



PERGAMON

Information Processing and Management 38 (2002) 381–400

**INFORMATION  
PROCESSING  
&  
MANAGEMENT**  
www.elsevier.com/locate/infoproman

# The cascade of interactions in the digital library interface

Marcia J. Bates \*

*Department of Information Studies, Graduate School of Education and Information Studies, University of California,  
405 Hilgard Avenue, Los Angeles, CA 90095-1520, USA*

Received 11 October 1999; accepted 17 June 2001

---

## Abstract

Each design element or layer in an information system interacts with every other design layer in a synergistic, neutral, or conflicting manner. This cascade of interactions culminates in the interface, where all the prior interactions have either worked to produce effective information retrieval or to produce a hodgepodge of system elements working at cross-purposes. Very large networked and World-Wide-Web-based online databases and online public access catalogs provide numerous illustrative examples of how the cascade works. Good design requires that these interactions be well understood and properly designed for the purposes of the information system in question. Interactions among metadata and indexing systems, information system front-ends, user search capabilities, and interface design are discussed and used as illustrations for the effective design of digital library access. © 2002 Elsevier Science Ltd. All rights reserved.

*Keywords:* Digital library design; Information retrieval systems; Interface design; Information system design; Online information systems

---

## 1. Introduction

The creation of means of access to large bodies of digitized information has rightly been viewed in our society as a major development in the history of human knowledge use. The elimination or reduction of prior constraints of distance, fragility of resources, or limited physical access to resources promises a freedom and flexibility in information access unprecedented in human history. However, we have a way to go yet in creating systems that are both easy to use and truly effective in retrieval. The very growth in computing power and design options for the latest systems also means that there are more opportunities to get it wrong – to design systems that have

---

\* Tel.: +1-310-206-9353; fax: +1-310-206-4460.

E-mail address: [mjbates@ucla.edu](mailto:mjbates@ucla.edu)<http://www.gseis.ucla.edu/faculty/bates/> (M.J. Bates).

some wonderful individual design features, but which do not comprise an effectively integrated information system that works well as a whole.

In the 1970s – the early days of mounting databases such as online catalogs and abstracting and indexing services for searching purposes – developers were generally happy to get a database functioning in any reasonable way possible. Information systems often consisted of a body of information, originally selected and indexed for the paper environment, being mounted for electronic access with a search engine and system front-end designed entirely independently from the body of information being retrieved. The resulting systems were often crude, but faster than the paper alternatives, and so represented an important step forward. Most computing in those days was directed toward running large numbers of operations on relatively small bodies of information. Appropriate technology and software were slow to develop where few operations and very large bodies of textual information were concerned.

Major amounts of money and attention are at last flowing into studying questions of information management and retrieval from large bodies of heterogeneous textual and image databases. The landscape of information retrieval is changing in fundamental ways. Much more power, variety of options, and flexibility are now available for each layer of an information system's design.

Deep expertises are developing in each of the design layers. In practice, that means that it is harder and harder for any one person to have sufficient knowledge of all the design layers to see the entire forest for the trees. That, in turn, means that systems are prone to being designed with many different objectives in mind and with little understanding between the several groups of professionals addressing the design.

The purpose of this article is to present a model of some of the typical important layers in the design of information systems, and to show how those layers interact in operation. It is argued that these layers are in the way of a cascade – what is done, well or poorly, at one layer affects the next, which then affects the next layer, and so on. Thus, if designers do not understand the ways in which the layers interact, and they design and develop the layers independently of each other, then these information system layers may end up actually working at cross-purposes with each other. The result will be ineffectual or sub-optimized information systems.

In this article, a model is presented of this cascade of interactions in information systems. Several examples – from networked databases, online catalogs, and a Web-based information system – are presented in detail to illustrate the subtle and complex ways in which the design layers may interact, for good or ill. Finally, it is argued that staff working on all design layers *must* work together, sharing their expertise and design decisions, or face the likelihood of producing a sub-optimal information system.

Though the discussion is largely about integrating designer expertises, a fundamental assumption throughout the article is that this is *user-centered design*. The choices made at each one of the design layers should draw on what is known of prospective user needs and interests, their capabilities in system use, and their attitudes. A good, user-centered information system is not just a pretty interface. Design choices based on knowledge of users should penetrate deep into the system, should shape each one of the design layers and affect how those layers can best be integrated. (See also Bates, 1990a, 1994, 1998; Butler, Bahrami, Esposito, & Hebron, 2000; Dillon, 1994; Hayman & Elliman, 2000; Kilker & Gay, 1998; Payette & Rieger, 1998; Zhang & Fine, 1996; Zhu, 2001.)

## 2. Background: information system design models

Researchers and system developers have used a variety of sorts of models to aid in the process of designing and developing information systems. Here, some of the types are described and contrasted with the approach to be taken in the Cascade Model.

In automated information systems, there are three major components that appear in many different forms and permutations: (1) the computer/network system, i.e., the technical infrastructure, (2) the information or “content” that is supported and transmitted by the system, and (3) the user and his/her interaction with the system. Many models of automated systems emphasize the technical infrastructure necessary to make the system work. For example, the seven-layer Open Systems Interconnection Reference Model was designed to clarify relationships and support standards for inter-networking between computer systems. (See description in ISO/IEC 7498-1, 1994; Shor, 1991.)

The emphasis in this discussion is on the non-technical components of the Cascade Model – the information and the user. The layers of the model are identified from the perspective of the person who must attend to user-centered design and to optimizing the effective delivery of information in the development of some sort of information retrieval system. In particular, integrating these two components – the user and the information – *together* in modern, particularly Web-based, information retrieval system design deserves, as I hope to demonstrate, much more attention than has generally been given.

The Cascade Model describes the several design layers that have to be taken into account in the process of developing and implementing an information system in an automated environment. *Starting with the interface, imagine turning that interface on its side and pulling it apart, like an accordion, to reveal the several design layers backing that interface and culminating in the interface.* Those dovetailing layers will appear as in Fig. 1. The body of this article will elaborate the Cascade Model in more detail, but in the meantime, to sharpen its profile, let us compare it to other common types of information system models.

One type of model is the *design process model*. The purpose of these models is to lay out the various steps in sequence that designers should take in creating their system. Examples of these are to be found in Damodaran (1991), Galitz (1997, p. 46ff), Gardner (1991), Shneiderman (1998) and Syan and Menon (1994). For example, Gardner describes six design stages, from project initiation through product phase-out. The Cascade Model depicts the layers of a system that are simultaneously operating in a functioning system. The presence of layers in the Cascade Model is *not* to imply that one should design first one layer, then the next, in sequence. All layers need to be designed for in relation to each other. How the work then proceeds sequentially is another question – for another model.

