

Need for a Systemic Theory of Classification in Information Science

Murat Karamuftuoglu

Department of Communication and Design, Faculty of Art, Design and Architecture, and Department of Computer Engineering, Faculty of Engineering, Bilkent University, Bilkent, Ankara 06800, Turkey. E-mail: hmk@bilkent.edu.tr

In the article, the author aims to clarify some of the issues surrounding the discussion regarding the usefulness of a substantive classification theory in information science (IS) by means of a broad perspective. By utilizing a concrete example from the High Accuracy Retrieval from Documents (HARD) track of a Text REtrieval Conference (TREC), the author suggests that the “bag of words” approach to information retrieval (IR) and techniques such as relevance feedback have significant limitations in expressing and resolving complex user information needs. He argues that a comprehensive analysis of information needs involves explicating often-implicit assumptions made by the authors of scholarly documents, as well as everyday texts such as news articles. He also argues that progress in IS can be furthered by developing general theories that are applicable to multiple domains. The concrete example of application of the domain-analytic approach to subject analysis in IS to the aesthetic evaluation of works of information arts is used to support this argument.

Introduction

The Debate—Do We Still Need a Classification Theory?

Recent articles by Hjørland and Nissen Pedersen (2005) and Spärck Jones (2005) instigated a dialogue regarding the need and usefulness of a substantive theory of classification for information retrieval (IR). Briefly stated, Hjørland and Nissen Pedersen (2005) argue that further progress in IR is impeded by a lack of a substantive classification theory, which would enable categorization of documents on the basis of goals, purposes, and values of given user communities. They further argue that such pragmatic interests (i.e., goals, purposes, and values) are always theory-laden, in other words, depend on particular theories, approaches, and world-views that exist in a given domain. In an article written in response to Hjørland and Nissen Pedersen's article,

Spärck Jones (2005) agreed on the importance of contexts and goals, however, raised doubts about usefulness of a priori classification of documents by human experts for the general retrieval task. She suggested that minimal and indirect operational mechanisms such as relevance feedback practically diminished the need for human involvement, and consequently, a general or substantive theory of classification in IR.

My purpose here is twofold: first, to clarify some of the issues surrounding the discussion regarding the usefulness or otherwise of a substantive classification theory in IR raised in the aforementioned articles; and second, to broaden the debate to a larger context by including in the discussion tasks other than IR and domains other than information science. In the first part of the article, the discussion carefully elucidates the pragmatic, theoretical, and metatheoretical levels in classification. By means of a concrete example from the High Accuracy Retrieval from Documents (HARD) track of a Text REtrieval Conference (TREC), the limitations of the state-of-the-art automatic document classification are documented. This suggests that there is room for improvement in the existing technology, and a substantive theory of classification may directly or indirectly help the goal of improving retrieval effectiveness. In the second part, I argue that a substantive theory of classification may have important uses in tasks and domains other than IR and information science, in particular, aesthetic theory in art. This final point is in line with the approach outlined in general systems theory (GST), and suggests that it may be beneficial for the development of information science as a discipline to elaborate theories that can be exported to domains that lay outside its traditional boundaries.

Spärck Jones (2005, p. 601) suggests,¹ broadly speaking, that a priori or “pre-” classification in IR is not needed in most circumstances because relevance feedback is an effective means of constructing classes (relevant, nonrelevant) automatically based on chosen exemplars. This view advocates

Received September 15, 2006; revised January 3, 2007, January 23, 2007, February 8, 2007; accepted February 8, 2007

© 2007 Wiley Periodicals, Inc. • Published online 26 September 2007 in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/asi.20678

¹There are a number of other related issues discussed in Spärck Jones' (2005) article; however, our focus here is limited to the sufficiency of a posteriori methods, specifically, relevance feedback, in classification.

a posteriori classification of documents, which does not require a substantive classification theory or a priori classification of documents. In relevance feedback, the user examines some of the top-ranked documents retrieved in response to the user's query, and assesses their relevance to his or her information need. User's initial query statement could then be enriched by expanding it with useful terms extracted from the user-assessed relevant documents. Query statements expanded in this way have been shown to yield higher retrieval effectiveness. However, there are several limitations of this approach, which will be discussed in detail later in this article. It suffices here to note briefly some of the major limitations of relevance feedback.

- On a practical level, most commercial systems, especially those on the Web do not employ relevance feedback for technical and other reasons. Therefore, application of relevance feedback in operational systems is limited, and it may be useful to investigate alternatives to improve retrieval performance of operational systems.
- Importantly, existing IR systems' reliance on the "bag of words approach" (i.e., treatment of search and index terms as independent, atomic units; cf. Jones, Vechtomova, & Dias, 2005) means that complex information needs cannot be expressed in terms of existing query languages. In such cases, there may not be any relevant document among the top-ranked ones for a user to feed back into the system for query expansion. Certain sophisticated information needs that depend on the specific goals, contexts, and purposes of the users could only be resolved if they can first be expressed in a structured and detailed way.
- There is a deeper level that underlies certain information needs, which cannot be adequately dealt with on an impromptu basis. In science and similar domains, there are philosophical (ontological, epistemological, and methodological) factors that shape and determine information needs in certain situations, for example research, which require a systematic approach to the classification of documents. In this article, it will be shown that a similar layer may underlie information needs in more mundane domains.

Hjørland (1992, 1997, 2002) and Hjørland and Albrechtsen (1995) argue for a philosophically informed pragmatic approach to subject analysis and classification. This approach, which is sometimes referred to as *domain* or *sociocognitive analysis*, can be summarized in the following way: Document classification should be based on (informed by) the theories/paradigms that exist in a given domain or discipline, as well as tasks and goals of specific user groups. Such analysis is usually applied to scientific or academic domains, however, examples from other disciplines, for instance, the arts (Ørom, 2003), can also be found. Here this kind of analysis is applied to a newswire story taken from the HARD track of TREC 2004. The choice of the example from journalism makes it easier to relate the discussion to ordinary, everyday uses of retrieval technology. By means of this example, some of the points raised in the discussion between Spärck Jones and Hjørland and Nissen Pedersen are explored.

The rest of this article is organized as follows: in the next section, a brief description of the TREC conferences and HARD track is presented. In the following two sections, the interdependence between pragmatic and metatheoretical bases of information needs are discussed by means of a realistic example taken from the HARD track of TREC 2004. In the remaining part of the article, I argue that a substantive theory of classification can be of value in domains other than information science. Specifically, I will show that a theory of classification developed in the context of IR can be adapted to an entirely new task, namely aesthetic criticism, in a different domain, art. My goal in this part of the article is to investigate the issue of the value of classification theory in information science from a wider perspective.

