Bringing People, Technology, and Systems Together Through Classification Research: Designing for Change, Learning, and Maintenance

Paul Solomon
School of Information and Library Science
University of North Carolina at Chapel Hill
solomon@ils.unc.edu

1. INTRODUCTION

This 13th SIG/CR Classification Workshop provides an opportunity for reflection on the challenges that we face as researchers in an applied field. In considering the papers of the previous 12 American Society for Information Science and Technology (ASIST) Special Interest Group (SIG) in Classification Research (CR) Workshops as well as those of the International Society for Knowledge Organization (ISKO) Conferences from the point of view of systems thinking (Emery, 1969), several themes (among others) emerge. These include 1) the tension that exists between stability and change in systems of knowledge organization, 2) the view of knowledge organization schemes as vehicles for learning and exploration as people engage in information search, and 3) the problem of maintenance of knowledge organization schemes as the world turns. These themes are all related through the challenge that we face in designing structures (e.g., metadata arrays, classifications, displays) that not only support people's tasks and problem solving, but which include self-referential actions to renew, update, or recreate the structures themselves as new technologies, ideas, and inventions arise and old ones fade from consideration.

The following explores these three themes after reflecting on the need for grounding knowledge organization and classification research in an understanding of people's information intensive tasks, problems, and situations.

2. WHY KNOWLEDGE ORGANIZATION AND CLASSIFICATION RESEARCH?

In our inclination to organize, it sometimes appears that we forget why the organization function is critical within the information field. After all we don't organize for organization's sake, but to provide a foundation for something else. What we do is part of a process that begins with the production, selection, and collection of artifacts. Given the work of organizers of information with regard to representation (e.g., specification of metadata), classification (e.g., construction of classification schemes, index languages, thesauri, ontologies), and enabling display (e.g., arrangement, ordering, adaptive views) through representation and classification decisions, the organization function provides a foundation for other activities. When we think about this at all we most often stop with information retrieval as our *raison d'être*. As Taylor (1986) notes there are a variety of value-added processes, including analytical, judgmental, and decision support, that go beyond retrieval of relevant items from a collection to helping people complete tasks and deal with life's problems.

Besides the formal functions of information storage and retrieval (ISAR) systems to retrieve as many relevant items as possible and as few non-relevant items as possible (recall and precision)

and whatever additional values such systems might provide beyond information retrieval, research on people's use of relevance criteria as they interact with ISAR systems, suggests that people's interaction with the mechanisms of an ISAR system and items retrieved as well as their place in their information search influences what an individual searcher views as relevant (Kuhlthau, 1993; Tang & Solomon, 1998). That is, people are seldom looking for everything that is relevant to a task, problem, or situation and their interaction with retrieved items helps people learn and, consequently, further qualifies what is relevant at that stage in their search processes. This has been labeled by Kuhlthau (1993) as a shift from relevance to pertinence. Hert (2002) offered the important insight that such relevance judgment studies are at the same time metadata studies, which often provide important insights into system requirements and, thus, influence systems design.

> It seems that the more that we know about the role that representation, classification, and display plays in people's tasks, problems, and information search, the better job we can do in grounding the products of knowledge organization in the problems of real people and not the traditional assumptions of ISAR systems (Solomon, 1992).

The Getty Online Searching Project work of Bates and her colleagues (Bates, Wilde, & Siegfried, 1993; Siegfried, Bates, & Wilde, 1993; Bates, 1994; Bates, Wilde, & Siegfried, 1995) is a significant example of a body of work, which strives to understand the challenges of humanities scholars in accessing online materials through mapping of search terminology, searcher profiles, and research practices in order to design or redesign knowledge organization tools to better serve researchers in this domain.

Research in progress by Greenberg and Pattuelli provides a current example of this grounding. With the growth of the World Wide Web, many have put their faith in schemes incorporating metadata, ontologies, and XML. These mixes of strategies very much depend on the availability of metadata as the basis for retrieval or other processing. Recent work by Yang (2002) suggests that such metadata is often absent from web pages or otherwise deficient, and is, thus, a poor source of information for retrieval. Greenberg and Pattuelli have been conducting a study to test a system for structured support for author generated metadata tags as a means of ensuring the availability of quality metadata. This is an active step to test basic assumptions regarding the availability of quality metadata to support retrieval and other purposes.

In short, there is a basic need for grounding classification research in an understanding of the tasks and problems of people who depend on our efforts to organize information/knowledge as we consider how organizing is a means to a variety of ends.

