

Clare Beghtol

Faculty of Information Studies, University of Toronto, Ontario, Canada

Naïve classification systems and the global information society

Abstract: Classification is an activity that transcends time and space and that bridges the divisions between different languages and cultures, including the divisions between academic disciplines. Classificatory activity, however, serves different purposes in different situations. Classifications for information retrieval can be called “professional” classifications and classifications in other fields can be called “naïve” classifications because they are developed by people who have no particular interest in classificatory issues. The general purpose of naïve classification systems is to discover new knowledge. In contrast, the general purpose of information retrieval classifications is to classify pre-existing knowledge. Different classificatory purposes may thus inform systems that are intended to span the cultural specifics of the globalized information society. This paper builds on previous research into the purposes and characteristics of naïve classifications. It describes some of the relationships between the purpose and context of a naïve classification, the units of analysis used in it, and the theory that the context and the units of analysis imply.

1. Introduction and background

The growth of electronic storage and retrieval, particularly through the Internet, necessitates the re-examination of classification systems for their usefulness in the global information society. Traditionally, the objects of classification have been books or textual documents of various kinds. Now, however, the objects that require classification for information retrieval can be representations of non-textual virtual documents such as musical scores, theatrical performances, films, or paintings. It is thus useful to consider the possibility of expanding the concept of literary warrant, upon which bibliographic classification systems are customarily based, to include the concept of “artefact” or “object” warrant. This expansion may help span divisions between information retrieval classifications and academic disciplines, both of which habitually cross national, cultural, and linguistic boundaries.

Scholars in all disciplines regularly use classification systems to analyse and organize the raw data of their research, and these data are often based on inventories of physical artefacts or objects. Previous research described some of the specific purposes and characteristics of these scholarly classification systems and used the term naïve classification to denote classification systems created by people who have no particular interest or training in classificatory design and implementation (Beghtol, 2004). The purpose of these naïve classifications is to advance the particular discipline by discovering new knowledge or by re-examining old knowledge. In contrast, the purpose of professional information retrieval classifications is to group classifiable entities so that information seekers can discover previously-known and previously-recorded knowledge. The goal of this paper is to examine naïve classifications further in order to analyse the relationships between the purpose and context of a naïve classification, the units of analysis it uses, and the theoretical foundation that these elements imply. These goals help advance research into common classificatory structures and content that may be globally useful.

2. The relationship between purpose and context for classification systems

Langridge (1969) used various organizations of knowledge (e.g., Adler's *The Great Ideas: A Synopticon of Great Books of the Western World*) to call his students' attention to the importance of the purpose of a classification system. One of the determinants of purpose for a classification system is generally the context from which it arose and in which it is to be used. In the case of both naïve and professional classifications, the purpose is directly related to the context for which the classification was developed. Many naïve classifications are developed in the context of scholarly research in a particular area, and professional classifications are developed in the context of information seeking and use by those who need to find information about some particular area. Previous research (Beghtol, 2004) found that naïve and professional classifications used the same four structural principles that had been identified by Kwasnik, i.e., hierarchies, trees, matrices, and facets (1999). Nevertheless, the contents of the naïve and professional classifications are strikingly different from each other. For example, the place of ship building and design in a discipline-based bibliographic classification differs markedly from the position of ship building and design in a scholarly investigation of Aegean ship design during the Bronze and Iron ages (Wedde, 1996).

Altman (1967) carried out an analysis of the specific points at which choices must be made during the development of classification systems of all kinds. Choices were required about: 1) the level at which to analyse the objects undergoing classification (i.e., unit of analysis and level of detail needed for that unit of analysis); 2) the overall conceptual structure of the system; 3) the number of dimensions (i.e., facets) to use in the system; and 4) the methods of testing the adequacy of the system (including, for example, assessment of mutual exclusivity, joint exhaustivity, hospitality, and acceptance by the community in which it is going to be used). These choices are all necessary in both professional and naïve classifications. This paper focuses on the first two elements (i.e., 1) unit of analysis and level detail at which to classify and 2) the overall conceptual structure of the system) in order to illuminate the relationships between purpose, context and units of analysis in relationship to boundary crossing in the globalisation of information. The similarities and differences of these two classificatory choices in naïve classifications can help illuminate issues for globally useful knowledge organization and information retrieval systems.

