Alexandra J. Vidas. The Relative Effectiveness of Text and Images in Image Search Result Listings. A Master's Paper for the M.S. in I.S degree. August, 2006. 70 pages. Advisor: Barbara M. Wildemuth

This study was conducted to determine the best type of image surrogate to use within search result sets: Text, Image Preview, or Text + Image Preview. Users' performance and satisfaction with the three different image surrogates within search result sets were evaluated. Data was collected from 28 participants via a web-based system of questionnaires and logs of their interactions with result set presentations.

Of the three image surrogate types, Image Preview and Text + Image Preview surrogates consistently outperformed Text surrogates on measures of the time required to make relevance judgments, the quality of those relevance judgments, perceived ease of use and perceived usefulness. While relevance judgment scoring with Image Preview and Text + Image Preview surrogates was identical, answers to the post-session questionnaire indicated that users may prefer the Text + Image Preview surrogate, as it was "liked best overall" by more people.

Headings:

Pictures

Information Services/Special Subjects/Pictures

Information Retrieval

Relevance Judgments

Use studies

Information Systems/Design

THE RELATIVE EFFECTIVENESS OF TEXT AND IMAGES
IN IMAGE SEARCH RESULT LISTINGS

by
Alexandra J. Vidas

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

August, 2006

Approved by

Barbara M. Wildemuth

1. Introduction

As the number of hours people spend online continues to increase, so do the number of ways in which they use the Internet to locate, gather, and distribute informational resources (UCLA, 2003; Rainie & Shermak, 2005). According to a recent Pew Internet and American Life Project memo, approximately 60 million American adults use search engines on an average day, placing that activity just behind email as the main daily internet activity for users: while 77 percent of 94 million American adults perform email related tasks during an average day, 63 percent use search engines during a typical day (Rainie & Shermak, 2005).

This demand for search engine use extends into the realm of multimedia search engines. Increasingly, photo-sharing websites have garnered significant traffic (13 million users in November 2002), indicative of a growing demand for image browsing and retrieval services among Internet users (NUA, 2003). With the advent of large-scale image collections, systems are required to not only store but to represent an increasing volume of images in such a way that users can locate images and make accurate judgments about the relevance of images to their retrieval needs.

One way in which this information is made available to users is within result sets, which are designed to support user browsing and recognition of relevant result listings. User interaction with a result set, where each result listing, or surrogate, in some way describes and/or represents the object to which it refers, is a constant in image retrieval. Due to the growing trend in image retrieval and large-scale image collections, users interact with image search result sets of increasing size with growing frequency (NUA, 2003). For example, a search via Google's image search engine retrieved 222,000 images for the search term, "gorilla".

Given the volume of results and possible inadequacy of surrogates (representations of objects) for relevance judgments, search abandonment or dissatisfaction with search results are risks within current search environments. To mitigate these risks, some means to enable quick and accurate browsing of result sets is needed.

Previous studies, such as Diadosz et al. (2002), Woodruff et al. (2001) and Hughes et al. (2003) indicate that the type of surrogates present in result listings affect document discrimination and user satisfaction within web-document search systems. As images are more complex to represent than textual objects, representing image objects both concisely and accurately entails difficult decisions as to which aspects of an image should be represented and how. In particular, image retrieval in the context of search tasks that target images with specific conceptual and literal content (subjects of, objects in, themes and topics within) raises issues as to how images should be described and indexed within a system, including how images should be represented in the context of result sets, how to increase discrimination between (and accurate relevancy decisions about) search results, and how to support satisfactory levels of performance.

Studying and determining possible improvements to the results-browsing stage of the retrieval process offers an opportunity to both improve user experience and enable higher service thresholds for image retrieval engines. This study addresses the problem of optimizing user speed in and satisfaction with locating relevant targets through the medium of result sets, where surrogates of possibly relevant items are presented to users. Examining user interactions with different types of surrogates in the context of result sets related to search tasks that yield images with specific conceptual and literal content and a familiar World Wide Web-style interface, this study seeks to determine the best way in which to list search result descriptions, or surrogates: both text and image-preview components, only text components, or only image-preview components.

2. Background

A driving force governing the usefulness of any collection of information is how easy it is to find what one wants within that collection, whether it is a target object, or an object that possesses target attributes that make it desirable to the searcher. Optimizing this search process relies heavily on how information about items in the collection is organized or modeled, the existence of a systematic means to search a collection for objects matching specific criteria, and a means to provide access to items in the collection. Taken in conjunction, these elements make up an information retrieval system. Whether a search is conducted in a physical or digital environment, the goals of an information retrieval system are: 1) to organize and present information that meets preliminary search criteria in such a way that users can both find and recognize information they seek, and 2) to minimize the amount of resources used to complete the retrieval and recognition process.

The information retrieval process involves an iterative cycle of user action and system response. While system response is bounded by consistent technical limitations, such as bandwidth, screen space, or the number of users that can be supported at one time, user action is limited only by the ability to make decisions based on information the system provides. The type, extent, and presentation of information to users that allows them to make these decisions depends on how information about collection items is modeled within that system; specifically, for an image collection, which information about a particular image would be collected, which information would be available for the system to use as search criteria, and how image surrogates would be presented to the users so that they are able to determine whether the image is relevant to a particular search. Determining the best way to model information about images in order to optimize the retrieval process is a topic of much discussion and research, and holds broad implications for both semantic retrieval and user interface design.

2.1 Indexing images

Modeling data about images relies on the information one can extract from images by available indexing methods, and the ways in which it is possible to store this extracted information. The type and extent of information collected during indexing determines the overall flexibility and usefulness of a search system, since it limits both the search access points and the information available to represent an object.

2.1.1 Indexing methodologies

Two dominant approaches to gathering information about images are concept-based and content-based indexing. Concept-based indexing ranges "from the purely descriptive ('Winston Churchill,' 'a duck on a pond') to the abstract or subjective ('poverty,' 'despair')," while content-based indexing focuses more on modeling of an object (an image), isolating features (i.e., color or texture) and, within those features, representations (i.e., for color: histogram and moments; for texture: tamura and wavelet).

Concept-based indexing collects semantic information about images, such as subject, objects, theme and topic. This method of indexing is an extension of successful methods of text retrieval. In practice, concept-based indexing implements controlled vocabularies and natural language descriptions (often containing contextual information from text surrounding an image or hypermedia links). Terms may be assigned manually to images, although natural language descriptions such as captions are often used to implement automatic interpretation of digital images. One particular concept-based approach is to associate the image with as many keywords as possible. This “keyblock” approach, represents an image’s semantic content with a grouping of keywords (Zhu et al., 2002).

