

SEARCH STRATEGIES IN DOMAIN-SPECIFIC IMAGE RETRIEVAL:  
A PILOT STUDY

by  
Sheila Denn

A Master's paper submitted to the faculty  
of the School of Information and Library Science  
of the University of North Carolina at Chapel Hill  
in partial fulfillment of the requirements  
for the degree of Master of Science in  
Information Science

Chapel Hill, North Carolina

July, 1999

Approved by:

---

Advisor

Sheila Denn. Search Strategies in Domain-Specific Image Retrieval: A Pilot Study. A Master's paper for the M.S. in I.S. degree. July, 1999. 55 pages.  
Advisor: Stephanie W. Haas

This paper describes a prototype World Wide Web-accessible database of slide images developed for the University of North Carolina at Chapel Hill School of Journalism and Mass Communication Library, and a pilot study conducted with prospective users of the slide images. The purpose of the pilot study was to observe information-seeking behaviors and strategies with reference to image data, and to investigate the role of domain expertise in determining these behaviors and strategies.

Five journalism faculty members and six journalism students completed retrieval tasks using the slide database. They then completed a questionnaire about their experiences. Results showed that keyword or full text searching is preferred over browsing for structured retrieval tasks, and that the students were more likely to use the thumbnail image to pare down a result set, while faculty were more likely to use the text descriptions for this purpose. Suggestions for further study are given.

Headings:

- Information retrieval – Evaluation

- Information systems – Design

- Use studies – End-user searching

- Use studies – Information systems

- Image Retrieval

## Table of Contents

Introduction . . . . .	.3
Chapter 1 . . . . .	5
Chapter 2 . . . . .	.16
Chapter 3 . . . . .	.25
Chapter 4 . . . . .	.28
References . . . . .	39
Appendix A . . . . .	42
Appendix B . . . . .	52
Appendix C . . . . .	54

*It is not so much that a picture is worth a thousand words, for many fewer words can describe a still picture for most retrieval purposes. The issue has more to do with the fact that those words vary from one person to another.*

Keister (1994), p. 17

## **Introduction**

This paper describes a prototype World Wide Web-accessible slide database (hereafter referred to as the JoMC Slide Database) built for the University of North Carolina at Chapel Hill School of Journalism and Mass Communication. The JoMC Slide Database contains 695 slides from 13 slide sets. These slide sets were originally designed for classroom presentation in the field of journalism and mass communication, and cover topics such as historical events depicted in newspaper front pages, African-Americans in the media, foreign correspondents, advertising art, themes, and slogans, and the image of women in film.

A pilot study was conducted, using this slide database, to determine if there are differences in search strategy formulation when searching for image data as opposed to textual data, and if there is a difference between the image search strategies of novices and experts in a subject domain using a domain-specific image collection. Chapter 1 of this paper provides background and a literature review in the areas of information-seeking in subject domains and end-user searching both in text and non-text environments; Chapter 2 describes the

design and implementation of the JoMC Slide Database; Chapter 3 describes the design of the pilot study; and Chapter 4 describes the results and conclusions gleaned from the pilot study and outlines suggestions for future research.

## **Chapter 1: Background**

The current activity in the field of image retrieval research is in developing methodologies for searching image collections using the image data itself (for a good overview of mid-1990s work in this area, see Cawkell, 1998; or Heidorn & Sandore, 1997). The most recent research in this area concerns the use of fuzzy logic to retrieve images based on their content (Wu, Narasimhalu, & Desai, 1998); similarity measures to retrieve images based on color features extracted from the images during the indexing process (Pei & Cheng, 1999); and knowledge-based visual query languages to retrieve images by feature and content (Chu & Hsu, 1998). Sophisticated image retrieval technology is now being developed by commercial ventures and is being integrated into relational database systems from Oracle, Sybase, Informix, and others (Virage, Inc., 1999). Indexing and retrieval of images based on the ability to match image features, colors, shapes, etc. is one of the most fruitful avenues of retrieval research in the late 1990s.

On the other hand, however, cutting edge image retrieval systems are currently out of the reach of many smaller libraries, museums, or other institutions with image collections. For these smaller entities, indexing based on subject

keywords or other textual material associated with the images is still the most viable strategy, and likely will be so for at least the next several years. Even when content- or feature-based image retrieval becomes more widespread, there will likely always be a role for textual descriptors to aid in retrieving images. While sophisticated image indexing algorithms can and will be developed allowing the user to use color, features, and shapes to retrieve images, such algorithms will never be able to extract the semantic content of the image, what the image is “about”, in anything approaching an adequate fashion. For example, consider the

image from our slide collection shown in Figure 1 at right. A visual feature-extracting indexing algorithm would be able to extract the dominant colors (orange, black, green and white) in the image, the fact that the image contains many curved shapes, perhaps the fact that there is a face depicted in the image, and the juxtaposition of the colors. What could not be extracted is that this image is an advertisement for Job cigarette papers from the late



**Figure 1**

19<sup>th</sup> century by a French artist named Alphonse Mucha, that it is representative of a general trend in advertising of that period to use fine art to sell products, and that it is also representative of the use of the images of women in advertising. These elements are a better indicator of what this image is “about” in the context of our collection. We can think of this “aboutness” in four layers: the medium, the creator, the visual objects

represented, and the context, in the sense of what the visual objects represented in the image “mean” or why they are significant. Of these, the most difficult layer (but arguably the most important) to represent is the context, because often this information is not directly represented in the image itself. To further complicate the issue, the context layer of the image is likely to vary depending on the viewer. This is where sophisticated algorithms are no substitute for a human indexer in providing the textual metadata that will capture an image’s “aboutness.” In addition, that indexer must be aware of the characteristics of the intended user base for the collection in order to provide contextual information that will be of significance to them.

It can be (and has been) argued, in fact, that the need for human indexing is even greater in dealing with images than in dealing with text, precisely because there is no “full text” in an image upon which a user can conduct a search. Despite the importance of proper textual indexing of images, however, very little research has been conducted to determine how this should be accomplished. For this reason, we will turn to a review of the literature on the relationship between subject indexing and user information-seeking behaviors in text collections for principles to guide us in applying textual metadata to images.

One of the first issues that must be addressed is that of the user population for the image collection. Will the users come from members of the general public? Or will they be experts in a particular subject domain? If the

latter, how will their domain expertise influence the ways in which they will search for images? Do they have particular needs in terms of the kinds of textual descriptors that should be provided?

These questions are important because there may be a great deal of difference between the way a collection of images designed to be accessed by the general public should be indexed versus the way that a collection designed for a small population of experts within a particular field should be indexed. With a large, publicly accessible collection, the indexer must contend with the fact that users will come to the collection with a myriad of different information needs and uses, and likely a diverse set of language constructs in which they will frame their information needs. For this reason, it is best to keep the indexing terms for such collections general and based on fairly objective descriptions of image features (such as the colors in the image; the objects represented in the image; or the creator of the image, whether it be a photograph, painting, or other kind of image). It is impossible to predict the purposes for which users will be searching for images in such a publicly accessible collection, and it is therefore unwise to index the images in such a way that they can only be retrieved by queries that deal with one facet of the image's semantic content. A good example of this kind of collection is the Prints and Photographs Collection of the National Library of Medicine. In studying the usage of this web-accessible collection of images, Keister (1994) found that a large number of the queries run against the collection were based on visual elements in the images themselves, and not necessarily on

the specifically medical contexts of the images. Keister refers to these kinds of queries as “image construct queries” and states that fully one-third of the requests made of the NLM’s Prints and Photographs Collection are image construct queries (p. 13). She gives as examples requests for “a warm picture of nurse, mother, baby;” “people racing in wheelchairs;” “surgeons standing;” and “the man sitting in the chair with a box on his head” (p. 13).

Keister also gives a striking example of how the context of an image can change with the passage of time. She describes an 1899 photograph of nurses that was originally taken to show the nurses’ professionalism and the progress of the field of nursing in the late 19<sup>th</sup> century. Now when patrons ask for the photograph, they are most likely to be looking for an image that captures the feeling of “nursing in the quaint old days” (p. 13). This example dramatically illustrates the difficulty in capturing contextual information for retrieval purposes, even for domain-specific collections: that context is a moving target and can change dramatically with the passage of time.

For collections that are designed to be used not by the general public but rather by users from a particular subject domain, the question of how to proceed with indexing images using textual metadata is perhaps a bit less complex than it is for public collections. For the domain-specific collection, one can utilize the terminology and concepts in the subject domain to categorize the images, on the assumption that these concepts will be integral to how the images will be used.

