The future of classification: knowledge organization in physical and digital space

Vanda Broughton, Senior Lecturer, UCL

Current views of classification, particularly from outside the LIS field, are that it is variously obsolete, unnecessary, redundant, an artificial construct from the nineteenth century which is falsely hierarchical, inaccurate, and founded on wrong ideas.[1] Proponents of such thinking may also argue that, in reality, the world and information about it is chaotic, 'miscellaneous', unstructured, and that relationships between topics are random and transient,[2] so that structured organization systems have no useful function in managing or retrieving information.

Counter to this, one may observe that many areas of knowledge, particularly in the sciences, do display a natural structure with regular patterns of relationships between topics based on their characteristics. The identification of such patterns and regularities provides much of the basis for theorising about the physical world and the consequent formulation of scientific laws and paradigms, and incidentally gives us the conceptual foundation for classification, both scientific and documentary. Often the purpose of the theory is to build a model of the world, to understand the way in which the world works, and to infer from the data you have knowledge about the world which is as yet undiscovered. This kind of thinking dates back to the ancient world and the origins of natural science and philosophy, and is exemplified throughout history in the work of such thinkers as Aristotle, Ramon Lull (whose logic diagrams are regarded as an early form of programming),[3] Roger Bacon, John Wilkins, John Ray, Carl Linnaeus, Francis Bacon, Gottfried Leibniz (a sometime librarian, and another precursor of artificial intelligence),[4] Dmitri Mendeleev (with his periodic table of the elements, a near perfect classification), and in the twentieth century by information theorists and designers of classification systems including Paul Otlet,[5] H. E. Bliss[6]and S. R. Ranganathan.[7]

It should be conceded that in the social sciences, arts and humanities such naturally occurring structures are harder to discern, and attempts to impose them can be seen as inherently artificial and occasionally fantastic, as in the essays of J. L. Borges,[8] himself a library practitioner in his early life. Man-made classifications (although not necessarily bibliographic ones) have more recently been regarded as potentially socially and politically dangerous, where they privilege certain viewpoints and perceptions of the world.[9]

---

1. David Weinberger. *Everything is miscellaneous; the power of the new digital disorder*. New York: Times Books, 2007

2. Clay Shirky. *Ontology is overrated: categories, links and tags*. http://shirky.com/writings/herecomeseverybody/ontology_overrated.html

3. Martin Gardner. "The ars magna of Ramon Lull" in *Logic machines and diagrams.* Chicago: University Press, 1958. pp. 1-27; and Anthony Bonner. *What was Lull up to?* http://www.ramonllull.net/sw_studies/studies_original/compbon.html

4.Jonathan Gray. "'Let us calculate'; Leibniz, Llul and the computational imagination" *Public domain review* https://publicdomainreview.org/2016/11/10/let-us-calculate-leibniz-llull-and-computational-imagination/

5. Boyd Rayward (Ed.) *International organization and dissemination of knowledge: selected essays of Paul Otlet.* Amsterdam: Elsevier, 1990

6. H. E. Bliss. *The organization of knowledge and the system of the sciences.* New York: Holt, 1929

7. S. R. Ranganathan. *Prolegomena to library classification.* London: Asia Publishing House, 1967

8. J. L. Borges. "The analytical language of John Wilkins" http://www.alamut.com/subj/artiface/language/johnWilkins.html

9. Geoffrey Bowker & Susan Leigh Star. *Sorting things out: classification and its consequences.* Cambridge, Mass.: MIT, 1999

Nevertheless, in the efficient management of information some systems of ordering must be employed in order to organize large information stores, whether physical or digital, and to enable information retrieval through the allocation of metadata, and the use of search and browse mechanisms. In that context, an important role of the classification system is to represent the subject of documents, either in terms of their content, for retrieval purposes, or in the relationships between concepts and the structure of the subject domain, to support browsing. In an age increasingly attached to the visualization of data, organized 'maps' of information play an important part in access to that information, and classificatory structures supply the underpinning of such tools.