Content-based indexing extracts (usually automatically) pixel-level data, such as color, shape, texture, spatial similarity, and text within an image. Through content-based indexing, the image is analyzed and compared to other images according to similarities in low-level image feature representations (i.e., histogram) (Rasmussen & Chen, 1999; Rui et al., 1999). Content-based indexing is valuable for certain types of retrieval, such as fabric matching, face retrieval and fingerprints, and is powerful for queries focusing on texture, color, and overall image similarity, but it may lead to a visually similar but conceptually disjoint match. In applied, narrow domains, information gathered by this indexing method is extremely useful, but for general use in semantic querying, the fields it captures may not be quite as useful in meeting user needs, especially when relevant patterns cannot be reliably predicted.

2.1.2 Difficulties in semantic indexing of images

Effectively indexing semantic content in images is a challenging prospect, partly due to the complex nature of images and partly due to the means we have available to represent semantic content. While both content and concept-based indexing offer some functionality for semantic searches, they are limited—content-based indexing to narrow applications, and concept-based indexing by the subjective nature of interpretation.

For specific applications, such as fingerprint and face recognition, images can be fully characterized by content-based indexing methodology, but effective approaches for fully characterizing, querying and browsing images in general remains elusive (Zhu et al., 2002). Concept-based indexing offers greater semantic functionality for general image searches. It allows a higher level of analysis than content-based indexing, since it includes information about the context and provenance of an image as well as varied interpretations of the picture that are useful in multiple domains and for heterogeneous uses. However, the indexer's notion of what the picture represents may not always match the user's perception. This is partly due to the interpretive process required when associating an image with a keyword and the subjective relevance judgments made in the course of that interpretation.

Unfortunately, while text document keywords are, fundamentally, units of a text document that have intrinsic semantic content related to the document, image units (pixels) do not share this semantic extensibility, and, even taken as segments, offer only object descriptions of questionable reliability that poorly emulate keyword functionality (Zhu et al., 2002). In and of themselves, pixels may not provide access point for queries, and, "in general, it is impossible to predict the particular pattern that would match an information need" even when querying by image similarity, since feature similarity (such as color) does not necessitate content similarity (Rasmussen & Chen, 1999, 292-3). Essentially, this puts the burden of gathering semantic information for general image searches on concept-based methods rather than content-based methods. This difference in the usefulness of components of text and image objects also means that semantic information about images is most frequently collected and stored in text format.

Due to the complex nature of images, many researchers question whether visual content and "meanings" of images can be accurately translated into text for indexing and surrogacy purposes and, if so, how? Jorgensen investigated this problem and found that, while assigned access points may be technically correct, they "do not capture the wide variety of information which may be associated with an image" and thus may not be useful to the searcher (Jorgensen, 1998, 163). Assigning more or different access points may remedy this problem. However, additional issues surround the assignment of values to these access points. Interpretation of images, first and foremost, is difficult to standardize or assign a single value for, since,

> Even where human indexing of the image is undertaken, it is difficult to reach agreement on the content and meaning of the image or on what aspects are appropriate for indexing. The same image may mean different things to different people and may be used to project different meaning at different times depending on the way it is used or the aspect that is the focus of attention or the context it is chosen to illustrate. (Rasmussen & Chen,1999, 293)

While indexers often rely on cognitive heritage and social conventions to offset interpretive differences and to provide a shared lexicon and shared natural and synthetic ontologies for expressing important information about images, this presumes that users share this cognitive similarity, an assumption that may not be realistic in an increasingly global environment (Heidorn, 1999). Even assuming a shared cognitive heritage and social conventions, once an interpretation is chosen, difficulties still remain in the choice of vocabularies to use when codifying these interpretations. The fact that, "The untrained person's vocabulary does not necessarily match that of a descriptive system," and interindexing difficulties occur both on the ideological and the terminological level among professionals, problems can arise in matching query criteria with access point values on a consistent basis (Jorgensen, 1998, 163). Use of a fuzzy retrieval system, as in Wu and Narasimhalu's study, where a query provides a subjective conceptual description of an object to retrieve, allows for differing vocabularies of image interpretation in the query modality and can offset the choice of vocabulary, but it is not a foolproof solution (Wu & Narasimahlu, 1998).

In addition to issues surrounding choice of vocabularies and differing interpretations of semantic content, some question whether visual information *can* be conveyed through language, since, for instance, many colors exist, but only some are named—communicating about these unlabeled colors without using visual cues is difficult if not impossible (Heidorn, 1999). Due to this lack of translatability, "there are those who feel that text-based systems are inappropriate for retrieval of visual materials" (Jorgensen, 1998, 163). Both Heidorn and Jorgensen are in agreement that, while many aspects of images "may be easily described linguistically… other aspects might best be described or communicated by example of image" (Heidorn, 1999, 312). The perceived "inadequacy of language as a recording medium for describing a work of art" and the "hypothesized disjunction in cognitive modalities which arises from searching for visual media through text" have led to proposed solutions ranging from pairing images and text to discarding text elements and implementing image-to-image methods (Jorgensen, 1998, 163).

Despite the difficulties inherent in semantic indexing of images, both concept-based and content-based indexing can extract enough information (however subjective) to permit general semantic searching and representations of images. While this information *can* be extracted, the question remains as to what information *should* be extracted to support particular types of semantic searches and to provide the best possible access to and recognition of images.

2.2 Access Points to Collect and Display

At its core, data modeling is a process that requires making decisions about which information is collected about data objects—in this case, images. These determinations result in descriptions of images in terms of access points. An example of an access point is "creator", where for each image in a collection a value is assigned to author, such as "Picasso". Creation of access points enables search by that access point (finding all images where the creator is "Picasso") and allows the possibility of displaying that information to users at the result set browsing stage of the search process.

As of yet, there is "no general agreement on what attributes of an image should be indexed" (Rasmussen & Chen,1999, 295). When information needs are predictable, what the user requires from a search system in order to achieve "intellectual access"[1] is often predictable as well, and that consistency of requirements can be used to optimize particular aspects of the search cycle. However, when information needs levied on an information system are unpredictable or complex and subjective in nature, as in the case of semantic searches, determining what users require from a retrieval system in order to complete their search scenarios is a difficult proposition. In this case, "the subject orientation of users and the information need that will lead them to pose queries to the

---

[1] Hastings defines 'intellectual access' as "the image searcher's ability to find and use (retrieve) the image that meets a stated need" (Hastings, 1999, 442).

collection cannot be anticipated, and hence the dimensions along which the collection should be indexed cannot be predicted" (Rasmussen & Chen, 1999, 295; Shatford, 1986).

