Basic-level concepts and the assessment of subject relevance: Are they really relevant?

Lala Hajibayova
Assistant Professor
School of Library & Information Science
Kent State University
314 University Library
Kent, OH 44242-0001
lhajibay@kent.edu
(330) 672-0026

Elin K. Jacob
Associate Professor
School of Informatics & Computing
Indiana University Bloomington
1320 E. 10th Street
Bloomington, IN 47405-3907
ejacob@indiana.edu
(812) 855-4671

Abstract

Studies of user-generated tagging vocabularies have suggested a preference for basic-level terms in tagging vocabularies (e.g., Golder & Huberman, 2006; Munk & Mork, 2007). A high proportion of basic-level terms has also been observed in systems of knowledge organization (Green, 2006). This study addresses the relevance of basic-level terms in knowledge representation and organization systems from the perspective of theories of relevance assessment proposed by Saracevic (1975, 2007a, 2007b) and Hjørland (2010) and argues that domain knowledge and expertise, which are thought to be central in the assessment of relevance, may be based on the subordinate level of conceptualization rather than the basic or superordinate levels seemingly favored in representational systems.

Basic Level Categories

Several studies of user-generated tagging vocabularies have suggested that taggers frequently use basic-level terms to represent resources (e.g., Munk & Mork, 2007; Yoon, 2012). Golder and Huberman (2006) argue that "earlier tags in a bookmark represent basic levels" (p. 202) because these terms are generally known to taggers and thus are the first terms they think of when tagging resources. Speculations on the widespread use of basic-level terms as descriptors (e.g., Green, 2006) are generally based on the argument that basic-level terms are "most cognitively efficient" (Mervis & Rosch, 1981 p. 92). In their report of a series of experimental studies, Rosch, Mervis, Gray, Johnson and Boyes-Braem (1976) describe basic-level category terms as most inclusive and most efficient because they "carry the most information, possess the highest cue validity, and are, thus, the most differentiated from one another" (p. 383; see also Rosch, 1978). Applying Wittgenstein's (1953/1963, p. 32) notion of family resemblance, which suggests that categories are defined not by formal definition but rather by a complex set of similarities, Rosch et al. (1976) proposed a theory of basic level categories as comprised of prototypes or best examples.

In contrast to Rosch et al.'s (1976) findings, some studies have suggested variations in preference for the assignment of superordinate, basic and subordinate levels of abstraction (e.g., Hajibayova & Jacob, in press; Mandler, Bauer & McDonough, 1991; Mandler & McDonough, 1993; Morris & Murphy, 1990). For example, Shaver,

Schwartz, Kirson, and O'Connor (1987) explored the prototypes and hierarchical taxonomy of five basic emotions. They conducted a hierarchical cluster analysis of categories representing 135 emotions generated by subjects that indicated three distinct category levels: general, intermediate and specific (p. 1067). Six categories of emotions grouped at the intermediate level were represented as LOVE, FEAR, JOY, SURPRISE,[1] ANGER and SADNESS. The most specific level consisted of 25 subordinate categories clustered under the six intermediate-level categories: For example, AFFECTION, LUST, CHEERFULNESS, PRIDE, and NERVOUSNESS were nested under the category LOVE. Shaver et al. suggest that the three category levels corresponded to the hierarchical representation of object categories, with the intermediate level representing the basic level of categories; and they argue that "within each basic level category, one subcluster, containing the basic-level term, appeared to designate a generic or core form of the emotion in question, whereas the other subclusters seemed to designate more specialized forms" (p. 1081). Additionally, Hajibayova and Jacob's (in press) analysis of user-generated tags assigned to four non-biological objects (FRUIT, CLOTHING, TOOL, VEHICLE) used in Rosh et al.'s (1976) study did not demonstrate significant preferences for a particular level of abstraction in the assignment of tags to resources.

Assuming that subject knowledge is central to the assessment of relevance, the question remains as to whether individuals consider subject relevance when assigning basic-level terms to resources and whether basic-level knowledge of the subject is sufficient for explaining assessments of relevance.