TREC and Hard Track

Text REtrieval Conference (TREC) series, cosponsored by the National Institute of Standards and Technology (NIST) and the Defence Advanced Research Projects Agency (DARPA), has been the major platform for evaluation of IR methods and systems under laboratory conditions since its inception in 1992. There are several "tracks" in TREC, which focus on different types of text classification and related tasks (e.g., ad hoc queries, filtering, question answering) on a variety of collections (Web documents, newswire articles, etc.). High Accuracy Retrieval from Documents (HARD) track of TREC (Allan, 2005) focuses on improving the retrieval performance by means of a limited interaction with the users. Unlike the ad hoc scenario, in HARD track a restricted one-time interaction with the users (assessors) is allowed. Topics (queries) in HARD track are defined in the standard TREC fashion, consisting of title, description, and narrative fields. The title field summarizes the information need in a few words. The description field is a one-sentence brief outline of the information need. The narrative field provides more details about the need and the relevance criteria (see Table 1).

Evaluation is done in TREC by pooling documents retrieved by the participating sites in a track (Voorhees, 2005). Each participating site submits 1,000 top-ranked documents retrieved by their system. Documents returned by the participants are then pooled and the top *n* (usually 100) documents submitted by each participant are evaluated by the users (assessors) invited by NIST. Standard metrics of

TABLE 1. An example topic from the High Accuracy Retrieval from Documents (HARD) track of a Text REtrieval Conference (TREC).

Topic number:	HARD-409
Title:	AIDS in Africa
Description:	What is the state of AIDS in Africa?
Topic-narrative:	Little attention has been given to the AIDS epidemic in Africa that has decimated an entire generation of Africans. What is being done to help prevent the spread of AIDS and to treat those already infected? What sorts of public education/health measures have African governments taken? What are the barriers?

recall, precision, and/or other similar measures are usually used to evaluate the effectiveness of the systems participating in the experiments.

Analysis of the Pragmatic Basis of an Information Need

The topic shown in Table I from HARD track of TREC 2004 will be used to illustrate the discussion in this section. In our discussion, we will carefully differentiate the pragmatic level from the philosophical level that underlies information needs. The narrative field of the topic given in Table I describes the information need of the user in some detail. Clearly, the user is not just interested in documents that broadly discuss the acquired immune deficiency syndrome (AIDS) epidemic in Africa, but in finding out what agent(s) (method, process, drug, etc.) can stop the spread of the disease and help treat its victims. The user is also interested in the barriers that may impede the fulfillment of these goals, namely prevention and treatment of the disease. The above-named requirements are related to the pragmatic aspects of the user's information need. The factors that may help or hamper the identification of documents that fulfill the pragmatic needs of the originator of this query are examined next.

Acquired Immune Deficiency Syndrome is a significant health concern and the subject of public discussion. The accepted explanation of the cause of the disease by the overwhelming majority of the medical profession is the Human Immunodeficiency Virus (HIV). However, there is a small number of so-called AIDS dissidents or AIDS denialists, who raise doubts about the HIV/AIDS connection. One well-known dissident is Peter Duesberg, who is an award-winning professor of molecular and cell biology at the University of California, Berkeley (Duesberg, 2006; Peter Duesberg, 2006). His objections to the HIV/AIDS link are quite detailed and technical. His alternative theory of causation of AIDS is, however, straightforward. First, he makes a distinction between the American/European AIDS epidemic and the African AIDS epidemic (Duesberg 2000, 2006; Duesberg & Rasnick, 1998). In the former case, he claims, recreational and pharmaceutical drug use (especially AZT, which is often prescribed to AIDS sufferers), and not HIV, is the primary cause of AIDS (Duesberg 2000, 2006; Duesberg hypothesis, 2006; Duesberg & Rasnick, 1998). The African AIDS cases, he claims, are also caused not by HIV, but by various conventional and widespread factors, such as malnutrition, parasitic infections, and poor sanitation (Duesberg 2000, 2006; Duesberg hypothesis, 2006; Duesberg & Rasnick, 1998). There are many supporters of his claims within the medical profession and among AIDS activists; however, the majority of experts in the subject vigorously oppose his views, even accuse him of harming the fight against AIDS. Regardless of the validity or not of his claims, clearly this is a radically different explanation of the cause of AIDS that puts the emphasis, in the case of the African epidemic, on nutrition, sanitation, and general health care, which are linked ultimately to economic, social, and political factors. The pragmatic goals of treatment and prevention of the spread of

AIDS are therefore, from this perspective, related to the improvement of economic and social conditions, which contrasts sharply with the mainstream view. The mainstream view links AIDS primarily to infection with a virus, therefore emphasizes drugs, medical practices, and education in the fight against AIDS, in contrast to broader economic, social, and political factors.

It is interesting to analyze one of the documents from the HARD TREC 2004 corpus retrieved in response to the above query; document XIE20030221.0307, entitled "Roundup: France, Africa Seek Mutual Benefit," which is judged as nonrelevant by the TREC assessor. This document reports the results of the 22nd Africa–France Summit held in Paris, February 2003. The second sentence in the news story underlines the overall outcome of the meeting: "At the two-day summit, France promised to use all its influence to muster concrete international support, especially that from developed countries, for African development and the New Partnership for Africa's Development (NEPAD)." The article continues with a brief comment on the consequences of France's commitment to African development, and summary of each of the main issues discussed at the summit. Among the issues discussed was agricultural development:

With 40 million Africans facing famine and 70 percent of the population tied up in the agricultural sector, agriculture emerged as a hot topic at the summit with the African countries being offered favorable proposals on African agricultural development. "Agricultural development is vital for sub-Saharan Africa to attain the average annual growth rate of seven percent, that will enable it to achieve the millennium goal," Chirac said. . . . He called for a long-term response to famine in Africa, encouraging African countries to develop ambitious agricultural policies.

The story concludes with United Nations Secretary General Kofi Annan's remarks on AIDS in Africa:

Annan, for his part, put agriculture on the list of three areas besides AIDS and governance that should be addressed urgently in Africa.

He called for the anti-AIDS war throughout the continent, revealing that the United Nations will set up a high-level commission on HIV/AIDS and governance in Africa, which will "come up with detailed recommendations for stemming the tide of the disease across Africa, and advise African policymakers on how to address the profound structural impact it is beginning to have on their ability to tackle their many development challenges."

France also confirmed its determination to increase its official development aid to 0.5 percent of the country's GDP, most of which being expected to flow to Africa.

