Aaron Loehrlein (a.loe@ubc.ca)
The University of British Columbia, Vancouver, BC Canada

An Examination of Interdisciplinary Theory Between Cognitive Categorization and Knowledge Organization

Abstract: This paper is a preliminary comparison of the theories used to represent concepts in the field of knowledge organization with similar theories that are used in the field of cognitive categorization. Both disciplines employ the same general types of conceptual representations, such as features, dimensions, labels and relationships. Both disciplines are also concerned with how people use these representations to organize and retrieve information. However, cognitive categorization is generally concerned with internal conceptual representations (ICRs), which are maintained in memory, while knowledge organization is generally concerned with external conceptual representations (ECRs), which are maintained outside of memory. This fundamental distinction has profound effects on how the representations are structured, accessed and maintained. This paper explores these differences and points out cases in which ICRs and ECRs may affect each other. This analysis will support future research that utilizes ICRs in the study and evaluation of ECRs.

1. Introduction

It is generally accepted that knowledge organizing schemes are derived at least in part from the categories and other conceptual structures that people maintain in memory. This paper makes a distinction between internal conceptual representations (ICRs), which are stored in memory, and external conceptual representations (ECRs), which are stored outside of memory. ICRs are studied extensively in cognitive psychology, particularly research involving categorization, similarity judgments, choice sets, conceptual combination, and inductive inferences. ECRs are studied extensively in knowledge organization and include classification schemes, subject heading systems, thesauri, ontologies, folksonomies, and standards for bibliographic description.

ICRs are often shared between people within a culture, domain, or discipline. For that reason, the study of ICRs is potentially of great value in the study of knowledge organization. Intriguingly, while there is a great deal of research into the categories or groupings that people form in memory, this research is still relatively underutilized in the field knowledge organization (KO). While models of categorization have many clear similarities to ECRs that are used in knowledge organization, such as classification systems and social tagging systems, there are also many subtle inconsistencies between ICRs and ECRs. In this paper, I propose to examine some of the irregularities between the two types of conceptual representations. This preliminary examination seeks to bolster the foundation of interdisciplinary research that brings together the approaches and concerns of existing research in categorization and KO.

In the past, cognitive approaches to KO have been criticized on the grounds that such approaches seek to discover the rules by which people identify the subjects of documents. Cognivitists are said to use these rules to create external KO systems that mimic people's internal conceptual representations, an approach that has been criticized as being ultimately futile. Specifically, these approaches are problematic because internal conceptual structures vary according to the social context in which they are applied, while KO systems are generally

designed to useable and sharable between different groups of people from a variety of social contexts (Frohmann 1990, Hjørland and Albrechtsen 1995, Jacob and Shaw 1998, Hjørland 2002).

While these criticisms are well advised, other cognitive approaches have the potential to be fruitful in the study of KO systems. In research involving categorization, it is generally believed that categories are subject to change depending on the context in which they are used (Jacob 2004). Barsalou (1982) found that the properties of some concepts were context dependent. Participants in his study verified that "can be walked upon" is a property of the concept ROOF, but could not easily associate that property with ROOF except in certain contexts. Medin, Lynch, Coley, and Atran (1997) found that professional landscape architects grouped trees differently depending on whether they were considering the trees in the context of landscaping or in the context of disease that affect trees. Nevertheless, theoretical models have effectively predicted many regularities in how people categorize (e.g., Smith and Medin 1981). As Hjørland (2009) points out, cognitive approaches that take into account scientific theories and broad cultural perspectives can be fruitful to the study of KO systems.

2. Comparison of Internal and External Conceptual Representations

Both research into cognitive categorization and research in knowledge organization are fundamentally concerned with how items are grouped together and the uses to which the resulting groups may be put (e.g., Svenonius 2000, Bates 2003, Murphy 2004). These groupings and their members may all be thought of as concepts. The essential nature of concepts is widely and vigorously debated (see Hjørland 2009 for a detailed analysis of this debate). For the purposes of this discussion, I suggest that it can be generally accepted that the word "concept," at the very least, applies to anything for which a representation can be made. Although concepts may be widely shared, a concept may be structured in a variety of ways. These structures are manifested in representations of the concept. In specific models of ICRs or ECRs, these structures can take a variety of forms, such as features, dimensions, relationships, labels, and free text descriptions. Specific structures might not be as widely shared as the concept itself.