Modern classification theory dates from the beginning of the twentieth century, although its roots can be found much further back in history, and in a variety of disciplines. Writers such as Bliss, Otlet and Ranganathan built on ideas from philosophy, logic, languages, science, and mathematics to explore the nature of classes, concepts, and terminology, to look at structures and relationships, and to consider the representation of subjects through coding and notation, most precisely formalised in Ranganathan's triple of the idea plane, the verbal plane, and the notational plane.[10] The close interrelationships between these different aspects of the information domain help to triangulate the theory and confirm its existence in different manifestations of knowledge organization systems, or KOS. The development of such a proper body of theory for classification gives us the philosophical rationale for knowledge organization, together with a methodology for building classificatory structures, whether they be classification schemes per se, thesauri, subject heading lists, or, more recently, taxonomies and ontologies. The common ground between these different kinds of tools reinforces the idea of a generally applicable fundamental theory, a sort of documentary 'theory-of-everything', and, in the twenty-first century, work on the semantic web and on web ontology has drawn on this earlier library and information science research to inform automatic indexing and retrieval and machine reasoning.

Although there are different schools of thought about classification theory, the most significant in the twentieth century has probably been faceted classification, initially conceived by Ranganathan, and further developed by the Classification Research Group in the UK,[11] and the Documentation Research and Training Centre in Bangalore.[12] Faceted classification has influenced nearly all modern KOS, in both traditional information organization and management, and in online information work.[13] We might ask what features of classification in general, and faceted classification in particular, make it especially suitable for contemporary knowledge organization. Firstly, it is very logical and predictable, with a clear, regular structure; this means that the classification data can easily be held in a database, as the relationships are explicit and identifiable. Secondly, the way in which the system works, its rules for building and linking (the system syntax) are also very regular and consistent, which makes it highly compatible with machine 'understanding' of the system. A faceted system is very like a natural language, except without all the irregularities and idiosyncrasies of natural language. When the rules for building are applied, as when classmarks are created for complex subjects as they occur in documents, a very elaborate structure can emerge, just as phrases and sentences are created in language. This is very much in line with the idea that Ranganathan was inspired by the engineering toy, Meccano, where very complicated models and machines can be created from a set of fundamental parts.

---

10. Frank Exner. "Ranganathan's three planes of work." http://associates.ucr.edu/306cexn.htm

11. *Classification Research Group* https://en.wikipedia.org/wiki/Classification_Research_Group

12. Documentation Research and Training Centre http://www.isibang.ac.in/

13. Vanda Broughton "The need for a faceted classification as the basis of all methods of information retrieval" *Aslib proceedings* 58 (1/2) 2006. pp. 49-72 10.1108/00012530610648671.

Whatever one's opinion of the correctness of classification schemes, there is not much evidence that physical collections of documents intend to abandon them, although it is true that in recent times additional methods of retrieval have been introduced to many libraries.

What does seem to be the case is that there are now fewer schemes in operation, with the big two, Dewey Decimal Classification (DDC) and the Library of Congress Classification (LCC), being dominant. In a recent survey of UK academic libraries,[14] of the first fifty institutions in the universities' league table, 20 used DDC, and 18 LCC. Six libraries used a local scheme, and the remaining six used Garside (2), Universal Decimal Classification (2), Bliss (1), and the National Library of Medicine classification (1). The age of the large scale re-classification project appears to be over, most of these having been from a local or less popular scheme to DDC or LCC, and few are reported in the professional press in the last five or six years. What is apparent is that many multi-site libraries, formed from the merger of institutions with a variety of schemes, are now more tolerant of that variation, with 15 libraries showing evidence in the main catalogue of multiple classification schemes. Partly this may be explained by the lack of financial or human resources for reclassification, but the role of other tools in retrieval removes dependence on the classification scheme for this purpose. Of the 50 libraries, 48 were using Library of Congress Subject Headings, doubtless because in many cases these came as standard in imported or outsourced bibliographic records. A more conscious choice would be for discovery tools, now used by all but one of the libraries surveyed.[15] The same convergence towards a small number of proprietary systems could be seen, with 33 libraries opting for Primo; Encore (9), Summon (5), and Ebsco (1) were the runners up, with one example of an in-house system. Such tools do not, of course, replace the more conventional means of representing subject content, since they rely on bibliographic data from the catalogue as the basis of search, a fact that seems sometimes to be missed by senior managers.

The classification scheme also remains essential for physical organization of books, and for browsing, and it is notable that many libraries continue to use classification for electronic resources such as e-books, and in some cases in institutional repositories.

The great majority of public libraries continue to use DDC in some shape or form (with the notable exception of LCC at Edinburgh), even where they may supplement this with a reader interest or categorization system in parts of the collection. In such cases the categories are often derived from publishers' classification data,[16] and a recent innovation has been the development of a more complex scheme, Thema,[17] which is being promoted as an international standard.[18] It has 2,700 classes, and can be used post-coordinately, making it amenable to 'faceted search'.