Another type of model is the *human–computer interaction model*. There is a huge literature on the subject of human–computer interaction. (See, for example: Bødker, 1991; Card, Moran, & Newell, 1983; Carroll, 1991; Dix, Finlay, Abowd, & Beale, 1993; Galitz, 1997; Norman & Draper, 1986; Shneiderman, 1998.) Many researchers and designers have attempted to develop general principles of human–computer interaction design. The interface is indeed the pivotal point where people interact with automated information systems, but it is by no means the only point of interest here. The interface is just one layer of the Cascade Model; user-centered design of information systems must be informed by a knowledge of general information seeking behavior on

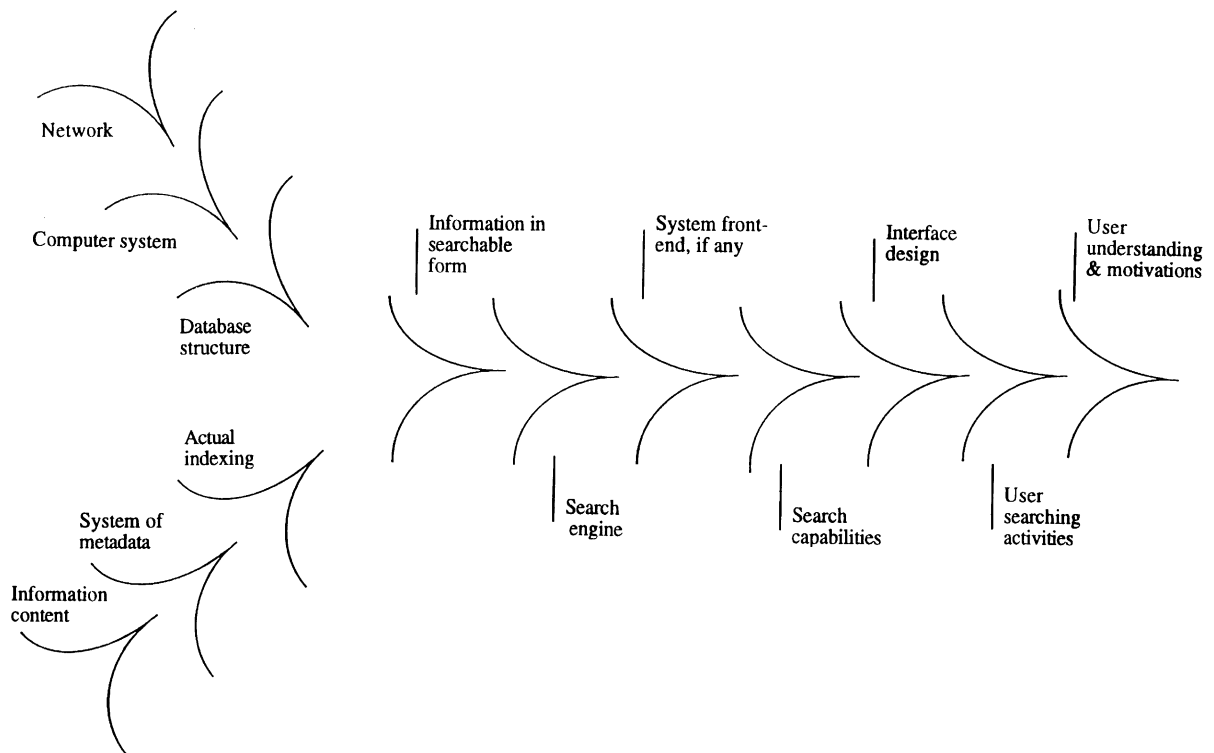


Fig. 1. The cascade of interactions in the digital library interface. The diagram presents example layers, even including use patterns and use environment at the right, that determine the overall functional effectiveness of a digital library information retrieval system. Design decisions at each layer have a cascading impact on subsequent layers; poor design at later layers can block products of earlier layers.

the part of the human users, and it must also be informed with a knowledge of the many layers behind the interface that go into creating an effective information retrieval system.

The Cascade Model is also not an *information retrieval model*. Decades of research and thousands of articles have been produced around various theories of optimal automatic information retrieval techniques. (See Ellis, 1996; Robertson, 1977; Salton & McGill, 1983). In these cases, the focus is on varying the design characteristics of the search engine and sometimes the search capabilities, and testing which system does best. These testing methods frequently control for the queries put to the system and the relevance assessments on the retrieved records. In other words, in order to compare the information retrieval systems, all the layers of the Cascade Model except the search engine are held constant in order better to test the comparative effectiveness of the searching algorithms. Here, with the Cascade Model, we are considering all the design layers that actually affect the real-world performance of an information retrieval system, and are giving particular attention to the way variations in the layers affect other layers' performance.

Next, the Cascade Model is not a model of *interaction in information retrieval research*. In recent years, the information retrieval research community has become increasingly interested in the importance of human interaction with the system, in addition to search algorithm performance. Attention is being paid to the patterns and characteristics of human–system interaction, and the

influence of these patterns on information system performance is being recognized (Beaulieu, 2000; Belkin, Brooks, & Daniels, 1987; Ingwersen, 1992; Saracevic, 1996, 1997; Spink, 1997).

Unquestionably, this perspective enriches both the information retrieval research and the human–computer interaction research. More of the full picture involved in the creation of a successful information searching experience is being incorporated into a single model.

I would argue, however, that still more needs to be incorporated into our thinking about means of creating effective online information search experiences. I will argue for each of the other layers in the Cascade Model as having crucial impacts on the resulting system success. These other layers must also be a part of the design and evaluation.

Finally, where digital resources and Internet-based resources are concerned, the Cascade Model is relevant in any case where the system user is expected to be able to search on any collection in excess of a few dozen items in number. Currently, there are countless commercial and non-profit efforts to create good, high-impact Web sites, Intranets, CD-ROMs, and other electronic resources. The Cascade Model is not a *Web site design model*, nor an *Intranet design model*, nor a model for any other specific technology. Rather, the Cascade Model refers to *any information retrieval component of an electronic resource*. So when a Web site has a dozen features, one of which is the ability to search on some collection or database of information, then the Cascade Model refers to that information retrieval component.

Currently, in the general discussion regarding Web site design, there is frequently a failure to differentiate the information retrieval component from the rest of the design. We know a lot of things about how people act and how systems need to be designed to be effective specifically in the context of information retrieval. This knowledge is distinct from the questions Web designers address when they consider advertising, chat rooms, screen design, and the other features of sites independent of a search capability. This is not to say that in design thinking for a site, information retrieval should be utterly independent of all other design considerations for the site. What is needed, however, is a recognition that when information retrieval capabilities are to be provided to users, the expertise we have about that area of design of electronic systems needs to “kick in” and not be drowned out by the other considerations that inevitably play a role in site design.

So, in sum, the Cascade Model is *not* a design-process model, a human–computer interaction model, an information retrieval model, an interaction in information retrieval model, nor a Web site design or other specific technology design model. The Cascade Model *is* a design model for operational online information retrieval systems. The model emphasizes the many design layers that should be considered in relation to each other in the process of designing and implementing an automated information system.

The model describes the layers in the design and is labeled “Cascade”, because the layers interact in a cascading manner. Design features of earlier layers inevitably affect the success of later design features. Later features, if poorly designed, can block the effectiveness of the earlier layers. Either way, without integrated good design across all layers, *and constantly considering the layers in relation to each other in design and development*, the resulting information system is likely to be poor, or at least sub-optimal. In the new world of information systems – those associated with digital libraries and databases on the Internet, particularly – we need a unified model, for design purposes, of the underlying network, hardware, information, database structures, search capabilities, interface design and social context.