To highlight the difficulties inherent in choosing access points for semantic indexing, consider a well-known search scenario: the Arthurian quest for the Holy Grail. In this search scenario, multiple agents sought a single target, a target none of them had seen before. Each seeker had a broad idea of the characteristics the object possessed (shape that can be held within hands and contain liquid), but not the particulars (substance it is made of, adornment). In fact, many of the characteristics they sought in their target were not physical characteristics at all, but referenced context and non-tangible qualities (e.g., provenance, metaphysical properties of healing). While each seeker possessed a hazy preconception of what the Grail might or should look like, their search was governed by the precept that recognition of the Grail was as important as the accomplishment of finding it.

For the seekers of the Holy Grail, as well as more modern searchers, this *Eureka!* moment of recognition depends on the existence of meaningful information about the object being considered for possible acceptance, and the communication of this information to the seeker. An information system, in order to aid this type of search, must consider what qualities or characteristics are necessary to convey in order to establish that the target (Holy Grail) is found, and how less readily apparent or communicable qualities can be relayed to the seeker. For instance, in the area of image retrieval, what information needs to be stored about an image in order for a user to be able to search for it and recognize it on the basis that it contains a "happy family"?

Examining precedents in art history, media (newspapers), search modeling, and result set discrimination offers some insight into access points that are useful for information modeling. The problem of how to best index images for specific domains and for the general domain is not a new development. Art historians have debated and suggested numerous approaches and levels of detail.

Currently, most image indexers build on the work of Panofsky (1939) who identified three levels of semantic significance in works of art. The first level, which he calls “pre-iconographic” designates subject matter as nonsymbolic, factual (“ofness”), or expressional (“aboutness”), including generic actions, entities and entity attributes in an image, interesting objects and events as through everyday experience. For instance, an image interpreted at this level may contain a stone bridge over a river (where “stone” is attribute, “bridge” is an entity and “river” is an entity). The second, “iconographic” level identifies particular instances of entities or actions, for example, “Golden Gate Bridge” and usually requires some cultural knowledge of concepts or themes (not “a sailor” but “Ulysses”). The third, “iconologic” level includes the symbolic meaning of an image. For instance, terms such as “peaceful” or symbology may be attributed to it (i.e., this image represents “simpler times”) by interpreting the image through the filter of world or cultural knowledge and an understanding of the history or background of a work (Pankofsky, 1939; Pankofsky 1955; Rasmussen & Chen, 1999; Heidorn, 1999).

Shatford (1986) expands on Panofsky’s work, using his framework to explore possible subjects of images, and proposing additional facets, such as “Generic Of,” “Specific Of,” and “About,” and subdividing them into facets to answer the questions Who? What? When? and Where? (Shatford, 1986; Rasmussen & Chen, 1999, 293).

Layne (1994) suggests the inclusion of contextual and interpretive attributes to provide access points to images. He "proposes four categories (1) "biographical" attributes that deal with the images' origin and provenance; (2) subject attributes (the "most problematic and least objective"[p.584]); (3) exemplified attributes that seem to be physical characteristics such as medium, and (4) relationship attributes (relationship to other images or texts)" (Rasmussen & Chen, 1999, 295; Layne, 1994, 584).

In the movement from specific applications of image retrieval in the art history domain, which offer insights into the depth of possible descriptions and levels of specificity, to more global applications, considering studies of newspapers may offer direction for the development of general-purpose image retrieval systems, as a way to anticipate changes in use and purpose in an image retrieval system over time. Markkula and Sormunen's field study concentrating on journalist users of a digital newspaper photo archive considered a faceted classification of image attributes developed by Shatford along the axes of "Objects (Who), Activities and Events (What), Place (Where), and Time and Space (When)". These facets could then be used to "represent both concrete and objective entities (*ofness*, e.g. objects, places, actions) and abstract and subjective entities (*aboutness*, e.g. feelings, concepts manifested or symbolised by objects)" as well as the simultaneous specific meaning (i.e., this house) and generic meaning (i.e., a house) of an image (Markkula & Sormunen, 2000, 261). These elements were present in image captions and image-accompanying text in this archive, allowing application of natural language processing techniques (Markkula & Sormunen, 2000). Results of caption analysis showed that those produced by photo agencies are suitable for some common needs and described specific objects, events and the news context of photos. However, accessing images with generic object types, themes, and restrictive structures was difficult via caption and accompanying text, and often required additional indexing (Markkula & Sormunen, 2000). Chen's (2001) study of a keyword-based retrieval system found that the level of success for search results increased with increases in the percentage of search keywords or phrases drawn from the topic title or topic description (Chen, 2001). This finding supports Markkula and Sormunen's determination of the importance of elements within captions and accompanying text.

Jorgensen diverges slightly from this framework, recommending the collection of color data and contextual (story) data as well. Her research (1998) examining descriptions of color images written by participants led to the suggestion of a minimal framework of four perceptual classes: objects, people, color and location, and identified content/story attributes as important for image description. Both a previous study (Jorgensen, 1995) and this study noted a need to include interpretive as well as perceptual attributes (Rasmussen & Chen, 1999). In her effort to classify attributes needed to address all facets of interest to those using pictorial images, Jorgensen (1998) extracted 47 image attributes, grouped conceptually into 12 higher levels of attributes, from participants' statements about image sets. This is a fair indication of how complex the translation from image to attribute sets could be in a given system, and provides a challenging contrast to traditional assumptions about image indexing systems and what constitutes an accurate representation of an image object.

Studies investigating how users choose to query image collections are also a valuable source for guidance about access points. Garber and Grunes found that descriptive terms, objects in, objects not in, and general characteristics of images were common search specifications (Garber & Grunes, 1992). Where searches implement query-by-visual example rather than linguistic matching, maintaining access points that express visual information, such as histograms in addition to Jorgensen's suggestion of color would be valuable (Enser, 1995).