Relevance

Because relevance is considered one of the key issues in the evaluation of information retrieval (IR) systems, it has been studied for the decades. As Saracevic (1975) notes, the first formal studies of relevance can be traced back to early studies of bibliometrics (e.g., Bradford, 1934; Lotka, 1926; Zipf, 1949). Analysis of the IR literature suggests there have been two waves in the study of relevance (Huan & Soergel, 2013). The first wave of studies, which covers the period from the 1950s to the early 1970s, attempted to elucidate and conceptualize the concept of relevance by proposing various aspects including logic (e.g., Cooper, 1971), situation (e.g., Wilson, 1973), pertinence (e.g., Foskett, 1972) and user relevance (e.g., Rees & Saracevic, 1966). The second wave of IR research represented a shift in the study of information retrieval to a more "naturalistic" or *"real user* perspective" (Huan & Soergel, 2013, p. 19; emphasis in origin), suggesting a very personal, dynamic and situational understanding of relevance.

Nonetheless, after decades of research, the definition of relevance and its operationalization remain nebulous. In an early paper, Saracevic (1975) defines relevance as a "measure of the effectiveness of a contact between a source and a destination in a communication process" (p. 321) and suggests that there are multiple approaches to assessing relevance: the subject knowledge view of relevance (i.e., the relation between the user's knowledge of the subject and a topic about the subject); the subject literature view of relevance (i.e., the relation between the subject and the literature); the logical

---

[1] Due to the small size and poor showing of the category SURPRISE in previous studies, Shaver et al. (1987) found it to be questionable as a basic level category.

view of relevance (i.e., the relation between premises and conclusions based on logical consequences); the system view of relevance (i.e., the contents of a resource or the processes of a particular information system and their relation to a subject or topic); the destination view of relevance (i.e., human judgment of the relation between a resource and a topic); the pertinence or destination of knowledge view of relevance (i.e., the relation between the stock knowledge of the user and subject knowledge); and the pragmatic view of relevance (i.e., the relation between the immediate problem of a user and the information provided, which involves utility and preference as the basis for inference) (p. 338). Saracevic (1975) emphasizes that, while "any specific consideration of relevance is tied in with systems of relevance" (p. 339), the "*subject knowledge view of relevance is fundamental to all other views of relevance,* because subject knowledge is fundamental to communication of knowledge" (p. 333, emphasis in original).

In 1996, Saracevic further developed his definition of relevance, proposing a stratified framework comprised of the following five levels of relevance (p. 210):

- System or algorithmic relevance: the relation between a posed query and an information resource in a collection as either retrieved or failed to be retrieved;
- Topical or subject relevance: the relation between the topic stated in a query and the topic covered in the retrieved resources;
- Cognitive relevance: the relation between the user's knowledge and cognitive need and the retrieved resources;
- Situational relevance or pertinence: the relation between the situation or task at hand and the retrieved resources;
- Motivational relevance: the relation between the user's motivation or intent and the retrieved resources.

Using Saracevic's (1996) framework of relevance, this paper addresses the question of whether basic-level knowledge of a subject can explain the assessment of system, topical (or subject), cognitive, situational and motivational relevance.

Although the system view of relevance is usually equated with topical relevance (Saracevic, 1996; Froehlich, 1994), Huang and Soergel (2013) argue that "topicality is not limited to system determined relevance, nor is the system limited to determining topical relevance" (p. 25). Huang and Soergel suggest that a retrieval system may incorporate a variety of non-topical "variables pertaining to the user, task, and situation, as well as to integrate reasoning based on *nonmatching* type of topical connections" (2013, p. 25, emphasis in origin). Green (1995; Green & Bean, 1995) investigated topical relevance as a set of complex relationships that occur between the topic(s) of the user's needs and the topic(s) of the relevant texts. Green (1995) argues that, along with simple topic matching, a user's topical relevance assessment is influenced by both syntagmatic and paradigmatic relationships. Green and Bean (1995) propose that the recognition of paradigmatic, hierarchical relationships improves search strategies and increases recall, whereas the recognition of syntagmatic relationships impacts precision (p. 662). Thus, taking into consideration the complex nature of topic relevance detection involving both deductive reasoning (Cooper, 1971) and inductive reasoning (Wilson, 1973), the applicability of a basic-level of knowledge appears questionable.