### 3. The cascade layers

Each design element or layer in an information retrieval system interacts with each other design layer in a synergistic, neutral, or conflicting manner. This cascade of interactions culminates in the interface, where all the prior interactions have either worked to produce effective information retrieval or to produce system elements working at cross-purposes. Each layer cannot be laid on top of the lower layers as if there were little or no relationship between those layers.

This point may be relatively obvious in some cases – a search engine cannot be used with a server that is not powerful enough to process records in real time, for example – but is much less obvious in other cases. In particular, there are problems of integrating the underlying information and its metadata with the user search capabilities and other features for the user in the interface. (For an extensive discussion of user search capabilities in the interface, see Bates, 1990a).

Fig. 1 illustrates a number of the layers in a typical information system. (The model builds on Hildreth's model for online catalog searching, Hildreth, 1982, p. 114, Fig. 12). These layers could be broken out in a different manner, or even put in different orders relative to each other, depending on the specific system being analyzed, but this figure will do as a general example.

There are actually four areas to the figure. The upper left wing presents the technical infrastructure supporting the system, the lower left wing represents the information, or content, combined with the metadata structuring necessary to make it accessible in an information system. The middle wing, from "Information in searchable form" to "Interface design", represents the information retrieval system itself – supported by the technical infrastructure and containing the structured information. Finally, the two far-right layers, "User searching activities" and "User understanding and motivation", are the fourth area, representing the human portion of the overall system. So we see the technology, the information, and the human being all brought together in the larger system that represents a person actually using an information system.

The various layers in Fig. 1 will now be described. The upper left wing will be relatively slighted in this discussion; the emphasis here is on the information, information system, and user portions of the figure. The upper left wing composes the computer infrastructure – communications network, hardware and software platforms, and database structures. The strengths and constraints of that infrastructure will, to some extent, determine what can be done with the superstructure of information and interface. In particular, there are often crucial decisions made at the stage of selecting software and designing database structure that promote or limit the power of the overall information retrieval system tremendously. Here, it is crucial that the "technical people" and the "information people" understand each other's domains enough to work out the best information system possible within resource constraints.

Starting from the lower left of the diagram, the "Information content" layer is of fundamental importance, though it is often taken for granted. Beautiful systems have been designed and built to access the wrong information for the intended users. We will return later to this matter of matching the information to the user.

"System of metadata" refers to the intellectual and organizational system for description of the digital information. As people have begun to recognize the wide variety of types of information, and contexts of information use, a corresponding variety of metadata schemes are being developed to meet those various needs (see, e.g., Digital Object Identifier, 2001; Encoded Archival Description, 2001; Text Encoding Initiative, 2001; Weibel & Lagoze, 1997). Where subject

description is concerned, it has long been recognized in information science that the character of the information to be indexed should be analyzed, and that targeted thesauri and classification schemes should be developed that are adapted to that information. This lesson has not yet been completely absorbed in the digital library world.

In particular, understanding that *systems* of metadata are used in assigning metadata is key to effective metadata use. It is not uncommon in the digital library literature to treat all descriptive terminology – especially terminology for subject content – as if all thesauri or term lists were simple aggregations of independent terms, and not part of an intellectual system whose integrity should be respected for optimal retrieval performance. Explanatory examples will be given in the following section.

The “Actual indexing” of objects and texts is distinguished from the “System of metadata” to draw attention to the importance of the quality and character of the indexing as actually carried out. Great variability is possible in the actual description done of digital objects, compared to what the rules of any one theoretical system allow or encourage. Variability in *depth of indexing* (average number of descriptive elements applied) and in quality of indexing can have an enormous influence on quality and power of results at the point of actual searching. That which is not indexed, either by algorithmic or human means, cannot be retrieved.

“Information in searchable form” refers to the first point where the information and database structure come together as a part of the actual information retrieval system. A library, for example, may use a particular computer platform, have a particular body of catalog or other data to put into the system, and buy or develop an online catalog, which functions as the information system. Almost invariably, bringing these three together – platform, data, and information system – requires local configuring of various types, as well as “massaging” of the data structure in order to make the information actually searchable.

The “Search engine” provides the core retrieval facility in the system. It determines how and whether fields can be searched, and how efficiently they are searched.

“System front-end” refers to any of a variety of types of systems interposed between the interface and the information system proper, whose purpose is to assist the user in searching in any of several ways. Such front-ends may be gateways employing a single search command language for accessing several different database vendor systems, interrogators helping searchers better identify and specify their queries (Vickery & Vickery, 1993) vocabulary support systems that suggest additional and related terms to use in searching (Bates, 1986, 1990b), or others. Efthimiadis (1990, 1996) reviewed some of the literature on such front-ends and gateways. In the digital library environment, the potential of such assistive systems has only begun to be exploited.

“Search capabilities” are the searching mechanisms made available to the user of the system. This term does not refer to whatever underlying methods are used by the search engine, but rather refers to what the user can do from his/her perspective. For instance, internal storage and indexing mechanisms may vary from system to system, but if they provide the same end result capability to the user, then the underlying differences do not matter for this discussion. For the discussion here, those internal mechanisms are considered to be a part of the search engine. Typical search capabilities offered to the user include keyword searching, proximity searching, Boolean combination, searching by field, and many more.

“Interface design” is pivotal to the effective use of an information system by users. As noted earlier, much of the interface literature represents an effort to identify general principles for

interface design for all kinds of systems. The application environment of information retrieval systems, however, also has its own distinctive needs and characteristics that need to be understood and addressed in design.

Finally, we come to the two right-hand layers of Fig. 1, those representing the user role in the overall system. Commonly, in information system design models, these two layers are left out. Efforts may be made to identify in advance what people need and what features they would use in an information system. But from that point, it is often assumed that the performance of the system is measured by how well actual use matches what the system was designed for. However, there is much unpredictability in even the best-designed and planned system. The final, final measure of performance will always be how people actually use a system, not how one expected them to.

Consider, for example, two organizations which have each introduced computer-based information systems that are intended for the same purpose and role in the organizations. In the one organization, use of computer systems is viewed as the prerogative of the senior management. In the other organization, the management consider computers to be secretarial devices and disdain using them. In these contrasting cases, system use, including types and content of information queries, will have dramatically different profiles – with possibly major consequences for the perceived and actual effectiveness of the information system in the organization. So we need to consider the “using person” or “using community” as a part of a larger system, which is the actual functioning unit consisting of system, information, and user. For these reasons, the two right-hand layers are included as functional elements of the information retrieval system in Fig. 1.

#### 4. Examples of cascade interactions

Now let us examine how the cascade of interactions works in practice. To date, the greatest expertise in the mounting and accessing of very large bodies of structured machine-readable information has accumulated in the areas of online public access catalogs and online and CD-ROM commercial databases. Online catalogs of major library systems routinely exceed 10 million records in size; some of the online databases, such as *BIOSIS* (1997) and *Chemical Abstracts* (Chemical Abstracts, 1907–) are also very large. Many of the digital library experiments to date, exciting and pioneering as they are, still do not yet fully challenge our skills in designing methods of retrieval from very large structured databases. We can thus learn from these earlier experiences. (At the same time, it is important not to be limited by earlier practices.)

Below, after discussing a simple, introductory example from the world of database searching, I present three further, more complex examples, one each from online catalogs, online databases, and a Web-based database and query expansion tool.

