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

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 19th 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 19th 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.

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

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, 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

Figure 2
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


Appendix A: Unique JoMC Keywords

18th Century
19th Century
20th Century
Abell, A.S.
abolition
Adams, John
Adams, Samuel
Addams, Charles
advertising
advertising layout
Advocate
Afghanistan
African-Americans
album cards
Alien
All Alaska Weekly
Allner
Allyson, June
Alta California
alternative press
American Broadcasting Company (ABC)
American Crisis, The
American Tobacco Company
anchors
Andersonville
Andrews, Dana
Andrews, Julie
Andropov, Yuri
animals
Antietam
Aquino, Corazon
Arabian Gulf
Arafat, Yassir
Archduke Ferdinand
Arias, Oscar
Arm & Hammer
Armenia
Arnett, Peter
Arnold, Benedict
art
Art Nouveau
Arthur Shoes
Arthur, Jean
artist correspondents
Artzybashef, Boris
Asbestos Limited
Asia
assassinations
Associated Press
associations
Astaire, Fred
Astor, John Jacob
athletes
Atlanta
Atlanta Constitution
Atlanta Gazette
automatic cylinder press
Averill, John
Avirgan, Tony
Avril, Jane
Ayatolleh Khomeini
Bacall, Lauren
Backstreet
Bahrain
Baker, Ray Stannard
balloons
Baltimore Afro-American
Baltimore Sun
Bancroft, Anne
Bara, Theda
Bardot, Brigitte
baseball
Battle of Gettysburg
beer
bees
Begin, Menachem
Beirut
Bendana, Alejandro
Bennett, James Gordon, Jr.
Bennett, James Gordon, Sr.
Bentsen, Lloyd
Bergen, Candice
Bernhardt, Sarah
Bernstein, Carl
Best Years of Our Lives, The
bibliographies
billboards
birth control
Birth of a Nation
black media
Black Media, Inc.
Black Monday
Blair, Frank
blaxploitation
Blondell, Joan
Cooper, Gary
Cooper, James Fenimore
Copley Newspapers
Copley, Helen
copy
copyright
Cornish, Samuel
Cosmic Revolutionist
Costa Rica
Courier Communications
Corporation
Crawford, Joan
Cream of Wheat
Creelman, James
Cronkite, Walter
Crown Prince Hassan
Cry Havoc
Cuba
Cuban missile crisis
Curtius, Mary
Czolgosz, Leon
D'Escoto, Miguel
daugerreotypes
Dali, Salvador
Dallas Times Herald
dames
Dana, Charles
dance hall girls
Davis, Bette
Davis, Richard Harding
Dawn Magazine
Day, Benjamin
Day, Doris
Dean, James
debates
DeBeers
Declaration of Independence
Democratic Party
demographics
Dempster, Carol
Destry Rides Again
Detroit News
diamonds
Diario de las Americas
Diary of a Mad Housewife
Dickinson, John
Dietrich, Marlene
Digital
disasters
Discovery
Disney hats
Dobson, Tamara
dogs
Dole, Robert
dolls
Double Indemnity
double sextuple press
double supplement stereotype
perfecting press
Douglas Shoes
Douglas, Stephen A.
Douglas, Frederick
Dow Jones Industrial Average
Downs, Hugh
drawings
Drew, Elizabeth
Drummond
Dubai
Dufy, Raoul
Dukakis, Michael
Dumont, Margaret
Dunaway, Faye
Dunlap, John
East Village Eye
Ebony
Eddy, Nelson
Edinburgh Courant
editing
editorials
editors
Egypt
Eisenhower, Dwight
El Pais
elections
Eli Lilly
Emporia Gazette
English Bill of Rights
English, Greg
engravings
entertainment
Equal Times
Esplanade
Essence
Essex Gazette
Essex Journal
everywoman
exotica
features
Feller, Bob
feminism
femme fatale
Penno, John
Ferraro, Geraldine
film and television
film noir
fingers
fire
fireside chat
Fisher, Dan
Fisk
flappers
Fleet Street
flour
Fonda, Jane
Fontaine, Joan
Ford
foreign correspondents
four-color process
France
Franklin, Ben
Franklin, James
Fredricksburg
Free Venice Beachhead
Freedom's Journal
freelance journalists
Friedman, Thomas
frogs
front pages
frontier printing
Funicello, Annette
galleys
gangster films
Garbo, Greta
Gardner, Alexander
Garfield, John
Garland, Judy
Garson, Greer
Gay Community News
Gaza Strip
Gazette Francoise
Gazette of the United States
General Electric
General Mills
Getting Straight
Gettysburg
Geyer, Georgie Ann
Gibbons, Floyd
girl-next-door
Gish, Lillian
Gismonda
Glen Miller Story, The
Glenn, John
Goddard, William
Godfather, The
Godkin, E.L.
Gold Coast Free Press
Gold Medal
gold-diggers
Goldstein, Al
Goldwater, Barry
Good-Bye Columbus
Gorbachev, Mikhail
Gottlob, T.
Gould, Elliott
Grable, Betty
Graduate, The
Grady, Henry
Graham, Katharine
Grant, Cary
Grant, Ulysses S.
graphic design

