December 2007

SYSTEMATIZING WEB SEARCH THROUGH A METACOGNITIVE,SYSTEMS-BASED, INFORMATION STRUCTURING MODEL (MCSIS)

Ayman Abuhamdieh
Indiana State University

SYSTEMATIZING WEB SEARCH THROUGH A META-COGNITIVE, SYSTEMS-BASED, INFORMATION STRUCTURING MODEL (MCSIS)

Ayman H. Abuhamdieh
Indiana State University
ayman@indstate.edu

ABSTRACT

With the exponential expansion of information availability on the Internet, users are locating needed information through search engines and relying on heuristics and trial-and-error to find what they need. This paper proposes a Meta-cognitive, Systems-based, Information Structuring (McSIS) model to systematize users' information search behavior. The General Systems Theory's (GST) prepositions, especially the concepts of systems and their hierarchical structure, serve as its framework. Factors influencing information seekers, such as the individual learning styles of Field Independence (FI) and Field Dependence (FD); Holist (H), Serialist (S), and (PF) Problem-Focused vs. Emotion-Focused (EF) problem-solving approaches, and individual or personal domain knowledge are discussed and incorporated in the model. The paper presents a walk-through example on finding information on the Web using the model, and concludes with the study's limitations and suggestions for future research.

KEYWORDS

Web Information Search, Meta-cognitive Model, Information Structure, Systems Theory

INTRODUCTION

Since the Web became a public domain in the mid-nineties, few mouse clicks or key strokes could retrieve mass amounts of informational resources (Chung et al. 2005). This poses a serious multidimensional informational overload problem attributable to personal factors, information characteristics, task and process parameters, and information technology (Eppler et al. 2004). Consequences range from mild to severe and include less systematic information search strategies (Swain et al. 2000), information omission (Edmunds et al. 2000), lower decision quality (Hwang et al. 1999), and a higher level of error tolerance (Sparrow 1999).

Several countermeasures on the personal, informational, procedural, organizational, and technological levels have been proposed to alleviate information overload (Eppler et al. 2004). For example, on a personal level, user training programs are suggested to enhance information literacy, including information handling, information classification, and information-processing skills (Koniger et al. 1995). Improving users' ability to screen information and become better selectors of relevant facts is another suggested strategy (Van Zandt 2001).

From an informational perspective, *categorizing* information and structuring it for comprehensibility is one approach to deal with information overload (Grise et al. 1999/2000). In fact, several online portals have directories that classify informational resources into different categories, such as arts, music, business, games, health, and so on, in an effort to structure a rather confusing information labyrinth.

This study proposes a meta-cognitive, systems-based, information structuring model founded on the open systems concept of the General Systems Theory (GST) to aid information seekers when using online search engines. It defines a set of hierarchical information structuring strata, or levels, to guide web information seekers into refining and organizing their web search. Although hierarchical information organization is not a novel idea (it goes back to Aristotle, 384 B.C., was later developed through the works of Dewey (1920) and Ranganathan (1957), and is widely-used in the library sciences), applying it to cognitive web information seeking scenarios is the proposed novel approach.

The following sections describe earlier studies of models of information seeking and search behavior, which are summarized in Table 1, followed by a discussion of the pertinent factors that influence information seekers. The study then introduces the Meta-cognitive, Systems-based, Information Structuring Model (McSIS) and presents a walk-through applicatory example, and concludes with the study limitations and suggestions for future research.

MODELS OF INFORMATION SEEKING BEHAVIOR

The information seeking behavior field could be organized into three categories and represented by a nested model, where information behavior acts as the overarching field of research, information-seeking behavior acts as a subset and embodies models and methods individuals use to access information resources, and finally, information search behavior, which studies information user interactions with information retrieval systems (Wilson 1999). Since this study is proposing a web information-seeking model, it falls under the information-seeking behavior category, as shown in Figure 1.