In the examples, the emphasis is on the interactions of indexing, searching, and interface features, as that is the author’s area of expertise. These correspond to the lower left branch, as well as the middle and right-hand parts of Fig. 1. In illustrating how decisions and design features at earlier stages influence later stages, and vice versa, the cascading of interactions through the layers of the system is demonstrated.



#### 4.1. Example #1. Database producer and database vendor designs conflict

We begin with a very simple example. Biological Abstracts, in the early days of producing its database of biological literature, *BIOSIS*, discovered that its users were having problems in searching terms such as “Vitamin B 6”. Searching on the Boolean statement “Vitamin AND B AND 6” would pick up the three elements from all over records and would yield many “false drops” (retrievals of unwanted senses of the search terms). Proximity searching (allowing the searcher to require that search terms be in a designated proximity to each other), which would solve most of the problems, was not initially available in database vendor systems, such as *DIALOG*<sup>R</sup>. Biological Abstracts solved their problem by hyphenating “B-6” and a select group of other subject terms in their records. In principle, searchers could then use the hyphenated terms and eliminate the many false drops they were otherwise retrieving.

At that same time, however, *DIALOG* had the practice of stripping hyphens, among other punctuation, to make searching simpler and more consistent across the many databases, including *BIOSIS*, that *DIALOG* offered to its customers. So Biological Abstracts went to the human and automatic effort necessary to add the hyphens, and *DIALOG* turned around and stripped them out again. Both organizations had good reasons for what they did, but, needless to say, the end result was a conflict between two laudable objectives that left searchers at a loss. In terms of Fig. 1, the transition between “Actual indexing” and “Information in searchable form” conflicted, work being done at the former stage (to assist users ultimately employing search capabilities in the interface) being undone at the latter stage.

Now, let us consider three more complex examples.

#### 4.2. Example #2. Boolean search system being superimposed on non-Boolean indexing system in online catalogs

The most common type of searching available in automated information systems of all kinds is some form or another of Boolean searching, that is, the combining of search terms with the Boolean operators AND, OR, and NOT. The use of Boolean operators is a “natural” in computer systems, because it echoes so much of the deeper structure of computer design. However, Boolean logic for information searching actually predates modern computer searching applications by decades. It was first promoted in a big way in the early 1950s (Taube, 1953), with cumbersome manual paper systems.

For our purposes here, the important point about Boolean searching in those manual and online systems, was that it was associated – *necessarily* – with a system of indexing called “concept” indexing. With concept indexing, each distinct important concept used in the record was to be indexed separately. These one-concept-only terms have come to be called “descriptors.” In the indexing process, as many as 30 or 40 such descriptors might be applied to a single short article or report. Then, at the time of the search, Boolean logic would be used to combine selected concepts in any way desired by the searcher. So, for example, using concept indexing, a Boolean search could be designed as follows:

Digital libraries AND Information retrieval AND Distance learning

Concept indexing contrasts with an older system of indexing that involves the use of what are known as “subject headings”. In particular, subject headings found in the Library of Congress Subject Headings (Library of Congress, 1999) and Sears Subject Headings (Miller, 1994) are used virtually universally by academic and public libraries in the US. As examples, two different children’s books on the US Civil War might be indexed as below. Dashes separate the main heading and each subsequent subdivision.

US – History – Civil War, 1861–1865 – Campaigns – Juvenile literature

US – History – Civil War, 1861–1865 – Naval operations – Juvenile literature

Online databases have generally been indexed by some form or other of descriptor. Boolean search systems became standard for accessing these databases. It seemed natural, then, to use implicit or explicit Boolean search capabilities as well for accessing library catalogs when they went online in the 1980s.

There was a major problem, however – still only partially solved – in using Boolean searching with library subject headings. The theory of the design of subject headings was that they were *not* to be single-concept terms. For lack of a better word, we might call them “whole-document” indexing terms. (See also Bates, 1988; Rowley, 2000; for more detail on differences in underlying indexing systems.) That is, the subject heading string (main heading plus subdivisions) applied to a book was to describe – all by itself – the entire book.

Consequently, the subject heading, even including the subdivisions, was generally quite broad, often much broader than individual concept index terms. Instead of the 30 or 40 terms per record used with concept indexing languages, just one or two subject headings were typically applied, as each heading was assumed to describe all or most of the book. The types of subdivisions allowed were strictly controlled, as was their order of appearance in the subject heading.

Under these circumstances, it makes no sense to search on subject-heading-indexed records with Boolean logic. If most of the records have only one subject heading, then the searcher will almost always come up with no hits on a search that combines more than one subject heading. Yet Boolean search systems have routinely been superimposed on subject-heading-indexed catalogs.

Not surprisingly, these systems did not work very satisfactorily (Markey, 1984; Matthews et al., 1983) until keyword searching, especially on titles, was allowed. Early on, users had more success searching by subject with such keyword searching than with subject-heading searching, and use of keywords shot up in the profile of online public access catalog use (Larson, 1991). But keyword searching has its own limitations. If a title does not use a word, then the entire document is missed, even if it is exactly on the subject desired. In other words, all the benefits of controlling and standardizing vocabulary in indexing is lost, if the searcher must rely only on automatic searches for keyword matches.

How, then, to solve these problems? There is no one best solution that has been employed, and a number of ideas are still floating around for ways to improve things. One of the responses is that the Library of Congress has taken to applying more, and shorter, subject headings, thus creating a hybrid indexing system more responsive to Boolean searching.

Keyword searching on subject headings can also be very helpful, though it is not a fully satisfactory solution, because the headings are typically quite broad, conceptually. With the

keyword-search capability the searcher can cherry-pick words and phrases from within the long strings of subject headings and combine them with Boolean logic. However, neither of these solutions has had the flexibility and power of the concept-indexing/Boolean logic combination.

Since the strength of the subject heading approach lies in its combination of elements to describe the whole document in a single string, some online catalog designers have made it possible for the user to see a segment of the alphabetized main headings plus subdivisions in the interface. In this way, the searcher can see a list of the main headings grouped alphabetically and then differentiate among different heading/subdivision combinations to find the most relevant search phrase. (The example above is such a list, just two headings long.)

This method brings its own limitations, however. For example, it can be seen that in a real catalog of any size, a person searching for juvenile Civil War literature would find that the two “Civil War” subject headings listed above would be separated by many other subject headings on *adult* literature of the Civil War. All the adult literature on the Civil War that uses subdivisions falling alphabetically between “Campaigns” and “Naval operations” would fall between the two juvenile literature headings.

These are just some of the problems encountered when an indexing system designed for a paper card catalog has imposed on it a search system that is not compatible with it. The consequences ripple through the layers of the entire system, all the way to, and including, the interface.

Considering it from the searcher and interface design perspective, should there be keyword searching on subject headings? Boolean searching? Searching on subdivision segments? Alphabetical listing of subject headings on the screen? Batching of subject heading levels so as to group main heading and subdivisions in successive hierarchical displays? Provision of searcher front-ends to assist with vocabulary and/or grouping of records? All these and many other solutions have been proposed and experimented with by designers of online public-access catalogs (Bates, 1977, 1986; Cochrane & Markey, 1983; Drabenstott & Weller, 1996; Hildreth, 1989; Kristensen, 1993; Larson, McDonough, O’Leary, Kuntz, & Moon, 1996; Schatz, Johnson, Cochrane, & Chen, 1996).