3. STABILITY AND CHANGE

Knowledge organization, in essence, involves the design and development of structures (i.e., resources and rules). These structures, in turn, enable some actions and disable or restrict others. Enabled actions reinforce the enabling structures and possibly create new or adapted structures (Giddens, 1984: Luhmann, 1995). These ideas provide a broad foundation for considering

knowledge organization and classification research as their products accommodate structure in action.

For example, thesauri enable assignment of descriptors; descriptors enable controlled retrieval; retrieval failures suggest new or revised descriptors. From another point of view, people's interactions with the conceptual structures of thesauri and their descriptors may result in their learning something about the terminology in a field that leads to new insights or ideas with regard to their task or problem. Lykke Nielsen's (2000) work on domain analysis as a foundation for thesaurus construction is an example of how a broader understanding of contextual influences can provide an empirical foundation for knowledge organization schemes.

The world is in motion. As people interact with the world they learn, develop new ideas, and think of things with their current personal labels. Yet, many classification systems value stability and, thus, actively resist change—perhaps because of their rules and resources regarding change. Structures developed in times preceding the computer, the Internet, and the World Wide Web remain as the foundation for current systems.

Efforts to bridge the past and the present are welcome, as are those that consider how to support people in their information intensive tasks given the current state of technology. People's labels and those of current classification schemes often don't match. This leads at times to frustration and the inability of such systems to well serve the functions for which they were designed (Solomon, 1993). Consider too the scope and scale of the Internet and the likelihood that there will be thousands of hits in response to a query. There is, thus, a considerable continuing investment in information systems that do not quite fit with the way people employ information in life and work.

A current example arises in the formation of registries for metadata schemes, ontologies, taxonomies, and the like (e.g., the SCHEMAS Project, http://www.schemas-forum.org/; Microsoft's BizTalk, http://www.microsoft.com/biztalk/default.asp). The contents of these registries are often not as current as those in use by their creators. Users of the registries may end up using an out-of-date version. Yet, having these various instantiations of these tools available might provide a vehicle for helping people move from an older term or conceptualization to those currently in use in the system (or the reverse).

There seems to be an opportunity for classification research, in a time where we are no longer necessarily burdened by the physicality of shelving materials to provide dynamic mappings that help people look forward and backward from their place in an evolving world. In other words, there is the possibility of allowing people to select or create their own views of an information space to support their learning through interaction with what the ISAR system can provide.

4. LEARNING THROUGH INTERACTION

The above suggests that information systems need not be static, but may be designed to allow people to shape the information space to fit their own tasks, problems and situations. One area, which has attracted the attention of classification researchers in recent years, has been approaches for specifying context (e.g., domain, information use environment, community of

practice, information ecology and boundary objects, language games) as a foundation for organizing knowledge. Together these and other approaches to contextualization suggest a variety of views, which may provide the 'stuff' we need to gain the insight that we need to design knowledge systems that help people learn.

Contextualization is an important aspect of how people come to know or make sense of their situations. Blackler (1995) suggests 5 dimensions of knowing, which provide a frame of reference for considering the design of knowledge organization schemes that support interaction and learning.

- Knowing is *situated* in patterns of work and life. This involves understanding what information is to people and how people struggle to gain, interpret, and shape knowledge.
- Knowing is *mediated* by structures of communication. This involves understanding that ISAR systems have a power to shape what is understood.
- Knowing is *provisional* in that it highlights people's movement through time and space. That is, what we know now is embedded in the past and will be influenced by what we learn in the future.
- Knowing is *pragmatic* as a variety of purposes may be supported. A current interaction for a particular purpose may also be employed for other purposes.
- Knowing is *contested* in that understanding the nature of conflicting viewpoints may provide insights into building new insights from that disparity.

These dimensions offer insights into what knowledge organization schemes must be and do in order to support learning and other activities beyond 'just' retrieval. In short, we have an opportunity to expand our vision of the role of knowledge organization and classification research to consider the role of such systems beyond information retrieval as supports for learning and other higher order cognitive activities.