AIDS is mentioned only in the above excerpt from the document, which is about 90 words long, and comprises roughly 7% of the whole document. There was no mention anywhere in the document, of the measures taken to prevent

the spread of AIDS, such as educating the public on the transmission channels of the HIV virus, allegedly the cause of the disease according the dominant view, or the use of drugs in the treatment of the sufferers. From the perspective of the dominant view, there is no reason to consider the above document relevant to the query given in Table 1. However, from the alternative theory of causality, which links AIDS primarily to poverty and malnutrition, the document may arguably be considered as relevant, given the French president's and the United Nations Secretary General's promises for economic aid to Africa reported in the article. From an alternate perspective, AIDS is no different from many other "conventional" diseases that are caused by or spread because of poor sanitary and environmental conditions, and malnutrition. Therefore, from this perspective combating the disease is likely to be seen as foremost a matter of development of Africa's overall economic and social conditions. It is arguable that the mention of the U.N. Secretary General's concerns about AIDS (although there was no explicit mention of measures to combat it), makes the connection between the document's content and AIDS in Africa stronger. This argument is strengthened by the presence of the South African President Mbeki at the summit as reported in the news story: "South African President Thabo Mbeki, who is also president of the African Union, said after the summit that the just-concluded meeting will enhance African cooperation with France and the developed world." Thabo Mbeki is known for his voiced support for the Duesberg hypothesis (Bethell, 2000; Peter Duesberg, 2006). He organized a conference in 2000 on AIDS "where it was announced that the HIV theory would get proper epidemiological testing . . ." (Bethell, 2000; Peter Duesberg, 2006). Given the presence of three related elements: promised aid for economic development in Africa, active participation in the negotiations of the South African president Mbeki who is a prominent supporter of the alternative view on AIDS, and the U.N. Secretary General's voiced concerns about AIDS, there is a strong case for considering this document relevant to the query from the alternative perspective.

Metatheoretical Basis of Document Classification: An Example from Medicine

In the preceding section, I demonstrated that relevance of a document to a query or to a particular enquirer depends on the ability of the document to provide answers to specific questions. This is the pragmatic aspect of relevance: whether the document helps the enquirer to achieve her specific goals. However, the above analysis of the TREC topic "AIDS in Africa" suggests that what kind of questions a document could help answer depends on the conceptual model(s) in the mind of the reader that explain, roughly speaking, the behavior of a certain natural or social phenomenon, i.e., theories that explain cause and effect relationships between things in the world. The question of whether computational means alone are sufficient in evaluating/classifying documents with respect to models or theories is discussed later in the article.

It would be useful first to consider the limits of the human classifier.

Human classifiers are usually expected to make themselves reasonably familiar with the concepts, terminology, and even the basic models/theories in the subject of classification; however, it is unreasonable to expect them to have a detailed understanding of all significant issues in a domain. It is arguable that even subject experts do not necessarily know everything relevant to a particular information need in the domain of their expertise. For instance, it may be unreasonable to expect a graduate in medicine to know about the Duesberg hypothesis discussed in the preceding section, which would be useful in classifying the document (XIE20030221.0307). It is this limitation of the human classifier that makes education in basic philosophy (epistemology, ontology, methodology, history of science, etc.) a necessity as shown below.

Theories are based on metatheoretical assumptions or metatheories. A *metatheory*, or a *paradigm*, is a set of principles that prescribes what is acceptable and unacceptable as theory in a scientific discipline (Overton, 1998). In epidemiology, for instance, one can distinguish between the dominant biomedical paradigm, and alternative emerging or past paradigms. The dominant biomedical paradigm focuses on the biology of disease. Therefore, research in the dominant paradigm investigates supposedly universal exposure-disease relationships in isolation from the contexts in which exposures occur (Wing, 1994). The populations of modern epidemiology are counts of individuals. Populations, in this context, are means of making comparisons of averages; therefore, they are not defined as groups with unique economic, social, and ecological histories (Wing, 1994). As Wing (p. 76) explains,

Their features of organization, in the epidemiological context, are not considered to have etiological consequences. Epidemiological studies that do address factors such as economic position or occupation generally treat them only as individual attributes or exposure markers rather than as aspects of social and economic organization that provide the context for biopsychosocial development.

The dominant paradigm in epidemiology has been criticized increasingly for ignoring the contextual factors: ". . . the fundamental object of inquiry in modern epidemiology, dose response, should be recognized as essentially contextual (developmental or historical) rather than universal, vastly complicating the reductionist program and, indeed, challenging the very de-contextualization on which it is based" (Wing, 1994). Recent attempts to expand the scope of modern epidemiology search for past examples of successful contextual epidemiological practices. Wing (1994, p. 79) gives an example from 19th century Germany:

The mid-19th century work of the young Rudolf Virchow has been revived as an early example of a quantitative approach to understanding disease in populations that, while recognizing the importance of specific agents or exposures,

did not reduce the explanation of disease to a matter of these isolated factors themselves. . . . Virchow, in investigating an epidemic of typhus in Silesia, was deeply moved by the suffering of the people, and his explanations stressed the conditions that fostered the epidemic: lack of agricultural land, malnutrition, poor housing, low wages, and language barriers for the large Polish minority. His report to the government advocated land reform, progressive taxation, establishment of agricultural communes, local political autonomy, and, lastly, creation of a system of public hospitals. Virchow's conclusion: "Medicine is a social science, and politics nothing but medicine on a grand scale."

Wing, cites a number of other examples from the history of epidemiology, as well as more recent attempts for a new contextual approach in Latin America. One particularly interesting example is the public health research that Friedrich Engels (1845/1958) conducted in England:

He documented the health problems that arose from crowding, lack of sanitation, malnutrition, and abuse of alcohol to alleviate chronic pain. . . . Unlike Virchow, however, Engels did not believe that reform of the political system would ever create the underlying conditions for adequate public health. Rather, he identified the capitalist economic system itself as the source of ill health (Wing, 1994, p. 79).

The Latin American example of contextual epidemiological practice concerns the attempts of Salvador Allende, former Chilean president, who was also a physician:

In Chile, the physician Salvador Allende came to believe that he would make the greatest contribution to the health of his people . . . by working against the devastating effects of under development . . . Latin America—where it is clear that the immediate public health problems of disease in populations have less to do with specific exposures than with a position in the international economic system that sustains a lack of decent jobs, housing, clean water, food, and democratic control of institutions—is now home to a number of alternative currents of development in epidemiology (Wing, 1994, p. 79).

The above examples of alternative paradigms in epidemiology provide strong support for the plausibility of classifying the document XIE20030221.0307 discussed in the preceding section as relevant to the query presented in Table 1, given the Duesberg hypothesis regarding the causes of AIDS.² It is interesting to note that Virchow in the above example specifically cites a lack of agricultural land among the factors that contributed the typhus epidemic in Silesia in the 19th century. Recalling that document XIE20030221.0307 reports on the promise of help for agricultural development in Africa,

²Although the Duesberg hypothesis may one day be proven false, the importance of the above example is that it demonstrates that different theoretical views may always be at play in interpreting documents, and document classification ultimately depends on the theoretical views used in the analysis of documents.

the relevance of this document to the topic of prevention of African AIDS epidemic becomes, from the alternative epidemiological paradigm, evident. It is also interesting to note the similarity of the political views of the two advocates of a broader contextual perspective in public health. Both Friedrich Engels and Salvador Allende were important socialist political figures. It is plausible, therefore, to imagine that an enquirer from an alternative, perhaps a Marxist or socialist perspective, could consider the information presented in document XIE20030221.0307 significant to his or her research on the state of AIDS in Africa (Table I). Marxism, along with Durkheimian and Weberian sociology, behaviorist psychology, symbolic interactionism, and structural functionalism is one of the main metatheoretical frameworks in the social sciences (Roberts, 1999). This last point brings to the fore the importance of a generalist approach to the training of human classifiers.