All of these have implications at the interface layer. Forms of assistance that could be provided to users, such as those mentioned in the preceding paragraph, may not be considered at all, if system designers do not understand the underlying conflict between the conceptual organization of the indexing language and the retrieval methods provided to the user.

#### *4.3. Example #3. Structure of online retrieval system designed for one searching style conflicts with needs of those with different searching style*

In the 1990s, the Getty Information Institute (now absorbed into the Getty Research Institute, a part of the Getty Trust, which also supports the Getty Museum in Los Angeles, CA, USA) carried out a program of study, in which the author was involved, on online database use by Visiting Scholars at the Getty Research Institute (Bates, 1994, 1996a,b; Bates, Wilde, & Siegfried, 1993, 1995; Siegfried, Bates, & Wilde, 1993). For two years of the Visiting Scholar program, scholars willing to participate were taught to do online searching through DIALOG. The entire protocols of their searches were captured by a computer program, and the scholars were interviewed extensively about their experiences. This abundant data formed the basis of the six articles that resulted from the project. In examining the scholars’ queries, their actual search strategies as

input to the system, their overall search histories, and their reactions and attitudes regarding their experiences, we learned, without initially intending to, the importance of the interactions at every step of the cascade. The full results are available in the article reports, but a number of the interactions are illustrated below in the bullet points:

- *Information content.* Databases for the humanities literature had been developed in parallel to the design of databases for the sciences. Emphasis in such databases is on recent secondary literature in each humanities subject field. The scholars consistently wanted more original (primary) materials – the sort likely to be mounted in future years as digital libraries – rather than secondary materials, and they wanted older materials. They felt they uncovered relevant references adequately other ways besides database searching, and wanted to use computer databases for research on things that were not easy for them to find. Database searching as currently constituted had some uses for them, though of a somewhat different nature than assumed in the design of the databases. (See especially discussion in Bates, 1994.)

So, all the way back at the very first design layer – selection of information content – database providers were not meeting the needs of the anticipated users of the information. Not surprisingly, then, no matter how well mounted for the user, secondary databases have proven to be of relatively less interest to humanities scholars than to scientists. It became apparent through this study and through examination of the literature on humanities information seeking that the mix of motivations and objectives of humanities scholars demands a different type of information be selected for mounting and searching than is common for the sciences. Once that problematic selection of information content had been made, the consequences cascaded through the rest of the system design and use.

- *System of metadata and actual indexing of objects.* The analysis of the scholars' natural language queries and the terms used in the actual search formulations demonstrated that the character of humanities queries and search terms are quite different from those in the social and natural sciences. (See especially discussion in Bates et al., 1993.) We found that whole classes of subject terms were routinely used in the humanities that were virtually non-existent in the sciences. Names of individuals, historical periods, and geographical locations are very common and important in the humanities, and seldom used in the sciences.

In the Getty study, commonly, a scholar would create a query out of several broad terms, one or more drawn from each of the different typical classes of subject term in the humanities. A query topic from the Getty study is illustrative: "Image of the tree in literature, art, science, of medieval and renaissance Europe" (Bates et al., 1993, p. 18). A typical science topic, by contrast, might use solely conventional common terms, few or none of which fall in the common humanities term categories. An example from a National Science Foundation funded study: "Occurrences, causes, treatment, and prevention of retrolental fibroplasia" (Saracevic & Kantor, 1988, p. 195).

The different character of humanities subject terminology has dramatic consequences for indexing of such material. A different design from either concept indexing or subject headings, known as faceted classification, lends itself very well to the terminology of humanities literature. In fact, the Getty Information Institute, finding the faceted approach effective, had already committed itself to that approach in the design of its *Art & Architecture Thesaurus* (Petersen, 1994).

In faceted classifications, recurrent types of terms, or facets, are developed independently of other classes of terms with which they might be combined in conventional hierarchical

classification. For example, in the arts, historical period might be one facet, geographical location another facet, style another, materials used still another, and so on. In this manner whole classes of terms are grouped together in facets, so it is easy for a searcher to find and use a relevant term of the desired type.

However, most of the database producers, again drawing on the concept-indexing/Boolean-logic-searching model, had developed descriptor-type indexing, rather than faceted. The scholars had a hard time using Boolean combination – its logico-mathematical character did not sit well with their very different mental strengths. Further, they had difficulty finding and combining the types of terms they needed to properly express their queries. (See Bates, 1996b; Bates et al., 1993.)

In fact, the conventions of descriptor indexing actually militated against successful searching for the scholars in some characteristic elements of their searching. The example above, “the image of the tree...”, uses some very broad terms – Medieval and Renaissance, Europe, literature, science. Descriptor indexing generally requires the use of specific concepts, because otherwise far too many hits are returned in a typical search. So, frequently, terms of such breadth either do not exist in descriptor thesauri or are used very sparingly in indexing practice. For the humanities scholar, however, it is the combination of perhaps a half dozen of such very broad terms that produces, in the end, a highly specific topic that should retrieve manageably few hits. So the development of indexing vocabularies along the science descriptor model had actually inhibited good indexing of humanities materials. The scholars’ discomfort with Boolean logic further hamstrung them, to the point where the searches were often only incidentally rewarding.

- *System front-end, search capabilities, and interface design.* Interactions among the above-mentioned layers and those in this bullet can be seen in several ways. With respect to the design of the search interface, two things from the above considerations become very evident. First, a way must be found to enable humanities searchers *not* to have to use and understand Boolean logic. Second, the underlying information should be facet indexed, and the searcher helped to use those facets explicitly in the search interface design.

These two needs could be met simultaneously by designing the search interface for humanities scholars so that it both presented the comfortable, familiar conceptual facets to the searcher, *and* enabled the searcher to use these facets without appearing to do Boolean searching. This could be achieved by presenting the searcher with labeled prompt boxes for each facet. Vocabulary assistance associated with each prompt box could enable the searcher to select good terms from within each facet to express the element(s) of the query relevant to that facet. The searcher chooses which facet prompt boxes to fill in – all would not be necessary. In this fashion, the entire query could be expressed effectively, and collectively, through the various prompt boxes actually used.

Once terms are selected and entered in the prompt boxes, searchers then simply click on “Search”, and the system does the Boolean combining for them. What the system could actually do in these circumstances, is conduct Boolean searching according to an algorithm, based on experience and testing, that optimally and sequentially uses the various logical combinations until a satisfactory set size is retrieved for the searcher. Low-priority facets could be left out entirely, if the retrievals were few from the use of initial core facets. Subsequent requests by the user to increase or reduce the retrieved set could lead to further Boolean searching behind the scenes for the humanities searcher.

#### 4.4. Example #4. Design of Web-based retrieval system conflicts with design of indexing system

A final example is drawn from an experimental Web-based vocabulary support and query expansion tool created by the Getty Information Institute. (The Information Institute has now been absorbed into the Getty Research Institute). The tool was known as “a.k.a.” (for “also known as”) and was available for a time at the Getty Web site.