3. Relationships in naïve classifications: purpose, context and units of analysis

Naïve classifications are often developed for physical objects, either virtual or actual (i.e., electronic or represented in some other way), and these classifications have specific purposes in their particular contexts. The two examples are only two of an apparently unlimited number of naïve classifications. These two examples are from the humanities, but similar examples from the social sciences are also available. Unlike the natural sciences, the humanities and social sciences are not constrained by the attributes of the physical world, so these two disciplinary divisions are appropriate for initial exploration of naïve classification systems and their characteristics.

First, Katritzky (1997) analysed the costumes worn by the Harlequin figure as depicted in Renaissance pictures during the lifetime of the actor Tristano Martinelli (1557 – 1630), who is generally credited with creating the Harlequin role. Katritzky developed the classification of Harlequins in costume as represented in pictures in order to help establish how the familiar patched Harlequin costume developed and whether or not only Harlequin figures wore it. This iconographic classification has four categories: 1) identified Harlequins; 2) dubious Harlequins; 3) probable Harlequins; and 4) possible Harlequins. The unit of analysis for this system was, thus, Harlequin costumes shown in pictures, and the level of detail at which the costumes were analysed was whether the costumes were patched or not. The purpose of the classification was to help trace the history of the Harlequin costume, and the context of the classification was theatre history as depicted in Renaissance paintings and drawings of Harlequins. The

conclusion Katritzky reached from this classificatory exercise was that not all Harlequins wore patches and that not all patched costumes were worn by Harlequins. In addition, other elements of Harlequin costumes were identified (e.g., bells on collar points). In the context of theatre history, this classification helped expand knowledge about one important figure in theatre history, the Harlequin. Relating visual and written documentation is one of the fundamental concerns of theatre iconography, and analysis of Harlequin costumes addressed this concern by the creation of a classification of pictures of the Harlequin figure.

Second, Belkin (1994) analysed drawings that Rubens (1577 – 1640) collected and kept in his own studio and argued that one can infer the organization Rubens used for these drawings by looking at groups of drawings that were not broken up after Rubens' death. The purpose of this speculative analysis was to show how Rubens organized his own drawings' collections. The objects of Belkin's examination were the drawings that have been kept together, and the unit of analysis was the elements (e.g., parts of the human anatomy) depicted in the drawings. From these classificatory elements, Belkin concluded that Rubens classified the drawings by subject, not by school, geography, or artist, which was the more common way of classifying works of art at that time. Thus, Rubens' "method of collecting and system of classification were primarily determined by their usefulness in the studio" (1994, 111) and the drawings were used to provide models for himself and his students. The example of Rubens' drawings furthered scholarly knowledge of artistic practice in Rubens' studio in the same way that classification of Harlequin costumes furthered scholarly knowledge of theatre history.

Both the Harlequin and Rubens examples thus highlight the typical concerns and methods of naïve classificatory activities within one academic discipline, in this case the history of the Renaissance arts of theatre and painting. These two naïve systems are concerned with objects represented through the medium of drawings and paintings, and their purpose is to discover new knowledge that will, in turn, enhance our understanding of the period they cover and advance our knowledge of the past world. In contrast, professional classification systems for information retrieval are customarily concerned with the classification of the world of pre-existing knowledge—classification of ideas and concepts that have appeared in the literature rather than of objects (either actual or virtual) that can be used for knowledge discovery.

