

Folksonomies, Syntagmatic Relationships, & Feminist Research: Alternative Knowledge Structures

HOPE A. OLSON

*School of Information Studies
University of Wisconsin-Milwaukee, USA*

INTRODUCTION

The traditional knowledge structures that we conventionally use for organizing information are effective in some contexts for some information seekers, but they are not universally applicable. This paper brings together elements from folksonomies, syntagmatic relationships, and epistemic models in feminist research as possible factors in the development of alternative knowledge structures.

In addition to identifying problems of inappropriate terminology and juxtapositions of topics in subject access standards, my earlier research is critical of existing knowledge structures because they are used too widely and do not serve all information seekers equitably.¹

¹ For example: Hope A. Olson, (2001). Sameness and difference: A cultural foundation of classification. *Library Resources & Technical Services* 45(3) 115-122; Olson, Hope A. (2002). Classification and universality: Application and construction. Review article. *Semiotica* 139 (1/4) 377-391; Olson, Hope A. (2002). *The Power to Name: Locating the limits of subject representation in libraries*. Dordrecht, Netherlands: Kluwer Academic.

They cannot serve as universal frameworks for organizing information. In particular, the elements of mutually exclusive categories, teleological progression from concrete to abstract, and hierarchical structure (discussed further below) are reflective of the logical tradition established in classical Athens and as developed in European and American culture they are a poor fit with the epistemic traditions of other cultures, such as indigenous cultures, that hold a more holistic worldview. Further, there are indications that there are gender differences in epistemology that may affect the type of structure most conducive to organizing information (more on this perspective below). Although it is still worthwhile to further understand this weakness in conventional structures, it is now sufficiently established to merit exploration of alternative knowledge structures. This paper is a first effort at combining work from two very different projects - one informetric and one philosophical. The informetric research, with Dietmar Wolfram, explores the syntagmatic relationships that occur in the folksonomies of social tagging.² The philosophical research uses example from feminist ethical and epistemological research to create a foundation for alternative knowledge structures.³

To develop this approach to alternative knowledge structures I will first briefly describe the relevant characteristics of conventional knowledge structures. I will then survey folksonomies in relation to those structures and to syntagmatic (and paradigmatic) relationships. Finally, I will relate these factors to knowledge structures that emerge from my examination of a specific stream of feminist empirical research.

2 Hope A. Olson, and Dietmar Wolfram. (2008). Syntagmatic relationships and indexing consistency on a larger scale. *Journal of Documentation* 64(4): 602-615 and Dietmar Wolfram and Hope A Olson. This research is continued in Wolfram, Dietmar, and Hope A Olson. (in press). Measuring Consistency for Multiple Taggers Using Vector Space Modeling. *Journal of the American Society for Information Science*.

3 Hope A. Olson, (2007). How We Construct Subjects: A Feminist Analysis. Special issue on: Gender Issues in Information Needs and Services, eds. Cindy Ingold and Susan E. Searing. *Library Trends* 56(2): 509-541.

CONVENTIONAL KNOWLEDGE STRUCTURES

The conventional knowledge structures that we use in classification and other controlled vocabularies can be traced back to the logic developed in classical Athens. I have traced these structures through foundational works on knowledge organization in mainstream western culture and found them consistently present even in widely varied times, values, and views. They emerge with startling regularity beginning with their Greek origins in the work of Aristotle and his predecessors; in the Medieval Scholasticism of the 13th century mystic Hugh of St. Victor; in the Renaissance logic of Francis Bacon; in the Enlightenment Encyclopædism of Jean Le Rond d'Alembert and Denis Diderot; in the 19th century militancy GFW Hegel and the Scottish Common Sense philosophers; and in the twentieth century understanding of classification and culture displayed in Émile Durkheim and Marcel Mauss's *Primitive Classification*.⁴