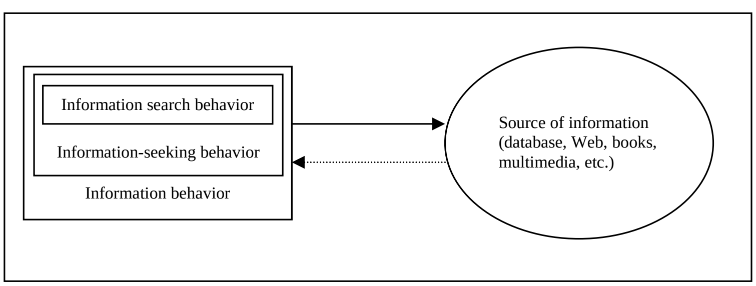

Figure 1. **A Nested Model Showing the Relationship Between Information Behaviors and Sources of Information. (Adapted from Wilson 1999).**

Table 1 summarizes several studies on information-seeking behavior. Two of these models discuss information-seeking behavior in a problem-solving context (Dervin 1983; Wilson 1999). Dervin (1983) suggests a meta-model that explains user information-seeking behavior in terms of an effort to close a gap between a current problematic situation in need of resolution and a desired outcome. Wilson's model, on the other hand, links the stages that an information seeker passes through when searching for information, beginning with problem identification, followed by problem definition, problem resolution, and ending with the solution statement. Each of these steps contributes to uncertainty reduction and resolution. Both models are general in nature. Dervin's model is explicatory in nature, and Wilson's model focuses on the problem per se, and not any steps or stages for information seekers to follow. Such general models are important in providing a broad overview of the rational and basis for information seeking behavior, but they present limited help for information seekers who might be looking for empirical steps and a cognitive framework to structure and organize information they retrieve.

Study	**Model**	**Emphasis**
Wilson (1981)	Information-seeking behavior	Need-based information seeking behavior
Dervin (1983)	Sense-making framework	Problem-solving approach
Ellis (1989)	Information-seeking features	Stages of information search
Kuhlthau (1991)	Information-seeking stages	Stages of information search
Wilson (1996)	Information behavior	Macro-information search behavior
Ingwersen (1996)	Information retrieval process	Cognitive information-seeking behavior
Belkin (1995)	Information activity process	Steps followed in information retrieval
Spink (1997)	Information search process	Interactions in seeking information
Wilson (1999)	Information-seeking process	Problem-solving steps/stages
Hodkinson and Kiel (2003)	Web information search	Personal and technical variables affecting web information search

Table 1. **Information-seeking and search models studies.**

The Web as an open source of information has taken center stage as a starting point, and sometimes as a destination for information need satisfaction. Some of the recent information-seeking models incorporate this new resource. Hodkinson and Kiel, (2003), for example, combine user personal variables (both demographic and behavioral), technical environment, web search task, and several web use and experience variables to be factors affecting external information search behavior on the Web. Variables in the web experience are the web usage level, experience, browser skills, web search skills, web purchase history, and expected benefits from searching the Web. All these variables arguably influence several web search factors, whether horizontal (broad) or vertical (depth).

Ingwersen (1996) explicitly included cognitive elements in the information retrieval process. These include user's work task/interest, current cognitive state, uncertainty, and the problem that needs resolution. These elements interact with the information objects or text/knowledge representations, and the information retrieval system settings to formulate a request which translates into a query carried out through an interface or intermediary. This model emphasizes the process aspect of information retrieval, and the pre-conditions, cognitive and otherwise, that work together to produce a query. The information or knowledge retrieved using this model reacts with the situational and contextual factors shaping the query, but that information is not refined or granulated to enable a sharper informational target pursuit, as suggested by the McSIS model.

FACTORS INFLUENCING INFORMATION-SEEKING BEHAVIOR

Earlier studies demonstrated that user cognitive abilities, such as spatial thinking, and cognitive styles, such as field dependence and independence, influence their information search performance in different information systems (Kim et al. 2002). A cognitive style refers to a person's individual approach of ordering and processing information (Goldstein et al. 1978). One of the most extensively investigated cognitive styles is Field Dependence (FD) vs. Field Independence (FI) learning styles. Field dependent persons are more influenced by their surroundings, while field independent persons overcome the immediate surrounding field and experience objects as separate from their backgrounds (Witkin 1973). These attributes seems to influence web searchers, where FD users searched the Web in a linear fashion, while FI users followed a nonlinear approach, which enabled them to find the information they need faster and more efficiently (Palmquist et al. 2000).