However, since there has been to date no work done on what impact domain knowledge has on how users search for image data in that domain, we must turn to the work that has been done on the impact of subject expertise on information-seeking behavior in the text environment.

One concept that continues to be brought out in research on user information-seeking behavior is that individual users bring widely disparate thinking processes to any retrieval system (Morgan, 1999). As Morgan states, "At the present time, there is no practical way for indexers to create intuitive indexes for every set of user populations, let alone compile an index reflecting the thinking processes of individuals" (p. 38). Nonetheless, conventional wisdom (and a number of studies) suggest that it is desirable to accommodate as many different styles of information-seeking as is feasible. In modern text retrieval systems, accommodation to different information-seeking styles generally takes the form of providing a) access based on *browsing* the collection in some fashion; b) access based on the ability to *search the full text* of the document in question; or b) access based on the ability to *search keywords* assigned to the document by a human indexer.

Each access methodology brings with it certain advantages and disadvantages. Browsing, while allowing the user "to bring nothing to the information-seeking process except the desire to know or learn something" (Morgan, p. 38), is notoriously inefficient when the user is looking for a particular

document or a specific piece of information. Free-text searching provides a great deal of flexibility to the user and does not require her to know or understand an indexing classification scheme, but at the same time has serious limitations due to its dependence on words or phrases occurring within the text itself. Finally, keyword indexing schemes provide a structured system for organizing and retrieving documents, but these schemes are more likely to represent the thinking processes of the indexer or the conventional knowledge structure of the subject area or domain than the cognitive constructs of the users of the system.

Since the system developed for this project was intended for users in an academic environment, we drew upon the work that has been done in characterizing the information-seeking behaviors of scholarly users. One of the major findings of such studies has been that scholars can gain a great deal of benefit from retrieval interfaces that are flexible and adaptive to their needs (Payette & Rieger, 1998). One of the guiding principles of Payette & Rieger's work is that "Library professionals must fully understand the nature of users' research and information-seeking behavior to be able to create useful systems and related services" (p. 121). We would add that retrieval system designers must understand these behaviors as well.

We also felt that our project was closely related to the ongoing work in digital libraries, based on the following definition of "digital library" given by the UCLA-NSF Social Aspects of Digital Libraries Workshop (UCLA, 1996):

Digital libraries are a set of electronic resources and associated technical capabilities for creating, searching, and using information. In this sense they are an extension and enhancement of information storage and retrieval systems that manipulate digital data in any medium (text, images, sounds; static or dynamic images) and exist in distributed networks. The content of digital libraries includes data, metadata that describe various aspects of the data (e.g., representation, creator, owner, reproduction rights), and metadata that consist of links or relationships to other data or metadata, whether internal or external to the digital library.

It is perhaps more precise to think of our slide collection database as a component piece of a digital library, one that will eventually be integrated with the rest of the resources available through the JoMC Library's web page. But the concepts are close enough that the experiences of researchers in the digital library field are of benefit to our project.

As might be expected given the relative novelty of the digital library concept, most of the studies of user information behaviors that have been conducted over the last 50 years have occurred in traditional, print-based library settings (Wilson, 1994). There is a great need for this kind of user study in the emerging digital library arena, but we must first answer the two questions Lancaster posed to the digital library research community (Lancaster, 1995):

1. Do users of digital libraries have objectives different from those of users of "traditional" libraries?
2. Do the evaluation criteria remain the same or do digital libraries demand use of different criteria?

Lancaster himself feels that despite the differences between the traditional library and the digital library, user goals and criteria for success are likely to be similar for the two settings. (Lancaster, 1995).

Wilson also found in his review of user studies that these studies have been undertaken primarily from a system-centered point of view; that is, they have been concerned with how a searcher uses a particular system and not with how a searcher formulates information needs in the absence of any system in particular (Wilson, 1994). Unfortunately such a study of the “pure” information-seeking behaviors of journalism and mass communication students and faculty is beyond the scope of this particular project, but as the JoMC Library’s digital library grows in size and diversity of materials, we hope to undertake such a study to inform the design of the digital library at an early stage of the process.

Payette and Rieger have rightly pointed out that evaluation in the digital library setting is complicated by the digital library’s position as the nexus point of a number of different research facets. Some of the many issues included under the digital library umbrella are interface design, usability, navigation and presentation of information, searching and retrieval, indexing, and collection development (Payette & Rieger, 1998). For the purposes of this particular study, we are going to concentrate on navigation, retrieval, and indexing as these relate to the information-seeking behaviors of a user group from the particular subject domain of journalism and mass communication. Marchionini (1995a) has argued

forcefully that a more integrated, holistic approach to digital library systems is critical at this point in the history of the digital library, but again this kind of holistic approach is beyond the scope of our project. We hope to take the conclusions we have reached in this pilot study and apply them to a more integrated research approach in future enhancements to the slide collection and to the JoMC digital library web page of which it is a part.

Another trend in digital library-oriented user studies is the use of qualitative research methods as opposed to quantitative ones (Payette & Rieger, 1998). These studies use techniques such as direct observation, interviews, and focus groups, as opposed to the traditional mass-mailing survey, for example. This kind of research methodology provides much more in-depth information about user information-seeking behavior, but brings with it the disadvantages of not being easily replicable or generalizable across user populations. However, generalizability is less of an issue since this kind of research is primarily undertaken to assist in the iterative design and refinement of the digital library by studying the specific population that library serves. The fact that these studies could be useful to other digital library designers is a secondary consideration.

There are a number of user studies that have been undertaken for systems geared toward an academic audience. These studies have found that scholarly work tends not to be structured and linear, but rather intuitive and cyclical. Their methods for information-seeking reflect this intuitive nature and

often include browsing, tracing footnotes, and consulting with peers (Miller & Tegler, 1986). It has been found that undergraduate students, in particular, utilize information-seeking strategies that involve a great deal of backtracking, reassessment, and reformulation of queries and search strategies (Fister, 1992). These findings suggested to us that our system would need to incorporate browsing as one of the options for gaining access to the slide images, and so we accommodated browsing in our prototype system design.

As useful as it is to construct a general picture of how a general user class, in our case academic users, tends to undertake the information-seeking process, we were also cognizant of the many studies that show that individual information-seeking behaviors can vary widely from the “average”, even in a fairly specific user group (Payette & Rieger, 1998). For example, significant differences have been found between academic users from humanities fields and those in other areas, such as the physical sciences or social sciences (Borgman, 1990). Cognitive style also plays a significant role in accounting for differences between individual information-seeking behaviors (Palmer, 1991). Payette and Rieger (1998) stress the importance of not regarding the user population as an “undifferentiated mass”, and one of the goals of this pilot study is to try to tease out the differences in user behavior influenced by such things as domain expertise, work role (student vs. faculty), and experience with other information-seeking environments.

## **Chapter 2: System Design and Implementation**

The impetus for the design of the JoMC prototype slide database was a desire on the part of the JoMC Librarian to increase the utilization of the JoMC Library's slide collections. At the present time, these slide collections are largely unknown to the School's faculty and students. On the other hand, the Library's web page is very well utilized, and contains a wide range of resources related to journalism and mass communication. Since the user population is already accustomed to gaining access to many of the Library's resources via the World Wide Web, devising a system whereby the Library's slides were Web-accessible was a natural course of action. One of the strengths of the World Wide Web is its ability to present text and graphics seamlessly, thus making it an appropriate medium for providing access to an image collection.

The fact that the slides had been such a little known and seldom used resource made the initial stages of design more difficult. We did not have any data about how faculty and students would be using the collection, because by and large they had never used the slides at all and in fact were unaware of their existence. In addition, converting the slides to digital format opens a broader range of potential future uses for the images. We felt it to be likely that users would develop new uses for the images as they became more familiar with the

collection. For this reason, we decided that the best course of action would be to build a prototype system based on some preliminary guesses about how the system might be used, and then to engage in an iterative design process to refine the system as we get more data on actual usage.

For the initial prototype design, we relied on the Librarian's knowledge of the sorts of tasks likely to be undertaken by faculty and students in the School with reference to the slide collections. We envisioned faculty as wanting access to slide sets for use in the classroom, or wanting access to individual items for use as illustrations in other types of presentations, for reference in their scholarly research, or for other purposes. We envisioned students as wanting to gain access to individual slides they might have seen in classroom presentations, or for use in their own presentations and classroom assignments, or as reference in their own research. In addition, because some of the School's students are pursuing concentrations in visual communications, we could foresee that they might want to refer to the images in the collections as a source of ideas for creating their own images, especially for those students concentrating their studies in the field of advertising.