⁴ For example: Hope A. Olson, (1999). Cultural discourses of classification: Indigenous alternatives to the tradition of Aristotle, Durkheim and Foucault. In Hanne Albrechtsen and Jens-Erik Mai (eds.), *Proceedings of the 10th ASIS SIG/CR Classification Research Workshop, October 13, 1999: Advances in Classification Research*, v.10 (pp.91-106) . Medford, NJ: Information Today, Inc. for the American Society for Information Science and Technology; Olson, Hope A. (2000). Reading "Primitive Classification" and misreading cultures: The metaphysics of social and logical classification. In Clare Beghtol, Lynne C. Howarth and Nancy J. Williamson, eds., *Dynamism and Stability in Knowledge Organization: Proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada* (pp.3-9). Würzburg, Germany: Ergon Verlag; Olson, Hope A., Juliet Nielsen, and Shona R. Dippie. (2002). Encyclopædist rivalry, classificatory commonality, illusory universality. In María J. López Huertas (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. 457-464). Würzburg, Germany: Ergon Verlag; Olson, Hope A. (2004) Bacon, Warrant, and Classification. In Breitenstein, Mikel, Eds. *Proceedings 15th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research*, Providence, Rhode Island. Available <http://dlist.sir.arizona.edu/1767/01/SIG-CR2004Olson.pdf> accessed May 31, 2009; Olson, Hope A. (2004). The ubiquitous hierarchy: An army to overcome the threat of a mob. *Library Trends*, 52(3): 604-616.

In examining these texts I have found three culturally specific characteristics. First is the presumption that categories must be mutually exclusive or that there must be defining boundaries around concepts. Second, these categories must form a linear, teleological progression, typically from the concrete to the abstract. Finally, this linear arrangement of categories culminates in hierarchy as categories aggregate into broader categories. These characteristics grow from logic as it was developed in 4th century BCE Athens⁵

Athenian logic began in a specific cultural context. In the 4th century BCE the Athenians who learned logic were elite males. They were taught logic—it did not come to them as some innate ability. Logical argument was their exclusive province. The rest of the population—women, workers, slaves, and “barbarians” (that is, foreigners)—did not have logic available to them.⁶ The importance of this Athenian context is to remind us of two things: logic is a powerful tool and logic is not innate; it began as a technique for argument constructed in a specific time and place for particular people. However widespread it has become, it is not central to all groups and cultures. Logic and the structures that have grown from it are then not universal. They are useful, but not necessarily to everyone.

Conventional knowledge structures are widespread in applications that organize information. Hierarchical relationships are integral to classification schemes in their arrangement of topics and in thesauri and subject headings in the form of broader and narrower terms (BT/NT). Even faceted classifications turn into hierarchies once a citation order of the facets is established. These standards are applied in a broad range of contexts partly because standardization enables sharing. However, at the same time, it limits contextualization—customization to specific groups of information seekers. Globalization has spread standardized bibliographic products worldwide. For example, OCLC now operates in 112 countries;⁷ the Dewey Decimal Classification is used in

5 Hope A. Olson, (1999). Exclusivity, teleology and hierarchy: Our Aristotelean legacy. *Knowledge Organization* 26(2): 65-73.

6 For a full development of this context see Andrea Nye. (1990). *Words of Power: A Feminist Reading of the History of Logic*. New York: Routledge.

7 WorldCat facts and statistics. Available: <http://www.oclc.org/us/en/worldcat/statistics/default.htm> accessed May 31, 2009.

135 countries, more than 60 of which use it in their national bibliographies, and is translated into more than 30 languages;⁸ Library of Congress Subject Headings are also used in many national bibliographies, sometimes adapted and/or translated.

FOLKSONOMIES AND SOCIAL TAGGING

Radically different from these information standards are the so-called folksonomies. They develop from social tagging —the naming of information by the user for the user. These tags are usually shared with other users. There is generally no controlled vocabulary and no hierarchy, or only a very shallow one. In fact, there is typically no structure at all imposed on tagging. So, for example, when a user uploads a photograph to flickr <http://www.flickr.com/> he or she may add tags without having to consult a list. Web pages bookmarked in del.icio.us are given tags defined as "... a little bit like keywords, but you choose them yourself and they do not form a hierarchy".⁹ However, users of some tagging sites can create hierarchies of a sort. For example, in flickr one can organize photographs into sets and sets into collections and collections into broader collections, naming each set and collection. In del.icio.us the user can combine multiple tags in a "bundle" and name it. These options are akin to the folders one uses to organize computer files. They are set up to be hierarchical. Flickr collections can be nested five deep, which means that since sets can be held within collections and photographs within a set, a seven-level hierarchy is possible. However, these are not strict hierarchies since a set can be in more than one collection (although a collection can be in only one broader collection).¹⁰ So some social tagging sites allow some

8 Dewey is the world's most widely used library classification system. Available: <http://www.oclc.org/us/en/dewey/about/translations/default.htm> accessed May 31, 2009.