Another cognitive style that polarizes individuals along two ends and influences their information-seeking behavior is holists (H) and serialists (S). Hs perceive their surroundings globally, and they tend to have a broad perspective of what they are engaging in to interact with it. They go from global to local. Ss, on the other had, build their perceptions of their surroundings in a piecemeal fashion, and tend to go from local to global (Pask 1976a; Pask 1976b). The two perspectives of S/H and FD/FI are matched by many researchers, where FD individuals possess cognitive attributes akin to Hs, and FI individuals and Ss behave in a parallel fashion when searching for information (Ford et al. 2000).

Problem-solving styles were also found to affect information-searching behavior (Brown 1991; Wilson 1999). Two problem-solving styles in particular were found to affect individual information-seeking behavior: emotion-focused (EF) and problem-focused (PF) (Zamble et al. 1994). PF individuals exhibit a nonlinear information-seeking behavior, broadening their search before delving deeper into a particular resource. This is in contrast to EF individuals who search linearly, and have a narrower and deeper initial search focus (Kim 1999). These two problem-solving styles share similarities with the learning cognitive styles of FD/FI, and Hs/Ss when seeking information. FI, S, and, PF individuals share similar information seeking behavioral patters, while FD, H, and, EF individuals are on the other side of this cognitive continuum. Figure 2 shows the relationship between these individuals' traits.

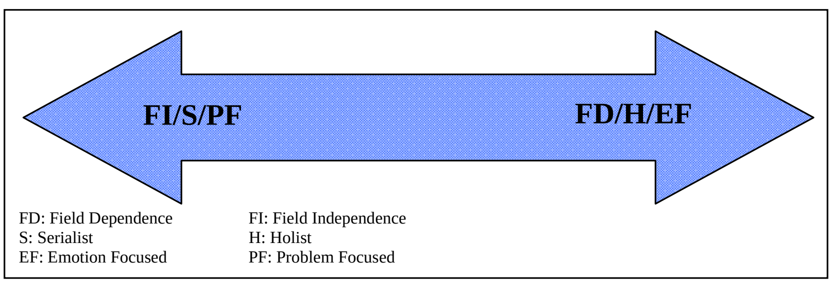

Figure 2. **Factors Influencing Information-seeking Behavior.**

Individuals with FI/S/PF cognitive characteristics could be summarized as preferring induction (from specific to general) rather than deduction. They have analytical perception of facts and events, and prefer to structure situations personally, rather than have them pre-structured. Their goals and motivations are internal, and they like more than one solution to the problems they encounter. FD/H/EF individuals, on the other hand, prefer deduction (from general to specific). Their perception of facts and events is global rather than analytical, and they prefer pre-structured situations. Their goals are defined for them externally, and they prefer external motivation. They also prefer a single solution to a problem (Rose et al. 2005).

An individual's subject or domain knowledge is one of the important factors related to information search behavior. It refers to the "knowledge of the subject area (domain) that is the focus or topic of the search." (Wildemuth 2004, p. 246). It is different from the knowledge of search tactics and strategies information-seekers deploy when searching for information (Bates 1977). Different terms are used to describe the knowledge about a particular issue or topic, such as subject knowledge (Hsieh-Yee 1993; Vakkari 2002), topic or factual knowledge (Allen 1991), and domain knowledge (Bhavnani 2002; Drabenstott 2003). This study will adopt the term domain knowledge to denote individuals' subject knowledge about a topic or issue.

The relationship between domain knowledge and information search behavior has been the focus of many earlier studies. Some studies reached an inconclusive result or found no relationship between domain knowledge and information search behavior (Wildemuth et al. 1995). However, several empirical and theoretical studies found a positive relationship (Bhavnani 2002; Vakkari et al. 2003; Wildemuth 2004). The variables of FI/FD, S/H, PF/EF, and personal domain knowledge are incorporated into the McSIS model. As will be discussed later, these variables influence the manner by which information-searchers traverse the model laterally and vertically.

A META-COGNITIVE SYSTEMS-BASED INFORMATION STRUCTURING MODEL (MCSIS)

The model relies on the concept of meta-cognition, or thinking about thinking (Wolf et al. 2003), where individuals become sentient of their thinking processes as they assimilate information. This cognitive ability ranges form a simple recall process to a complex meta-cognitive introvert dialogue. Hence, when individuals search for information, they are actively cognizant that they are using the model to organize the information retrieved.

The McSIS model proposes a hierarchical conceptual structure of an individual's information or conceptualization about topics or issues. The model has six assumptions, four of which are based on the GST's prepositions and concepts.

