

Centripetal and Centrifugal Forces in Bibliographic Classification Research

D. Grant Campbell
Faculty of Information and Media Studies
University of Western Ontario
gcampbel@uwo.ca

1. INTRODUCTION

In the middle of *Weaving the Web*, Tim Berners-Lee's (2001) account of the growth of the World Wide Web, the author makes a startling admission. After pages of advocating the benefits of non-hierarchical, associative linking technologies that can facilitate the "tentative connective relationships that [define] much of our society's workings" (p. 5), Berners-Lee describes his early efforts to keep track of the different servers appearing on the nascent Web:

When the list became larger, it needed to be organized, so I arranged it in two lists, by geography and by subject matter. As more servers arrived, it was exciting to see how the subjects filled out. ... [A student] set up the lists into what we called the Virtual Library, *with a tree structure that allowed people to find things.*" (2000a, p. 55) [Emphasis added.]

This account suggests two things about classification. First, formal systems based on widely-accepted standards of grouping and hierarchy continue to have their place in schemes of information and knowledge organization, even in cultures that resist the constrictions of such hierarchies. Second, the information revolution brought about by the Internet and the World Wide Web is settling into a counterpoint relationship familiar to classification researchers: a relationship between top-down organization, in which a definable knowledge domain is divided into enumerated hierarchies, and bottom-up synthesis, in which a clearly-defined facet structure orders data into whatever combinations are required, whether or not they were previously foreseen.

Bibliographic classification brings a rich history to this counterpoint relationship, and therefore stands both to give and to receive benefits from the developments in information and knowledge organization that lie ahead. This paper discusses what I consider to be two fundamental forces at work as bibliographic classification seeks to redefine itself in a new information environment: the *centripetal* force by which the field draws in concepts and methods from other fields in order to redefine and answer its recurring questions, and the *centrifugal* force by which concepts from bibliographic classification detach themselves from our field and move outwards to play an important role in a broad range of fields.

2. CENTRIPETAL FORCES: DEFINING OUR FIELD

Classification research today negotiates two mutually-dependent but antithetical approaches. One approach argues that classifications are cultural productions of diverse information ecologies: "within such [an] information ecology," argue Albrechtsen and Jacob, "a classificatory structure cannot follow a one-size-fits-all paradigm, but must evolve in cooperative

interaction between librarians and their user groups” (Albrechtsen & Jacob, 1998, p. 1). The other argues that even the most culturally-determined classification depends on certain “universal” concepts that form the foundation of any system for any culture (Beghtol, 2000, p. 313).

Both sides of the antithesis are drawing on other domains to enhance their work. Scholars defining the universal concepts, working as they are in a complex area of meaning and signification, draw on epistemology, linguistics and semiotics. Indeed, much of the work in this area resembles structuralism in its treatment of human artefacts, such as classification systems, as self-contained meaningful signs, “with an underlying system of distinctions and conventions which makes this meaning possible” (Culler, 1975, p. 4). Philosophy, literature, anthropology, sociology and linguistics all provide rich ground for concepts and methods that can be used to isolate and understand the principles of grouping and differentiation that underlie classification schemes of all types and contexts. As a result, knowledge organization has drawn on the writings of text linguistics (Beghtol 1986), mentalism (Frohmann, 1993), text types (Beghtol, 1997), structuralist poetics (Campbell, 1998), pragmatist philosophy and the social studies of science (Hjorland & Albrechtsen, 1999) semiotics (Mai, 2001) and Bakhtinian chronotope theory (Campbell, 2002) in order to clarify, and sometimes to challenge, the central concepts of the field.

Scholars who focus on the social context of classification draw on concepts from other societies and other domains of knowledge, both to broaden the applicability of existing classifications, and to release classification from some of its limitations. Theorists have drawn on indigenous cultures in North America for alternative principles of ordering (Olson, 1999); others have used mathematics and computer science to break the hold of the hierarchical tree structure and develop new methods such as concept lattices (Priss, 2000, p. 171).