9 Delicious Frequently Asked Questions. Available: <http://delicious.com/help/faq#tags> accessed May 28, 2009.

10 Flickr Help / FAQ / Collections. Available: <http://www.flickr.com/help/collections/> accessed May 28, 2009.

semblance of hierarchy, but it is supplied by the users themselves and is specific to the individual. Other users will set up their own sets and bundles. In general then, social tagging is different from the culturally specific characteristics of western classification: mutual exclusivity, teleology, and hierarchy.

PARADIGMATIC AND SYNTAGMATIC RELATIONSHIPS

What teleology and hierarchy actually do in classification is to display the relationships between the mutually exclusive categories and the concepts that those categories represent. The relationships that are thus reflected are primarily paradigmatic relationships. Paradigmatic relationships are those that always apply regardless of context. They are the relationships represented in controlled vocabularies such as the narrower terms (NT), broader terms (BT), and related terms (RT) in thesauri. They are also reflected in the hierarchical structure of classification. Paradigmatic relationships hold standards together by giving them an infrastructure that create the shapes into which concepts are organized. An example of a paradigmatic relationship is that “cats” are always in the same broader class as “lions,” “tigers,” “panthers,” etc; they are all always felines. When we are talking about the animals this will hold true no matter what the context may be. A cat is a feline while curled up asleep and while out stalking prey. This hierarchical relationship is paradigmatic. Non-hierarchical relationships can also be paradigmatic. “Birds” are always related to “bird watching” because “bird watching” always involves “birds.” “Birds” are always related to “ornithology” because it is the study of birds. These relationships are also true regardless of context.

On the other hand, syntagmatic relationships are linked to specific instances.¹¹ They are generally reflected in surrogates for entities, that is the terms assigned to a given catalog record, index entry, or metadata record. For example, if I write a book about my three cats

11 See Olson and Wolfram 2008 for a fuller explanation of paradigmatic and syntagmatic relationships, especially the potential for overlap between the two.

in suburban Milwaukee it might have a subject heading for: “Cats—Wisconsin—Wauwatosa.” “Cats” and “Wauwatosa” are linked only because of this book. The relationship is dependent on context. If I write a book about the birds that come to the birdfeeder that I have set up in my backyard and how my cats and I watch them together through the window it might have the subject headings:

Bird watching—Wisconsin—Wauwatosa
Cats—Behavior—Wisconsin—Wauwatosa
Suburban animals—Wisconsin—Wauwatosa
Cats—Effect of human beings on

These precoordinated subject heading strings embody syntagmatic relationships. Postcoordinate indexing implies syntagmatic relationships by assigning terms to the same surrogate. That is, in databases co-occurrences are manifestations of syntagmatic relationships although they do not name those relationships as paradigmatic relationships are often named (e.g. BT, NT, RT). Syntagmatic relationships can, in this sense, form patterns that reflect underlying structure.

RELATIONSHIPS IN SOCIAL TAGGING

In a recent study, Dietmar Wolfram and I simulated a folksonomy by asking students (n=64) to tag a conference paper on problem-solving as a cognitive approach to explain the ubiquitous lack of interindexer consistency.¹² The distribution of terms followed a Zipf-like pattern (see *figure 1*) demonstrating a core of terms. That core suggests something like M.E. Maron’s idea of retrieval aboutness or R-about.¹³ Maron identified three kinds of aboutness: subjective aboutness or S-about which is the individual’s internal understanding of what a document is about; objective aboutness or O-about which consists of the terms

12 Olson and Wolfram 2008.

13 Maron, M.E. (1977). On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science*, 28(1), 38-43

an individual would search to retrieve a particular document; and retrieval aboutness or R-about which consists of the terms that a group who would find a document relevant would search to retrieve it. The most frequently searched terms, the Zipfian core, are arguably the R-about terms.