Such innovations are interesting, because, while they indicate dissatisfaction with older classification schemes, they demonstrate the continuing need for subject organization, metadata to facilitate search, the means of representing of complex content, and the enabling of more sophisticated search techniques. So while specific systems may have gone out of favour, the general concept of classification continues to be relevant.

---

14. Vanda Broughton "Classification and subject organization and retrieval" in J. H. Bowman (Ed.) *British librarianship and information work, 2011-15* www.lulu.com

15. V. Spezi, C. Creaser, A. O'Brien and A. Conyers, *Impact of library discovery technologies: A report for UKSG*. http://www.uksg.org/sites/uksg.org/files/UKSG_final_report_16_12_13_by_LISU.pdf

16. A. Hopkinson, "BIC and the E4libraries accreditation scheme" *SCONUL focus* 52 2011, 39-41 https://www.sconul.ac.uk/sites/default/files/documents/13_2.pdf

17. *Thema and UKSLC* http://www.bic.org.uk/86/THEMA/

18. Howard Willows. (2015) "Book description standard goes worldwide" http://www.bookbrunch.co.uk/article_free.asp?pid=book_description_standard_goes_worldwide

The same phenomenon can be observed in the wider, and wilder, unmanaged environment of the web. All the features of a knowledge organization system that demonstrate the good conceptual basis and sound theory developed in the last fifty or so years are highly relevant to subject organization and retrieval in non-library contexts. They can be seen in equivalent tools, and in a parallel strand of literature, and some particular features of classification schemes can be observed as key in various types of tool.

One of these is what is commonly referred to as faceted browse, in which a website interface is set up as a browsing tool by the organization of its content as a series of distinct 'facets' characterized by different attributes. This type of structure is very common in e-retail sites where the combination of the different facets allows the customer to formulate a search for quite exact products, such as a variety of wine specifying the colour, degree of dryness, country of origin, and so on.[19]

At a more complex level, the same methodology enables the building of a theoretical model for a domain for use in, for example, an organizational taxonomy, an educational website, or a modern multi-purpose knowledge organization tool.[20] Faceted structures and faceted techniques are evident in a range of 'non-classifications' that all have underlying classificatory functions. Standards and specifications for thesauri, topic maps, search software packages, taxonomies and ontologies all show evidence of the influence of faceted classification theory.[21] Of particular interest is the Semantic web application SKOS (Simple Knowledge Organization System) which aims to represent controlled vocabularies using a web ontology language.[22] SKOS identifies a range of structures and relationships inherent in classifications (notably faceted classifications) such as hierarchy, broader, narrower, and associative relationships, membership of facets and sub-facets, and values of concepts in array. Use of the SKOS version not only enables straightforward search and retrieval, but like other ontology work supports intelligent search, since the search software includes an inference engine which allows the system to make logical deductions about relationships in the domain on the basis of its knowledge of the system rules. Like Lull's logic diagram, and Mendeleev's Periodic Table before it, the ontology allows the system to use the knowledge it has to predict as yet unknown knowledge.

Overall, the need for careful analysis of concepts in the domain, and the rigorous and logic structuring of these concepts is the key to effective access to resources in a way that can be easily understood by both humans and machines. The theory that has been developed for the design and construction of the traditional bibliographic classification scheme is a vital foundation for the creation of these more contemporary tools, and to that end the thinking of library and information scientists over the last hundred years, and the lessons learned in classification lectures in library school, remain relevant to a new century and a new generation.

---

19. Peter Merholz, 'Innovation in classification' 23.09.2001 http://www.peterme.com/archives/00000063.html and https://www.wine.com/list/wine/7155

20. Mike Atherton. "Domain modelling at the BBC" (http://www.slideshare.net/reduxd/domain-modelling-at-the-bbc?related=1)

21. Stella G. Dextre Clarke, 'ISO 25964: a standard in support of KOS interoperability', in Alan Gilchrist & Judi Vernau (eds.), *Facets of Knowledge Organization; 2011 Jul 4-2011 Jul 5; London*. London: Emerald; 2012. Also available at: http://www.iskouk.org/conf2011/papers/dextreclarke.pdf

22. SKOS (Simple Knowledge Organization System) http://www.w3.org/2004/02/skos/