The results of research in these two areas will undoubtedly lead to greater facility in the development of local and specific classification schemes. Equally important, the World Wide Web, with its capacity to present information in two dimensions rather than the single dimension of shelf space, has encouraged experimentation with system navigation and displays. More work is needed, particularly in concert with scholars in information architecture and information visualization, to explore how classification schemes can be adapted to new and vivid visual metaphors, made possible by enhanced graphics capability and tailored to a variety of information uses. Some metaphors will undoubtedly be unfamiliar to classificationists at this time; others, however, may emerge from our own field. It may be time, for instance, to revisit Bliss’s theory of the three-dimensionality of classification in a digital medium that has released classification from the constrictions of linear shelving and printed schedules.

Furthermore, this research into interface design and navigation will also contribute to the ongoing revision of our universal classification schemes such as Dewey Decimal Classification (DDC), Library of Congress Classification (LCC) and the Universal Decimal Classification (UDC). In particular, we need to explore further how the rise of graphically-sophisticated tools such as web-based thesauri, WebDewey and even the Cataloguer’s Desktop are affecting the classifying process in conventional library settings. Do graphical interfaces facilitate or impede the activity of scanning schedules? And is the process of number building conceptually and

practically enhanced by the multiple windows, lists and drag and drop capabilities of electronic schedules?

Hopefully, the end of this trend in research will be the collapse of the antithesis between social specifics and conceptual universals. Barthes considered the crossing of an antithesis to be one of the ultimate transgressions in human social behaviour (Barthes, 1974, p. 65), and this transgression could be a deeply productive one for classification research. Sooner or later, our increasingly sophisticated grasp of conceptual constants will open up a new level on which the biases and hidden narratives of classification can be explored. This breakthrough will enable us to query our classification systems, and other pervasive systems of meaning, in new and exciting ways.

3. CENTRIFUGAL FORCES: DISSEMINATING OUR FIELD

Bibliographic classification theory has traditionally concentrated on the process of ordering documents: discrete objects with information-bearing properties that can be described, ordered and stored, whether in physical or digital form. Recently, this concentration has come under scrutiny: Clare Beghtol has suggested that current information retrieval techniques “call into question whether a whole document (e.g. book, article) is the most appropriate unit of analysis in online retrieval systems” (1998, p. 8).

If we allow our classification techniques to move beyond the document, we find evidence of a centrifugal tendency that carries our classification concepts outward into new areas and fields that conceptualize information according to different base units. Classification activities are central to data management systems of all kinds, and the principles of bibliographic organization can create a number of exciting synergies.

3.1 Object-Oriented Computer Programming

Object-oriented programming is inherently classificatory, working as it does on the principle of objects, each of which is an instance of a class. Furthermore, this class system is hierarchical, and uses principles of inheritance to invest subordinate objects with the attributes of superordinate objects. Furthermore, with the rise of Java, object-oriented concepts are becoming central to Web authoring, with its emphasis on reusing programming modules in different contexts. Bibliographic classification theory is rich with concepts of hierarchy—consistent principles of subdivision, mutual exclusivity, and joint exhaustivity—which could significantly enhance the quality and efficiency of programming and of Web design. In particular, the distinctions between whole-part, genus-species and class-instance relationships, central to thesaurus design could be a useful lens through which to analyze the efficiency, coherence and limits of object-oriented paradigms. Does the principle of inheritance in object-oriented programming imply that only genus-species and class-instance relationships can be represented in object-oriented systems?

3.2 Relational Database Design

The relational database structure is ideally suited for facet analysis, providing as it does a series of rows and columns that can be ordered and sorted in a variety of ways. Facet analysis could conceivably supplement or join with normalization techniques to enhance the efficiency and conceptual clarity of database structures and the usefulness of different data displays. A comparative analysis of facet analysis techniques and entity-relationship and semantic object modeling might lead to new relational modeling techniques. Other studies could analyze the grouping and ordering capabilities of database queries and reports according to principles of facet order. These might lead to improvements in report wizard interfaces that enable the user to group and order data in more flexible and useful ways.

Furthermore, domain analysis in classification research has relevance for database design. Every attribute in a database relation has a defined domain, which articulates the range of all possible values. These domains, by defining what can and cannot be "said" in a particular field, have considerable intellectual, administrative and ethical significance. Classification research has worked extensively on the idea of "warrant":

The authority a classificationist invokes first to justify and subsequently to verify decisions about what classes/concepts to include in the system Warrant covers conscious or unconscious assumptions and decisions about what kinds and what units of analysis are appropriate to embody and to carry the meaning or use of a class to the classifier (Beghtol, 1986, p. 110-111).