The purpose of the system was to provide the searcher access to the rich vocabulary resources of two Getty databases, the *Art & Architecture Thesaurus* (Petersen, 1994), and the union list of artist names (ULAN) (Bower & Baca, 1994). The intent was that the user could identify alternative or additional search terms in the vocabulary databases, either subject terms or artist names or both, and then use those terms to search several Getty databases, such as the *Bibliography of the History of Art* (1999–), also accessible from within a.k.a.

The system was designed in such a way as to give the searcher, at the interface, the option of using vocabulary assistance or not. Further, the searcher who chose vocabulary assistance could turn the task over to the system for automatic query expansion or explore the vocabularies on his or her own and select desired terms.

This was a powerful idea and was well implemented in many ways. At the behest of the Information Institute, I led a group to evaluate the system extensively. We studied actual use by 36 users with a variety of backgrounds, using a carefully designed activity protocol that combined unstructured and directed searching on a.k.a. We studied both local and remote users. Yet with all this power, a.k.a. did not work very well in the study cases. We ultimately discovered that the problems were due, to a substantial degree, to conflicts between design layers. Here, I will describe two high-impact design layer conflicts that we discovered.

First, as noted earlier, the *Art & Architecture Thesaurus* is a faceted system; in fact, it probably represents the best-developed faceted indexing system in the world. So the general system of indexing used was well suited to the information content and to the needs, particularly, of arts researchers.

Yet with all this power, a.k.a. did not work well in our study. Frequently, users did better without using the vocabulary enhancement than with. Why? We found that the most serious problem concerned the interaction between the system of indexing and the structure of the interface.

The power of faceted classification lies, jointly, in (1) its grouping of terms of a similar type into facets, and in (2) its ability to combine terms from different facets (usually with Boolean AND) to create a total query. As noted earlier, humanities queries characteristically draw terms from several facets. So, for the “image of a tree” example mentioned earlier, a searcher might want to draw a term or terms for Europe from a geographical facet, terms for “Medieval” and/or “Renaissance” from a historical period facet, and so on through the various facets touched on in the query statement.

The online a.k.a. system, however, as originally designed, only permitted the searcher to explore and select terms from one facet at a time. Once terms from a facet had been selected, it was not possible to move to another facet and select additional terms to AND with the terms selected from the first facet. The searcher could only move on to a completely new search. Clearly, in the image of the tree example, and in many other humanities queries, it is not enough to search on just one facet of a subject. Of what use would it be to search to bring up every record in the database on the Renaissance, without also limiting that search by image of the tree, “Europe,” and so on?

This was a crucial flaw in the original design of a.k.a. The main purpose of the system was to enable people to expand queries with the rich and carefully researched vocabulary resources of the Getty. But the query expansion capabilities were designed without adequately incorporating a core feature of the intellectual structure of faceted vocabularies – facet combination, a feature that was crucial to effective operation for many real queries. We learned that there had been inadequate communication between technical staff and vocabulary staff at critical design phases of the project.

The a.k.a. system was a daring experimental design (a design, incidentally, similar to what I have been recommending (Bates, 1986) – and implementing (Bates, 1990b) – for a number of years), and, as such, was bound to have start-up problems that needed to be ironed out. But this particular new-system glitch might not have happened with a design process that required frequent and detailed communication between the representatives of the various design layers.

The second example problem with a.k.a. reflected what was probably another communication mix-up. In this case the problem arose with the ULAN (Bower & Baca, 1994). Artists may be known by literally dozens of names – names in different languages, nicknames, and the like. We found, for example, that the painter Titian had more than 40 name variants in his entry. The material available on name variants in the ULAN represents extensive research by the developers of the vocabulary at the Getty, and can be of great value to scholars who may not realize at the beginning of a scholarly project on an artist how many names they may need to use in their research.

It therefore seemed to make sense to offer ULAN to searchers as a part of the a.k.a. system. However, the evaluation found that ULAN-enriched queries did not seem to help people find more references in the Getty bibliographic databases, such as the *Bibliography of the History of Art* (1999–) (BHA). We soon discovered why. The indexers who prepared the entries for the BHA database, for example, themselves used ULAN to identify the standard preferred name for the painters referenced in the database. So, even if Titian was described by the variant name “Ticiano” or “Ziano” in the record indexed, the catalogers would also always add the standard name “Titian”. They thus eliminated the need for the end-user to identify the alternative names for searching purposes. That was an excellent use of ULAN, but it defeated the point, for the most part, of including ULAN in a.k.a. Here, in this second example from the a.k.a. evaluation, we see a conflict, or misunderstanding, between the people at the “actual indexing” layer and those designing the “system front-end” and “search capabilities” layers.

After the evaluation project, a number of changes were made in a.k.a. to solve these layer conflicts, and before it was taken down from the Web, a.k.a. performed closer to its potential.

## 5. Implications for digital libraries

The elaboration of these various examples and their proposed solutions in the previous section has not been done principally to suggest solutions to those specific problems. Others may have different solutions. Rather, the purpose was to illustrate how the mix of circumstances in a particular information system design situation can lead to implications and interactions that cascade throughout the layers of the structure of an information system.

In the first example, two entities that manipulated the database contents – the database producer (Biological Abstracts) and the database vendor (DIALOG) – worked at cross-purposes

to each other. In that case, the database producer and vendor had different motivations: the producer to create a unique resource, shaped to the distinctive needs of its audience, and the vendor to create as uniform a set of databases as possible, to make it easier for searchers moving from one database to another.

It is not hard to imagine these same sorts of conflicts arising again and again in the digital library world. The producing institution shapes and develops its own distinctive mix of types of metadata, which some search engine or interoperability mechanism overrides or ignores.

This is not (or, at least, not always) an insoluble problem, however. Ultimately, proximity searching capabilities added to DIALOG made it possible for searchers to require that the components of these multi-element biological concepts be found together or not be retrieved.

In the second example, the use of Boolean search mechanisms with subject headings, the important lesson to learn is that the conflict between user search capability and underlying indexing principles can go very deep. Because subject headings were designed for a different environment, they did not mesh well with the Boolean searching requirements of typical online catalogs. This is one case where the solutions have been halting, expensive, and not wholly satisfactory.

Yet much of the discussion in the digital library world regarding metadata seems to presume that the intellectual structure guiding the design and application of the metadata is of little import. In the Dublin Core, “subject” is just one generic category among 15 (Weibel & Lagoze, 1997). If there are four different types of subject description applied to each document – a situation not at all unlikely – each designed on different principles, they will all be lumped together in the repeating fields of that category. For good retrieval, however, they may each need a different kind of retrieval mechanism and presentation in the interface.

In the third example in the previous section, given what was learned about the character and structure of the information in the humanities, and about the nature of query formation by humanities scholars (elements in the far left and far right of Fig. 1), conclusions were drawn about the nature of the indexing that should be done in the humanities (faceted), and, consequently, about the nature of presentation of search capabilities (prompt boxes with hidden Boolean searching) in the interface. Choices made at every layer of the system design culminated in the final look of the information system interface.

For the end user attempting to use subject access in the system interface, it is desirable, first, to provide search capabilities to the searcher that are appropriate for the type of intellectual structure contained in each type of subject metadata, and second, to represent those capabilities in the interface so that their function is as self-evident and as easily usable as possible.