Enser and McGregor's (1993) Hulton Study provides additional insight into level of specificity about objects within images and contextual information. They mapped requests into four categories along two dimensions: *unique* ("Kenilworth Castle") or *non-unique* ("dinosaurs") and *refined* (e.g., specified by activity, time period, and so on) or *nonrefined*. An example of a query in the unique refined category is "Edward VIII looking stupid" and in the non-unique refined category is "couples dancing the Charleston" (Rasmussen & Chen, 1999, 294; Enser & McGregor, 1993)

In addition to investigating queries made by art historians, Hastings (1995) also reviewed access points used in a Caribbean art collection. She established four levels of query complexity and showed a direct correlation between type of query and access points. Access points for different types of queries included: (for Least Complex searches) text fields and image in general, (for Complex searches) sorted text information and images, (for More Complex searches) style, keywords and complex images, and (for Most Complex searches) the addition of subject as well as style. Queries relying on these access points targeted similar access points to those described by Panofsky and Shatford, with the addition of "What are?", "Why?", style, and "How?".

Hastings' four levels of query complexity differed in the information targeted via access points: Least Complex queried "Who", "Where", "When"; Complex addressed "What are?"; More Complex searched on style, subject, "How?", and the identifier for objects or activities; and Most Complex looked for meaning, subject, and “Why?”. Elevations in search complexity led to changes in both access points used and computer manipulations employed by the user. Hastings (1999) found that Web searches fell into two categories: a combination of Least Complex and Complex, as nearly 60 percent of collected queries asked for artist, place, or activities, and Most Complex, as the remaining 40 percent targeted the subject of the image (Hastings, 1999). This finding supports Markkula and Sormunen’s addition of “Activities” to Panofsky’s and Shatford’s framework, and the inclusion of “Where?” and Layne’s (1994) suggestion to include “biographical” information about images.

In order to support the diversity of searches people conduct on image retrieval systems, many access points are needed to describe images in such a way that users can reach target images, even when their search needs are unpredictable. Access points collected should include the results of both concept-based and content-based indexing to provide the most accurate representation of semantic and useful data about images.

2.3 Surrogates

Once information is gathered about images, and a means is enabled to search image collections, one more determination is necessary to enable system performance: a decision about which information is displayed to users to enable relevance judgments, and how that information is displayed. The goal of using surrogates is to represent images as well as or better than the representation achieved by viewing the image itself. This is a complicated prospect, since a wide variety of information about images is stored in text form, and the effectiveness of text as a representation for the complexity of images is questionable (Jorgensen, 1998). Representing as complex an object as an image even by multiple access points that combine visual features with textually represented data may not describe the "aboutness" of images as well as seeing the original image would, and so would not allow the user to discriminate between surrogates of images with the same accuracy as if the user were viewing the images those surrogates represent (Chang et al., 1997; Hirata et al., 2000). At the same time, simply seeing an image may not impart information about its provenance or context necessary for resolving some information needs, leading to questions of what information needs should be supported, and what information (including context) should be included in search results in order to assist relevance judgments for particular information needs. Moreover, how do we encapsulate information about images as concisely as possible, isolating the essentials necessary for users to make relevance judgments, thus conserving storage, bandwidth, screenspace, and processing resources and enabling more efficient result set processing by users?

Surrogates generally contain either text alone, image preview or thumbnail alone, or a combination of text and thumbnail. How these components are arranged within a surrogate depends on the individual system. Access points reflected in text or thumbnail content vary according to the information collected about the image and information the data modeler decides the user should see in order to make an accurate relevance judgment. Often, a combination of access points can offer more semantic information than the image alone—for instance, information about the author or resolution is not evident when viewing the original image without associated text.

The key aspect of a surrogate is recognition, which relies on matching descriptive components with a defined target criterion or criteria (Aslandlogan et al., 2000). This recognition depends on the presence of information within the surrogate that the user needs in order to make a judgment about the relevance of the image in the context of their search. Whether a surrogate is the image itself, textual descriptions of the image or associated information, an altered image such as a thumbnail, or a combination of textual and visual access points, the access points that indicate relevance to a user must be present in the surrogate for recognition to occur. On an access point level, values that determine relevance must be reflected to users. However, the choice as to which access points are reflected varies across systems and with projected use of systems and the images it references. Examples of common access points within image surrogates are: author, title, thumbnail, resolution, caption, and hierarchical category information. In particular, Dumais and her colleagues found that including context in the form of hierarchical category classifications aided performance (speed) as well as satisfaction (Dumais et al., 2000).

The reasons users search for images and how they intend to use these images have implications for the access points presented and emphasized in surrogates, how that information is represented within surrogates, and the organization of those surrogates in a result set in such a way as to aid discrimination and processing time on the part of a user (Efthiamiadis & Fidel, 2000). For instance, Efthiamiadis and Fidel investigated the effect of query type on searching behavior in image databases, using a spectrum developed by Fidel, beginning with the Data Pole (retrieval based on inclusion of information, targeted by queries such as, “Find a castle in Ohio. What is the color of its window frames?”) and terminating at the Objects Pole (retrieval based on the image’s own merit, targeted by queries such as, “Find a picture for the homepage of the homeowner’s association of Seattle.”). The study found that searchers tended to use subjective or abstract criteria and judgments more frequently for Object Pole queries. Only 18 percent of participants felt that selected images for Data Pole queries required interpretation to locate the answer while 67 percent felt that interpretation was required for the Object Pole (Efthimiadis & Fidel, 2000, 328-9).

The type of query also affects whether relevant criteria can be predicted. Fidel's earlier study (1997) noted a difference between Data Pole and Object Pole relevance criteria, in that for the Data Pole queries, relevance criteria can often be determined in advance, while for Object Pole queries browsing the entire answer set is necessary to determine the "best" match (Fidel, 1997).

Additionally, the type of search task affects preferences for the presentation of relevant information, whether in visual or linguistic form, or a combination of both. For instance, in Markkula and Sormunen's study, elements of ofness, aboutness and specific and generic meaning were present (to varying degrees) in image captions and image-accompanying text, so their role in presenting semantic information in search results was evident (Markkula & Sormunen, 2000). Markkula and Sormunen's results also indicate that "some criteria used in selecting photos seemed to be difficult to express by words but were easily applied when the photo was seen," which may have implications for the pairing of image previews with textual listings (Markkula & Sormunen, 2000, 281). The study displayed query results as thumbnail images that linked to enlarged versions with captions included in the bottom left corner, but did not address the pairing of image previews and captions on the primary search result page (Markkula & Sormunen, 2000).