In many domains, multi- or interdisciplinary research becomes increasingly prominent. For instance, Peterson and Martin (2000) underline the importance of interdisciplinary approaches in the medical domain:

Multidisciplinary and interdisciplinary approaches to research have developed fairly rapidly as ways of addressing more complex issues in health research. Much has been gained from models of health research in developing countries, which have forced creative approaches to difficult problems. These have essentially covered the need to combine health research approaches between medical researchers (operating with both biomedical and epidemiological approaches), and those from other disciplines (i.e., economics, sociology, psychology, anthropology and demography). They produce very useful approaches and models for researching a variety of illness, disease, service provision, health management and treatment outcomes, more so than many single discipline research programs.

It is unreasonable to expect human classifiers to be experts in various disciplines that contribute to interdisciplinary research in today's world of complex problems. Even if a human classifier could be trained to a satisfactory level in a particular discipline, in-depth knowledge of several disciplines is obviously not feasible given the number of disciplines potentially relevant to the solution of a complex problem. This fact intensifies the need for a generalist approach to the training of classification professionals, reflecting the increased need for an interdisciplinary approach to complex problems encountered in the world. Although attaining a level of expertise in many disciplines is not feasible for most people, acquisition of reasonable knowledge of main metatheoretical positions seems realistic, especially given the fact that certain schools of thoughts in philosophy and social sciences play a fundamental role in shaping research frameworks or paradigms in various disciplines. The discussion of recent approaches in epidemiology earlier illustrated the role played by metatheories from social sciences and political philosophy in understanding the alternative paradigms. Knowledge of such fundamental schools of thought is indispensable in

uncovering implicit assumptions that underlie research in many domains. Returning to the example of a search for documents about AIDS in Africa, it may be unreasonable to expect a human classifier to be familiar with the Duesberg hypothesis, given its marginal position in the medical community,³ or the emerging alternative paradigms in epidemiology, within which the Duesberg hypothesis makes sense. However, it is arguable that a human classifier familiar with major schools of thoughts in political philosophy and social sciences would be able to uncover or evaluate the potential value of a document such as XIE20030221.0307 in relation to the subject of prevention of AIDS in Africa for someone or a group from a different background of race, gender, or class, or with atypical political views. Such factors and sociopolitical contexts of information needs are usually overlooked by the positivist and mentalist (or methodological atomist, cf. Albrechtsen & Hjørland, 1997) traditions in IR and IS. Hermeneutically and critically oriented approaches in IS, such as the domain-analytic (sociocognitive) approach, tend to incorporate such factors in the interpretation of documents, and therefore, could provide a basis for a substantive classification theory that helps users with different theoretical perspectives in identifying documents relevant to their needs.

Evaluation of document relevance with respect to different theories (e.g., the Duesberg hypothesis) that may exist in a domain could become practical if the human classifiers are trained in metatheories in different domains (e.g., the dominant and alternative metatheoretical perspectives in epidemiology) and/or basic philosophical and social/political theories. Continuing with the same example, the Duesberg hypothesis is linked to alternative epidemiological paradigms, which are informed by one or more sociopolitical theories/movements (e.g., Marxism, environmentalism, etc.). From this perspective, information science is an interdisciplinary/generalist discipline, and education in information science should be based on general knowledge of various social, political, and philosophical theories, movements, and schools of thought. This point will be taken up later in the penultimate section of this article.

Automatic Document Classification: Pragmatic and Metatheoretical Requirements

The document XIE20030221.0307 is ranked by the Okapi retrieval system, one of the standard best-match systems, at 609th place in one of our experiments in HARD TREC (Vechtomova & Karamuftuoglu, 2005) in response to query terms “AIDS, Africa”⁴ derived from the title field of topic

³It is claimed that Duesberg’s research funding has been cut-off and major publishers of medical literature have been unwilling to publish Duesberg’s writings because of his controversial claims (Duesberg, Koehnlein, & Rasnick, 2003; Peter Duesberg, 2006).

⁴The actual query terms were “aid” (stemmed form of “aids”) and “afric” (stemmed form of “africa”). Stemming of the terms in documents are done typically to increase recall.

HARD-409 given in Table 1. Clearly, the search statement AIDS, Africa is a poor representation of the request given in Table 1. However, it is typical that most state of the art IR systems work within what is called as the *bag of words paradigm*, in which the user query is represented as a weighted list of independent terms. The document mentioned above, in fact, would be ranked lower than the 609th place if the correct senses of the word “aid” in the document were identified by the system.⁵ Although phrases, special terminology, and acronyms could be handled in ways different from ordinary terms in most retrieval systems, including Okapi, such mismatching is difficult to completely avoid in most cases. The above example illustrates the difficulty of dealing with even relatively simple query formulations, such as “AIDS, Africa,” in systems based on the bag of words approach. It is, obviously, much more difficult, if not impossible, to express the actual pragmatic information need given in Table 1 in terms of a list of independent keywords, unless of course, documents are indexed by appropriate terms, such as “alternative view on AIDS in Africa,” that represent their contents as seen from different theoretical positions.

The common way to deal with complex information needs in most state-of-the-art retrieval systems is to exploit the relevance feedback mechanism, which is a means of a posteriori classification of documents, mentioned in Spärck Jones’ (2005) article referred to earlier. The usual approach of systems that implement relevance feedback is to retrieve a ranked list of documents in response to a simple search statement, such as “AIDS, Africa.” It is hoped that in the top n ⁶ documents retrieved there are one or more relevant documents identified by the user. If relevant documents are identified in the first pass, an expanded query (search statement) is constructed automatically or semiautomatically by choosing useful terms⁷ from relevant documents. The expanded query, similarly, is composed of a weighted list of independent terms, but contains additional useful terms that were not entered by the user. The more there are useful terms in a query, the better are the chances of retrieving relevant documents. However, as shown in the example about AIDS in Africa, a relevant document could be ranked below the point an average user would normally look. Query expansion by means of relevance feedback based on a few known relevant documents, therefore, may not be practical in such cases.

Undoubtedly, relevance feedback is a powerful tool for increasing retrieval effectiveness for certain types of queries. There are, however, practical technical reasons, which

⁵There were seven instances of the stem “aid” in the document, of which only three correctly refer to AIDS, the others refer to “financial/material help or assistance” sense of the term. BM25 matching function of Okapi (Spärck Jones, Walker, & Robertson, 2000) used in our experiments ranks documents based on the frequency of the query terms in the document and in the database index, as well as document length and other factors.