graphics
Great Depression
Great Northern Railroad
Great Wall of China
Greeley, Horace
Green Dragon Tavern
Green Giant
Grenada
Griffith, DW
Grit
Group, The
Guardian
Guisti, George
Gulf War
gun molls
Gutenberg Bible
Gutenberg printing press
Gutenberg, Johannes
Haberman, Clyde
halftone
Halloween
Hamm's
hand press
hands
Harlow, Jean
Harper's Weekly
Hart, Gary
Harvest of Death
Harvey Girls, The
Hayward, Susan
headlines
health
Hearst, William Randolph
Hellman, Lillian
Henry, Patrick
Henry, Ragan
Hepburn, Audrey
Hepburn, Katherine
Higgins, Marguerite
Hill, George Washington
Hinckley, John
Hindenburg
Hindus, Maurice
Hirohito
His Girl Friday
Hispanics
Hitchcock, Alfred
Hitler, Adolf
Hoe's cylinder press
Holt, John
homosexuality
Honduras
Honey, Martha
Honeywell
Hong Kong
horses
House and Garden
North, Oliver
Northern Pacific Railroad
Northwest Airlines
O'Sullivan, Timothy
Ochs, Adolph
offset printing
Old Dutch
Olympics
Oomph Girl, the
opinion pages
Oppenheimer, Andreas
Ortega, Daniel
Oswald, Lee Harvey
Output
Packard
Paine, Thomas
Palestine
Palestine Liberation Organization (PLO)
Palmerola
Pan Am 103
Paris Herald
Parrish, Maxfield
party girls
Pathe Newsreel
Pathe Newsreels
Payne, Lewis
Pear's Soap
Pearl Harbor
Pears Soap
Pennsylvania Chronicle
Pennsylvania Gazette
Pennsylvania Journal
Pennsylvania Magazine
Pennsylvania Packet
Penny Press
pharmaceuticals
Philadelphia Inquirer
Philip Morris
Philippines
photo opportunities
photocomposition
photoengraving
photographers
photographic studios
photographs
photography
photojournalism
Pickford, Mary
Pillsbury
pin-up girls
Pittsburgh Courier
Plains of Abraham
planes
platen power press
Poe, Edgar Allan
Poindexter, John