Given these preliminary assumptions about how faculty and students might use the images in our collection, we wanted to provide as many access points as possible to the collection to facilitate as many of these disparate needs as possible. The prototype system we built is designed to be just that – a

prototype that is meant to be redesigned iteratively as we get more feedback from the user population about how they are likely to use the images contained within the system. We believe strongly in the principle of iterative design centered around the needs of the user, as has been consistently advocated in the human-computer interaction literature (see, for example, Adler, 1992; or Norman & Draper, 1986). However, we have also found from our own anecdotal experience that end users have an easier time getting involved in the design process *when there is a prototype that they can see*. Users are not always good at articulating what kinds of information needs they have, what their information-seeking behaviors are, or what kinds of things they are looking for in an interface to a retrieval system. If you show them a prototype, however, they can often speak more intelligently to what they like or dislike about the prototype, and this in turn can spur them to be able to articulate what features they would like to see in later iterations. On the flip side, however, there is always the danger that the user will not be able to see beyond whatever mental model informed the design of the original prototype. In other words, if it so happens that the initial model for the prototype is really not appropriate for the kinds of tasks being performed, the user may have difficulty articulating that the model itself is faulty, as opposed to the details of how that model is implemented. Realizing that this was a danger, we attempted to make our first-pass prototype as simple as possible so that we could use it as an aid in studying how JoMC faculty and student users interact with image data. This in turn will inform the further development and refinement of the system.

Based on the ways we initially envisioned the JoMC faculty and students using the slide collections, we decided that the initial prototype should contain three access methodologies: browsing the slides by collection; browsing or searching by assigned keywords; and searching through the full text of the descriptions provided with each individual slide. We had initially hoped to implement a fourth kind of access that would allow the user to query the slides by image content, but we did not have the resources either in terms of money or staffing to be able to accomplish this for the initial prototype system.

Once we had settled upon the functionality we wanted to provide, we needed to assess the resource constraints we faced. We did not have access to any funding for additional hardware or software, so we had to work within the constraints of hardware and software already owned by the Library. In addition, the School of Journalism and Mass Communication does not have its own web server, so we would either have to serve the slide database pages from another unit within the University or devise a way that these pages could be served over the School's Local Area Network. Finally, since we had no funding for temporary help to scan the slides and perform data entry, we were limited to the number of slides that one graduate student could process while simultaneously building the database system.

We decided early on that the image metadata should be stored in a relational database that could be accessed via ODBC drivers over the World

Wide Web. Our first thought was to use the MySQL relational database system installed on Metalab at UNC-CH, which is the site that hosts the Library's main web page. However, we ran into complications that caused us to abandon this platform. The MySQL database engine at Metalab resides on a different physical machine than the Metalab Apache Web server, which made interacting with the database via the Web exceedingly difficult. In order to use Java, we would have needed to use Remote Method Invocation, an area of Java with which the author is not at all familiar and which would have required months to learn to use. In fact, we also had a great deal of difficulty using PHP with the database because of this physical setup, and at the time of this initial exploration (Fall 1998), there was not enough documentation available about how to use PHP across more than one physical machine to make building a system with PHP feasible.

So finally we decided to do our development and prototyping work using Microsoft Access as the database back end, Personal Web Server on one of the library workstations accessible to the JoMC LAN as the web server, and Microsoft Active Server Pages (ASP) as the middleware. This decision had both advantages and disadvantages. The advantages were the author's prior experience with Access, Personal Web Server, and ASP; control of the hardware on which the collection was to reside; and the ease of use of ASP as a tool relative to other middleware solutions like Java and PHP. The primary disadvantage was speed. Personal Web Server is not designed to be a heavy duty web server; nor is the workstation on which the database now resides

designed for high bandwidth, graphics-intensive web-serving. As it stands now, the response time for queries against the database can be quite slow, owing in large measure to the amount of time it takes to download the thumbnail images from the database server machine. We will be looking for a more permanent and robust home for future versions of the slide database.

Once we had decided on a platform, we built a very simple relational database consisting of two tables: one with the slide filename, the title of the collection from which it came, and the text description of the slide; and one with the slide filename and the keywords associated with that slide. We did not attempt to store the images themselves in the database, but rather stored the filenames as references to them. The ASP scripts we wrote for the slide database web pages use the filenames to pull up thumbnails of the images for retrieval sets, so there was no need to consume the extra overhead of trying to store the image data within the database itself. Data entry forms for both the slide and keyword tables were created to make the data entry proceed a little more quickly and smoothly.

Once the database was set up, we set about the task of scanning the slides and entering data into the database. This turned out to be extremely time-consuming and took the better part of eight months of part-time work. In all, we scanned 695 slides from 13 collections, which were associated with 4,369 keyword entries for an average of six keywords per slide (see Table 1). Each

slide took between 15 and 25 minutes to process, including scanning, image editing, converting to JPEG format, assigning keywords, and entering the appropriate data into the database.

<b>Collection Title</b>	<b>Description</b>	<b>Number of Slides</b>
Advertising Themes and Slogans	The use of themes and slogans in advertising	76
Alternative Press	Examples of non-mainstream newspapers and magazines	35
Art in Advertising	The use of fine art in advertising	68
Black Media	History of African-American media	36
Foreign Correspondents Around the World	A look at what it's like to be a foreign correspondent	60
History in Front Pages	U.S. history as revealed in newspaper front pages	55
Media Technology: From Gutenberg to Videotex	Innovations in media technology from the printing press to Videotex	70
People in Journalism	Important people in the media world	50
Pictures of the Past	Photography in the Civil War era	60
Revolution and Newspapers, 1759-1789	The role of newspapers in the American Revolution	53
Updates I	Assorted slides to be used as updates for the collections above	36
Updates II	More updated slides	36
Women in Film	The image of women in film	71

**Table 1**

The keywords that were assigned to each slide were taken from the list of subject keywords that the JoMC Library uses for its internal cataloging system. A total of 1,016 unique keywords were assigned to the slides in the database (see Appendix A). This proprietary subject heading system grew out of the Librarian's frustration with the inadequacy of the Library of Congress Subject Headings for cataloging materials in journalism and mass communication. We felt that it would

be advantageous to utilize a keyword system with which the user population already has some familiarity.

Once all of the data had been entered into the database and spot-checked for correctness, we set about the task of constructing the web pages and ASP scripts that would be used to access data from the database. As stated before, the three access methodologies we wanted to offer were browsing by slide collection, browsing/searching by keyword, and searching the full text of the slide descriptions (See Appendix B for screenshots of the browsing, search keyword, and search full text interfaces). Implementing this in ASP simply required building HTML forms that would then be preprocessed by ASP, sending SQL queries to the Access database and returning results to be formatted using HTML.

The keyword searching function consists of three drop-down menus containing all the keywords that have been assigned to slides in the database in alphabetical order. In order to form a query, a user simply chooses a keyword from the list. This eliminates one source of uncertainty in searching, by allowing users to recognize a keyword that might be appropriate for their search rather than having to guess. This follows the principles of relying on recognition rather than recall and eliminating as much keying as possible (Nielsen, 1994).

For this prototype version, the full text search can search only for one word or phrase from within the description text; future enhancements will contain more sophisticated searching options for searching the full text slide descriptions.

When a user runs a search on the database, a table of results is returned that contains a thumbnail image of the slide, the title of the collection it is in, and a textual description of what is depicted in the slide. Keeping all the information about an image together with the image is an important point in image retrieval system design (Van House, Butler, Ogle, and Schiff, 1996).

### **Chapter 3: Study Design**

Having built a prototype slide database system, we wanted to do some user testing to try and determine whether our subject indexing of the slides was adequate and to get some idea of how our user population was likely to formulate information needs for image data. We were also interested to see if subject expertise would have an impact on the ways in which the users formulated their information needs and went about the information-seeking process. This in turn might impact the ways in which segments of our user population would want to search through the slide collections, and the kinds of online help, tutorials, or examples we would provide, so we felt it was an appropriate avenue of inquiry.

We decided to design a pilot study to obtain qualitative data from the subjects about the ways they formulated queries on the image database, to see how they would formulate queries when looking for image data as opposed to documents and whether there were any differences in query formulation between different segments of the user population. We invited all journalism and mass communication faculty (approximately 40 faculty members) and sophomores and juniors who had declared themselves to be journalism and mass communication members and who were attending summer school at UNC-Chapel Hill to participate (approximately 70 students). Five faculty members and

five students agreed to participate in the study. We created three tasks for them to run on the database, and then asked them to do one search of their own based on some topic of interest to them. The three tasks we assigned them were as follows:

1. Find a headline about the end of World War I.
2. Find a cigarette advertisement.
3. Find a Mathew Brady portrait of Abraham Lincoln.