Hastings (1995) also found that in some retrieval situations, searchers use a combination of both visual and linguistic features. In her study of the Collection of

Caribbean Art, Hastings found that simpler queries (such as who, what, where) could be answered with only text elements, while some more complex queries (such as meaning, subject, why) could not be answered with primary textual information or image alone, but needed secondary subject resources (Hastings, 1999; Rasmussen & Chen,1999). In this system, “the problem of ‘relativity’ or queries of ‘why’ is largely unsolved” (Hastings, 1999, 449). However, when experimenting with surrogates that included visual information, the study found that almost 60 percent of the queries collected were answered with the use of the thumbnail images (Hastings, 1999).

Comparisons between text, image thumbnail, and combination surrogates found that the type of descriptors present in surrogates affects relevance discrimination, performance, and user satisfaction (Dziadosz et al., 2002; Woodruff et al., 2001; Hughes et al., 2003). Dziadosz discovered that when given web-document search results in three formats (text only, thumbnails only, and text + thumbnails) that the combination case yielded more accurate decisions about the potential relevance of results (a page) than text-only or thumbnail-only. However, this was an examination of web documents rather than images (Dziadosz et al., 2002).

A similar study by Woodruff et al. compared task performance (searching webpages for information) when using textually enhanced thumbnails (of the webpage), image thumbnails (of the webpage), and text summaries. They found that, while text outperformed plain thumbnails and vice versa for some questions, the enhanced thumbnail combined the advantages of image thumbnails and text summaries to provide consistently best (or indistinguishable from the best) performance across tasks. Finding the answer took an average of 67, 86, and 95 seconds to find the answer with enhanced thumbnails, plain thumbnails, and text summaries, respectively, with a strong effect of question category (Woodruff et al., 2001). Text summaries performed better when textual information in the page was required (since text was difficult to read in thumbnails), but plain thumbnails provided support for layout, objects within the page, genre and style of the page (especially if the user has seen the page or a similar page before) (Woodruff et al., 2001). Several participants perceived the enhanced thumbnails as more intuitive and requiring less work than the alternatives. Sixteen of 18 participants used genre cues present in thumbnails and 14 used callouts, search term relationships, search term location, and term frequency when using enhanced thumbnails. Nine rated the enhanced thumbnails as their favorite summary type overall, while most others preferred the enhanced thumbnails for certain types of tasks (Woodruff et al., 2001). Woodruff et al. mentioned that the human visual system processes images more quickly than text as a possible contributing factor to thumbnail performance for certain types of questions (Woodruff et al., 2001).[2]

Hughes et al. agreed that multimedia retrieval is dependent on metadata that "stands for" the full object, providing context and clarity during the retrieval process that enable accurate relevance judgments. Their study used representations of video objects, and employed eye-tracking methods to explore interactions, including the length of time participants spent looking at text or pictures. Hughes et al. found that participants looked at textual representations 22 seconds longer per search, on average, than pictorial surrogates. Participants began scanning the middle section of the page and focused on elements in that position first. Time spent looking at text varied with search task, and

---

[2] Possible extensions of this textually enhanced visual surrogate include video indexing, since captioning on videos can also supply index terms and aid searches (Hastings, 1999, 440).

results indicated that most users seemed to use the text as an "anchor from which to make judgments about the search results" (Hughes et al., 2003,6). Scan path analyses and interview responses indicated that participants felt most comfortable performing video retrieval tasks using textual metadata surrogates (Hughes et al., 2003).

The type of surrogate users are more comfortable with, feel suits their searching needs, and offers the best performance may vary depending on the searching task users perform. While the "best" surrogate may change, any surrogate can be aided by functionality within result sets that supports relevance discrimination activities. Since user interaction with surrogates takes place primarily within result sets, the effectiveness of surrogates is influenced by the result set environment, much as any picture's overall appearance is affected by its frame and matting. Manipulating the result set environment to improve presentation of surrogates and accomplish preliminary sorting offers an opportunity to further optimize the relevance judgment process.

2.4 Result Sets

Structuring result sets to provide useful information in addition to or about surrogates is an additional element contributing to optimization of the relevance judgment process. Jorgensen (1995) identified three tasks or modes involved in modeling image retrieval: describing, searching, and sorting. The result set's function is to support this sorting, or scanning task associated with image browsing. Marchionini (1995) defines this sorting activity as "a perceptual recognition activity that compares sets of well-defined objects with an object that is clearly represented in the information seeker's mind" (Marchionini, 1995, 111). This activity also extends to comparing objects within a result set, finding out more information about those objects in order to determine whether an image may be relevant, and making a decision as to what objects are relevant or possibly relevant (Hastings, 1999; Hastings 1995).

Users view result sets with “a mental model” or a collection of ideas (“image concept”) about the target image (Garber & Grunes, 1992, 159; Heidorn, 1999, 308). This image concept guides the sorting task of image retrieval, which Marchionini (1995) characterizes as an interacting task (Garber & Grunes, 1992; Jorgensen, 1995; Chen, 702; Marchionini, 1995, 110).

Result sets support sorting actions by enabling interactive functionality for both browsing and manipulation of surrogates. At this sorting stage, browsing is essential to enable discrimination between surrogates and in order to further refine the image concept (Markkula & Sormunen, 2000; Garber & Grunes,1992). During browsing, users search, select, sort, display, enlarge, compare, mark, view resolution and style, and may even perform additional refining searches or access secondary subject resources (Hastings, 1995; Hastings, 1999). Frost et al. found that image retrieval is uniquely suited to result set browsing due to an image’s identity as a whole rather than a sum of its perceived components (Frost et al., 2000). Their focus group emphasized the importance of image quality (high resolution) as a discrimination factor within result sets. Two-thirds of their participants judged the thumbnail resolutions as acceptable, which indicates that thumbnails can be considered a viable surrogate or surrogate component (Frost et al., 2000).

Depending on system functionality, result sets also provide ways of setting aside and organizing images of interest. They can also supply both exact and close matches to facilitate broadening of the image concept (Garber & Grunes, 1992). Result sets also aid relevance judgments by providing contextual information in the form of hierarchical categories for individual surrogates, narrow result pools, and pools that may encompass both approximate and exact matches (Dumais et al., 2000; Chang et al., 1997; Yee et al., 2003). Yee et al. particularly emphasized the usefulness of faceted metadata, or categories, in narrowing the search set, organizing results, and expanding the image concept by browsing by additional category (Yee et al., 2003). Their study found that providing a faceted-category functionality within result sets garnered more positive ratings than use of thumbnails. While thumbnails were preferred for use in simple single-facet tasks (finding images of roses) by 50 percent of participants, the faceted category interface was preferred for every other type of search (Yee et al., 2003).