Studies of warrant—whether literary, scientific, philosophical or educational—could provide a meaningful vocabulary in understanding the limits, intentional and unintentional, which we place on the representative power of our database systems. Following the lead of Bowker and Star, classification theorists could use the concept of warrant to analyze the domain specifications of supposedly neutral and value-free databases, uncovering the subtle rules and restrictions which give privileged voices to some concepts and speakers at the expense of others. Such studies would be consistent with existing work on the social context of classification, and would help us to see the hidden values that order our entrenched, partially-invisible information infrastructure (Bowker & Star, 1999, p. 33).

3.3 Markup Languages

Markup languages, whether format-based languages such as HTML or semantic markup languages based on SGML or XML, conceive documents as structures of hierarchically-arranged elements in whole-part relationships. These documents generally use element configurations defined in a Document Type Definition or XML Schema, and the range of possible element values are often defined by XML namespaces. Many digital libraries, therefore, are based on meaningful hierarchies that are based on document genre, rather than on document content. Classification theory, with its principles of subdivision and its distinction between parent-child and whole-part relationships, can clarify our understanding of these hierarchical paradigms (Campbell, 2000). Such principles should enable us to enhance the work of XML parsers, enabling them to recognize and check for principles of semantic validity beyond the mere

presence of well-formed XML. In addition, a comparison of XML hierarchies and database data structures could provide guidance in developing programs for migrating data into XML from other hierarchically-based systems. Does data, for instance, stored in an object-oriented database based on genus-species hierarchies translate easily into XML based on whole-part relationships?

3.4 Hypertext and Linking Technologies

Faceted classification has always been concerned with linking concepts through notation, and systems such as the Colon Classification and UDC use sophisticated and informative facet indicators to indicate the nature of the link between one facet and another. While completely faceted notation is not common in North American libraries, the rise of hypertext has created a fresh need for facet indicators. As our linking languages grow more sophisticated, with the rise of Extensible Linking Language (XLL), we need new methods for indicating complex relationships between the nodes. The work of Ranganathan, the Classification Research Group and others on facet indicators may provide a relevant foundation for technologies that indicate those relationships. Future studies need to explore the feasibility of expressing document relationships through links, in a method similar to the punctuation in the Colon classification. They also need to investigate whether such expressions could be used in search engine design to rank and filter multiple pages retrieved from one domain, perhaps highlighting introductory and contact pages at the expense of bibliographic or secondary pages.

3.5 Information Architecture

The field of information architecture has recently become enamoured of faceted classification techniques as a means of solving recurrent problems with site organization. Initiatives such as the Flamenco Project have shown that principles of facet analysis can be incorporated into web design to facilitate browsing. Flamenco uses faceted metadata on sites such a digital library of architectural images: metadata by which users can browse for architectural images by isolating and organizing the specific categories into any order desired. When these categories are applied to a search, the designers have found that users prefer dynamically-organized faceted metadata to the conventional search interface (English et al., 2002). Such a practice, if implemented on a widely directory such as Yahoo! could revolutionize Web browsing.

3.6 Metadata and the Semantic Web

In the ambitious visions of the World Wide Web Consortium, a new information society is approaching, one in which Web information will be machine-understandable as well as human-understandable, transforming the current World Wide Web into a "Semantic Web."

The Semantic Web, as envisioned by Tim Berners-Lee (2000b), will emerge as a series of levels, working from the bottom up: building on the standard Internet protocols, as well as XML, XML schema and XML namespaces, individuals can create "self-describing" documents using the Resource Description Framework (RDF). This RDF metadata will be designed to identify and make simple statements about resources, statements which will help machines disambiguate between different items of the same name, place information and context, and create more meaningful information interactions. These RDF statements will be combined at a higher layer

of ontology vocabularies, enabling multiple meaning systems to interact with each other with some degree of accuracy (Berners-Lee, 2000b).