For the first and third task, there was only one slide in the database that would satisfy the query. There were seven slides that would satisfy the second task. We structured the tasks so that the answer could be found using any of the three access methods available, although an expert searcher would be most likely to use keywords (“World War I” or “headlines” or “front pages” for the first task, “cigarettes” or “advertising” for the second, and “Mathew Brady” or “Abraham Lincoln” or “portraits” or “photography” for the third (assuming that the expert would realize that Mathew Brady was a Civil War era photographer).

We tried to design these tasks keeping in mind that even within journalism and mass communication there are widely varying specialties, such as advertising, public relations, visual communications, newspaper reporting, etc., and tried to be as representative of these different specialties as we could with a limited number of tasks.

We asked them to think aloud during the tasks so we could get some idea of the kinds of thought processes that were going on as they were formulating their queries and conducting their searches. We were especially interested to see what kinds of queries the subjects would come up with on their own, as this would give us a better picture on the kinds of information needs users might bring to the system later on. Some typical examples of user-defined queries are the following:

- Show me items about the Titanic disaster. (2 users)
- I want to see if there are any pictures of Madonna.
- Is there anything else about the Kennedy assassination? (This query came about because the image on the database home page is of the front page the day Jack Ruby shot Lee Harvey Oswald.)

Once the subjects finished with the four tasks, they were asked to complete a short questionnaire that gathered some demographic information as well as information about their expertise with other online retrieval systems (see Appendix C). In addition, they were asked their opinion about the search options provided. The whole process took about 30 minutes for each subject.

## **Chapter 4: Results, Conclusions, and Areas for Further Study**

Because this was a small pilot study done early in the development process of our slide database system, we did not expect to get conclusive results, but rather some general ideas for formulating a larger scale, more formal user study in the future, as well as ideas for enhancements and improvements to the prototype system.

The participants in the study were evenly divided between faculty and students. Of the five faculty, three were male and two were female. All of the faculty owned IBM-compatible PCs. There was some variance in their level of computer sophistication; two characterized themselves as unsophisticated, two as sophisticated, and one as neither. Similarly, three faculty participants described themselves as sophisticated searchers, while two characterized themselves as unsophisticated searchers. None of the faculty reported having used either Lexis/Nexis or Dialog, but all of them reported having had at least six months' experience with online library catalogs or web search engines. Only one faculty user reported having had over 24 months' experience using web search engines.

The faculty participants reported being satisfied with the results they got from their queries. However, there were cases where they received results (or retrieved empty sets) that were unexpected, especially in full text search mode. The faculty were split between keyword search and full text search as far as which access method they liked best (three opted for keyword search and two chose full text search), as well as which they liked least (two chose browse by collection, two chose keyword search, and one chose full text search). For their own searches, two faculty participants chose browse by collection, and three chose to search by keywords. There was only one instance where a task was not successfully completed by a faculty participant. In this case, the faculty member was using the full text search for Task #3.

Of the student participants, two were seniors, and three were juniors. They were all Journalism and Mass Communication majors, and there were four females and one male. They all owned computers, with three owning IBM-PCs and two owning Macintoshes. They all considered themselves to be sophisticated computer users, and most characterized themselves as sophisticated searchers (one felt she was neither sophisticated or unsophisticated). All of them had used Lexis/Nexis for at least six months (the seniors for over 24 months), none had used Dialog, and all had used online library catalogs and web search engines for 24 months or more. Again, there was one instance where a user could not successfully complete a query task. The student participant in this case was using the full text search for Task #1.

All of the student participants characterized themselves as very satisfied with the results, although some students also reported that they retrieved results they did not expect. Most preferred keyword search (four students); only one preferred full text search, and none preferred browsing the collections. The four students who liked the keyword search best liked the full text search least, and the one who liked the full text search best disliked the keyword search most. All of the students used the access method they stated as their favorite for their free query.

One finding that was surprising was the lack of enthusiasm for the browse by collections feature. One faculty member noted that it was “too much ‘wasted’ effort”, and none of the participants could use it as efficiently for answering a query as they could the other methodologies. The tasks were not timed, but we noted that the browse collection tasks took, in general, about twice as long as the other two, no matter which query was being used with the browse collections feature.

This finding is somewhat surprising given previous studies that have found browsing to be beneficial for users in academic disciplines because of the often non-linear and intuitive routes they take to arriving at the information they need (Marchionini, 1995b; Miller & Tegler, 1986). The reasons for this disinclination toward browsing should be investigated in further studies. It could perhaps be related to the time it takes for all of the thumbnails for a particular collection to be

downloaded for viewing on the user's machine. For a 70-slide collection, the download time can be anywhere from 45 seconds to a minute. The download time for the results of keyword and full-text searches can be almost as long, but perhaps there is an expectation for these kinds of searches to take time, while there is an expectation for browsing to be instantaneous, relatively speaking. This point should certainly be investigated in further studies.

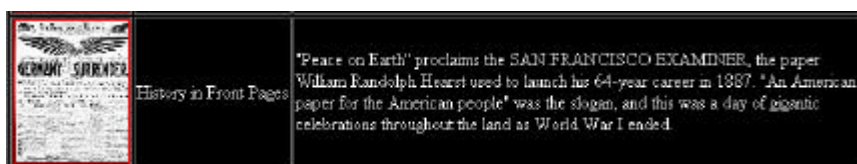
Another possibility is that the browse by collection methodology has its place in the information-seeking process, but that it is not well suited to the kinds of structured query tasks that we gave the study participants. One clue that this might be the case came from a faculty member who chose to use the browse by collections feature for her free query, despite it not being her favorite searching mode, because she was not sure what kinds of things were contained in the database, and wanted to browse through the collections to get an idea of what was there. This suggests that, despite the lack of enthusiastic endorsement of the browsing feature, it may still have a place in the information-seeking process, especially when a user is unfamiliar with what is available in the system.

Another interesting finding is that the participants did not seem to fully grasp the difference between keyword searching and full-text searching. They seemed much more comfortable with the keyword searching, and were much more successful with it than with the full-text searching (most had to reformulate their initial full-text query at least once, while only two had to reformulate the

initial keyword query). From what the participants said as they were formulating their full-text queries, it was clear that they were still thinking in terms of keywords. For instance, one faculty member who was using the full-text search for Task #2 made as his first attempt “advertising, cigarette.” This kind of a query makes sense in a keyword search, but not in a full text search unless for some reason you believe that there will be a sentence that contains that exact phrase. Similarly, a student participant tried to use “Lincoln and Brady” in the full text search for Task #3, while the faculty member who was unsuccessful in completing Task #3 used “Lincoln, Abraham” as the initial search. Either of these would have been an excellent strategy for the keyword search, but each retrieved no results from the full text search using these phrases. One difficulty is that the participants were not aware that the full text search can match only exact words or phrases at this point. We emphasized this to each user verbally in the introduction to the system, but we must do more to call the user’s attention to this constraint. This is something that needs to be pointed out to the user at the top of the full text search screen, and later versions should give the user the ability to search the full text in a more natural way by allowing for Boolean operators and multiple search terms.

Our observation of the subjects’ interaction with the result sets obtained from their queries corroborated Keister’s (1994) finding of a “dynamic” between the textual metadata and the image surrogates for selecting images from the result sets (p. 17), although we did find some difference between the faculty and

student subjects in this regard. Keister found that the user would gather a selection of images by searching the text in the catalog description of the image, then review the images until finding one that “worked” visually, and then re-verify the selection based on the textual metadata associated with the image. We detected the same phenomenon with our student subjects, but interestingly our faculty subjects employed a slight variation on this theme. The student subjects would submit an initial query (usually one word), and after the retrieval set had loaded, they would scan quickly through the thumbnail images to find likely candidates. Only after identifying a candidate slide based on the thumbnail would they scan the text description to confirm their choice. If there was no clear candidate based on the initial review of the thumbnail images, then they would go back and scan the text descriptions looking for keywords or appropriate dates. They would not read the descriptions line by line except as a last resort, and this in fact led one student subject astray on Task #1. The text description for the slide that answers this query (see Figure 2 below) references the background of the newspaper in which the headline appears, which was started in 1887. Since the student was scanning dates in the descriptions, she saw 1887 and moved on,



**Figure 2**

because she believed (rightly) that 1887 was the wrong year for the end of World War I. After trying a number of different searches, she came back to her first

result set (from the keyword entry “World War I”) and realized what she had done.