In both kinds of classification, however, the use of classificatory activity is informed by the purposes, context, and units of analysis from which they arise. In both cases, too, we can demonstrate "the interdependence between theory and method by showing how the methods underlying even a relatively simple analysis...are imbued with theoretical import" (Schiffirin, 1997, 75). The same point was made by Kwasnik (1993). Like naïve classifications, professional information retrieval classifications reveal intricate relationships with the cultures, the purposes, and contexts in which they operate. These interrelationships show how naïve and professional classifications systems are similar to each other in some ways. Both kinds of classification start with a basic inventory of the classifiable entities. As previous research has shown, however, naïve and professional classifications differ from each other in that naïve classifications are commonly based on shallow hierarchies developed from a small number of examples, while professional classifications commonly have complex hierarchies based on large numbers of examples (Beghtol, 2004).

4. Conclusion

Classificatory activity is a cognitive universal (Beghtol, 2000), but creating a classification serves different purposes in different contexts. Nevertheless, the two kinds of classification discussed here are can be seen to depend on each other. Communication among scholars is initially facilitated by means of naïve classificatory activity. This communication starts at a basic level using available evidence in the field. Eventually, the accumulation of classificatory activity and evidence enters the literary warrant of the field and in some cases presumably leads to consensus among scholars in that field. Literary warrant and consensus lead in turn to

incorporation of the scholars' original ideas in a professional information retrieval classification. Eventually, the influence of retrieving relevant documents via the information retrieval classification helps other scholars develop their own naïve classifications. Thus, the ability to retrieve the evidence of classificatory activity in naïve classifications allows these classifications to join the world of public knowledge because they have been recorded, communicated to others and disseminated through a professional classification. This cyclical relationship between the two kinds of classification has not been studied extensively, but it deserves closer examination. In particular, we need to study the effects of expanding the concept of literary warrant to include classificatory examination of artifact or object warrant. Understanding the different purposes, materials, and contexts of naïve and professional classifications helps advance our knowledge of how classification can span boundaries between cultures and between disciplines in a globalised information society.

Acknowledgements

This research was partially supported by the Social Sciences and Humanities Research Council of Canada grant 410-2001-0108. Research assistance was provided by Ann Simonds.

References

- Altman, I. (1967). Choicepoints in the classification of scientific knowledge. In: *People, groups, and organizations*, ed. Bernard P. Indik & F. Kenneth Berrien. New York: Columbia University, Teachers College Press. Pp. 47-69.
- Beghtol, C. (2004). Classification for information retrieval and classification for knowledge discovery: relationships between 'professional' and 'naïve' Classifications. *Knowledge Organization* (in press).
- Beghtol, C. (2000). A whole, its kinds, and its parts. In: *Dynamism and stability in knowledge organization. Proceedings of the Sixthth International ISKO Conference, Toronto, Canada, 10-13 Jul 2000*. Würzburg: Ergon. Pp. 313-319.
- Belkin, K.L. (1994). The classification of Rubens's drawings collection. *Wallraf-Richartz-Jahrbuch, Jahrbuch für Kunstgeschichte*, 55, 105-114.
- Katritzky, M.A. (1997). Harlequin in Renaissance pictures. *Renaissance Studies*, 11(4): 381-419.
- Kwasnik, B.H. (1999). The role of classification in knowledge representation and discovery. *Library Trends*, 48(1) Summer: 22-47.
- Kwasnik, B.H. (1993). The role of classification structures in reflecting and building theory. In: *Advances in Classification Research 3: Proceedings of the 3rd ASIS SIG/CR Classification Research Workshop*. Medford, NJ: Learned Information. Pp. 63-81.
- Langridge, D.W. (1969). Investigation 4. Kinds of knowledge. In: *The universe of knowledge: essays by members of the special seminar held during the fall semester, 1967*. College Park, MD: School of Library and Information Services. Pp. 90-96.
- Schiffrin, D. (1997). Theory and method in discourse analysis: what context for what unit? *Language & Communication*, 17(2): 75-92.
- Wedde, M. (1996). From classification to narrative: the contribution of iconography towards writing a history of early Aegean shipbuilding. *Mediterranean Historical Review*, 11(2): 117-164.