The concepts of the Semantic Web have attracted a sceptical reaction from some in the Web design community, who argue that the ideas are idealistic rather than realistic, similar to the dreams of artificial intelligence in the 1960s (Dumbill, 2000). Classification research may have a role to play in determining just how possible these ideas are. As a layered system of meaning and communication, in which each layer builds on the layer beneath it, the Semantic Web resembles classification procedures that are founded on particulate concepts combining to create new ones. In this case, the theory of integrative levels may provide a frame for defining how these layers of the Semantic Web could interact, particularly the thorny problems of combining the RDF layer with the ontology layer. As a theory which defines meaningful aggregation, and in particular the moment when parts come together to synthesize something that is greater than their collective sum, integrative levels theory could play a significant role in the development of the Semantic Web.

In addition, our existing knowledge of bibliographic relationships, both in terms of works and authors, could play a large role in the development of ontologies that enable RDF statements to be harvested across different domains. These ontologies will be based on the collection of categories from individual namespaces into "super-categories." Such collection may be clarified by Smiraglia's definition of the bibliographic entity as a combination of document and work, as well as Tillet's work on the primary entities of bibliographic catalogues, and the relationship devices, such as cross-references, that are used to link entities together (Smiraglia, 2001; Tillet, 1989). A comparative analysis of prototype RDF statements and authority files might suggest ways in which the massive entity definitions of the library community, particularly its authority records for names, titles and subjects, could be adapted to the Semantic Web.

4. CONCLUSIONS AND FUTURE RESEARCH

This brief survey has isolated the following potential areas for future classification research:

- Translating Bliss's concept of 3-dimensional classification into graphical interfaces;
- Studying the effect of graphical interfaces for classification schedules on the classification process;
- Analyzing the hierarchical relationships in object-oriented programming;
- Comparing facet analysis to relational database modeling techniques, such as entity-relationship and semantic object modeling;
- Using facet analysis to develop and test new database design wizards for queries and reports;
- Using the concept of warrant as the basis of a discourse analysis of classification infrastructures;
- Using concepts of hierarchy and subdivision to analyze the migration of data structures based on whole-part and genus-species relationships into XML structures based on whole-part relationships;
- Using classification concepts to enhance XML parsers;

- Applying the theory of integrative levels to facilitate the integration of namespaces and RDF statements into ontologies on the Semantic Web;
- Applying facet analysis to the design of large-scale Web directories and portals;
- Adapting the existing infrastructure of bibliographic entities for use on the Semantic Web.

The challenges facing bibliographic classification research, then, involve a double movement. On the one hand, researchers need to continue to refine and entrench the core field, by bringing in concepts from other disciplines and using them to negotiate the politically-charged issues of creating classification systems for document access for diverse domains, communities and cultures. This process will involve finding new metaphors and paradigms for information organization, and by heightening our awareness of voices and groups that are excluded by supposedly universal schemes, particularly through the sophisticated understanding and use of universal classification concepts.

On the other hand, researchers need to move outward from the field to put classification concepts to use in other fields of information organization and use. The results of such research will be more diffuse, as classification integrates with such areas as database theory, markup languages and website design. In this centrifugal activity, classification research, positioned as it is at the juncture of human and system, will be able to bring a necessary awareness of philosophy, ethics, sociology and linguistics to the development of information systems.

5. REFERENCES

- Albrechtsen, H. & Jacob, E.K. (1998). Classification systems as boundary objects in diverse information ecologies. In Efthimiadis, E. (Ed.) *Advances in classification research, Vol. 8. Proceedings of the 8th ASIS SIG/CR Classification Research Workshop, Washington D.C., Nov. 1-6, 1997* (pp. 1-13). Medford, N.J.: Information Today.
- Barthes, R. (1974). *S/Z: An essay* (Miller, R. Trans.). New York: Hill and Wang.
- Beghtol, C. (1986). Bibliographic Classification Theory and Text Linguistics: Aboutness Analysis, Intertextuality and the Cognitive Act of Classifying Documents. *Journal of Documentation* 42(2): 84-113.
- Beghtol, C. (1986). Semantic validity: Concepts of warrant in bibliographic classification systems. *Library Resources and Technical Services*, April/June 1986: 109-125.
- Beghtol, C. (1997). Stories: Applications of narrative discourse analysis to issues in information storage and retrieval. *Knowledge organization*, 24(2): 64-71.
- Beghtol, C. (1998). Knowledge domains: Multidisciplinarity and bibliographical classification systems. *Knowledge organization*, 25(1/2): 1-12.
- Beghtol, C. (2000). A whole, its kinds, and its parts. In Beghtol, C., Howarth, L. & Williamson, N. (Eds.) *Dynamism and stability in knowledge organization: Proceedings of the Sixth International ISKO Conference, Toronto, Canada, 10-13 July 2000* (pp. 313-319). Wurzburg: Ergon Verlag.
- Berners-Lee, T. (2000a). *Weaving the Web: The original design and ultimate destiny of the World Wide Web*. With Mark Fischetti. New York: HarperBusiness.
- Berners-Lee, T. (2000b). Semantic Web on XML. PowerPoint presentation delivered to the *XML 2000 Conference, December 12, 2000, Washington, D.C.* Retrieved from <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide1-0.html>