⁶The actual number depends on how many documents the user is willing to examine for a given search, and assumed normally to be between 10 and 100.

⁷Useful terms are those that are expected to be also present in other relevant documents.

prevented the widespread adoption of relevance feedback by the Web search engines so far. Relevance feedback requires keeping track of the search history, especially, recording explicit user relevance judgments, and processing of the documents on the fly, which limit the use of the method on the Web, where search engines serve thousands of users simultaneously.

There have also been attempts to develop more powerful query languages that can express complex information needs. The Indri search engine's structured query language (Metzler, 2005) enables the user to take advantage of the document structure and term proximity information. The search terms could be limited to certain parts of the document (e.g., title field), or a bounded unit of text (e.g., sentence or paragraph). More interestingly, it is possible to search by part of speech and named entity tags in Indri. For example, the following query "Where was George Washington born?" is expressed as *#combine[sentence](#1(george washington) born #place(?))* in the Indri query language. The result of this search would be a ranked list of sentences that contain the phrase "George Washington," the term "born," and a snippet of text tagged as a "place" named entity (Croft, 2005). The query, *#filreq (#band (NewsFeed.doctype #date:between(1991 2000)) #combine[paragraph](#any:person #any:money InfoCom))* would, on the other hand, retrieve paragraphs from news feed articles published between 1991 and 2000 that contain snippets of text tagged as a "person" and "monetary amount," and the term "InfoCom"—name of a company (Croft, 2005).

Although, the above examples demonstrate that Indri's query language is a more expressive medium for representing complex information needs, it is not certain whether the topic given in Table 1, or another sufficiently complex information need can be translated into this query language, and thus successfully resolved. However, in the above examples we see the basic requirements for a query language that can help express complex queries and resolve pragmatic information needs: the ability to limit the search to certain parts of a document by making use of document structure, and the ability to search by part of speech and other automatically identifiable textual elements (e.g., named entities).

Regardless of the speculation about how much of the meaning expressed in a natural language query can be translated into an artificial query language of a retrieval system, there is a more fundamental problem of metatheoretical nature that affects the evaluation and classification of documents with respect to queries and information needs. The discussion of the topic "AIDS in Africa" (Table 1) earlier makes clear that evaluation and classification of documents are nontrivial problems that are ultimately dependent on philosophical and metatheoretical assumptions. For this reason, the problems related to evaluation and classification of documents cannot be solved by computational methods alone. The need for an element of human reasoning in classification, therefore, seems to remain despite the advances made in computational methods of processing natural language texts. In the next section, we will discuss the relevance

of work on a substantive theory of classification in information science to another domain, namely arts, and the significance of this to the development of information science as an interdisciplinary subject.

Systemic Theory of Classification⁸

The aim of this section is to show that there is at least one more potential use of a substantive theory of classification other than information retrieval. The discussion will add another dimension to the debate on the use of classification theory in information science. Information science as an academic discipline is relatively new. As a young discipline, it is characterized by multidisciplinary traits,⁹ such as the use of disparate theoretical perspectives borrowed from various disciplines without a high-level of integration between them. One could list computer science, cognitive science, human-computer interaction, psychology, sociology, linguistics, philosophy as some of the disciplines that have lent concepts, tools, and methods to information retrieval researchers, and more generally, information scientists. The reverse cross-disciplinary fertilization is, however, relatively rare. Arguably, not many disciplines or subjects have felt the need to import ideas from information science. In this section, it will be shown that one of the endogenous pieces of work on classification theory in information science could become an example of the ability of information science to export ideas to other disciplines, which could be seen as a sign of gradual maturation of the field.

Maturity of an academic discipline is arguably measured in terms of its ability to export ideas to other disciplines. Various studies (Cronin & Pearson, 1990; Meyer & Spencer, 1996; Sanborn, 2002; So, 1988) suggest that information science literature receives relatively few citations from other disciplines. Different studies surveyed different range of journals and time periods found that about 8%–13% of all citations to information science literature come from other disciplines, whereas this is around 25% in developed disciplines (So, 1988). A similar pattern is true for citation of information science theory by other disciplines. Pettigrew and McKechnie (2001) looked at citations to major theories in IS. Their results show that about 20% of citations to IS theory are from other disciplines; however, this is an overestimation, and the actual figure is found to be about 9% when citations by IS authors publishing in non-IS journals are removed.

⁸The term *systemic* is used in many disciplines, such as linguistics, systems science, medicine, as well as information science and information systems, to refer to a view that sees things as systems, that is, wholes consisting of interdependent and interacting parts. In this article, it is used to refer to a holistic or integral view of natural and social phenomena. From this perspective, it is argued that a full discussion of the need or otherwise of a substantive theory of classification in information science should take place in a broader multidisciplinary context, in which dynamic processes of interaction and exchange between seemingly independent fields are emphasized.

⁹See, for example, Arms (2005) for an explicit view of information science as a multidisciplinary field.

The importance of cross-disciplinary fertilization, especially in the face of increasingly complex real-world problems that require multidisciplinary work is generally recognized. The systems movement and the General Systems Theory (Checkland, 1993; Rapoport, 1986; Skytner, 2005; Von Bertalanffy, 1968), which conceive reality as composed of interdependent and interacting parts, can be seen as explicit attempts to systematically investigate the applicability of a set of ideas developed in one discipline in another to tackle complex problems encountered in different realms of life. The aims of General Systems Theory (GST) are given as (Checkland, p. 93):

- To investigate the isomorphy of concepts, laws, and models in various fields, and to help in useful transfers from one field to another.
- To encourage the development of adequate theoretical models in areas which lack them.
- To eliminate the duplication of theoretical efforts in different fields.
- To promote the unity of science through improving the communication between specialists.

Information is an ambiguous term. Capurro and Hjørland (2003) state that “Almost every scientific discipline today uses the concept of information within its own context and with regard to specific phenomena.” (p. 356) “There are many concepts of information, and they are embedded in more or less explicit theoretical structures” (p. 396). There is little doubt that information science is only one of the many disciplines where the concept of information occupies a prominent position. Many disciplines, including sociology, economics, political science, biology, medicine and life sciences, as well as computer science and management sciences, deal with problems that are articulated around the concept of information. However, the concept of information is founded upon different theoretical premises in different disciplines, thus understood differently. Real-world problems, on the other hand, are usually complex and their solution often requires multi- or interdisciplinary perspectives, i.e., mobilizing knowledge/information from different disciplines and organizing them around a common goal of solving particular problems. Many traditional disciplines, i.e., those with well-defined boundaries and identities, for this reason have sought in recent years to broaden their disciplinary boundaries by appending information-related labels to their names. Examples of this tendency include, but are not limited to, disciplines such as computer and information sciences, library and information science, social/organizational/business informatics, bioinformatics, and health informatics.