5. DESIGN FOR MAINTENANCE

Classification research to support the design of tools of knowledge organization is often a fun thing to do. On the other hand, dealing with the mundane aspects of operations and maintenance is often not fun. Yet, this is what makes or breaks the systems we design. Thus, there are seldom reports of research that focus on operational issues of maintenance of such tools. Yet the costs of maintenance are key issues in both the implementation and continuing operation of systems of all sorts. All knowledge organization schemes have costs associated with them. A challenge for classification researchers is to design with the cost of operation and maintenance in mind. This is something, which is often taken for granted in the euphoria of exploring a new idea for representing or classifying information. Yet, it may be the single most important stumbling block in dealing with the challenges of stability and change as well as creating systems that fit with people's tasks and problems by creating an environment for exploration and learning.

If the structures of knowledge organization are seen as being composed of resources and rules, design attention needs to be given to resources to support operation, maintenance, and update, and rules, which encourage self-reflection and adaptation as systems are employed. Self-reflection, evaluation, and update are often implicit in the thinking of designers. They must be

made explicit and their importance recognized by providing dedicated resources and system rules, which are part of the way the system operates.

6. CONCLUSION

Knowledge organization and classification research is not just the design of metadata or classification schemes, index languages, and the like but seeing that such products fit with people engaged in work and life today and in the future.

- Grounding such products in an understanding of people's tasks, problems, and situations is fundamental and a relatively untouched area for classification research.
- Figuring out how to accommodate the dynamic nature of the world in these products is another area ripe for classification research.
- Understanding how people's interactions with representations, classifications, and displays influence their knowing and understanding needs attention too (Jacob, 2001).
- Considering maintenance as an element of design is an oft forgotten operational detail.

Yet, if an information system's rules and resources do not encourage actions, which are reflective in their provision of feedback and evolution in rules and resources, evolution is unlikely. Research into the ultimate actions of maintenance and update are critical for effective treatment of all of the other research areas.

ACKNOWLEDGEMENTS

REFERENCES

Bates, M. J. (1994). The design of databases and other information resources for humanities scholars: The Getty online searching project report No. 4. *Online and CD-Rom Review, 18(December)*: 331-340.

Bates, M. J., Wilde, D. N., & Siegfried, S. (1993). An analysis of search terminology used by humanities scholars: The Getty online searching project report No. 1. *Library Quarterly, 63(January)*: 1-39.

Bates, M. J., Wilde, D. N., & Siegfried, S. (1995). Research practices of humanities scholars: The Getty online searching project report No. 3. *Library and Information Science Research, 17(Winter)*: 5-40.

Beghtol, C., Howarth, L.C., & Williamson, N.J. (2000). *Dynamism and stability in knowledge organization: Proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada.*

Blackler, F. (1995). Knowledge, knowledge work and organizations: An overview and interpretation. *Organization Studies, 16*(6): 1021-1046.

Emery, F.E. (1969). *Systems thinking.* Baltimore, MD: Penguin Books.

Hert, C.A. (2002). *Taking a user perspective on metadata use: Examples from the domain of statistical information.* A lecture delivered on August 28, 2002 at the School of Information and Library Science, University of North Carolina at Chapel Hill.

Jacob, E. K. (2001). The everyday world of work: Two approaches to the investigation of classification in context. *Journal of Documentation, 57*(1): 76-99.

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration.* Berkeley, CA: University of California Press.

Kuhlthau, C. C. (1993). *Seeking meaning: A process approach to library and information services.* Norwood, NJ: Ablex.

Luhmann, N. (1995). *Social systems.* Stanford, CA: Stanford University Press.

Lykke Nielsen, M. (2000). Domain analysis, an important part of thesaurus construction: Methodologies and approaches. *Advances in Classification Research, 11.*

Siegfried, S., Bates, M. J., & Wilde, D. N. (1993). A profile of end-user searching behavior by humanities scholars: The Getty online searching project report No. 2. *Journal of the American Society for Information Science, 44(June)*: 273-291.

Solomon, P. (1992). User-based methods for classification development. *Advances in Classification Research, 2*: 163-170.

Solomon, P. (1993). Children's information retrieval behavior: A case analysis of an OPAC. *Journal of the American Society for Information Science, 44*(5): 245-264.

Tang, R., & Solomon, P. (1998). Towards an understanding of the dynamics of relevance judgment: An analysis of one person's search behavior. *Information Processing and Management, 34*(2/3): 237-256.

Taylor, R. S. (1986). *Value-added processes in information systems.* Norwood, NJ: Ablex.

Yang, K. (2002). *Combining Text-, Link-, and Classification-based Retrieval Methods to Enhance Information Discovery on the Web* (Dissertation). Chapel Hill, NC: University of North Carolina.