An additional, related way in which result sets assist relevance judgments is by organizing surrogates by similarity. Grouping items in a result set by similarity to other items in the result set as well as by commonality of or "closeness" to a criterion or a category assists discrimination on the basis of specificity. Rodden et al.'s study on image browsing and organization of thumbnails by similarity to each other found that grouping by similarity seemed useful to graphic designers searching for images to accompany text, especially when they wished to narrow their target pool to a subset containing a more specific commonality. The addition of broad captions to label these subsets (superimposed on groupings of thumbnails) also aided in discriminating between subsets. However, the usefulness of the captions depended on the level of detail available and, likely, how well the task's target overlapped with the level of detail in the caption. Rodden et al. stated that, "Labels may be necessary to help the user understand" the structure of result sets, and what lines of commonality determine arrangement (Rodden et al., 2001,197).

Other aspects of presentation that can help or hinder user interactions with result sets include pagination and the number of listings per page. Systems that require thumbnail browsing over several pages “are not suitable for large visual information retrieval systems” (Chang et al., 1997, 67). In the case of AMORE, Mukherjea et al. found that most people only looked at the first page of results. For this system, presentation of nine thumbnails at a time established a useful ratio of presentation (surrogate to result page) given their success (Mukherjea et al., 1999, 118, 131). Woodruff et al. concurred with Chang et al., stating that reading lists of search results is tiring and that the average user will not read more than a few pages of listings (Woodruff et al., 2001, 198). Heuristics agree with these statements, as long lists of results displayed as very long pages are noted as a blooper by Johnson (2000).

Choices made about how to model image data both on the result set level and on the surrogate level offer an opportunity to optimize the semantic search process and supply improved services to users. This study contributes to current literature by investigating the effect of surrogate type on the sorting stage of the image retrieval process.

3. Research Questions

The goal of this study is to advance the design of image retrieval systems through the development of recommendations for ways in which to display search results for image objects in order to contribute to user satisfaction in searching and browsing image collections, speed and accuracy of processing, and perceived ease of use and usefulness.

Through ascertaining which metadata elements and surrogate types are preferred for particular types of search scenarios within a hierarchical system, this study will augment existing studies, metadata indexing schemas, and user interface prototypes, within both academic and commercial applications and domains. Specifically, the research question to be investigated is: What is the best way to convey information about image objects in result sets: by using surrogates composed of a) only text components, b) only image preview components, or c) both text and image-preview components?

4. Study Methods

4.1 Participants

Twenty-eight participants were recruited for this study from among students, faculty and staff at the University of North Carolina at Chapel Hill and from people in the community via email and flyers. This recruitment included people identified as having an interest in image collections. Participation was voluntary and not associated with any class benefit or requirement. Care was taken to include both genders as well as people with a broad range of ages (over 18). Participants included people who are familiar with both taking and viewing pictures. Children, blind or legally blind people and computer-naïve people were excluded from the participant pool. Participants were paid $5 for their participation, and entered into a drawing for a $20 Amazon gift certificate. Any personal information gathered for payment purposes was destroyed after the incentive was given to the participant.

4.2 Research Design

The focus of this study is on the influence of surrogate type on people's browsing of search results. A within-subjects design was used to evaluate the three different surrogate types (Text, Image Preview, and Text + Image Preview) in terms of time, "correctness" of relevance judgments, perceived usefulness, perceived ease of use, and other affective responses.

This research study was conducted remotely, via the Internet, over broadband Internet connections. Twenty-eight participants completed a set of online questionnaires (pre-session, post-system for each surrogate type, and post-session) and study trials featuring images and metadata from a commercial image database site (http://www.webshots.com). Study trials consisted of interacting with result sets for content and concept-based search scenarios. The type of surrogate used within a given result set varied between subjects. All subjects interacted with all three types of surrogates during the course of this study.

Each participant interacted with a training search scenario result set and four study search scenario result sets for each surrogate type. Scenarios were organized in three equivalent blocks. Each block consisted of:

- one high specificity, close-ended training scenario,
- one high specificity, open-ended study scenario,
- one high specificity, close-ended study scenario,
- one low specificity, open-ended study scenario,
- one low specificity, close-ended study scenario, and
- a post-system questionnaire.

All blocks and scenarios were encountered in the same order, but the type of surrogate participants interacted with during that block was counterbalanced among participants. Ordering of scenario types within blocks was also counterbalanced: high and low specificity scenarios were alternated throughout. All blocks began with a high specificity training scenario, then alternated high specificity, low specificity, high specificity, low specificity. The combination of high or low specificity with close-ended or open-ended searches varied over blocks as shown in Table 1.

Table 1. Counterbalancing/order of scenarios within block

	Training Scenario	First Scenario	Second Scenario	Third Scenario	Fourth Scenario
Block 1 ordering	Close-ended, high specificity	Open-ended, high specificity	Close-ended, low specificity	Close-ended, high specificity	Open-ended, low specificity
Block 2 ordering	Close-ended, high specificity	Close-ended, high specificity	Close-ended, low specificity	Open-ended, high specificity	Open-ended, low specificity
Block 3 ordering	Close-ended, high specificity	Close-ended, high specificity	Open-ended, low specificity	Open-ended, high specificity	Close-ended, low specificity

The order in which participants encountered surrogate types was also counterbalanced: participants were assigned to one of six groups, as shown in Table 2.

Table 2. Counterbalancing of surrogate types

	Surrogate Type for Block 1	Surrogate Type for Block 2	Surrogate Type for Block 3
Group 1	Text	Image Preview	Text + Image Preview
Group 2	Text	Text + Image Preview	Image Preview
Group 3	Image Preview	Text	Text + Image Preview
Group 4	Image Preview	Text + Image Preview	Text
Group 5	Text + Image Preview	Text	Image Preview
Group 6	Text + Image Preview	Image Preview	Text

Note: For a detailed breakdown of surrogate orderings by participant, see Appendix E

4.2.1 Search Scenarios and Result Sets

This study used 15 search scenarios; three as training scenarios, and 12 as study scenarios. Each block of scenarios includes two open-ended scenarios (with a target of more than one matching result) and two close-ended scenarios (with a target of one matching result). Scenarios targeted specific literal or conceptual content, following Borlund's (2003) suggestion that researchers assign simulated search task situations to increase study validity. An example of an open-ended search scenario would be the Forest Path search scenario: "You are designing some flyers for a weekend hiking trip and need a picture of a forest with a path running through it. After selecting a category called "Forests" you see listings for a number of results featuring forests. Pick a few pictures of a forest with a path running through it to review later for use in your flyers." While this search scenario asks the participant to "Pick a few pictures" a close-ended scenario would ask the participant to "pick a picture" or "pick the best picture."