1. *Information about topics, objects, and issues are organized in levels/domains.*

A *domain* is a set of related objects, topics, or issues that share common characteristics. These characteristics enable a nomenclature for labeling the domain. Many objects and issues that individuals interact with could be members of a single domain. Examples are ample: computers, automobiles, education, healing, and so on. *This assumption is based on the systems and the production concepts of the GST.*

2. *When dividing a topic or an object into smaller or more detailed pieces, one moves into a sub-level/domain.*

For example, a computer system has several sub components or subsystems necessary for its operation, such as the different software and hardware pieces of cables, case, capacitors and transistors, and the screen. Each of these also could be divided further into subsystems and so on. *This assumption is based on the subsystems and supra-system concepts of the GST.*

3. *When aggregating topics or objects into a broader context, one moves into a supra-level/domain.*

Following the same example above, the set of wires, cables, boards, and transistors compose a whole unit, which is the computer. The set of computer hardware, software, people, procedures and other components compose the information system domain. *This assumption is based on the subsystems and supra-system concepts of the GST.*

4. *Information about topics and objects within a level/domain is related linearly and across levels/domains vertically (supra- and sub-levels/domains). In other words, intra-domain objects and issues are related to other sub- and supra-levels/domains through inter-level/domain relationships.*

Since topics and objects are encompassed within one domain, and share common characteristics, they are related. And since the GST proposes that systems are part of other supra systems, and they are composed of smaller components or subsystems, then there is also a relationship between domains, or supra- and subsystems. *This assumption is based on the subsystems and supra-system concepts of the GST.*

5. *Users are cognizant of their thinking process when interacting with information.*

Individuals are aware of the topic of interest and are actively seeking more information about it. When using this model, users are also aware of its levels/domains and how to traverse them. This awareness is called *meta-cognition* (Wolf et al. 2003).

6. *Users possess some knowledge (personal domain knowledge) about the topic searched, ranging from trivial to extensive expertise.*

The McSIS model is composed of five levels/domains (*meta, macro, meso, micro, and* n*ano*), and each represents a level/domain in varied degrees of detail or aggregation. As a searcher moves upwards in the model, topics, issues, and objects are integrated or aggregated, and the level of detail is decreased, while moving downwards increases the level of detail and narrowness. The following general discussion presents each level/domain.

1. *The meso-level/domain*. This is the model's mid-level entry point, where an individual has an idea or a topic which represents the current *domain/topic* of interest. Topics range from an abstract issue or concept to physical material and objects. The degree of determination or specificity of the topic depends on the individual's *personal domain knowledge*, which falls on a continuum ranging from minor or superficial to extensive and detailed.

2. *The macro-level/domain*. This level/domain lies above the meso-level/domain and is broader, comprising a higher-order, or supra-system, of the initial topic. At the macro-level/domain, the topic introduced in the earlier meso-level/domain would be a derivative or subsystem. At this level, ties to other subsystems and systems are identified.

3. *The meta-level/domain*. This is the highest level/domain in the model, and the most encompassing and broadest for the topic that was introduced at the meso-level/domain. The person would be at a global perspective and at the highest domain of abstraction. The macro-level/domain is a subsystem of the meta-level/domain.

4. *The micro-level/domain*. This is a subsystem of the meso-level/domain. The topic is broken into smaller pieces or components for further examination. One of these components or aspects becomes the focus of interest.

5. *The nano-level/domain*. At this level, the idea or topic considered in the micro-level/domain is further broken down into smaller components. At this level, the degree of specificity is much higher, and the topic is a subsystem of the micro-domain.