For example, the concept BIRD may be widely shared within an organization, discipline, or culture. However, different structures of BIRD are seen in different representations. These include the representation of BIRD that appears in the scientific taxonomy of living creatures, representations of BIRD that are used in bibliographic classification schemes, and representations that a person stores in memory. In fact, it is likely that people store in memory multiple representations of the same concept, each of which is useful in a different context. Because BIRD is a widely shared concept, representations of BIRD have emerged that are easily sharable (Freyd, 1983). Nevertheless, not all representations of BIRD are necessarily easily sharable. A person may maintain one representation of BIRD for the purpose of communicating with others, and another representation that represents the person's own idiosyncratic experiences with birds.

The means by which ICRs and ECRs are structured is broadly similar. For example, they both consist of groupings that are populated by instances or types. Membership in a grouping is typically based on the rules, characteristics, and/or exemplars that are associated with the grouping. These groupings facilitate information organization and retrieval as well as the transmission of knowledge. ICRs are used when people organize and retrieve information from memory. ECRs are used with information systems to organize and retrieve information from sources external to memory. Nevertheless, ICRs and ECRs are subject to many of the same

constraints. For example, in both cases hierarchical structures facilitate retrieval, since properties are naturally transferred between superordinate and subordinate concepts.

However, there are many fundamental differences between ICRs and ECRs. Changes made to ICRs are often at least partially involuntary and can be completely subconscious. ECRs usually change only as the result of a conscious action taken by the person, committee, or community that is responsible for maintaining them. Some ECRs, such as social tagging systems, resemble ICRs in that they continuously evolve (Golder and Huberman 2006). However, social tagging systems lack many elements of ICRs, such as relationships between tags and representations of the properties by which membership in a tag is determined.

ECRs are often grouped by the information system of which they are a part. For example, a controlled vocabulary is a single scheme of terms, each of which is an ECR. The controlled vocabulary is designed so that each term can be applied to a set of documents. Each document has its own representation, which is also an ECR. However, the representations of documents are not considered to be a part of the controlled vocabulary that is used to organize the documents. In contrast, while ICRs may form context-independent categories in some cases, they are not usually grouped into well-defined systems.

Because ECRs exist outside of memory, they are often too large and complex to be maintained in memory. In fact, in many cases they serve as a memory aides (Jacob 2004). Furthermore, it is possible to access only part of an ECR, especially if the ECR consists of a large scheme or system. In contrast, models of ICRs often emphasize the intension of the concept, which are the properties by which membership in the concept is determined. Models such as prototype and exemplar models assume that it is easier for people to remember the properties of a concept like BIRD than it is to remember all the instances of BIRD. ECRs rarely include depictions of a concept's intension. However, because the limits of short term memory do not constrain ECRs, many ECRs include the concept's entire extension, that is, the complete set of the concept's instances.

ICRs are stored in memory and cannot be directly shared. In order to be shared, a description of the ICR must be articulated. When other people receive the description, they may create or modify their own ICRs in response. In contrast, because ECRs exist outside of memory, they may be directly shared by anyone who has access to them. However, each person is likely to interpret the ECR in their own way, based on their own knowledge and situational needs. Because ICRs cannot be perceived directly, it is not possible to be certain of their structure. Interviews or think-aloud protocols may be used to record a person's ICRs, but a person may only report on an ICR of which he or she is aware. Cognitive science has developed different models to predict the constraints that act on ICRs. These include classical, prototype, exemplar, goal-derived, and theory-based models (Smith and Medin 1981, Murphy and Medin 1985, Barsalou 1991). These models generally focus on the means by which membership in a category is determined.

Because ECRs can be perceived directly, it is not necessary to develop theories that predict their structure. Instead, an ECR such as a classification system often simply declares its structure and the constraints that apply to it. In addition, from a user's perspective membership in an ECR is a given. For example, not only is the structure of the Dewey Decimal Classification (DDC) widely available, but an OPAC that uses the DDC explicitly designates each resource in the collection as either a member or a non-member of a given class. It is important to remember that ECRs such as knowledge organizing systems are based on theories and did not spring into being out of nothingness. However, once the knowledge organizing system is created and

implemented, theories are not needed to predict what constrains the structure of concepts, since those constraints are a given.