Since many image archives are organized according to subject category, granularity of categories becomes a factor when browsing result sets. To accommodate this, one scenario of each type (open-ended or close-ended) addressed a result set at a higher level of specificity in the subject-category hierarchy (e.g,. Animals>Bears), and one addressed a result set at a lower level of specificity (e.g., Animals) in order to examine effects of variance in level of homogeneity in result sets. The Forest Path scenario above is a high-specificity search, due to the granularity of the category (Forests). For a complete list of scenarios and their descriptions, see Appendix D.

Search scenarios were presented to participants at the top of each result set page, followed by directions for performing relevance judgment tasks for that result set page.

Training scenarios were marked with “Training scenario” at the top of the screen. Result sets were homogenous in level of specificity, all items belonging to the same category of things. Each result set contained a constant number of potential targets (one scrollable page of 15 items) and presented listings in the same order for each surrogate type. Additionally, each result set had a constant number of “correct” listings to choose from, regardless of whether the scenario asked the participant to look for one or multiple targets: training sets had five possible “correct” listings to choose from, and each study result set had six possible “correct” listings. The position of “correct” or relevant items in each scenario result set was randomized and was not duplicated across scenarios. Each scenario set featured three “correct” listings in the first 5 items (listings 1-5), two “correct” items in the second 5 items (listings 6-10), and one “correct” item in the last 5 items (listings 11-15).

4.2.2 The Images and their Representations

Images and associated metadata used in the study were obtained from a popular commercial image database site, http://www.webshots.com, a pool of approximately six million items. In a few cases, pictures were located through other search engines and paired with matching metadata from http://www.webshots.com to promote clear relevance judgments. Category information was standardized for search scenarios that specified a category within the search scenario text. Captions were limited to the specified field length.

The specific sample of images used in the study was selected based on category or search results returned for keywords used in the study search scenarios (e.g., cactus). Of these results, 15 listings were selected for inclusion in each scenario’s result set. Both relevant and non-relevant listings were collected for use in the study. The relevance of a listing was determined by whether the image preview and text metadata featured criteria matching the search scenario goals. Non-relevant listings did not feature the relevant information in either metadata or image preview. For example, when choosing listings for a scenario that specified the Washington Monument as a target, a relevant listing's image needed to have an identifiable image of the Washington Monument in it, *and* the text "Washington Monument" somewhere in the text element of the listing. In non-relevant listings, the Washington Monument would be absent from the image and "Washington Monument" would not appear in the text element of the listing.

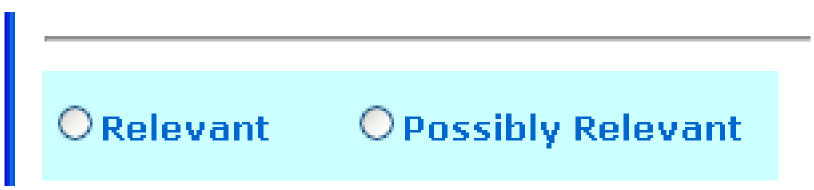
Three types of image surrogates were used in this study: Text, Image Preview, and Text + Image preview. The Text surrogate may have contained: author, copyright owner, copyright date, location (optional), title (optional, up to 50 characters), caption (optional, up to 50 characters), location in subject-category hierarchy, as entered in a commercial image database site, and resolution. Image Preview surrogates included a thumbnail image with associated filename (e.g., img099.jpg) of at least 1.92 effective pixel resolution (matching the current baseline for Web pictures in digital camera technology). Dimensions of image previews were either 110 x 82 pixels (horizontal orientation) or 75 x 100 pixels (vertical orientation), as in the http://www.webshots.com database site. The combination, or Text + Image Preview, surrogate consisted of both text data (as for text surrogate), and visual data (as for image preview surrogate). An example of each type of surrogate is shown in Figure 1. The result sets were all laid out the same way, in a grid with five rows and three columns of surrogates.

Figure 1: Examples of each type of surrogate

4.2.3 Relevance Judgments

Participants were asked to make relevance judgments as part of their interaction with the result set (see Appendix H). They were asked to, "Select whether an item is Relevant or Possibly Relevant to the scenario below," and asked to, "Skip over the item if it isn't relevant to the scenario." They used the interface shown in Figure 2 to mark their relevance judgments.

Figure 2. Interface used for making relevance judgments



4.3 Study Procedures

As mentioned earlier, each participant completed the study procedures remotely via high-speed Internet connections. Information was gathered and displayed through PHP pages and stored in a MySQL database.

Each participant was given a userid and information for accessing the website. When the participant first accessed the study website, s/he was given an overview of the project and its rationale and the study procedures were explained. The participant was provided with an online consent form (Appendix F). After giving informed consent via a webpage and having the opportunity to view more detailed information prior to consent (Appendix G), each person provided information to enable remuneration, and then participated in an individual evaluation session via interaction with webpages consisting of: a pre-session questionnaire (Appendix A), a training trial for the first surrogate type, four study trials on that system, an evaluation of that surrogate type (Appendix B), a training trial for the second surrogate type, four study trials on that system, an evaluation of that surrogate type, a training trial for the third surrogate type, four study trials on that system, an evaluation of that surrogate type, and a post-session questionnaire (Appendix C). Between each section (approximately every two pages) the participant was advised to pause at that time, or continue. Participants were asked not to backtrack or select the Back button for the duration of the study.

The pre-session questionnaire (Appendix A) collected information on general participant characteristics (age, sex), experience with computers and searching (years of experience with online searching; frequency with which they use a computer; frequency with which they conduct a search on any system; and experience with particular systems or interfaces: point-and-click interfaces, searching on computerized library catalogs locally or remotely, searching on CD ROM systems, searching on commercial online systems, searching on World Wide Web search services, and searching on other systems); information about image use and searches (frequency with which they take or view pictures; frequency with which they search for images; where they go to search for images: online, newspaper or magazine, image archives or collections, or other; how they search for images: by title, by author, by topic, by preview/thumbnail, or other; and for what purposes they usually search for images)..