Faculty subjects, on the other hand, relied more strongly on the text than did the students. They would submit an initial query, scan through the text descriptions first to select a likely candidate, and then review the thumbnail to verify their choice. This finding verifies the importance of displaying both a thumbnail image and the text description in the initial retrieved set. It also, in Keister’s words, “validates the suspicion that still image research is a different critter entirely from standard search systems, which essentially retrieve words with words” (p. 17). In addition, the participants’ preference to scan rather than read through the text descriptions suggests that it might be helpful to highlight the search terms within the description text (when they occur) so that the user can go right to the part of the description that is relevant to their query.

Completing the pilot study pointed up a number of areas that need further study. First of all, the pilot study itself suffered from the fact that the slide database resource is essentially new so far as the user population is concerned. They have never had good access to the slide materials before, and perhaps not even to other kinds of image data in electronic format, and so there was very little sense of how these materials would actually be used by the faculty and students as part of their scholarly work. It would be helpful to have more naturalistic information about how the users integrate the slide collections – if at all – into

their scholarly lives. This will require increasing awareness on the part of students and faculty that these resources exist and capturing data about who is accessing them, what kinds of queries they are conducting, and how they are using the materials after they have found them. The presence of this collection on the JoMC Library web page may help foster this awareness.

A concern that arose during the pilot study was that providing the subjects with a verbal description of the query task they were supposed to perform was in some sense “stacking the deck” in terms of their retrieval success. Indeed, all of the subjects were successful at finding answers to all of the queries, and most of them used words in their searches that came directly from the task itself. Would the results have been different had we shown them an image of the slide they were to retrieve? Would the novice students then have been at more of a disadvantage using the domain-specific keywords because they could not identify the proper domain terms to use when searching for the image? Even more importantly, when the users of the slide database have an information need for images in our collection, will they manifest themselves as verbal descriptions first or as “mind-pictures” first? If they first experience an information need in terms of a “mind-picture”, how do we get at this mind-picture? How can it be used to improve retrieval performance? These are questions that we hope to explore in further studies.

Another area that should be explored is the information-seeking behavior of the user population with respect to text materials. This would serve as a necessary baseline from which to determine if there are differences in the way information needs are expressed for image data as opposed to textual data, and may provide greater insight into the kinds of textual metadata that should be provided for such images. This would also provide a basis for building a more integrated digital library from the existing materials on the Library's web page, so that eventually a single query could be used to retrieve both text and image data from the Library's electronic collections.

On the user interface side, the most important future enhancement is to obtain a home for the collection such that speed will not be so much of an issue. It is well documented that the speed of response time (or lack thereof) contributes a great deal to the user's perception of the usability of the system, so that is an issue that needs to be resolved as quickly as possible. Once that is accomplished, it would be possible to test more innovative search options, such as the query by image content facilities that are being developed by researchers now.

If, after moving the database, we find that the access times for results pages are still slow, we will implement the system in such a way that the set of retrieved documents is broken up into smaller sets per page. We had initially decided not to break up the retrieved sets in order to keep the user from having

to page through the results, but the speed advantage may easily outweigh the disadvantage of having to page, especially given Nielsen's (1994) emphasis on giving the user the opportunity to quickly determine whether the results she is getting from a particular search are adequate without having to wait for the entire retrieved set to load. In addition, we discovered in the course of the user study that the faculty computers were considerably less powerful than we had anticipated, so breaking the result sets into smaller pieces would better adhere to the principle of matching the retrieval system to the user population's technology (Van House, Butler, Ogle, & Schiff, 1996).

As the slide database grows and becomes more tightly integrated with the other Web-accessible resources the Library offers, we may want to design and implement a user interface that is more configurable and allows the user to customize the user interface to serve her needs and to accommodate her cognitive style. Payette and Rieger (1998) make the point that this will begin to be an expectation that the user brings to using a digital library due to the integration of such customization features into word processing, web browsing, and other types of software that are widely used.

Finally, on a larger scale, it is perhaps time that we took a closer look at how non-textual data is actually used within scholarly disciplines. Scholarship being a primarily textual (or at least verbal) pursuit, it is fairly easy to see how text materials are used in academic work, but as yet we do not have as good a

grasp at how other, non-textual, multimedia materials are being used in the scholarly community. With the rapid rise in the number of academically oriented digital libraries offering non-textual data, it would seem prudent to have some idea of how such materials are, in fact, likely to be used by the target audience.

## References

- Adler, P., & Winograd, T. (Eds.). (1992). *Usability: Turning Technologies into Tools*. New York: Oxford University Press.
- Borgman, C. L. (1990). All users of information retrieval systems are not created equal: An exploration into individual differences. *Information Processing & Management* 25(3), 237-251.
- Cawkell, A. E. (1998). Checking research progress on “image retrieval by shape-matching” using the Web of Science. *Aslib Proceedings*, 50(2), 27-31.
- Chu, W. W., & Hsu, C.-C. (1998). Knowledge-based image retrieval with spatial and temporal constructs. *IEEE Transactions on Knowledge & Data Engineering* 10, 872-888.
- Fister, B. (1992). The research process of undergraduate students. *Journal of Academic Librarianship* 18, 163-169.
- Heidorn, P. B., & Sandore, B. (Eds.). (1997). *Digital Image Access & Retrieval*. Urbana, IL: University of Illinois at Urbana-Champaign Graduate School of Library & Information Science.
- Keister, L. H. (1994). User types and queries: Impact on image access systems. In R. Fidel, T. B. Hahn, E. M. Rasmussen, & P. J. Smith (Eds.), *Challenges in Indexing Electronic Text and Images* (pp. 7-19). Medford, NJ: Learned Information, Inc.

- Lancaster, F. W. (1995). Are evaluation criteria applied to “traditional” libraries equally applicable to digital libraries. In *How We Do User Centered Design and Evaluation of Digital Libraries: A Methodological Forum, 37<sup>th</sup> Allerton Institute*. Retrieved June 24, 1999 on the World Wide Web:  
<http://edfu.lis.uiuc.edu/allerton/95/s1/lancaster.html>
- Marchionini, G. (1995a). *Information Seeking in Electronic Environments*. London: Cambridge University Press.
- Miller, C., & Tegler, P. (1986). Online searching and the research process. *College & Research Libraries* 47, 370-373.
- Morgan, E. L. (1999). What’s more important: The questions or the answers? *Computers in Libraries*, 19(5), pp. 38+. Retrieved June 24, 1999 from EBSCO database (MasterFILE Premier) on the World Wide Web:  
<http://www.ebsco.com>
- Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability Inspection Methods* (pp. 25-62). New York: Wiley.
- Norman, D.A., & Draper, S.W. (1986). *User Centered System Design*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Palmer, J. (1991). Scientists and information: II. Personal factors in information behaviour. *Journal of Documentation* 47, 254-275.
- 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), pp. 121+. Retrieved June 24, 1999 from EBSCO database (MasterFILE Premier) on the World Wide Web: <http://www.ebsco.com>

- Pei, S.-C., & Cheng, C.-M. (1999). Extracting color features and dynamic matching for image data-base retrieval. *IEEE Transactions on Circuits & Systems for Video Technology*, 9, 501-512.
- UCLA-NSF Social Aspects of Digital Libraries Workshop. (1996). Final report to the National Science Foundation. Retrieved June 28, 1999 from the World Wide Web: [http://www-lis.gseis.ucla.edu/DL/UCLA\\_DL\\_Report.html](http://www-lis.gseis.ucla.edu/DL/UCLA_DL_Report.html)
- Van House, N. A., Butler, M. H., Ogle, V., Schiff, L. (1996, February). User-centered iterative design for digital libraries: The Cypress experience. *D-Lib Magazine*. Retrieved June 24, 1999 from <http://www.dlib.org/dlib/february96/02vanhouse.html>
- Virage, Inc. (1999). *Visual information retrieval (VIR) image engine*. Retrieved June 28, 1999 from the World Wide Web: [http://www.virage.com/products/image\\_vir.html](http://www.virage.com/products/image_vir.html)
- Wilson, T. (1994). Information needs and uses: Fifty years of progress? In B.C. Vickery (Ed.), *Fifty Years of Information Progress* (pp. 15-51). London: Aslib.
- Wu, J. K. , Narasimhalu, A., & Desai, A. (1998). Fuzzy content-based retrieval in image databases. *Information Processing & Management*, 34, 513-534.