The effectiveness of multidisciplinary research could however be increased, if the integration of knowledge and theoretical perspectives from different disciplines is achieved by exploring deeper (structural) similarities between them. The GST is founded on this premise, however, information science could also be seen as a generalist/interdisciplinary subject, which explores knowledge structures in diverse

disciplines. Arguably, the main task of information science is the production of new knowledge (Karamuftuoglu, 1998) through creation of secondary information sources (databases, indexes, thesauri, ontologies, knowledge maps, and so on), which organize the concepts created by other sciences, and through which we comprehend the world. Information scientists do not passively represent knowledge or transfer information from one source to another. Indexes, ontologies, databases, and other representations and sources of information they create make it possible for us to see new connections between concepts and thus understand the world in new ways.¹⁰ The artifacts that information scientists create, in particular the secondary sources of information, organize and transform our knowledge of the world by making certain connections between concepts, terms, and therefore disciplines visible, while inadvertently making others invisible. Information science, from this perspective, is a discipline that seeks to increase communication between specialties by uncovering similarities between them, and constructing theories/models of how knowledge is produced, recorded, structured, and represented across different domains. In the next section, an example of the application of a theory of subject analysis exported from information science to another field and adapted for a new task is illustrated, fulfilling the systemic ideals of transfer of knowledge from one field to another, and increased communications between specialties. Here, an example of transfer of Shannon’s theorem 17 (Shannon & Weaver, 1949), a fundamental theorem of communications systems developed for estimating the information capacity (in bits) of a physical communications channel, to an entirely new task of modeling human motor behavior in Human–Computer Interaction (HCI) will be discussed briefly to give a concrete example of how the exploration of isomorphy of concepts, laws, and models in different fields could fruitfully yield to the progress of knowledge.

Fitts (1954) suggested that the time required to rapidly move the hand from a starting position to a target area (movement time) could be modeled based on Shannon’s theorem (Shannon & Weaver, 1949) of transfer of information (data) in physical communication systems (MacKenzie, 1989). The proposed model is known as the *Fitts’ law*, and is used in HCI to predict the time it takes a human operator to point to a graphical object on a computer screen using a mouse or a similar input device. Fitts reasoned that movement amplitude (distance to the target) is analogous to the signal power transmitted over a communications channel, and width of the region within which the move must terminate (target width) is analogous to the noise power that perturbs the transmitted signal (MacKenzie, 1989). Fitts’ equation has proven to be highly accurate; it is used extensively to model the time to complete a given task in interactive systems. The axiological model for the aesthetic judgment of artworks proposed by Karamuftuoglu (2006) is discussed in the next section to argue that the domain-analytic theory of subject analysis

¹⁰See for example Ørom (2003) for a detailed analysis of how classification systems shape our knowledge of the arts.

proposed by Hjørland (1992) could similarly be exported and used fruitfully in another domain for another purpose.

From Classification Theory to Aesthetic Theory: An Axiological Model

Information arts is a relatively new term applied to a specific kind of conceptual art that combines art with science and technology (Karamuftuoglu, 2006). In a general sense, *conceptual art* or *concept art* is any work of art that gives primacy to the ideational/information content or the idea behind the artwork over its formal (presentational/visual) aspects. Although the term *information arts* has been in use for some time, it has received attention more recently with the publication of Stephen Wilson's (2002) compendium entitled, *Information Arts: Intersections of Art, Science, and Technology*. The volume is a survey of works of art created in recent years that integrate art with science and technology. In *Information Arts*, Wilson notes the changing role of art and artists in today's world dominated by technological and scientific research, and observes the increasing information content in contemporary works of art.

Based on Wilson's pioneering work, Karamuftuoglu (2006) argued that whereas earlier conceptual artworks in the 1960s informed the spectator about the ideas and opinions of their creators, the contemporary works of information art inform one about problems pertinent to scientific and technological issues. In recent years, artwork has become a "research document" that communicates technological and scientific information. There are many examples of artwork created recently that contribute directly or indirectly to science (Wilson, 2002). A particularly interesting and currently intensive area of artistic research is the *bioarts*, which uses living biological organisms, instead of traditional media such as paint or bronze, as its primary medium of expression in creating art. Based on the above observation, Karamuftuoglu (2006) suggested that the aesthetic values of a new generation of artwork could be evaluated in similar ways to documents evaluated in the domain-analytic approach proposed by Hjørland and his colleagues (Hjørland, 1992, 1997; Hjørland & Albrechtsen, 1995).¹¹ In this interpretation, artworks are seen as analogous to research documents, therefore, the purpose of a work of information art is seen "to facilitate communication of pertinent scientific and technological information" Karamuftuoglu (2006). This view leads to the following statement regarding the aesthetic value of a work of information art, which is analogous to the

¹¹Works of art like other objects are traditionally seen as documents in classification theory. However, stylistic (e.g., Impressionist, Cubist), historical (e.g., Renaissance, Baroque), physical (medium size, etc.), and other attributes of works of art—rather than their information content—provide the basis of classification. Detailed analysis of classification systems used in the arts domain can be found in Ørom (2003). In contrast, works of art here are seen as analogous to research documents; therefore, their information content is emphasized. It is thus suggested that the domain-analytic approach used for analyzing scholarly texts could be applied to aesthetic evaluation of works of information arts, which is a distinct task unique from classification.

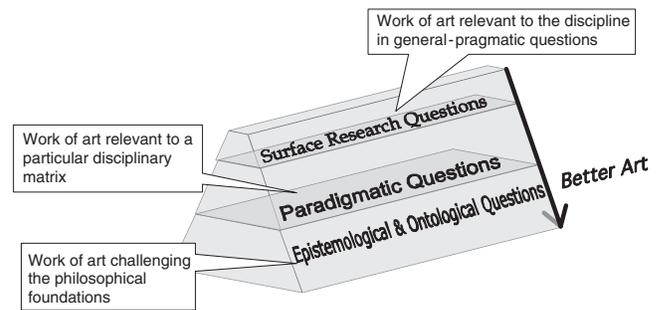


FIG. 1. The Axiological framework for information arts (Karamuftuoglu, 2006).

theory of subject analysis proposed by Hjørland (1992): Aesthetic value of a work of information art is a function of its potential to address significant questions regarding fundamental epistemological and ontological assumptions relevant to the scientific discipline(s) from which the work derives its methods, concepts or tools (Karamuftuoglu, 2006).