Given the situated nature of relevance assessment, Schamber, Eisenber and Nilan (1990) argue that relevance is based on dynamic human judgment processes that employ

both internal (cognitive) and external (situational) factors. They point out that studies of relevance grounded in the traditional source-to-destination model of information retrieval are "too linear, mechanistic, and static to serve as a valid conceptual framework for exploring the human relevant judgment process" (p. 770). Applying Dervin's (1983) notion of sense making (which suggests the centrality of the user in the assessment of "meaning-ness," Schamber et al. contend that the "locus of *relevance* is within individuals' perceptions of information and [the] information environment – not in information as represented in a document or some other concrete form" (p. 771). Based on the assumption that cognitive relevance implies a relation between the state of knowledge and the user's cognitive ability to effectively understand a retrieved resource, we contend that the cognitive state of the individual is necessarily unstable in that it is closely related to other parameters of Saracevic's (1996) relevance framework such as situation and motivation. Furthermore, the situational nature of relevance assessment undermines assumptions of consensus in the application of basic-level knowledge.

Application of grounded cognition in assessment of the relevance of basic-level concepts

In light of Saracevic's (1975) emphasis on the importance of the subject view of relevance, Hjørland (2010) contends that relevance is always "human" and that "determin[ing] which items are relevant in relation to a given goal/task requires subject knowledge" (p. 231). With respect to the subject knowledge view of relevance suggested by Foskett (1972) and Saracevic (1975), Hjørland (2010) adopts a pragmatic view of relevance, arguing that, even though expert assessment of relevance may be necessary, "experts may have different interests, goals and values ... and they are not seen as neutral or objective assessors" (p. 232). And, while he allows that users might not be "automatically competent to judge relevance" (p. 231), he also points to the unstable nature of assessments of relevance by subject experts, noting that "the opinions of experts change when they change theories" (p. 232). A user's relevance assessment is an inherently cognitive process that may be explained based on theories of situated or embodied cognition. Barsalou (in press – a) argues that cognition does not simply reside in a set of cognitive mechanisms but emerges from these mechanisms "as they interact with sensory-motor systems, the body, the physical environment, and the social environment" (p. 3). Moreover, in contrast to the Cartesian view that cognition is centered in the brain, emergent cognition is "ultimately grounded in (and emerge[s] from) a variety of bodily, affective, perceptual, and motor processes" (Pezzulo, Barsalou, Cangelosi, Fischer, Rae, & Spivey, 2013, p. 612). Barsalou (in press – a) suggests that "the construct of situated conceptualization integrates cognition across these domains" (p. 3). For Barsalou (in press – b) cognitive process should be understood not only in "constant entwinement with action, but also how action contributed to creating it" (p. 12). Similarly, Suchman (2007) emphasizes the importance of the "relation of knowledge and action to the particular circumstances in which knowing and acting invariably occur" (pp. 178-179). She argues that "coherence of action is not adequately explained by either preconceived cognitive schema or institutionalized social norms" but, rather, by the "organization of situated action …[as] an emergent property of moment-by-moment interactions between actors and between actors and the environments of their actions" (p.