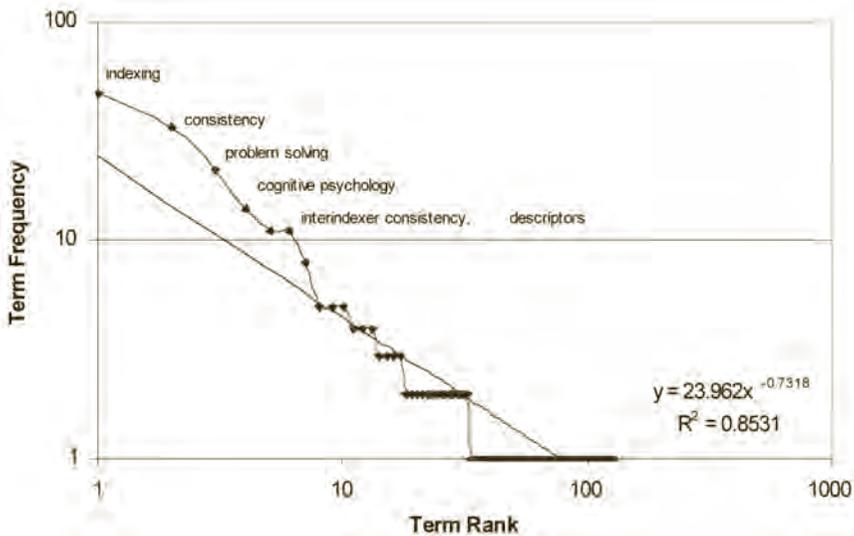


Figure 1: Rank-Frequency Plot of the Frequency of Occurrence of Different Indexing Terms¹⁴

However, frequency alone does not demonstrate relationships between terms, but terms assigned by the same student suggest syntagmatic relationships – relationships that apply in the instance of the article being indexed. In our data analysis, this co-occurrence of terms was measured in pairs. Over 350 term pairs were used only once. The most frequent co-occurrence was “indexing” and “consistency” which were used together 30 times —that is, nearly half of the students used both of these terms— followed by “indexing” and “problem solving” 17 times and “consistency” and “problem solving” 13 times. These reflect a core of syntagmatic relationships as illustrated

14 Figures 1-3 are from Olson and Wolfram 2008.

in *figure 2*. The terms in these relationships largely coincide with the terms most frequently used overall (*figure 1*).

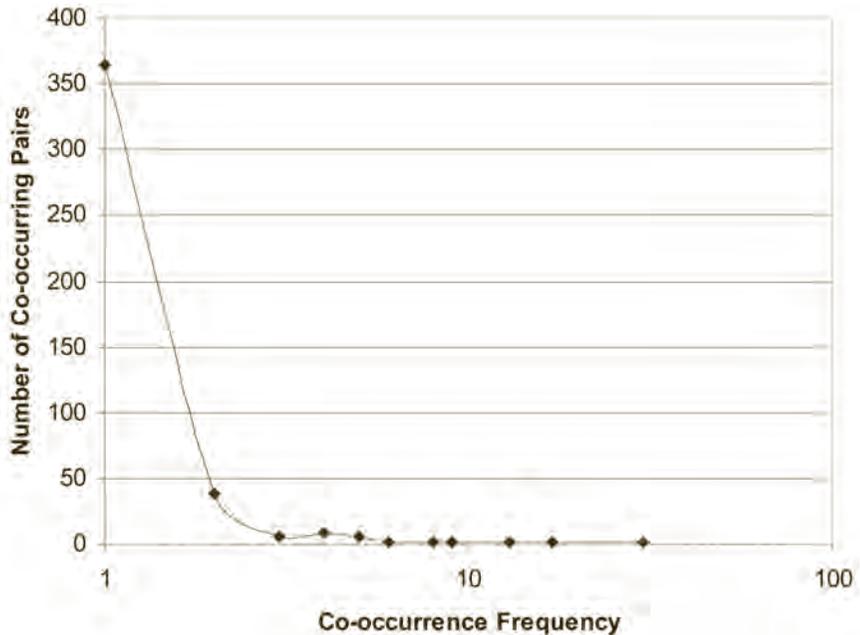


Figure 2: Size-Frequency Plot of the Frequency of Co-occurrence of Different Pairs of Terms