Comment [HAC1]: Word choice

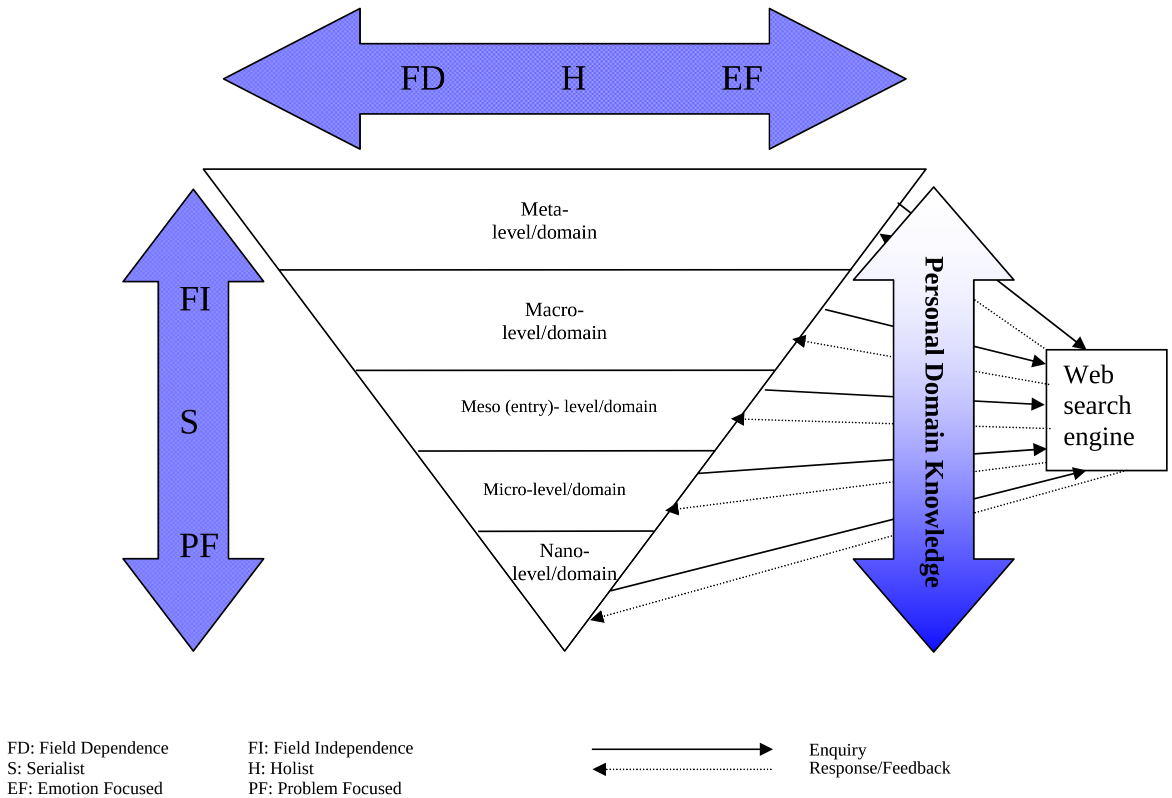

Figure 3. **A Meta-cognitive Systems-based Information Structuring Model (McSIS).**

The McSIS model shown in Figure 3 depicts the personal characteristics or variables that influence users when seeking information. These characteristics include the individual's learning style (FD/FI), the problem-solving approach (H/S, PF/EF), and the personal domain knowledge about the topic or issue sought. Figure 2 depicts the FI/S/PF individuals on one end of a continuum, and FD/H/EF individuals on the other end. In the McSIS model in Figure 3, however, these two ends are displayed on two sides, vertically and laterally. This separation is made to emphasize the influence of these variables on individuals traversing the model. Individuals whose cognitive characteristics are closer to the FD/H/EF end are more likely to stay within a single domain and find it more difficult to move vertically in the model. On the other hand, FI/S/PF individuals are more likely to be able to move in either direction (laterally and vertically) when using the model.

Because of the infinite number of topics and issues that could be sought on the Web, a rather simple and common topic will be demonstrated. Suppose an instructor is interested in knowing about the use of participatory technology and its effect on student enthusiasm and class participation. This would be the entry point to the model at the Meso level/domain, and an initial inquiry is initiated on the Web (solid arrow in the model). Keywords used might simply be "class participation *and* technology," as a Boolean search, for example. About fifteen million links are returned from this search from Google.com. If the information retrieved (dashed arrow in the model) is sufficient and answers the question or inquiry, searching simply terminates. This will be the case where the intent of the search is to explore this topic and read preliminary articles about technologies used that encourage student class participation, or review other's classroom experiences.

However, if more information is needed, an inquiry would be initiated at either the macro- or the micro- levels/domains, or it could stay at the meso- level/domain, depending on where the investigation is heading. If information is needed about the relationship between the class participatory technology and class size for example, then the investigation is still at the meso- level/domain, because the search is about the initial topic and no fragmentation (subsystems) or aggregation (supra-systems) takes place to take the inquiry to a different level. One can go up to supra-systems or down to subsystems, depending on the personal domain knowledge, and the type of personal learning style (FD/FI; S/H; or PF/EM).