One the primary motivations of studies involving ICRs is to determine how well different models of concepts seem to fit the actual ICRs that people use. For example, many studies over the past twenty years have sought to determine whether prototype models or exemplar models make more accurate predictions of the nature of ICRs across different circumstances. The goal of this research is to predict the nature of the conceptual representations that people hold in memory. There are fewer studies that explore whether a person's own ICR is accurate or appropriate to the concept. Because ECRs are directly observable, there is no need to predict what ECRs look like. Furthermore, ECRs can be created as people see fit. They are not subject to our subconscious whims. For these reasons, there is no pressing need for knowledge organization research to import and adapt prototype models, exemplar models, and other models used to study ICRs.

However, by the same token, there is often not a consensus in knowledge organization for how to best evaluate the quality of an ECR. Different studies use a variety of methods and dependent variables. These include analysis of transaction logs (e.g., Asunka, Chae, Hughes, and Natriello 2009), subjective ratings of satisfaction and/or variables measuring efficiency in completing tasks using ECRs (e.g., Larson and Czerwinski 1998), and domain analysis, which compares ECRs to conceptual structures that appear in the literature (e.g., Hjørland 2002). Each of these approaches has varying degrees of merit. In general, however, knowledge organization does not seem to be unified by a common research question to the extent that is seen in the research of ICRs. As result, the problem area in knowledge organization often seems less well defined than the problem area in cognitive categorization. This is not necessarily a problem, but it may impact the general focus of knowledge organization research.

3. Comparison of Bibliographic Hierarchical Classification with the Hierarchical Classification of Other Concepts

In addition to differences between ICRs and ECRs, KO is separated from categorization research in another important respect. In many hierarchical classification systems, each member of a class is also a member of each superordinate class under which it is nested. For example, in the taxonomy of living things, if an animal is classed under the species T. migratorius (North American Robin), we may conclude that the animal is also a member of the genus Turdus, the family Turdidae, the order Passeriformes, the class Aves, and so on. Any of the properties observed in the class Aves, such as WINGS and FEATHERS, will be inherited by each of these subclasses, as well as by the individual animal. However, bibliographical hierarchical classification systems differ from this general model. Bibliographic classification systems organize books and other resources not by the properties that the resources possess, but by the properties that they represent. Clearly, a book about the class Aves (i.e., Bird) does not have wings, but it presumably represents a concept—Bird—that does possess the property WINGS. The species North American Robin inherits the property WINGS from Bird. However, unlike the animal, a book classed under North American robin would not also be classed under Turdus, Turdidae, etc.

This is partly due to practical considerations. Traditionally, books could not occupy more than one physical location, which implied that they should not occupy more than one class. However, even if the books existed in a digital environment—which would allow them to be

placed into more than one class—a book about North American robins is clearly not about birds in general. The book represents a concept that has wings, feathers, and all the other properties of Bird. However, the book is likely to assign those properties specifically to North American robins as opposed to birds in general. It may do this in the form of a statement, such as "robins have wings," or it may simply include a picture of a robin in which the wings are visible. Therefore, we may conclude that, theoretically, information resources represent properties of a concept, but are likely to assign those properties to a concept at only one level of specificity. We may further speculate that information resources are likely to emphasize the properties of a concept by which it may be distinguished from concepts that are similar—such as the properties by which robins may be distinguished from blue jays. For example, in addition to being types of the concept BIRD, North American robins are also types of the concept OBJECT. That concept may have the property TANGIBLE OR VISIBLE. However, it is intuitively unlikely that a book about North American robins will emphasize that property. The book may be more likely to emphasize the property BROWN-GRAY UNDERPARTS, since there are many concepts that are similar to North American Robin that do not possess this property, such as Blue Jay.

These considerations affect how inferences are made by readers about the resources that are organized in the hierarchy. All members of North American Robin inherit each of the properties of Bird. Generally speaking, each of the books in the bibliographic class North American Robin describe a concept that inherits each of the properties of the concept that is described by books in the bibliographic class Bird. However, it is not possible to read a book about birds and simply transfer its description of birds to books about North American robins. First, the reader must mentally "unassign" the properties from Bird and reassign them to North American Robin. The reader must also consider the possibility that books about birds are more likely to emphasize certain properties than books about North American robins, such as properties by which birds are distinguished from mammal, fish, etc.