The above statement regarding the value of a work of art is illustrated in Figure 1. The figure depicts the different levels at which the value of a work of art could be evaluated (Karamuftuoglu, 2006). At a pragmatic level, a work of art may pose general questions or inspire new ideas that are valid in a field of inquiry, regardless of different paradigms that may exist in that field. At a deeper theoretical level, it may pose questions or inspire new ideas relevant to a particular research paradigm(s) in a field. At a metatheoretical level, questions it poses or ideas it inspires may have relevance to a particular set of epistemological, ontological and methodological assumptions that underlie one or more paradigms or theories in a field. According to this axiological framework (Figure 1), the aesthetic value of a work of information art issues primarily from its information content and its capacity to contribute to research in a given domain directly or indirectly.

The evaluation of works of art has become extremely problematic in recent years. One of the reasons for this is the extreme *laissez-faire* attitude: "If someone calls it art, it is art" dictum prevalent in postmodern art. The potential value of the above axiological framework elaborated by adapting the domain-analytic theory of subject analysis developed in the context of IR to the task of aesthetic judgment of artworks is that, it proposes a new and objective basis for evaluation of a certain kind of artwork, specifically, artwork with high information content.¹²

Regardless of whether or not the proposed axiological model will be accepted by the arts community at large, it could prove to be useful to the development of information science (IS) as an interdisciplinary subject. In this article,

¹²In a personal correspondence (September 10, 2006), Stephen Wilson agreed that the analogy drawn between documents and works of information arts, and the proposed axiological framework based on this analogy, are novel and interesting, thus providing further evidence to the claim that a theory developed in one domain or for a particular task can be fruitfully exported to another independent domain and adapted for a new task.

information science is seen as an interdisciplinary and generalist subject that aims to improve communication between specialties by uncovering similarities between them and constructing theories/models of how information/knowledge is produced, recorded, represented, and used in different knowledge domains.¹³ As noted above, information and knowledge are ambiguous terms founded upon different theoretical premises in different disciplines, and thus understood differently. However, increasingly complex real-world problems necessitate mobilization of knowledge and information from different disciplines to synthesize new solutions. If information science is to fulfill the ideal of making available information and knowledge produced in different specialties as diverse as biology and the arts for use in inter- and multidisciplinary teams for the solution of complex real-world problems, it needs to interpret, frame, and unite disparate conceptualizations of information and related phenomena in terms of common theoretical models and terminology. This could help IS to achieve a higher degree of internal consistency and coherence and build up its identity:

... disciplines require theories that originate from within to attain recognition as an independent field of scientific inquiry. In other words, if fields such as information science are to delineate their disciplinary boundaries and build a central body of knowledge, then they require their own theoretical bases for framing research problems, building arguments, and interpreting empirical results (Pettigrew & McKechnie, 2001, p. 62)

By learning from theories, concepts, and findings of other disciplines and constructing theories applicable to different types of information problems in multiple fields, information science could evolve from its current status of a multidisciplinary program to an interdisciplinary field. Developed disciplines or fields evolved to the interdisciplinary level, in contrast to multidisciplinary subjects, display coherent identity and integrated theoretical and methodological orientation (Besselaar & Heimeriks, 2001).

Conclusions

Here the usefulness of a substantive and general classification theory in IR are discussed in a broad perspective. Two significant uses of a classification theory are identified. First, complex information needs cannot be easily resolved by means of operational approaches such as relevance feedback. By means of a concrete example from the HARD TREC 2004 corpus, I argue that in certain situations relevant documents may not be retrieved in the top positions of a ranked list. In such cases, the relevance feedback mechanism is of

¹³Knowledge and information are difficult concepts to define. To discuss the issues surrounding this problem is outside the scope of the present article. A worthwhile attempt to define them can be found in Capurro and Hjørland (2003). Karamuftuoglu (2006) argued from the postmodernist knowledge regression position described in Hakken (2003) that the difference between knowledge and information is a matter of degree of decontextualization or deterritorialization.

little practical value. It is suggested that in such situations sophisticated query languages that make use of document structure and various linguistic features of texts could provide additional leverage to increase retrieval effectiveness. However, it is arguable that even a sophisticated query language may not be enough to discover complex meanings of texts. Certain types of information needs may require a deeper philosophical and metatheoretical analysis to explicate important information that may be relevant to a particular user or group of users. The first significant use of a substantive classification theory identified in this article is therefore related to the philosophical and metatheoretical analysis of contents of texts. The domain-analytic approach developed by Hjørland and his colleagues is a good example of pragmatic metatheoretical subject analysis, which could serve as a basis for a substantive and general classification theory.

The second use of a substantive classification theory is discussed in the context of the maturation of IS as a discipline. Today's increasingly complex problems necessitate transfer and integration of knowledge from isolated disciplines. Information science is seen, from this perspective, as a generalist/interdisciplinary subject, which aims to improve communication between specialties. To fulfill the aim of improving communication between specialties to facilitate the solution of complex multidisciplinary problems, information science needs to interpret, frame, and unite disparate conceptualizations of information by constructing general theoretical models applicable to different tasks in different disciplines. This approach is compatible with the aims of the systems movement, which also seeks to promote communication between specialties through exploration of the isomorphy of concepts and laws in different disciplines. It is argued that IS could achieve a higher degree of internal consistency and coherence by following a similar course, and constructing theories that can be applied to information-related phenomena in diverse disciplines.

Acknowledgments

I would like to thank Birger Hjørland for his useful comments on an earlier version of this paper, and Steven Wing for sending me his manuscript "Limits of Epidemiology."

References

- Albrechtsen, H., & Hjørland, B. (1997). Information seeking and knowledge organization: The presentation of a new book. *Knowledge Organization*, 24, 136–144.
- Duesberg hypothesis. (2006). Wikipedia, the free encyclopedia. Retrieved April 12, 2006, from http://en.wikipedia.org/w/index.php?title=Duesberg_hypothesis&oldid=45414817
- Allan, J. (2005). HARD track overview in TREC 2004. High accuracy retrieval from documents. In E. Voorhees & L. Buckland (Eds.), *Proceedings of the 13th Text Retrieval Conference*. Gaithersburg, MD: National Institute of Standards and Technology.
- Arms, W.Y. (2005). Information science as a liberal art. *Interlending & Document Supply*, 33(2), 81–84. Retrieved April 12, 2006, from <http://www.emeraldinsight.com/Insight/ViewContentServlet?filename=Published/EmeraldFullTextArticle/Articles/1220330202.html>