177). For Suchman, then, emergent properties are neither predetermined nor random but arise from relationships between "structures of action and the resources and constraints afforded by material and social circumstances" (p. 177). Action is contingent on the "complex world of objects, artifacts, and other actors, located in space and time" (p. 177); however, this "complex world" is not an "extraneous problem" but the "essential resource that makes knowledge possible and gives action its sense" (p. 177). Furthermore, as Barsalou (in press – a) observes, situated conceptualization not only "interpret[s] a situation" but also plays important roles in "coupling the individual with their physical and social environment, managing the interface between them, and controlling their situated actions" (p. 6). The notion of situated conceptualization provides an organic account of differences across individuals (e.g., Papies, Pronk, Keesman, & Barsalou, 2014), both in their assessments of relevance and in changes and modifications in their relevance assessments. Understanding the assessment of subject relevance as situated conceptualization echoes a large corpus of studies of basic-level categories suggesting that subject knowledge and domain expertise may actually precipitate a shift from basic-level categories to more specialized, subordinate-level concepts (e.g., Jolicoeur, Gluck, & Kosslyn, 1984). For example, an investigation of whether basic-level categories were modified by experience revealed that "expertise enhances the speed at which subordinate-level categories are accessed, making them at least as accessible as basic-level categories" (Tanaka & Taylor, 1991, p. 475; see also Belke, Leder, Harsanyi, & Carbon, 2010).

Conclusion

Even though cognitive efficiency may underlie the large number of basic-level terms found in user-generated vocabularies, basic-level categories may not be the most appropriate consideration in the analysis and representation of subject relevance. Theories of embedded cognition and situated action suggest that the use of superordinate, basic-level and subordinate category terms in the assessment of subject relevance (Hjørland, 2010) is inherently situational and depends on a variety of "bodily, affective, perceptual, and motor processes" (Pezzulo et al., 2013, p. 612) as well the user's level of domain knowledge. Thus, an individual's particular assessments of the subject relevance of articles on ethics will be based not only on her knowledge and/or understanding of the concept of ethics and the particular task or goal at hand (Hjørland & Christensen, 2002) but also on her situated conceptualization of the concept of ethics within her current affective state (e.g., emotions, time, space), which will favor the use of vocabulary at a particular level of abstraction.

References

Barsalou, L.W. (in press-a). Situated conceptualization: Theory and application. In Y. Coello & M. H. Fischer (Eds.), *Foundations of embodied cognition*. East Sussex, UK: Psychology Press. Retrieved from http://psychology.emory.edu/cognition/barsalou/papers/Barsalou_chap_in-press_SC_theory_application.pdf
Barsalou, L.W. (in press-b). Can cognition be reduced to action? Processes that mediate

stimuli and responses make human action possible. In A. K. Engel, K. J. Friston, & D. Kragic (Eds.), *Where's the action? The pragmatic turn in cognitive science* (Strüngmann Forum Reports, Vol. 18. J. Lupp, Series Ed.). Cambridge, MA. MIT Press. Retrieved from http://psychology.emory.edu/cognition/barsalou/papers/Barsalou_Strungmann_chap_in-press_mediated_action.pdf

Belke, B., Leder, H., Harsanyi, G., & Carbon, C. (2010). When a Picasso is a "Picasso": The entry point in the identification of visual art. *Acta Psychologica, 133*(2), 191-202.

Bradford, S.C. (1934). Sources of information on specific subjects. *Engineering: An Illustrated Weekly Journal (London), 137*, 85–86.

Cooper, W.S. (1971). A definition of relevance for information retrieval. *Information Storage and Retrieval, 7*(1), 19–37.

Dervin, B. (1983). An overview of sense-making research: Concepts, mehods and results to date. Paper presented at the *International Communication Association Annual Meeting*, Dallas, TX.

Foskett, D.J. (1972). A note on the concept of 'relevance.' *Information Storage and Retrieval*, 8(2), 77–78.

Froehlich, T.J. (1994). Relevance reconsidered–Towards an agenda for the 21st century: Introduction to special topic issue on relevance research. *Journal of the American Society for Information Science, 45*(3), 124–134.

Golder, S., & Huberman, B. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science, 32*(2), 198-208.

Green, R. (2006). Vocabulary alignment via basic level concepts. OCLC/ ALISE research grant report published electronically by OCLC Research. Retrieved from OCLC website: http://www.oclc.org/research/grants/reports/green/rg2005.pdf

Green, R. (1995). Topical relevance relationships. I. Why topic matching fails. *Journal of the American Society for Information Science, 46*(9), 646–653.

