects that a section in the library contains all bird books. […] To the extent that they do not (e.g. bird books are spread out into many different places), the system is awkward and irrational from the patron’s perspective, regardless of its logical
consistency from the librarian’s” (Donovan 1991, 26).

Whether or not universals in human mental activity are strong, the very practice of KO can reduce the problems of ambiguity. Placing a concept within a hierarchical classification establishes clearly what sort of thing it is and what sort of thing it is not, and often the sorts of subsidiary elements of which it may be comprised. Wittgenstein famously argued that the best way to define a concept was to provide examples of it (game: chess, soccer, poker). He did not appreciate that a classification that provided an exhaustive set of examples would provide a very non-ambiguous definition. We in the KO field should not take ambiguity for granted but should appreciate that we have some significant ability to reduce it through our own efforts. In other words, the existing level of ambiguity in cross-disciplinary communication (which deserves to be measured, not assumed to be huge) is not inevitable, but reflects the lack of a truly universal classification.

The degree to which ambiguity can be lessened by classification is hidden from us by the simple fact that most library users do not understand how library catalogs are organized (in part because these systems, especially the Library of Congress Classification (LCC) I must deal with most often, are not very logically organized). If library users understood the logic behind some universal classification, they would find the terminology used in that classification much less ambiguous. Nor do users need to master the entire classification: as long as a computer can be programmed with the logical structure of an entire universal classification, it can take a user from any search item to an outline of how the user’s interests are reflected within the organization of knowledge (DeRidder 2007). As we will see, one valuable entry point for a user may be from the terminology employed in a domain-specific classification.

2.2: The practical need for conversion. If scholars will from time to time wish to consult works from other domains, then information scientists will want to provide conversion tables or thesauri or some other device to facilitate users in moving from one domain-specific classification to another. Moreover in developing a domain-specific classification, information scientists may often find it useful to declare how a concept is not used (“this group treats concept X differently from common usage elsewhere”). Domain-specific classifications, then, neither will nor should be developed in isolation. While it is possible to imagine a set of conversion devices between any pair of domain-specific classifications, the practicality of such an approach is suspect. There would thus appear to be advantages in linking any domain-specific classification to a universal classification. Gnoli (2009) discusses how domain-specific classifications could be linked to the Integrative Levels Classification.

2.3: The need for both induction and deduction. The best classification systems are developed using both deduction and induction. Deduction provides a logical structure; induction ensures that everything has a place. A system such as LCC that relies too heavily on finding a place for every work to fit will deviate from a logical structure; a system that does not look at how scholars are actually utilizing concepts will inevitably lump together quite different types of work. Especially (but not exclusively) with respect to the most detailed entries in a hierarchical classification, there is simply no substitute for examining how scholars organize their thinking. The recent Nuovo
Soggettario, or Italian subject indexing system, pursued exactly this strategy of using induction but placing all elements within logical hierarchies (Cheti & Paradisi 2008). Developers of a universal classification can thus usefully try to integrate the insights of domain-specific classifiers (and vice versa).

2.4: The need for both epistemology and ontology. Gnoli (2007) argued that classification systems are best grounded in both ontology (an understanding of the essence of things) and epistemology (an understanding of how scholars study things). The domain-specific approach leans heavily on epistemology: it seeks to ground classifications in an understanding of how scholars in that domain operate. Universal classifications can and should have an ontological base. Most ostensibly universal classifications rely on disciplines as a classificatory device and are thus to a considerable extent largely epistemological in approach. The sort of universal classification advocated in the León Manifesto is grounded in ontology. Yet as noted above the details of such a system are worked out with careful attention to how scholars study things.

3: The empirical agenda
Developers of both universal and domain-specific classifications can and should work together.