Another way of looking at these syntagmatic relationships is through clustering which, in this instance using multi-dimensional scaling, showed that in the three clusters that emerged the concepts overlapped in terms of semantic meanings (see *figure 3*). The top two clusters in particular shared “indexing”/“indexing behavior” and “consistency”/“interindexer consistency” while the two clusters on the right shared “cognitive psychology”/“cognitive”/“cognitive approach.” “Problem solving” was identified as a cognitive activity in the article, so it is semantically linked as well. And “descriptor” and “controlled vocabulary” might well, if they were in a thesaurus, be related to “indexing.” In fact, in the ERIC thesaurus “indexing” is used for “descriptors.”

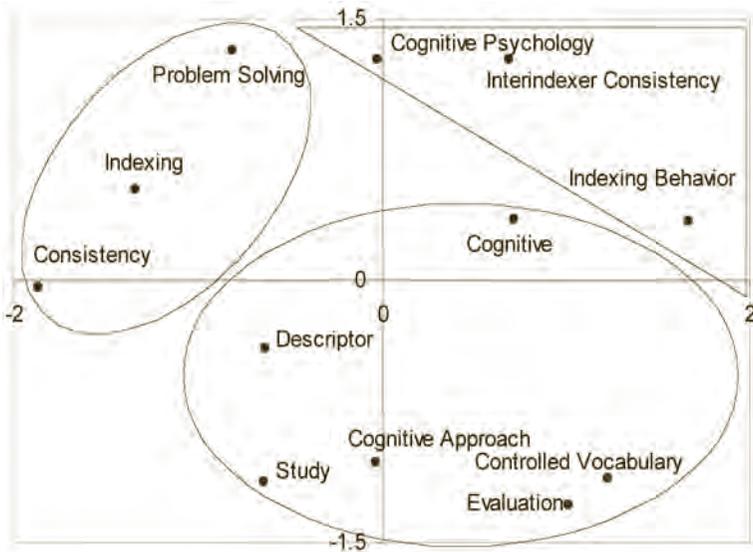


Figure 3: Two-dimensional MDS Plot of the Top Twelve Co-occurring Term Pairs

So there is some consistency in the students' choice of syntagmatic relationships as well as in their choice of terms. That consistency represents a commonality that reflects a user group such as that of Maron's R-about.

In a more recent study, we used data from CiteULike, a site for social tagging references to scholarly publications, and found similar distributions as did Kipp and Campbell in their 2006 study.¹⁵ Could something as conceptually simple as tag frequencies or the somewhat more complex clusters of co-occurrences offer the basis for an alternative knowledge structure? Is there evidence that such a structure would be meaningful to someone? For insight into these questions I turn to a well-established body of research.

15 Wolfram and Olson (in press); Kipp, Margaret E.I, & Campbell, D.Grant. (2006). Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. In Proceedings of the ASIST Annual Meeting, 43. Available <http://dlist.sir.arizona.edu/1704/01/KippCampbellASIST.pdf> Accessed May 31, 2009.

WOMEN'S KNOWLEDGE STRUCTURES

Understanding paradigmatic and syntagmatic relationships puts us into a good position to find a theoretical basis for the type of relationship that emerges in folksonomies. The view of relationships taken in some feminist research provides a useful perspective for understanding the potential of syntagmatic relationships and the structure that emerges from them. In particular, a stream of feminist empirical research has developed that suggests women may develop knowledge structures different than the conventional structures characterized by mutual exclusivity, teleology, and hierarchy. Two major studies originally developed this idea and considerable research has been grown from them. The first is Carol Gilligan's 1982 study, *In a Different Voice*,¹⁶ which addressed the development of ethical positions in women. Gilligan was working in reaction to Lawrence Kohlberg's then widely accepted conclusion that justice is the highest level of ethical development. Rather, she found that for many women, the pinnacle of ethics is an ethic of care. She found that women often value their connections with others more than they value justice. As Gilligan puts it, women replace "a hierarchy of rights with a web of relationships."¹⁷