- Besselaar, P. van den, & Heimeriks, G. (2001). In M. Davis & C. Wilson (Eds.), *Proceedings of 8th International Conference on Scientometrics and Informetrics* (pp. 705–716). Sydney: UNSW. Retrieved April 12, 2005, from http://www.niwi.knaw.nl/en/maatschappijwetenschappen/staf/peter_s_home/new/2001issi/toonplaatje
- Bethell, T. (2000). Mbeki takes on the aids industry: South African President queries epidemic, *AZT. Reappraising AIDS*, 8(3), 1–4. Retrieved April 12, 2005, from <http://www.duesberg.com/subject/mbekaitakes.html>
- Capurro, R., & Hjørland, B. (2003). The concept of information. *Annual Review of Information Science & Technology*, 37, 343–411.
- Checkland, P. (1993). *Systems thinking, systems practice*. Chichester: Wiley.
- Cronin, B., & Pearson, S. (1990). The export of ideas from information science. *Journal of Information Science*, 16, 381–391.
- Croft, W.B. (2005). Phrases and other structures in queries. Invited talk at the Workshop on Methodologies and Evaluation of Lexical Cohesion Techniques in Real World Applications—Beyond Bag of words (ELECTRA 2005), in association with the 28th Annual International ACM SIGIR conference. New York: ACM.
- Duesberg, P. (2000). The African AIDS epidemic: New and contagious—or—old under a new name? Retrieved April 12, 2006, from <http://www.duesberg.com/subject/africa2.html>
- Duesberg, P. (2006). Deuserg on AIDS. Retrieved April 12, 2006, from <http://www.duesberg.com/>
- Duesberg, P., Koehnlein, C., & Rasnick, D. (2003). The chemical bases of various AIDS epidemics: Recreational drugs, anti-viral chemotherapy and malnutrition. *Journal of Biosciences*, 28, 383–414. Retrieved April 12, 2006, from <http://www.duesberg.com/papers/chemical-bases.html>
- Duesberg, P., & Rasnick, D. (1998). The AIDS dilemma: Drug diseases blamed on a passenger virus. *Genetica*, 104, 85–132. Retrieved April 12, 2005, from <http://www.duesberg.com/papers/pddrgenetica.html>
- Engels, F. (1958). *The condition of the working class in England*. Stanford, CA: Stanford University Press. (Original work published 1845)
- Fitts, P.M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381–391.
- Hakken, D. (2003). *The knowledge landscapes of cyberspace*. London: Routledge.
- Hjørland, B. (1992). The concept of “subject” in information science. *Journal of Documentation*, 48(2), 172–200.
- Hjørland, B. (1997). *Information seeking and subject representation: An activity-theoretical approach to information science*. Westport, CT/London: Greenwood Press.
- Hjørland B. (2002). Epistemology and the socio-cognitive perspective in information science. *Journal of the American Society for Information Science and Technology*, 53, 257–270.
- Hjørland, B., & Albrechtsen, H. (1995). Toward a new horizon in information science: Domain analysis. *Journal of the American Society for Information Science*, 46, 400–425.
- Hjørland, B., & Nissen Pedersen, K. (2005). A substantive theory of classification for information retrieval. *Journal of Documentation*, 61(5), 582–597. Retrieved April 12, 2006, from <http://www.db.dk/bh/Core%20Concepts%20in%20LIS/Hjorland%20%20Nissen.pdf>
- Jones R., Vechtomova O., & Dias, G. (Eds.). (2005). *Proceedings of the Workshop on Methodologies and Evaluation of Lexical Cohesion Techniques in Real-World Applications—Beyond Bag of Words (ELECTRA 2005)*, in association with the 28th Annual International ACM SIGIR Conference. New York: ACM.
- Karamuftuoglu, M. (1998). Collaborative information retrieval: Toward a social informatics view of IR interaction. *Journal of the American Society for Information Science*, 49(12), 1070–1080.
- Karamuftuoglu, M. (2006). Information arts and information science: Time to unite? *Journal of the American Society for Information Science and Technology*, 57(13), 1780–1793.
- MacKenzie, I.S. (1989). A note on the information-theoretic basis for Fitts’ law. *Journal of Motor Behavior*, 21, 323–330.
- Metzler, D. (2005). Indri query language quick reference. Retrieved April 12, 2006, from <http://www.lemurproject.org/lemur/IndriQueryLanguage.html>
- Meyer, T., & Spencer, J. (1996). A citation analysis study of library science: Who cites librarians? *College & Research Libraries*, 57(1), 23–33.
- Ørom, A. (2003). Knowledge organization in the domain of art studies—History, transition and conceptual changes. *Knowledge Organization*, 30(3/4), 128–143.
- Overton, W.F. (1998). *Metatheory & methodology in developmental psychology*. Retrieved April 12, 2006, from <http://astro.ocis.temple.edu/~overton/metatheory.html>
- Peter Duesberg. (2006). Wikipedia, the free encyclopedia. Retrieved April 12, 2006, from http://en.wikipedia.org/w/index.php?title=Peter_Duesberg&oldid=44661197
- Peterson, C., & Martin, C. (2000). A new paradigm in general practice research—Towards transdisciplinary approaches the utilisation of multiple research methodologies in general practice research. Retrieved April 12, 2006, from <http://www.priory.com/fam/paradigm.htm>
- Pettigrew, K., & McKechnie, L. (2001). The use of theory in information science research. *Journal of the American Society for Information Science and Technology*, 52(1), 62–73.
- Rapoport, A. (1986). *General system theory*. Kent, UK: Abacus Press.
- Roberts, A. (1999). *ABC of thinking*. Retrieved August 2, 2005, from <http://www.mdx.ac.uk/www/study/glothi.htm>
- Sanborn, E.C.J. (2002). Other-field citation rates of library and information science literature. Unpublished master’s thesis, School of Information and Library Science of the University of North Carolina at Chapel Hill.
- Shannon, C.E., & Weaver, W. (1949). *The mathematical theory of communications*. Urbana, IL: University of Illinois Press.
- Skyttner, L. (2005). *General systems theory* (2nd ed.). London: World Scientific Publishing.
- So, C. (1988). Citation patterns of core communication journals: An assessment of the developmental status of communication. *Human Communication Research*, 15(2), 236–255.
- Spärck Jones, K. (2005). Revisiting classification for retrieval. *Journal of Documentation*, 61(5), 598–601.
- Spärck Jones, K., Walker S., & Robertson S.E. (2000). A probabilistic model of information retrieval: Development and comparative experiments—Parts 1 and 2. *Information Processing and Management*, 36(6), 779–840.
- Vechtomova, O., & Karamuftuoglu, M. (2005). Approaches to high accuracy retrieval. In E. Voorhees & L. Buckland (Eds.), *Proceedings of the 13th Text Retrieval Conference*. Gaithersburg, MD: National Institute of Standards and Technology.
- Von Bertalanffy, L. (1968). *General system theory: Foundations, development, applications*. New York: Braziller.
- Voorhees, E.M. (2005). Overview of TREC 2004. In E. Voorhees E. & L. Buckland (Eds.), *Proceedings of the 13th Text Retrieval Conference*. Gaithersburg, MD: National Institute of Standards and Technology.
- Wilson, S. (2002). *Information arts: Intersection of, art, science, and technology*. Cambridge, MA: The MIT Press.
- Wing, S. (1994). Limits of epidemiology. *Medicine and Global Survival*, 1, 74–86. Retrieved August 10, 2007, from <http://www.ippnw.org/Publications/MGS/V1N2Wing.html>