The challenges of designing a system so that all the layers work together effectively is amply illustrated by the fourth example, the a.k.a. system. The system deserves to be revived and refined, to take full advantage of the rich and deeply researched information available in the Getty vocabularies and databases. But for it to work in the end, that revival must be done in a way that truly integrates the complex expertises of the personnel working at each layer of the system. The kind of information, the systems of description used for that information, the search mechanism(s) used on that information, and the manifestation of search capabilities in the user interface all have to dovetail for effective retrieval to result.

The four examples provided above – as well as many others encountered in a long career of consulting on the design of real-world information systems – illustrate two key points of this article. First, excellent design at individual layers of a system can nonetheless work completely at



cross-purposes and thwart the purpose of the overall system. Second, deep expertises are developing at every design layer in Fig. 1. Work necessarily must be compartmentalized to some degree to achieve the objectives of a project. People working at each layer cannot become experts at every other layer. Therefore, genuine, ongoing communication must take place between designers working on the several layers throughout the development and implementation process. That communication must go deeper than a memo passed around or an occasional meeting. Cooperative interaction needs to take place frequently, and be built into the administrative and social structure of the organization.

That information exchange, by the way, should take place on an equal footing. Human nature being what it is, it is not uncommon for competitiveness and domination struggles to interfere with genuine openness between the representatives of the various layers.

## 6. Conclusions

Changing even seemingly small things at one information system design layer can have huge implications for the other layers. Since the layers themselves interact in a cascade from system and information content chosen (left-hand side of Fig. 1) all the way through interface design and characteristics of use (right-hand side of Fig. 1), the design of such systems must also manifest mutual knowledge between those layers.

In the development of digital libraries, the layers are proliferating and the potential for conflict between layers multiplying (Kramer, Nikolai, & Habeck, 1997; Lynch & Garcia-Molina, 1995). It is not uncommon nowadays for information systems to be designed collectively by several different individuals or groups, some of which either never talk to each other or talk past each other. Now that so much computing power and sophistication is entering the information world, especially in the development of digital libraries, deep expertises are developing in each of the layers displayed in Fig. 1. One person cannot know all that is needed to put such a system together effectively.

Unfortunately, people working at each layer may do an excellent job in their own area of expertise but may fail to recognize or influence the design issues interacting between their layer and the other layers. As a consequence, individual layers may function very well, but work at cross-purposes with the functions of other layers.

Digital libraries cannot be fully effective as information sources for users until the entire design process is done in a manner that involves genuine conceptual and practical coordination among the people working on the system layers. The information content, its database structure, and retrievable elements, should not be selected without full consultation with experts in the subject domain and in the information seeking behavior and context of use of the proposed digital library information.

The interface design should meet not only general criteria of good interface design, but should also draw on expertise in *information system* interface design. That expertise will include understanding of various options in the provision of search capabilities for the user, including front-ends, as well as understanding of the underlying indexing and metadata structure, and how that structure can best be represented and used in the interface.

In sum, all layers of the system for accessing and displaying digital library information should be simultaneously designed with knowledge of what is going forward in the other layers. It takes

only one wrongly placed layer to thwart all the clever work done at every other layer. For effective information retrieval to occur, all layers of a system must be designed to work together, and the people doing the designing must genuinely communicate.

## References

- Bates, M. J. (1977). System meets user: Problems in matching subject search terms. *Information Processing and Management*, 13(6), 367–375.
- Bates, M. J. (1986). Subject access in online catalogs: A design model. *Journal of the American Society for Information Science*, 37(6), 357–376.
- Bates, M. J. (1988). How to use controlled vocabularies more effectively in online searching. *Online*, 12(6), 45–56.
- Bates, M. J. (1990a). Where should the person stop and the information search interface start? *Information Processing and Management*, 26(5), 575–591.
- Bates, M. J. (1990b). Design for a subject search interface and online thesaurus for a very large records management database. In D. Henderson (Ed.), *Proceedings of the American Society for Information Science* (Vol. 27, pp. 20–28). Medford, NJ: Learned Information.
- Bates, M. J. (1994). The design of databases and other information resources for humanities scholars: The Getty Online Searching Project report no. 4. *Online and CDROM Review*, 18(6), 331–340.
- Bates, M. J. (1996a). Document familiarity, relevance, and Bradford's Law: The Getty Online Searching Project report no. 5. *Information Processing and Management*, 32(6), 697–707.
- Bates, M. J. (1996b). The Getty end-user online searching project in the humanities: Report no. 6: Overview and conclusions. *College and Research Libraries*, 57(6), 514–523.
- Bates, M. J. (1998). Indexing and access for digital libraries and the Internet: Human, database, and domain factors. *Journal of the American Society for Information Science*, 49(13), 1185–1205.
- Bates, M. J., Wilde, D. N., & Siegfried, S. (1993). An analysis of search terminology used by humanities scholars: The Getty Online Searching Project report no. 1. *Library Quarterly*, 63(1), 1–39.
- Bates, M. J., Wilde, D. N., & Siegfried, S. (1995). Research practices of humanities scholars in an online environment: The Getty Online Searching Project report no. 3. *Library and Information Science Research*, 17(1), 5–40.
- Beaulieu, M. (2000). Interaction in information searching and retrieval. *Journal of Documentation*, 56(4), 431–439.
- Belkin, N., Brooks, H., & Daniels, P. (1987). Knowledge elicitation using discourse analysis. *International Journal of Man–Machine Studies*, 27(2), 127–144.
- Bibliography of the History of Art* (1999–). Los Angeles, CA: J. Paul Getty Trust, Vandoeuvre-les-Nancy, France: Centre National de la recherche scientifique. See description at: <http://library.dialog.com/bluesheets/html/b10190.html> (Accessed: April 27, 2001).
- BIOSIS Search Guide* (1997). Philadelphia, PA: Biological Abstracts, Inc..
- Bødker, S. (1991). *Through the interface: A human activity approach to user interface design*. London: Lawrence Erlbaum Associates.
- Bower, J. M. Baca, M. (1994). *Union list of artist names*. Computerfile Version 1.0. New York: G.K. Hall/Macmillan. Also: <http://www.getty.edu/research/tools/vocabulary> (Accessed April 27, 2001).
- Butler, K. A., Bahrami, A., Esposito, C., & Hebron, R. (2000). Conceptual models for coordinating the design of user work with the design of information systems. *Data and Knowledge Engineering*, 33(2), 191–198.
- Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human–computer interaction*. London: Lawrence Erlbaum Associates.
- Carroll, J. M. (Ed.). (1991). *Designing interaction: Psychology at the human–computer interface*. New York: Cambridge University Press.
- Chemical Abstracts* (1907–). Easton, PA: American Chemical Society.
- Cochrane, P. A., & Markey, K. (1983). Catalog use studies – before and after the introduction of online interactive catalogs: Impact on design for subject access. *Library and Information Science Research*, 5(4), 337–363.