3.1: The universal classificationist. I am beginning a large research project in which I will flesh out in detail how social science phenomena could be classified in a non-disciplinary universal classification such as the Integrative Levels Classification (ILC). If I take the study of the economy as a starting point, I would envision two major kinds of economic phenomena (facets). The first are the components of economic output (the types of expenditure, such as investment and consumption; the types of income, such as wages and profits; and finally the output of particular goods and services). The second are types of economic institution (financial, trade, labor, and so on). This logical structure can claim an ontological basis: output must either be used for consumption today or invested to enhance consumption tomorrow; the several types of institution can be defined in terms of their function or purpose. Yet in its details it also reflects epistemology: “investment” is defined in a particular way by economists. Yet the very place of investment in the classification reinforces this meaning: it involves the devotion of real output to increases in the capital stock (and thus can be distinguished from paper transactions such as buying a money market fund which may just reallocate funds among people without any effect on the size of the capital stock). This logical structure, then, simultaneously reflects an ontological understanding of the world, applies terminology in a way that all economists can readily appreciate, and yet is amenable to understanding by non-economists. Whereas it is often thought that these goals are in conflict, this logical structure arguably performs each of these functions better than existing classifications in which related phenomena like consumption and investment are often dispersed (LCC separates the study of economic growth from the study of fluctuations; economists increasingly appreciate that these are related) while distinct phenomena are commingled (the institutions of property are treated by LCC in
the same section as income from property; likewise capital is treated together with the institution of capitalism).

Of course, the devil is in the details, and any classification can seem excellent if only small parts are scrutinized. Yet if economic phenomena can be classified in a way that is useful to both economists and non-economists, then this may prove true for other phenomena as well. While the literature on social divisions is fraught with controversy across many disciplines, a fair bit of agreement might be achieved on the different types of gender, family, and occupation/class that might be conceived. Most political phenomena are institutional in nature, and can thus be classified in terms of the purpose of various types of institution (decision-making, enforcement, and so on). Even cultural values can be classified in a manner where their meaning is fairly clear (think of “attitude toward punctuality”, for example). Indeed it might be argued that most (though not all) scholarly dispute concerns how phenomena relate to each other rather than the definition of particular phenomena (those terms that are most ambiguous, such as “culture” itself are generally aggregative terms that subsume a variety of subsidiary elements: precise definition in such cases can only come from precisely classifying these subsidiary elements).

And then it becomes possible to use synthetic notation on a far wider scale than any existing classification. Existing classifications can be improved on but also simplified by appreciating that entries such as “value” reflect a philosophical evaluation of economic behaviors (at present, LCC needlessly mixes philosophical and economic treatments of such phenomena as price), that utility is a psychological evaluation of economic output, and that most subdivisions under management would be better characterized by synthetic notation (management of economic phenomena in many cases but also non-economic phenomena such as technology or gender relations). Note that even faceted classifications such as the Bliss Classification (BC2) provide entries for the same term under multiple (disciplinary) main classes, and thus eschew much of the potential of synthetic notation.

3.2: The domain-specific classificationist. The analysis above suggests a set of questions that could usefully guide the developers of domain-specific classifications. In each case, these questions would serve both to enhance the quality of a domain-specific classification while facilitating its connection to a universal classification:

- Can the meanings attached to terms by members of a domain be placed in some logical structure? That is, can one distinguish each term from others that have different meanings? If so, can these various terms then be placed in some sort of classification that serves to clarify their meaning? (I would suggest that if this cannot be done then the domain in question suffers from needlessly ambiguous terminology. If so, information scientists might move toward a less familiar advocacy of clarity within the domain.) Of course, the very idea of a domain-specific classification denotes some effort at classification (though a domain-specific thesaurus might not). Yet existing classification systems differ greatly in their adherence to any logical structure.

- Do the terms reflect some ontological view of the world? If so, how does this compare to ontological views in other domains?
• Do the terms reflect some epistemological view of how scholarship should proceed? If so, how does this compare to epistemological views in other domains?
• If terms can be placed within a classification, how does this classification compare to some universal classification? Note that such a comparison may serve to clarify the domain-specific classification. It will certainly aid in identifying what are the key terminological characteristics of a particular domain. This will be useful information for scholars both within and beyond the domain in question.
• (How) can the terminology and structure of a domain be translated into the terminology of a universal classification?

More generally, whatever sort of classification we develop or investigate, we should ask what are the sources of linguistic ambiguity and how can these be lessened by classification.

References
Iyer H., 1995, *Classificatory structure: concepts, relations, and representations*, Indeks, Frankfurt M.