- Bowker, G.C. and Star, S. (1999). *Sorting things out: Classification and its consequences*. Cambridge: MIT Press.
- Campbell, D.G. (1999). Discourse analysis and document analysis: SGML, scholarly publishing and structuralist theory. *Information Science at the Dawn of the Next Millennium: Proceedings of the 26th Annual Conference of the Canadian Association for Information Science*. Halifax: CAIS: 198-208.
- Campbell, D.G. (2000). The relevance of traditional classification principles to the development and use of semantic markup languages for electronic text. In Beghtol, C., Howarth, L. & Williamson, N. (Eds.) *Dynamism and stability in knowledge organization: Proceedings of the Sixth International ISKO Conference 10-13 July 200, Toronto, Canada* (pp. 345-351). Wurzburg: Ergon Verlag.
- Campbell, D.G. (2002). Chronotope and classification: How space-time configurations affect the gathering of industrial statistical data. In López-Huertas, M. (Ed.) *Challenges in knowledge representation and organization for the 21st century: Integration of knowledge across boundaries: Proceedings of the Seventh International ISKO Conference, 10-13 July 2002, Granada, Spain* (pp. 318-323). Wurzburg: Ergon Verlag.
- Culler, J. (1975). *Structuralist poetics: Structuralism, linguistics and the study of literature*. London: Routledge and Kegan Paul.
- Dumbill, E. (2000). Berners-Lee and the Semantic Web vision. *XML.com*, Dec. 6, 2000. Retrieved from <http://www.xml.com/pub/a/2000/12/xml2000/timbl.html>.
- English, J., Hearst, M., Sinha, R., Swearingen, K., & Yee, K. (2002). Hierarchical faceted metadata in site search interfaces. *CHI 2002 Proceedings*. Retrieved from <http://bailando.sims.berkeley.edu/flamenco.html>.
- Frohmann, B. (1990). Rules of indexing: A critique of mentalism in information retrieval theory. *Journal of documentation* 46(2): 81-101.
- Hjorland, B. & Albrechtsen, H. (1999). An analysis of some trends in classification research. *Knowledge organization* 26 (3): 131-139.
- Mai, J. (2001). Semiotics and indexing: An analysis of the subject indexing process. *Journal of documentation* 57(5): 591-622.
- Olson, H. (1999). Cultural discourses of classification: Indigenous alternatives to the tradition of Aristotle, Durkheim and Foucault. In Albrechtsen, H. & Mai, J. (Eds.), *Proceedings of the 10th ASIS SIG/CR Classification Research Workshop, October 13, 1999*.
- Priss, U. (2000). Comparing Classification Systems using Facets. In Beghtol, C., Howarth, L. & Williamson, N. (Eds.) *Dynamism and stability in knowledge organization: Proceedings of the Sixth International ISKO Conference 10-13 July 200, Toronto, Canada* (pp. 170-175). Wurzburg: Ergon Verlag.
- Smiraglia, R. (2001). *The nature of "a work": Implications for the organization of knowledge*. Lanham: Scarecrow Press.
- Tillett, B. (1989). Bibliographic structures: The evolution of catalog entries, references, and tracings. In Svenonius, E. (Ed.) *The conceptual foundations of descriptive cataloging*. San Diego: Academic Press.