The second key work is the 1986 study, *Women's Ways of Knowing* by Mary Field Belenky, Blythe McVicker Clinchy, Nancy Rule Goldberger, and Jill Mattuck Tarule,¹⁸ which examined the stages of how women gain knowledge. As women became more sophisticated in their epistemologies, one significant factor that the researchers uncovered was what they termed procedural knowing: a stage at which the women in the study were very aware of how they sought knowledge. These women took two approaches: separate knowing and connected knowing. Separate knowing calls for autonomy and logic

16 Carol Gilligan, (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.

17 Gilligan, p. 57.

18 Belenky, Mary Field, Clinchy, Blythe McVicker, Goldberger, Nancy Rule, & Tarule, Jill Mattuck. (1986). *Women's ways of knowing: The development of self, voice, and mind*. New York: Basic Books.

and is especially prevalent for women heavily influenced by those traditional ways of knowing such as are typically stressed in our educational system. The other approach is “connected knowing.” Connected knowing privileges experience and relies on connections with others to discover knowledge. In connected knowing, knowledge is gained through empathy and the experience of oneself and others. In further work following *Women’s Ways of Knowing*, Clinchy proposes the idea of “knowing communities” made up of unique individual knowing subjects who focus on belief rather than doubt or skepticism to achieve connected knowing.¹⁹ These communities have common interests, but go further to rely on their collective experience. This brings us back to consistency and core concepts of aboutness expressed by communities akin to Maron’s R-about, but with an added dimension of interaction.

CONCLUSION

This valuing of connection as a means of knowing as found in Gilligan and *Women’s Ways of Knowing* and considerable research that has been done in the years since their publication can suggest a theoretical basis for a very different kind of knowledge structure than we now practice. This is one that is potentially compatible with social tagging and may be more appropriate for some communities of users.

In real social tagging as opposed to our experiment with students in a class, the connection does not end with tagging a photograph or a bookmark or a reference. The virtual communities of a social tagging site continue to interact and build their “folksonomies.” Various dynamics come into play such as tagging for emotions²⁰ or tagging as

19 Blythe McVicker Clinchy, (1996). Connected and separate knowing: Toward a marriage of two minds. In Nancy Rule Goldberger, Jill Mattuck Tarule, Blythe McVicker Clinchy, & Mary Field Belenky (Eds.), *Knowledge, difference, and power: Essays inspired by Women’s Ways of Knowing* (pp. 205--247). New York: Basic Books.

20 Margaret E.I. Kipp, (2007). @toread and cool: Tagging for time, task and emotion. In *Proceedings Information Architecture Summit 2007*, Las Vegas, Nevada. Available: <http://eprints.rclis.org/archive/00010445/01/mkipp-iasummit2007.pdf> Accessed May 31, 2009

performance²¹ as well as tagging for topic —approaches that seem compatible with connected knowing. So there is a knowing community similar to the relevancy group of Maron's retrieval aboutness (R-about). The terms are those used by a group who would find a document relevant to describe the document (or photo or website or article) plus terms that evaluate it or that reflect the tagger's point of view. The patterns of the tags can identify the core topics (tag frequency distribution) and the syntagmatic relationships (co-occurrence frequency distribution) and whether or not they have a core that reflects one or more knowing communities.

Together these concepts can form a theoretical framework —or at least the beginning of one— to be explored and tested. This incipient framework in turn suggests future research questions to develop and evaluate these ideas:

- What is the nature of the knowing communities that social taggers constitute?
- Do the syntagmatic relationships that can be derived from social tagging add a dimensions that searchers want?
- Do they foster connectedness?
- What kind of systems and interfaces would facilitate this approach?
- Would a combination of the paradigmatic relationships from controlled vocabularies and the syntagmatic relationships from tagging be the ideal form of indexing?

Ultimately, we need to try out this framework to discover whether or not we can build alternative knowledge structures using it as a foundation.

21 Emma Tonkin, Tourte, Gregory J.L. and Zollers, Alla. (2008). Performance tags -- who's running the show? In Lussky, Joan, Eds. Proceedings 19th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research, Columbus, Ohio. Available <http://dlist.sir.arizona.edu/2456/> Accessed May 31, 2009