CONCLUSION

One of today's dilemmas for information-seekers is not the availability of information as much as finding the needed information among the voluminous sources retrieved by online search engines. This study introduces a McSIS model that is based on the GST. The model is proposed to systematize information seeking when searching online. Several end-user factors influencing information-seekers are incorporated in the model, such as the continuums of FI/FD, PF/EF, S/H, and personal domain knowledge. The model is dynamic and treats information retrieval from an end-user perspective, where the information-seeker interacts with an information source using this model as a frame of reference to find the sought information. The outcome of using the model is expected to be faster, more accurate, and more relevant search results.

An analogy of the benefit from using the McSIS model is using different tools and methods to draw a map that a hunter would use to navigate a forest. The different levels of the McSIS model represent the different tools that could be used to draw the map. For example, the macro- and meta- levels are like using an airplane or helicopter to look at the forest from above to aid in drawing the elevations and the general topography of the terrain, while the micro- and nano- levels represent a walk inside the forest to gather information about the different species of plants, insects, and animals. The outcome is a rich map for hunters to use when seeking prey. With the map, quicker and more accurate finds are expected, and sans the map, hunting is more haphazard and serendipitous than targeted and focused.

STUDY LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

The study has several limitations. First it is theoretical in nature, whereas the McSIS model has not been empirically tested. Testing the model could be accomplished through an experimental design study for example, with two groups: a treatment group that uses the model and a control group that does not. Search accuracy, speed, and relevance of both groups is compared to determine the effectiveness of the model in assisting online searchers in finding information.

A second limitation of the study is that it is normative rather than explicatory, as the other models of information-seeking discussed in the study are designed to accomplish. The McSIS model does not explain why individuals seek information online as much as it proposes a cognitive model for organizing the retrieved information and recreating a better set of terms or keywords to use in an online search engine.

The third limitation pertains to the variables influencing information-seeking behavior. Environmental variables were not included in the model to avoid model complexity. Although situational or contextual variables play an important role in influencing individuals' information-seeking behavior, such as the task or subject searched, the time constraint, and the urgency of the information need, the model focuses on the cognitive factors closely related to the searcher, such as the cognitive learning styles and the problem-solving approaches. Future studies could incorporate different permutations and combinations into the model to examine their individual and collective effect on information-seeking behavior.

REFERENCES

Allen, B. "Topic Knowledge and Online Catalog Search Formulation," *Library Quarterly* (61:2), 1991, pp. 188–213.

Bates, M. J. "Factors Affecting Subject Catalog Search Success," *Journal of the American society for Information Science* (28:3), 1977, pp. 161–169.

Bhavnani, S. K. "Domain-specific Search Strategies for the Effective Retrieval of Healthcare and Shopping Information," *SIGCHI, Extended Abstracts*, 2002, pp. 610–611.

Brown, M. E. "A General Model of Information-seeking Behavior," *Proceedings of the 54th Annual Meeting of the American Society for Information Science*, USA, 1991, pp. 9–14.

Chung, W., Chen, H., and Jr, J. F. N. "A Visual Framework for Knowledge Discovery on the Web: An Empirical Study of Business Intelligence Exploration," *Journal of Management Information Systems* (21:4), 2005, p. 57.

Dewey, M. "Decimal Classification Beginnings," *Library Journal* (45), 1920, pp. 151–154.

Drabenstott, K. M. "Do Nondomain Experts Enlist the Strategies of Domain Experts?," *Journal of the American Society for Information Science and Technology* (54:9), 2003, pp. 836–854.

Edmunds, A., and Morris, A. "The Problem of Information Overload in Business Organizations: A Rreview on the Literature," *International Journal of Information Management* (20), 2000, pp. 17–28.

Eppler, M. J., and Mengis, J. "The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines," *The Information Society* (20), 2004, pp. 325–344.

Ford, N., and Chen, S. "Individual Differences, Hypermedia Navigation, and Learning: an Empirical Study," *Journal of Educational Multimedia and Hypermedia* (9:4), 2000, pp. 281–311.