## Appendix A: Unique JoMC Keywords

18th Century	Atlanta
19th Century	Atlanta Constitution
20th Century	Atlanta Gazette
Abell, A.S.	automatic cylinder press
abolition	Averill, John
Adams, John	Avirgan, Tony
Adams, Samuel	Avril, Jane
Addams, Charles	Ayatolleh Khomeini
advertising	Bacall, Lauren
advertising layout	Backstreet
Advocate	Bahrain
Afghanistan	Baker, Ray Stannard
African-Americans	balloons
album cards	Baltimore Afro-American
Alien	Baltimore Sun
All Alaska Weekly	Bancroft, Anne
Allner	Bara, Theda
Allyson, June	Bardot, Brigitte
Alta California	baseball
alternative press	Battle of Gettysburg
American Broadcasting Company (ABC)	beer
American Crisis, The	bees
American Tobacco Company	Begin, Menachem
anchors	Beirut
Andersonville	Bendana, Alejandro
Andrews, Dana	Bennett, James Gordon, Jr.
Andrews, Julie	Bennett, James Gordon, Sr.
Andropov, Yuri	Bentsen, Lloyd
animals	Bergen, Candice
Antietam	Bernhardt, Sarah
Aquino, Corazon	Bernstein, Carl
Arabian Gulf	Best Years of Our Lives, The
Arafat, Yassir	bibliographies
Archduke Ferdinand	billboards
Arias, Oscar	birth control
Arm & Hammer	Birth of a Nation
Armenia	black media
Arnett, Peter	Black Media, Inc.
Arnold, Benedict	Black Monday
art	Blair, Frank
Art Nouveau	blaxploitation
Arthur Shoes	Blondell, Joan
Arthur, Jean	
artist correspondents	
Artzybashef, Boris	
Asbestos Limited	
Asia	
assassinations	
Associated Press	
associations	
Astaire, Fred	
Astor, John Jacob	
athletes	

Blue Angel, The	censorship
Bly, Nellie	Central America
Blyth, Ann	cereal
Bob and Carol and Ted and Alice	Challenger
Bon Ami	Chancellorsville
Bonhomme Richard	Chappelle, Dickey
Bonnie and Clyde	Cheret, Jules
book publishing	Chernenko
Booth, John Wilkes	cherubs
Borax	Chesterfield
Boston Gazette	Chicago Daily News
Boston Globe	Chicago Defender
Boston Massacre	Chicago Sun-Times
Boston News-Letter	Chicago Tribune
Boston Phoenix	chicks
Boudreaux, Richard	child-woman
Bow, Clara	children
Bowles, Samuel	China
Boys in the Band	China Reconstructs
Brady, Mathew	Christian Science Monitor
Brando, Marlon	cigarettes
Brassi, Rosanno	civil rights
Brewer, Edward	Civil War
Brinkley, Joel	Clairol
broadcast journalism	Claypoole, David
broadcasting	cleaners
broadsides	Cleopatra Jones
Brooklyn Dodgers	Cleveland Press
Broun, Heywood	Cleveland, Grover
Browning, Michael	clocks
Budweiser	clothing
Buffalo News	Clymer Press
Buick	Coca-Cola
bulls	Colbert, Claudette
Burger, Warren	Cold War
Bush, George	colonial newspapers
Butt, Archibald	color
bylines	Columbia Broadcasting System
cable television	(CBS)
Cadillac	Columbus Citizen-Journal
calendars	column rules
Camel	columnists
cameras	Common Sense
Camp David Accords	competition
Campbell Soups	composing room
careers	composing table
carpets	computer editing
Carrier Air Conditioners	computer graphics
Carroll, Diahann	Computer User
cars	computerized layout
Carter's Ink	computers
Carter, Jimmy	Connecticut Courant
cartes de visite	Constitution
cartoons	Constitutional Convention
Cartright, Marguerite	Container Corporation of America
Cat on a Hot Tin Roof	Continental Journal
cats	Contras

Cooper, Gary	double supplement stereotype
Cooper, James Fenimore	perfecting press
Copley Newspapers	Douglas Shoes
Copley, Helen	Douglas, Stephen A.
copy	Douglass, Frederick
copyright	Dow Jones Industrial Average
Cornish, Samuel	Downs, Hugh
Cosmic Revolutionist	drawings
Costa Rica	Drew, Elizabeth
Courier Communications	Drummond
Corporation	Dubai
Crawford, Joan	Dufy, Raoul
Cream of Wheat	Dukakis, Michael
Creelman, James	Dumont, Margaret
Cronkite, Walter	Dunaway, Faye
Crown Prince Hassan	Dunlap, John
Cry Havoc	East Village Eye
Cuba	Ebony
Cuban missile crisis	Eddy, Nelson
Curtius, Mary	Edinburgh Courant
Czolgosz, Leon	editing
D'Escoto, Miguel	editorials
daguerreotypes	editors
Dali, Salvador	Egypt
Dallas Times Herald	Eisenhower, Dwight
dames	El Pais
Dana, Charles	elections
dance hall girls	Eli Lilly
Davis, Bette	Emporia Gazette
Davis, Richard Harding	English Bill of Rights
Dawn Magazine	English, Greg
Day, Benjamin	engravings
Day, Doris	entertainment
Dean, James	Equal Times
debates	Esplanade
DeBeers	Essence
Declaration of Independence	Essex Gazette
Democratic Party	Essex Journal
demographics	everywoman
Dempster, Carol	exotica
Destry Rides Again	features
Detroit News	Feller, Bob
diamonds	feminism
Diario de las Americas	femme fatale
Diary of a Mad Housewife	Fenno, John
Dickinson, John	Ferraro, Geraldine
Dietrich, Marlene	film and television
Digital	film noir
disasters	fingers
Discovery	fire
Disney hats	fireside chat
Dobson, Tamara	Fisher, Dan
dogs	Fisk
Dole, Robert	flappers
dolls	Fleet Street
Double Indemnity	flour
double sextuple press	Fonda, Jane

Fontaine, Joan	graphics
Ford	Great Depression
foreign correspondents	Great Northern Railroad
four-color process	Great Wall of China
France	Greeley, Horace
Franklin, Ben	Green Dragon Tavern
Franklin, James	Green Giant
Fredricksburg	Grenada
Free Venice Beachhead	Griffith, DW
Freedom's Journal	Grit
freelance journalists	Group, The
Friedman, Thomas	Guardian
frogs	Guisti, George
front pages	Gulf War
frontier printing	gun molls
Funicello, Annette	Gutenberg Bible
galleys	Gutenberg printing press
gangster films	Gutenberg, Johannes
Garbo, Greta	Haberman, Clyde
Gardner, Alexander	halftone
Garfield, John	Halloween
Garland, Judy	Hamm's
Garson, Greer	hand press
Gay Community News	hands
Gaza Strip	Harlow, Jean
Gazette Francoise	Harper's Weekly
Gazette of the United States	Hart, Gary
General Electric	Harvest of Death
General Mills	Harvey Girls, The
Getting Straight	Hayward, Susan
Gettysburg	headlines
Geyer, Georgie Ann	health
Gibbons, Floyd	Hearst, William Randolph
girl-next-door	Hellman, Lillian
Gish, Lillian	Henry, Patrick
Gismonda	Henry, Ragan
Glen Miller Story, The	Hepburn, Audrey
Glenn, John	Hepburn, Katherine
Goddard, William	Higgins, Marguerite
Godfather, The	Hill, George Washington
Godkin, E.L.	Hinckley, John
Gold Coast Free Press	Hindenburg
Gold Medal	Hindus, Maurice
gold-diggers	Hirohito
Goldstein, Al	His Girl Friday
Goldwater, Barry	Hispanics
Good-Bye Columbus	Hitchcock, Alfred
Gorbachev, Mikhail	Hitler, Adolf
Gottlob, T.	Hoe's cylinder press
Gould, Elliott	Holt, John
Grable, Betty	homosexuality
Graduate, The	Honduras
Grady, Henry	Honey, Martha
Graham, Katharine	Honeywell
Grant, Cary	Hong Kong
Grant, Ulysses S.	horses
graphic design	House and Garden