Green, R., & Bean, C.A. (1995). Topical relevance relationships. II. An exploratory study and preliminary typology. *Journal of the American Society for Information Science, 46*(9), 654–662.

Hajibayova, L., & Jacob, E.K. (in press). Factors Influencing user-generated vocabularies: How basic are basic level terms? *Knowledge Organization*.

Hjørland, B. (2010). The foundation of the concept of relevance. *Journal of the American Society for Information Science and Technology, 61*(2), 217-237.

Hjørland, B., & Christensen, F.S. (2010). Work tasks and socio-cognitive relevance: A specific example. *Journal of the American Society for Information Science and Technology, 53*(11), 960-965.

Huang, X., & Soergel, D. (2013). Relevance: The improved framework for explicating the notion. *Journal of the American Society for Information Science and Technology, 64*(1), 18-35.

Jolicoeur, P., Gluck, M., & Kosslyn, S. (1984). Pictures and names: Making the connection. Cognitive Psychology, 16(2), 243- 275.

Lotka, A.J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences, 16*(12), 317–324.

Mandler, J., & McDonough, L. (1993). Concept formation in infancy. *Cognitive Development, 8,* 291-318.

Mandler, J., Bauer, P., & McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology, 23,* 263-298.

Mervis, C., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32,* 89-115.

Morris, M., & Murphy, G. (1990). Converging operations on a basic level in event taxonomies. *Memory & Cognition, 18*(4), 407 – 418.

Munk, T., & Mork, K. (2007). Folksonomy, the power law & *significance* of the least effort. *Knowledge Organization, 34*(1), 16-33.

Papies, E.K., & Barsalou, L.W. (in press). Grounding desire and motivated behavior: A theoretical framework and review of empirical evidence. In W. Hofmann & L.F. Nordgren (Eds.), *The psychology of desire. New* York: Guilford Press. Retrieved from http://psychology.emory.edu/cognition/barsalou/papers/Papies_Barsalou_chap-in_press-grounded_theory_desire_motiv.pdf

Pezzulo, G., Barsalou, L.W., Cangelosi, A., Fischer, M.A., McRae, K., & Spivey, M. (2013). Computational grounded cognition: A new alliance between grounded cognition and computational modeling. *Frontiers in Psychology*, *3*(612), 1-11.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch, E., Mervis, C, Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3)*,* 382-439.

Saracevic, T. (2007a). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology, 58*(13), 1915-1933.

Saracevic, T. (2007b). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *Journal of the American Society for Information Science and Technology, 58*(13), 2126-2144.

Saracevic, T. (1996). Relevance reconsidered. In *Proceedings of the Second Conference on Conceptions of Library and Information Science* (pp. 201–218).

Saracevic, T. (1975). Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, *26*(6), 321-343. Retrieved from https://comminfo.rutgers.edu/~tefko/Saracevic_relevance_75.pdf

Schamber, L., Eisenberg, M., & Nilan, M.S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. *Information Processing and Management: An International Journal, 26*(6), 755–776.

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, *52*(6), 1061-1086.

Suchman, L. (2007). *Human-machine reconfigurations: Plans and situated actions.* New York, NY: Cambridge University Press.

Tanaka , J., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23*(3), 457-482.

Wilson, P. (1973). Situational relevance. *Information Storage and Retrieval, 9,* 457–471.

Wittgenstein, L. (1953/1963). *Philosophical Investigations (G.E.M* Anscombe, Trans.). New York, NY: Macmillan Publishing (Original work published 1953).

Yoon, J. (2009). Towards a user-oriented thesaurus for non-domain-specific image collection. *Information Processing and Management, 45*, 452- 468.

Zipf, G.K. (1949). Human behavior and the principle of least effort. Cambridge, MA: Addison-Wesley.