- Damodaran, L. (1991). Towards a human factors strategy for information technology systems. In B. Shackel, S. J. Richardson (Eds.), *Human factors for informatics usability* (pp. 291–324). New York: Cambridge University Press. Digital Object Identifier. (Home page) <http://www.doi.org> (Accessed: April 27, 2001).
- Dillon, A. (1994). *Designing usable electronic text: Ergonomic aspects of human information usage*. London: Taylor & Francis.
- Dix, A., Finlay, J., Abowd, G., & Beale, R. (1993). *Human–computer interaction*. New York: Prentice-Hall.
- Drabenstott, K. M., & Weller, M. S. (1996). Failure analysis of subject searches in a test of a new design for subject access to online catalogs. *Journal of the American Society for Information Science*, 47(7), 519–537.
- Efthimiadis, E. N. (1990). Online searching aids – A review of front ends, gateways and other interfaces. *Journal of Documentation*, 46(3), 218–262.
- Efthimiadis, E. N. (1996). Query expansion. In M. E. Williams (Ed.), *Annual review of information science and technology* (Vol. 31, pp. 121–187). Medford, NJ: Information Today.
- Ellis, D. (1996). *Progress and problems in information retrieval* (2nd ed.). London: Library Association. *Encoded Archival Description*. (Home page) <http://www.loc.gov/ead> (Accessed: April 27, 2001).
- Galitz, W. O. (1997). *The essential guide to user interface design*. New York: Wiley.
- Gardner, A. (1991). An approach to formalised procedures for user-centred system design. In B. Shackel, & S. J. Richardson (Eds.), *Human factors for informatics usability* (pp. 133–150). New York: Cambridge University Press.
- Hayman, A., & Elliman, T. (2000). Human elements in information system design for knowledge workers. *International Journal of Information Management*, 20(4), 297–309.
- Hildreth, C. R. (1982). *Online public access catalogs: The user interface*. Dublin, OH: OCLC.
- Hildreth, C. R. (1989). *Intelligent interfaces and retrieval methods for subject searching in bibliographic retrieval systems*. Cataloging Distribution Service, Library of Congress, Washington.
- Ingwersen, P. (1992). *Information retrieval interaction*. London: Taylor Graham.
- ISO/IEC 7498-1: (1994) Information technology – Open systems interconnection–Basic reference model: The basic model. Available: <http://www.iso.ch/cate/d20269.html> (Accessed Feb. 7, 2000).
- Kilker, J., & Gay, G. (1998). The social construction of a digital library: A case study examining implications for evaluation. *Information Technology and Libraries*, 17(2), 60–70.
- Kramer, R., Nikolai, R., & Habeck, C. (1997). Thesaurus federations: Loosely integrated thesauri for document retrieval in networks based on Internet technologies. *International Journal on Digital Libraries*, 1(2), 122–131.
- Kristensen, J. (1993). Expanding end-users' query statements for free text searching with a search-aid thesaurus. *Information Processing and Management*, 29(6), 733–744.
- Larson, R. R. (1991). The decline of subject searching – long-term trends and patterns of index use in an online catalog. *Journal of the American Society for Information Science*, 42(3), 197–215.
- Larson, R. R., McDonough, J., O'Leary, P., Kuntz, L., & Moon, R. (1996). Cheshire II: Designing a next-generation online catalog. *Journal of the American Society for Information Science*, 47(7), 555–567.
- Library of Congress. Cataloging Policy and Support Office (1999). *Library of Congress subject headings* (22nd ed., 5 Vols.). Washington: Cataloging Distribution Service, Library of Congress.
- Lynch, C., Garcia-Molina, H. (1995). Interoperability, scaling, and the digital libraries research agenda: A report on the May 18–19, 1995 IITA Digital Libraries Workshop, August 22, 1995. Online: <http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.html> (Accessed: April 30, 2001).
- Markey, K. (1984). *Subject searching in library catalogs: Before and after the introduction of online catalogs*. Dublin, OH: OCLC Online Computer Library Center.
- Matthews, J. R., Lawrence, G. S., & Ferguson, D. K. (Eds.). (1983). *Using online catalogs: A nationwide survey: A report of a study sponsored by the Council on Library Resources*. New York: Neal-Schuman.
- Miller, J. (Ed.). (1994). *Sears list of subject headings* (15th ed.). New York: Wilson.
- Norman, D. A., & Draper, S. W. (1986). *User centered system design: New perspectives on human–computer interaction*. London: Lawrence Erlbaum Associates.
- Payette, S. D., & Rieger, O. Y. (1998). Supporting scholarly inquiry, Incorporating users in the design of the digital library. *Journal of Academic Librarianship*, 24(2), 121–129.
- Petersen, T. (dir.) 1994. *Art & architecture thesaurus* (2nd ed.). New York: Oxford. Also: <http://www.getty.edu/research/tools/vocabulary> (Accessed: April 27, 2001).

- Robertson, S. E. (1977). Theories and models in information retrieval. *Journal of Documentation*, 33(2), 126–148.
- Rowley, J., & Farrow, J. (2000). *Organizing knowledge: An introduction to managing access to information* (3rd ed.). Aldershot, Hampshire: Gower.
- Salton, G., & McGill, J. M. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.
- Saracevic, T. (1996). Modeling interaction in information retrieval (IR), A review and proposal. In *Proceedings of the 59th annual meeting of the American Society for Information Science* (Vol. 33, pp. 3–9).
- Saracevic, T. (1997). The stratified model of information retrieval interaction: Extension and application. In C. Schwartz, M. Rorvig (Eds.), *Proceedings of the 60th ASIS annual meeting* (Vol. 34, pp. 313–327). Medford, NJ: Information Today.
- Saracevic, T., & Kantor, P. (1988). A study of information seeking and retrieving. II. Users, questions, and effectiveness. *Journal of the American Society for Information Science*, 39(3), 177–196.
- Schatz, B. R., Johnson, E. H., Cochrane, P. A., Chen, H. (1996). Interactive term suggestion for users of digital libraries: Using subject thesauri and co-occurrence lists for information retrieval. In *Proceedings of the 1st ACM international conference on digital libraries* (pp. 126–133). New York: Association for Computing Machinery.
- Shneiderman, B. (1998). *Designing the user interface: Strategies for effective human–computer interaction* (third ed.). Reading, MA: Addison-Wesley Longman.
- Shor, R. (1991). A uniform graphics front-end. *Computers in Libraries*, 11(11), 48–51.
- Siegfried, S., Bates, M. J., & Wilde, D. N. (1993). A profile of end-user searching behavior by humanities scholars: The Getty Online Searching Project report no. 2. *Journal of the American Society for Information Science*, 44(5), 273–291.
- Spink, A. (1997). Information science: A third feedback framework. *Journal of the American Society for Information Science*, 48(8), 728–740.
- Syan, C. S., & Menon, U. (1994). *Concurrent engineering: Concepts, implementation and practice*. London: Chapman & Hall.
- Taube, M. (1953) *Studies in coordinate indexing*. Washington: Documentation, Inc.
- Text Encoding Initiative. (Home page) <http://www.tei-c.org> (Accessed: April 27, 2001).
- Vickery, B., & Vickery, A. (1993). Online search interface design. *Journal of Documentation*, 49(2), 103–187.
- Weibel, S. L., Lagoze, C. (1997). An element set to support resource discovery. *International Journal on Digital Libraries*, 1(2), 176–186. Also: <http://dublincore.org> (Accessed April 27, 2001).
- Zhang, J., & Fine, S. (1996). The effect of human behavior on the design of an information retrieval system interface. *International Information and Library Review*, 28(3), 249–260.
- Zhu, Z. (2001). Towards an integrating programme for information systems design: An oriental case. *International Journal of Information Management*, 21(1), 69–90.