Houston Chronicle	Kilgallen, Dorothy
How to Marry a Millionaire	Killing of Sister George, The
I Was a Male War Bride	Kim Dae Jung
illustrations	King Hussein
images	Kinzer, Stephen
Impact!	Klute
In These Times	Kodak
Indians	Korea
International Herald Tribune	Korean War
Intifada	Kuwait
investigative reporting	Langtry, Lily
Iran	Lasky, Jessie
Iran Hostage Crisis	Last Tango in Paris
Iran-Contra Scandal	layout
Iraq	Le Monde
Israel	Le Style Mucha
It Girl, the	Lebanon
It Happened One Night	Lee's Carpets
Ivory Soap	Lee, Robert E.
Iwo Jima	legs
J&P Coats	LeMoyné, James
Jackson, Andrew	Lescoulie, Jack
Jackson, Bob	Leslie's
Jackson, Jesse	Letter from a Pennsylvania
Jackson, Maggie	Farmer
Jackson, Stonewall	Lexington and Concord
Jansen, Dan	Lincoln, Abraham
Japan	Lindbergh, Charles
Jaws	linotype
jazz	listening posts
jazz journalism	Listerine
Jeffries, Jim	lithography
Jell-O	logos
Jennison, Edwin	London
Jerusalem	London Gazette
Jet	London, Jack
Job	Lord Calvert
John Deere	Los Angeles Herald
Johnson Publications	Los Angeles Herald Examiner
Johnson, Jack	Los Angeles Herald-Examiner
Johnson, John H.	Los Angeles Times
Johnson, Lyndon	Louisville Courier-Journal
Johnson, Van	Louisville Times
Jones, John Paul	Lovie, Henri
Jordan	Loy, Myrna
Journal of Occurrences	Lucky Strike
journalism history	Ludlow, George
Julia	Lustig, Ray
Kaltenborn, HV	MacArthur, Douglas
Kaufman	MacDaniel, Hattie
Kaufman, Marc	Macfadden, Bernarr
Kennedy, Edward	MacIntosh, Craig
Kennedy, John F.	Mackin, Cathering
Kentucky Fried Chicken	MacLaine, Shirley
Kerner Commission	Madison, James
Khrushchev, Nikita	magazines
KIIS	Magna Carta

Mahmoud, Aly  
 Malvern Hill  
 Man Who Loved Cat Dancing, The  
 man's film  
 Mao Zedong  
 maps  
 Marburg Brothers  
 Marcos, Ferdinand  
 Mars  
 Massachusetts Spy  
 Mata Hari  
 McCarthy, Francis  
 McCormick, Anne O'Hare  
 McCormick, Robert R.  
 McDonald, Jeanette  
 McGraw, Ali  
 McKinley, William  
 McMurray, Fred  
 McPherson, James  
 media advocacy  
 media criticism  
 media technology  
 Medill, Joseph  
 Meese, Edwin  
 men  
 Mergenthaler, Ottmar  
 Merrill Lynch  
 Miami Herald  
 Middle East  
 Midnight Cowboy  
 Mildred Pierce  
 Miles, Sarah  
 Millais, John Everett  
 Miller Beer  
 Milwaukee Journal  
 Milwaukee Sentinel  
 Minneapolis Star  
 Minneapolis Star and Tribune  
 Minneapolis Star Tribune  
 Minneapolis Tribune  
 minorities  
 Moffett, George D., Jr.  
 Mondale, Walter  
 Monroe, Marilyn  
 Morning Chronicle  
 Morrison, Herbert  
 Morton's Salt  
 Mother is a Freshman  
 Moulin Rouge  
 mouthwash  
 moveable type  
 Mr. Deeds Goes to Town  
 Mr. Smith Goes to Washington  
 Mrs. Minniver  
 Mucha, Alphonse  
 muckrakers  
 mules  
 Murphy, Carl J.  
 Murphy, John H.  
 Murrow, Edward R.  
 musicals  
 Mussolini, Benito  
 My Fair Lady  
 Nation  
 National Black Media Coalition  
 National Black Monitor  
 National Broadcasting Company  
 (NBC)  
 National Newspaper Publishers  
 Association  
 National Observer  
 National Recovery Act  
 National Security Council  
 Nazi Germany  
 NBC Today Show  
 Neighborhood  
 New Deal  
 New Haven Railroad  
 New Journalism  
 New Mexico Independent  
 New York  
 New York American  
 New York Daily News  
 New York Evening Graphic  
 New York Gazetteer  
 New York Herald  
 New York Herald Tribune  
 New York Herald-Tribune  
 New York Journal  
 New York News  
 New York Post  
 New York Sun  
 New York Times  
 New York Tribune  
 New York World  
 New York World-Telegram and Sun  
 New York Yankees  
 New Yorker  
 New-England Chronicle  
 New-England Courant  
 New-York Gazette  
 Newman, Paul  
 news  
 news conferences  
 Newsday  
 newspaper design  
 Newspaper Guild  
 newspaper layout  
 newspapers  
 newsracks  
 newsreels  
 Nicaragua  
 Nixon, Richard  
 Norfolk Virginian-Pilot

North, Oliver	political conventions
Northern Pacific Railroad	political reporting
Northwest Airlines	politicians
O'Sullivan, Timothy	politics
Ochs, Adolph	Polk, James K.
offset printing	Pond's
Old Dutch	Portland Oregonian
Olympics	portraits
Oomph Girl, the	post-feminism
opinion pages	posters
Oppenheimer, Andreas	Postman Always Rings Twice, The
Ortega, Daniel	power of the press
Oswald, Lee Harvey	Powers, Francis Gary
Output	presidential campaigns
Packard	presidential debates
Paine, Thomas	presidents, US
Palestine	press and politics
Palestine Liberation	press conferences
Organization (PLO)	press pool
Palmerola	press room
Pan Am 103	Princess Diana
Paris Herald	print shop
Parrish, Maxfield	printing
party girls	printing plates
Pathe Newsreel	printing press
Pathe Newsreels	producers
Payne, Lewis	propaganda
Pear's Soap	prostitution
Pearl Harbor	protests
Pears Soap	Providence Journal
Pennsylvania Chronicle	Public Enemy, The
Pennsylvania Gazette	publishers
Pennsylvania Journal	Pulitzer, Joseph
Pennsylvania Magazine	Pyle, Ernie
Pennsylvania Packet	Quaker Oats
Penny Press	Quayle, Dan
pharmaceuticals	Quinn, Sally
Philadelphia Inquirer	racism
Philip Morris	radio
Philippines	Radio Free Jazz
photo opportunities	radio journalism
photocomposition	Ramirez, Sergio
photoengraving	Rand, Paul
photographers	rape
photographic studios	RCA
photographs	RCA Victor
photography	Reagan, Ronald
photojournalism	Real Paper
Pickford, Mary	Rebel without a Cause
Pillsbury	Regan, Donald
pin-up girls	Rehnquist, William
Pittsburgh Courier	Reid, Helen Rogers
Plains of Abraham	reporters
planes	reporting
platen power press	Republican Party
Poe, Edgar Allan	Reuters
Poindexter, John	Revere, Paul

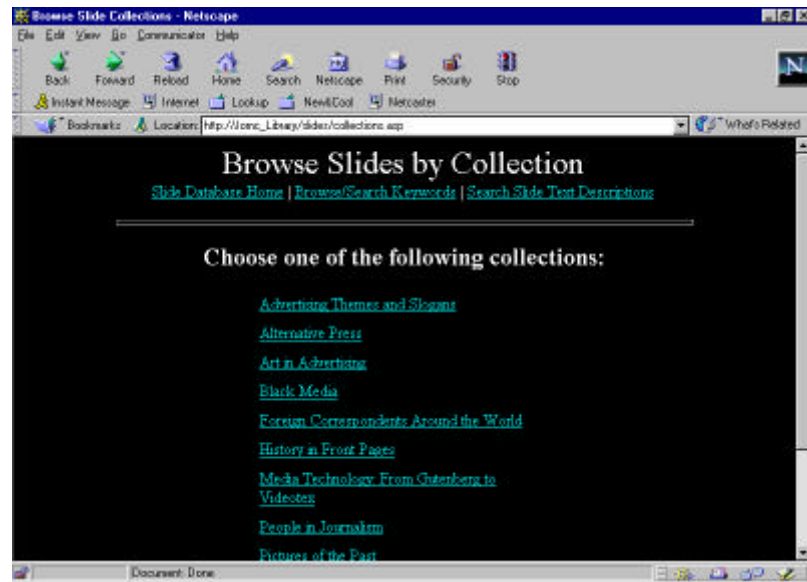
Revolutionary War	shipwrecks
Rhode Island	Shirer, William L.
Rice, Donna	shoes
Richmond	Shop Talk
Richter, Paul	Silence Dogood
Rickard, Tex	Simons, Lewis
Rivington, James	slavery
RJ Donnelly Company	slogans
Rockwell, Norman	slugs
Rocky Mountain Journal	Smith, Merriman
Roderick, John	Snodgrass, Carrie
Rogers, Ginger	Snyder, Ruth
Roh Tae-woo	soap
Roosevelt, Franklin D.	socialism
Roosevelt, Theodore	socialites
Ross, Michael	soft drinks
Rowan, Carl	SoHo News
Royal Gazette	Sojourner
Ruby, Jack	soldiers
Russell, AJ	Sound of Music, The
Russell, Rosalind	sound technicians
Russworm, John	South Korea
Sadat, Anwar	Soviet New Times
Saigon	Soviet Union
Salem, MA	space shuttles
San Francisco Bay Guardian	Spain
San Francisco Chronicle	Spanish-American War
San Francisco earthquake	sports
San Francisco Examiner	sports reporting
San Francisco News	Springfield Republican
San Salvador	Squibb
Sandinistas	squirrels
satellites	SS McClure
satire	St. Louis Globe-Democrat
Scheider, Roy	St. Louis Post-Dispatch
Schlitz	St. Paul Dispatch
Schneider, Maria	St. Paul Pioneer Press
Schultz, George	St. Paul Pioneer Press Dispatch
Scotland	Sta Emot
Scott, Harvey	Stahl, Lesley
Screw	Stalin, Josef
Scripps, Edward Wyllis	Stamp Act
Seagram's	Stanwyck, Barbara
seasons	Star Wars
Seattle Times	States Rights
Sengstacke, John H.	Statue of Liberty
sensationalism	Steffens, Lincoln
Seoul	Steinberg, Saul
sex papers	Steinlen, Theophile
sex symbols	Steppin' Out
sexual revolution	stereo cards
sexuality	stereoscope
Shah	stereotype plates
Shahn, Ben	stereotypes
Shearer, Norma	stereotyping
Shell Oil	Stewart, James
Sheridan, Ann	stock market