Consideration of specific books and classification systems reveal other complications. The Dewey Decimal Classification system (DDC) includes the following schedule:

745	Decorative Arts
745.5	Handicrafts
745.59	Making specific objects
745.592	Toys, models, miniatures, related objects
745.592 2	Dolls, puppets, marionettes
745.592 21	Dolls
745.592 24	Puppets and marionettes
745.592 3	Dollhouses and furniture
745.592 4	Soft toys
745.592 43	Teddy bears
745.592 8	Models and miniatures
745.592 82	Military models and miniatures

(Dewey et al. 2002)

Based on this schedule, we may conclude that books in the class 745.592 describe toys in general, while books in more specific classes describe specific types of toys. While this is broadly true, in practice most books that describe toys in general are very likely to devote much of their content to describing specific types of toys. For example, Ketchum (1981) is a book

entitled "Toys & Games." In WorldCat, it is assigned the DDC class 745.592. Its table of contents is as follows: "The Earliest Toys -- Wooden Playthings -- Dolls, Dollhouses and Playthings -- Tin and Cast-Iron Toys -- Soldiers -- Paper Playthings -- Card, Table and Board Games -- Optical Devices -- Banks -- Advice for the Collector." It is clear from this table of contents that the book spends little time describing toys in general. Certain chapters are devoted to types of toys that have a more specific subclass in the DDC, such as Dolls, Dollhouses, and Soldiers. Other chapters describe wooden and paper playthings, which are clearly subtypes of toys, but which do not have classes in the DDC. In fact, it is possible that there is more information about dollhouses in books classed under toys in general than there is in books classed specifically under dollhouses, especially if there are many more books in the former category than in the latter. Other books classed under 745.592 include Simon and Barton (1971) "The Paper Airplane Book" and Stevenson (1971) "The Art of Making Wooden Toys." The DDC schedule states that class 745.592 includes paper airplanes. Although the class description makes no specific mention of wooden toys, the introduction to the DDC states that books may be moved to a higher class if more specific classes are not used. Because there is no class in this section of the DDC that is specifically designated for wooden toys, 745.592 is likely to be the most appropriate location for Stevenson (1971).

This example illustrates several discrepancies between bibliographic classification and other types of classification. If the hierarchy above were applied to objects as opposed to bibliographic resources, it would be reasonably safe to conclude that dolls inherit all the properties of toys. But because the DDC is a bibliographic classification, we may conclude, with less certainty, that books in the class "Dolls, puppets, marionettes" describe a concept that inherits the properties of the concept described by books in the class "Toys, models, miniatures, related objects." We may infer that books in the former class probably assign those properties to specific concepts like DOLL, while books in the latter class probably assign the same properties to the more general concept TOY. We may further infer that the books in the former class may tend to emphasize different properties than the books in the latter class. However, we must also take into account the possibility that that books classed under "Toys" contain chapters that explicitly describe dolls. Other books in same class may explicitly describe types of toys for which there is no class in the DDC. For these reasons, it can be very difficult to use the hierarchy in the DDC to make inferences about the contents of the books that are organized in the hierarchy. This conclusion is not meant to be a critique of the DDC or an implication that the DDC contains serious flaws. Rather, it illustrates an important limitation of bibliographic hierarchical classification in general. These limitations are partly due to the fact that ECRs are often designed to be much more complex than conceptual representations that are stored in memory. Indeed, one of the benefits of ECRs is that they can be considerably more complex than ICRs. However, as ECRs become more complex, they become less apt to support certain processes involving conceptual representations, such as inferences by which people transfer properties from a concept to a related concept.

4. Conclusion

Based on these preliminary findings, research in categorization is often confounded when directly applied to KO schemes. Many of the external conceptual representations are too complex to fully support the types of processes that are possible with internal conceptual representations. In addition, the two disciplines seek to answer fundamentally different

questions. Much of the research into ICRs seeks to make the most accurate predictions possible about the nature of ICRs. Research into ECRs does not need to make these kinds of predictions, but studies in knowledge organization often struggle to identify methods for evaluating models of ECRs.