Goldstein, K. M., and Blackman, S. *Cognitive Style: Five Approaches and Relevant Research*, John Wiley, New York, 1978.

Grise, M., and Gallupe, R. B. "Information Overload: Addressing the Productivity Paradox in Face-to-face Electronic Meetings.," *Journal of Management Information Systems* (16:3), 1999/2000, pp. 157–185.

Hsieh-Yee, I. "Effects of Search Experience and Subject Knowledge on the Search Tactics of Novice and Experienced Searchers," *Journal of the American Society for Information Science* (44:3), 1993, pp. 161–174.

Hwang, M. I., and Lin, J. W. "Information Dimension, Information Overload and Decision Quality," *Journal of Information Science* (25:3), 1999, pp. 213–218.

Kim, K. S. "Searching the Web: Effects of Problem-solving Style on Information-seeking Behavior," *World Conference on Educational Multimedia, Hypermedia & Telecommunications*, USA, 1999, pp. 1541–1542.

Kim, K. S., and Allen, B. "Cognitive and Task Influences on Web Searching Behavior," *Journal of the American Society for Information Science and Technology* (53:2), 2002, pp. 109–119.

Koniger, P., and Janowitz, K. "Drowning in Information, but Thirsty for Knowledge," *International Journal of Information Management* (15), 1995, pp. 5–16.

Palmquist, R., and Kim, K. S. "Cognitive Style and Online Database Search Experience as Predictors of Web Search Performance," *Journal of the American Society for Information Science* (51:6), 2000, pp. 558–566.

Pask, G. "Conversational Techniques in the Study and Practice of Education," *British Journal of Educational Psychology* (46), 1976a, pp. 12–25.

Pask, G. "Styles and Strategies of Learning," *British Journal of Educational Psychology* (46), 1976b, pp. 128–148.

Ranganathan, S. R. *Prolegomena to Library Classification*, Second Edition, Library Association, 1957.

Rose, E., Le Heron, J., and Sofat, I. "Student Understandings of Information Systems Design, Learning and Teaching: A Phenomenography Approach," *Journal of Information Systems Education* (16:2), 2005, pp. 183–195.

Sparrow, P. R. "Strategy and Cognition: Understanding the Role of Management Knowledge Structures, Organizational Memory and Information Overload," *Creativity and Innovation Management* (8), 1999, pp. 140–149.

Swain, M. R., and Haka, S. F. "Effects of Information Load on Capital Budgeting Decisions," *Behavioral Research in Accounting* (12), 2000, pp. 171–199.

Vakkari, P. "Subject Knowledge, Source of Terms, and Term Selection in Query Expansion: An Analytical Study," *Advances in Information Retrieval: 24th BCS-IRSG European Colloquium on IR Research*, Glasgow, UK, 2002.

Vakkari, P., Pennanen, M., and Serola, S. "Changes of Search Terms and Tactics while Writing a Research Proposal," *Information Processing & Management* (39), 2003, pp. 445–463

Van Zandt, T. "Information Overload in a Network of Targeted Communication," (2001), http://ged.insead.edu/fichiersti/inseadwp2001/2001-36.pdf, INSEAD Working paper 2001/36/EPS.

Wildemuth, B. M. "The Effects of Domain Knowledge on Search Tactic Formulation," *Journal of the American Society for Information Science and Technology,* (55:3), 2004, p. 246.

Wildemuth, B. M., de Bliek, R., Friedman, C. P., and File, D. D. "Medical Students' Personal Knowledge, Searching Proficiency, and Database Use in Problem Solving," *Journal of the American Society for Information Science*, (46:8), 1995, pp. 590–607.

Wilson, T. D. "Models in Information Behavior Research," *Journal of Documentation,* (55:3), 1999, pp. 249–270.

Witkin, H. A. *The Role of Cognitive Style in Academic Performance and in Teacher-student Relations,* Educational Testing Service, Princeton, NJ, 1973.

Wolf, S., Brush, T., and Saye, J. "Using an Information Problem-solving Model as a Metacognitive Scaffold for Multimedia-supported Information-based Problems," *Journal of Research on Technology in Education,* (35:3), 2003, p. 321.

Zamble, E., and Gekoski, W. L. "Coping," in *Encyclopedia of Human Behavior,* V. S. Ramachandran (ed.), Academic Press, New York, 1994, pp. 1–10.