stone	US Correspondence Schools
Straus, Isidor	US Supreme Court
street reporting	USA Today
Sturbridge Village	USS Maine
suicide	USS Stark
Sulzberger, Arthur Hays	USS Vincennes
Summertime	vamp
Suro, Roberto	Van Anda, Carr
Suspicion	VDT
Swift Butter	Victrola
Syracuse Post-Standard	videotapes
Syracuse University	Videotext
Syzk, Arthur	Vietnam War
tabloid newspapers	Viewtron
Tammy Tell Me True	Village Voice
Tarbell, Ida M.	violence
Taylor, Elizabeth	Virginia
teenagers	Virginia Slims
telecommunications	Wall Street Journal
telegraph	Walters, Barbara
television	war films
Temple, Shirley	War of the Worlds
terrorism	war photography
Texas Observer	war reporting
Thaw, Harry K.	Warner's Safe Kidney and Liver
Third World	Cure
Thomas, Helen	Washburn Crosby
Thomas, Isaiah	Washington hand press
Times and District of Columbia	Washington Post
Advertiser	Washington Post Weekly
tires	Washington Times
Titanic	Washington, George
Toni	Watergate
Toulouse Lautrec, Henri	Watterson, Henry
Tower Commission	Waud, Alfred
Tower, John	Weaver, Sigourney
Townsend Acts	web-plate process
Tracy, Spencer	weeklies
trade cards	Welch, Raquel
travelogues	Welles, Orson
Treadwell, Daniel	Wells, H.G.
Trout, Robert	West, Mae
Truman, Harry S.	wet-plate process
Turner, Lana	WGPR-TV
twins	what-is-it wagon
Two for the Seesaw	Wheaties
typecasting	Wheatley, Phillis
typesetting	White House correspondents
typography	White, Paul
Tyson, Cicely	White, Stanford
United Kingdom	White, William Allen
United Press International	Wilson, Edward
United States Congress	windows
United States Marines	Woman's Day
United States Military	woman's film
United States Navy	woman-as-neuter
United States Supreme Court	woman-as-victim

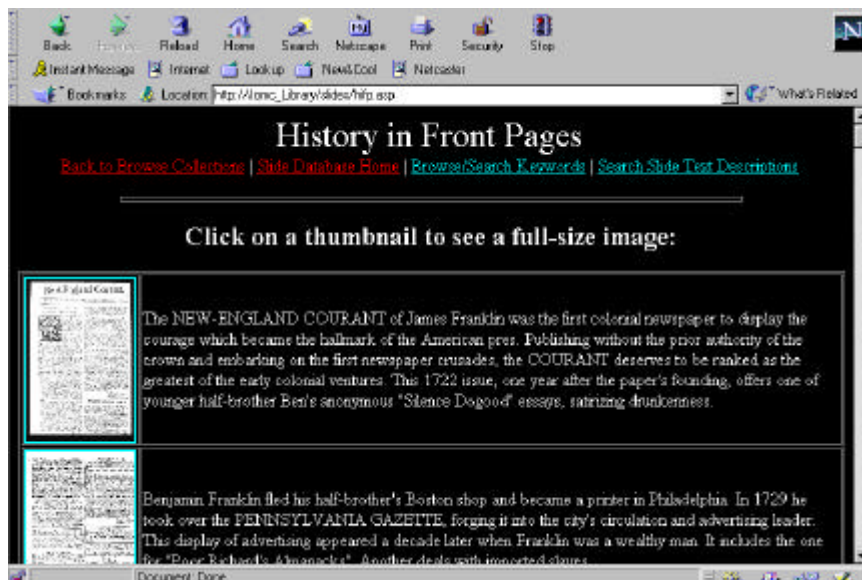
women  
Women, The  
woodcuts  
Woodward, Bob  
working girl  
World Series  
World Turned Upside Down, The  
World War I  
World War II

World, The  
Wright Brothers  
Wright, Theresa  
Yankee Doodle Dandy  
Yankey's Return to Camp, The  
yellow journalism  
Yellow Pages  
Yorktown  
Young, Loretta

## Appendix B: Screenshots



Browsing Collection Titles



Browsing the History in Front Pages collection

Back Forward Reload Home Search Netscape Print Security Stop

Instant Message Internet Look up News Cool Netcast

Bookmarks Location <http://monc.library/slides/keyword2.asp> What's Related

## Search Slides by Keyword

[Slide Database Home](#) | [Browse Slide Collections](#) | [Search Slide Text Descriptions](#)

---

Find all slides containing (choose up to three):

African-Americans

AND OR

20th Century

AND OR

Find Slides

Document Done

**Keyword Search Screen**

Back Forward Reload Home Search Netscape Print Security Stop

Instant Message Internet Look up News Cool Netcast

Bookmarks Location <http://monc.library/slides/fulltext.asp> What's Related

## Search Slide Description Text

[Slide Database Home](#) | [Browse Slide Collections](#) | [Browse Search Keywords](#)

---

Find all slides containing the following word or phrase

Find Slides

Document Done

**Text Description Search screen**

## APPENDIX C

### JoMC Slide Database Questionnaire

#### Demographic Information

1. Are you a:  
☐ Faculty member    ☐ Student (If student, are you:  
☐ Freshman   ☐ Sophomore   ☐ Junior   ☐ Senior  
☐ Graduate Student)
2. If student, what is your major?
3. How many credit hours of journalism courses have you completed?

#### Computer Experience

1. Do you own a computer?  
☐ Yes    ☐ No
2. If yes, what kind of computer do you own?  
☐ IBM-PC or compatible    ☐ Macintosh    ☐ Other (Specify: \_\_\_\_\_)
3. On the whole, how sophisticated a computer user do you consider yourself to be:  
☐ Very sophisticated   ☐ Sophisticated   ☐ Neither sophisticated   ☐ Unsophisticated   ☐ Very unsophisticated  

nor unsophisticated

#### Prior Searching Experience

1. On the whole, how sophisticated a searcher do you consider yourself to be:  
☐ Very sophisticated   ☐ Sophisticated   ☐ Neither sophisticated   ☐ Unsophisticated   ☐ Very unsophisticated  

nor unsophisticated
2. For how long have you been using each of the following:

	Never	0-6 months	6-12 months	12-24 months	24+ months
Lexis/Nexis					
Dialog					
Library Catalogs					
Web Search Engines (Excite, Alta Vista, etc.)					

#### JOMC Slide Database Results

1. On the whole, were you satisfied with the results you got from the database?  
☐ Very satisfied   ☐ Satisfied   ☐ Neither satisfied   ☐ Unsatisfied   ☐ Very unsatisfied  

nor unsatisfied

2. Did any of your queries return results you did not expect? ☐ Yes ☐ No  
If you answered yes, please explain:
3. Which searching mode did you like best:  
☐ Browse by collection ☐ Keyword ☐ Search description text  
Why?
4. Which searching mode did you like least:  
☐ Browse by collection ☐ Keyword ☐ Search description text  
Why?
5. Which searching mode did you choose for your free search?  
☐ Browse by collection ☐ Keyword ☐ Search description text  
Why?