However, by accounting for these confounding factors, interdisciplinary research between categorization and KO may become less fraught. A more precise understanding of the similarities and differences between the two disciplines sheds light on opportunities for interdisciplinary research. One such possibility is investigation into how interaction with an ECR such as a classification system affects a person's own internal conceptual structures. Interaction with an ECR produces or modifies an ICR in the mind of the person using the ECR. Therefore, it is possible to use methods employed in the study of ICRs to evaluate the effects that occur when people use ECRs. If studies are able to identify effects that occur consistently in the use of ECRs, we may come to better understand which aspects of ECRs best promote information retrieval. Detailed results of one such study will be reported in a future publication.

References

Asunka, Stephen; Chae, Hui S.; Hughes, Brian; and Natriello, Gary. 2009. Understanding academic information seeking habits through analysis of web server log files: The case of the Teachers College library website. *The Journal of Academic Librarianship* 35: 33-45.

Barsalou, Lawrence. 1982. Context-independent and context-dependent information in concepts. *Memory & Cognition* 10: 82-93.

Barsalou, Lawrence. 1991. Deriving categories to achieve goals. In Bower, G. H. ed., *The psychology of learning and motivation: Advances in research and theory* 27:1-64. San Diego, Ca.: Academic Press.

Bates, Marcia J. 2003. Task force recommendation 2.3 research and design review: Improving user access to library catalog and portal information. *Library of Congress Bicentennial Conference on Bibliographic Control for the New Millennium*. Department of Information Studies University of California, Los Angeles.

Dewey, Melvil; Mitchell, Joan S.; Beall, Julianne; Martin, Giles; Matthews, Winton E. Jr.; and New, Gregory R. 2002. *Dewey decimal classification and relative index*. Dublin, Oh.: OCLC Online Computer Library Center, Inc.

Freyd, Jennifer J. 1983. Shareability: The social psychology of epistemology. *Cognitive Science* 7: 191-210.

Frohmann, Bernd. 1990. Rules of indexing: A critique of mentalism in information retrieval theory. *Journal of Documentation* 46: 81-101.

Golder, Scott and Huberman, Bernardo. 2006. Usage patterns of collaborative tagging systems. *Journal of Information Science* 32: 198-208.

Hjørland, Birger. 2002. Epistemology and the socio-cognitive perspective in information science. *Journal of the American Society for Information Science and Technology* 53: 257-270.

Hjørland, Birger. 2009. Concept theory. *Journal of the American Society for Information Science and Technology* 60: 1519-1536.

Hjørland, Birger and Albrechtsen, Hanne. 1995. Toward a new horizon in information science: Domain-analysis. *Journal of the American Society for Information Science*, 46: 400.

Jacob, Elin K. 2004. Classification and categorization: A difference that makes a difference. *Library Trends* 52: 515-540.

Jacob, Elin K. and Shaw, Deborah. 1998. Sociocognitive perspectives on representation. *Annual Review of Information Science and Technology* 33: 131-185.

Ketchum, William C. 1981. *Toys & games.* New York: English Book

Larson, Kevin and Czerwinski, Mary. 1998. Web page design: Implications of memory, structure and scent for information retrieval. *Proceedings of the SIGCHI conference on Human factors in computing systems*: 25-32. Los Angeles, California, United States: ACM Press/Addison-Wesley Publishing Co.

Medin, Douglas. L.; Lynch, Elizabeth. B.; Coley, John. D.; and Atran, Scott. 1997. Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology* 32: 49-96.

Murphy, Gregory L. 2004. *The big book of concepts*. Cambridge, Mass.: MIT Press.

Murphy, Gregory L. and Medin, Douglas L. 1985. The role of theories in conceptual coherence. *Psychological Review* 92: 289-316.

Simon, Seymour. 1971. *The paper airplane book.* New York: Viking Press.

Smith, Edward. E. and Medin, Douglas. L. 1981. *Categories and concepts*. Cambridge, Mass.: Harvard University Press.

Stevenson, Peter. 1971. *The art of making wooden toys*. Philadelphia, Pa.: Chilton Book Co.

Svenonius, Elaine. 2000. *The intellectual foundation of information organization*. Cambridge, Mass.: MIT Press.