political conventions
political reporting
politicians
politics
Polk, James K.
Pond's
Portland Oregonian
portraits
post-feminism
posters
Postman Always Rings Twice, The
power of the press
Powers, Francis Gary
presidential campaigns
presidential debates
presidents, US
press and politics
press conferences
press pool
press room
Princess Diana
print shop
printing
printing plates
printing press
producers
propaganda
prostitution
protests
Providence Journal
Public Enemy, The
publishers
Pulitzer, Joseph
Pyle, Ernie
Quaker Oats
Quayle, Dan
Quinn, Sally
racism
radio
Radio Free Jazz
radio journalism
Ramirez, Sergio
Rand, Paul
rape
RCA
RCA Victor
Reagan, Ronald
Real Paper
Rebel without a Cause
Regan, Donald
Rehnquist, William
Reid, Helen Rogers
reporters
reporting
Republican Party
Reuters
Revere, Paul
Revolutionary War
Rhode Island
Rice, Donna
Richmond
Richter, Paul
Rickard, Tex
Rivington, James
RJ Donnelly Company
Rockwell, Norman
Rocky Mountain Journal
Roderick, John
Rogers, Ginger
Roh Tae-woo
Roosevelt, Franklin D.
Roosevelt, Theodore
Ross, Michael
Rowan, Carl
Royal Gazette
Ruby, Jack
Russell, AJ
Russell, Rosalind
Russworm, John
Sadat, Anwar
Saigon
Salem, MA
San Francisco Bay Guardian
San Francisco Chronicle
San Francisco earthquake
San Francisco Examiner
San Francisco News
San Salvador
Sandinistas
satire
Scheider, Roy
Schlitz
Schneider, Maria
Schultz, George
Scotland
Scott, Harvey
Screw
Scripps, Edward Wyllis
Seagram's
seasons
Seattle Times
Sengstacke, John H.
sensationalism
Seoul
sex papers
sex symbols
sexual revolution
sexuality
Shah
Shahn, Ben
Shearer, Norma
Shell Oil
Sheridan, Ann
shipwrecks
Shirer, William L.
shoes
Shop Talk
Silence Dogood
Simons, Lewis
slavery
slogans
slugs
Smith, Merriman
Snodgrass, Carrie
Snyder, Ruth
soap
socialism
socialites
soft drinks
SoHo News
Sojourner
soldiers
Sound of Music, The
sound technicians
South Korea
Soviet New Times
Soviet Union
space shuttles
Spain
Spanish-American War
sports
sports reporting
Springfield Republican
Squibb
squirrels
SS McClure
St. Louis Globe-Democrat
St. Louis Post-Dispatch
St. Paul Dispatch
St. Paul Pioneer Press
St. Paul Pioneer Press Dispatch
Sta Emot
Stahl, Lesley
Stalin, Josef
Stamp Act
Stanwyck, Barbara
Star Wars
States Rights
Statue of Liberty
Steffens, Lincoln
Steinberg, Saul
Steinlen, Theophile
Steppin' Out
stereo cards
stereoscope
stereotype plates
stereotypes
stereotyping
Stewart, James
stock market
stone
Straus, Isidor
street reporting
Sturbridge Village
suicide
Sulzberger, Arthur Hays
Summertime
Suro, Roberto
Suspicion
Swift Butter
Syracuse Post-Standard
Syracuse University
Syzk, Arthur
tabloid newspapers
Tammy Tell Me True
Tarbell, Ida M.
Taylor, Elizabeth
teenagers
telecommunications
telegraph
television
Temple, Shirley
terrorism
Texas Observer
Thaw, Harry K.
Third World
Thomas, Helen
Thomas, Isaiah
Times and District of Columbia Advertiser
tires
Titanic
Toni
Toulouse Lautrec, Henri
Tower Commission
Tower, John
Townsend Acts
Tracy, Spencer
trade cards
travelogues
Treadwell, Daniel
Trout, Robert
Truman, Harry S.
Turner, Lana
twins
Two for the Seesaw
typecasting
typesetting
typography
Tyson, Cicely
United Kingdom
United Press International
United States Congress
United States Marines
United States Military
United States Navy
United States Supreme Court

US Correspondence Schools
US Supreme Court
USA Today
USS Maine
USS Stark
USS Vincennes
vamp
Van Anda, Carr
VDT
Victrola
videotapes
Videotext
Vietnam War
Viewtron
Village Voice
violence
Virginia
Virginia Slims
Wall Street Journal
Walters, Barbara
war films
War of the Worlds
war photography
war reporting
Warner's Safe Kidney and Liver Cure
Washburn Crosby
Washington hand press
Washington Post
Washington Post Weekly
Washington Times
Washington, George
Watergate
Watterson, Henry
Waud, Alfred
Weaver, Sigourney
web-plate process
weeklies
Welch, Raquel
Welles, Orson
Wells, H.G.
West, Mae
wet-plate process
WGPR-TV
what-is-it wagon
Wheaties
Wheatley, Phillis
White House correspondents
White, Paul
White, Stanford
White, William Allen
Wilson, Edward
windows
Woman's Day
woman's film
woman-as-neuter
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
Keyword Search Screen

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:

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>0-6 months</th>
<th>6-12 months</th>
<th>12-24 months</th>
<th>24+ months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexis/Nexis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dialog</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Library Catalogs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web Search Engines (Excite, Alta Vista, etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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?