After completing the pre-session questionnaire, participants went on to the first of three equivalent (not identical) blocks of scenarios. Each block began with a training scenario. The training trial was intended to acquaint the participants with study procedures and the type of surrogate used in each result listing for that trial (Text, Image Preview, Text + Image Preview) and data-entry interfaces they would interact with throughout the study. In each training trial (three per participant), the participant viewed one search scenario and accompanying result set featuring the surrogate type (Text, Image Preview, Text + Image Preview) used in the four subsequent study trials.

Study trials and training trials asked the participant to make relevance judgments about result set items. In the first page of result sets, participants looked at each page of non-hyperlinked results and rated items they deemed relevant or possibly relevant to the search scenario, selecting “Possibly relevant” or “Relevant” or leaving the rating level unselected (designating it “Not relevant”).

After completing a study trial for a particular surrogate type, the participant answered a questionnaire about the surrogate type. This post-session questionnaire asked participants to indicate their level of agreement with six statements, three to determine perceived ease of use, and three to determine perceived usefulness. Statements to determine perceived ease of use included: “I found this listing type to be flexible to interact with”; “I found this listing type easy to use”; and My interaction with this type of listing was clear and understandable.” Statements to determine perceived usefulness included: “Using this listing type enables me to find images more quickly”; “This listing type makes it easier to find images"; and "Using this listing type enhances my effectiveness in finding images." Participants rated their agreement with these statements on a scale of 1 to 5, where 1 is Strongly Agree, and 5 is Strongly Disagree. Statements used in this questionnaire were adapted from those used in Davis' (1989) study of technology acceptance. Since perceived usefulness and perceived ease of use affect system adoption, these measurements were important to consider in this study. The questionnaire concluded by asking what participants liked most and least about this surrogate type. After completing the questionnaire, the participant went on to the next study trials.

After completing the study trial and evaluation of the third surrogate type, the participant completed a final questionnaire comparing all surrogate types within the context of this study (Appendix C). This questionnaire asked for direct comparisons of the study tasks to their typical searching tasks and of how different they found the systems from one another, and then asked which of the three listing types they found easier to use, liked the best overall, and liked the least overall. This questionnaire also asked open-ended questions of participants, including: "The search results were displayed with both text and images. Which aspect of the display was most useful to you, and why?" as well as what they liked and disliked about each of the listing types. It concluded by asking for any additional comments.

Immediately after the session, the participant was contacted about obtaining the incentive and asked to fill out a receipt online by returning to the login page and entering their userid. Each participant chose whether or not to be remunerated for the study ($5) and whether to enter the drawing for a $20 Amazon gift certificate.

4.4 Data

4.4.1 Data Collected

During the study, information was collected about participant interaction with result sets, both from PHP and HTML functions and from questionnaires. Unobtrusive information collection included time for each user per page linked with the page-identifying information and surrogate type information (can be extrapolated to per scenario), participant's study start and end times, relevance judgments for result set listings (marking of listing as Relevant, Possibly Relevant or Non-relevant), markers for completion of scenario pages, for consent, and for completion of the study. Userids and group assignments were used as references for page navigation and routing through surrogate types, but were only collected during the login process.

Before beginning the study scenarios, consent or non-consent information was gathered for users, as were requests for remuneration, drawing entry, study result information, contact information and additional comments regarding payment. This information was stored in the MySQL database. The participant was then forwarded to a pre-session questionnaire (Appendix A), then on to the scenarios and their accompanying result sets. After entering relevance judgment selections for a training search scenario and four study scenarios, participants were asked to fill out a post-system questionnaire (Appendix B) addressing their experience with that surrogate type. After completing all search scenarios and post-system questionnaires, the participant was asked to complete a post-session questionnaire (Appendix C). All questionnaire responses were captured and stored in the MySQL database.

4.4.2 Data Analysis

Quantitative data analysis considered participant characteristics, speed of completion, "correctness" of relevance judgments, perceived ease of use, perceived usefulness, and individual preferences for particular surrogate types. Qualitative data analysis considered answers to open-ended questions on the questionnaires.

Data from the pre-session questionnaire was examined as means (age, years of experience in online searching) and frequencies (all other information). Additional analysis was conducted to determine whether differences existed between the participant groups, including: one-way analysis of variance (ANOVA) with descriptive statistics, using group (order in which encountered surrogate types) as an independent variable and age and computer experience as dependent variables. A crosstabulation analysis with group as the independent variable was run against each additional questionnaire value, yielding Fisher's Exact Test values and significances for other pre-session values.

The time it took for participants to judge the relevance of the result sets (Appendix H) was analyzed in a one-way ANOVA with descriptive statistics, using surrogate type as the independent variable and the time to complete the relevance judgments as the dependent variable. Post hoc analysis included least square difference (LSD) analysis.

"Correctness" of relevance judgments was determined by a scoring algorithm. The scoring algorithm compared the perceived relevance judgment entered for listings in scenario to the actual relevance value (determined during the researcher's initial selection for the result set list – either Relevant or Non-relevant). Scores for listings within a scenario were totaled, and that total scenario score was used in data analysis. Scoring for each listing was determined as follows, with a higher number score indicating a greater degree of correctness than a lower number score, as shown in Table 3.

Table 3. Basis of scoring of relevance judgments

Actual Relevance	Perceived Relevance (Judgment of Participant)	Score assigned
Relevant	Relevant	2
Relevant	Possibly Relevant	1
Relevant	Non-relevant	0
Non-relevant	Non-relevant	2
Non-relevant	Possibly Relevant	0
Non-relevant	Relevant	0

This scoring algorithm is biased in favor of a participant reaching a Relevant result listing, and recognizing a Relevant item as possessing some level of relevance and a Non-relevant item as possessing no relevance—in other words, making a "correct" assessment of an item's relevance. This algorithm did penalize participants for marking Non-relevant items as Possibly Relevant, since these were initially evaluated as possessing no relevance.

The correctness scores were examined using a one-way ANOVA with descriptive statistics, with surrogate type as the independent variable and a participant's total score (for a scenario) as the dependent variable, accompanied by a post hoc LSD analysis.

In analysis of perceived ease of use, responses to the three ease of use items in post-system questionnaires were averaged and then examined in a one-way ANOVA with descriptive statistics and post hoc LSD test, with surrogate type as the independent variable, and the ease of use score as the dependent variable.

Analysis of perceived usefulness was accomplished the same way, instead averaging the three usefulness items to form the dependent variable, using surrogate type as the independent variable, and examining one-way ANOVA, descriptive statistics, and post hoc LSD results.

Post-session questionnaire responses were analyzed in terms of frequencies of each level of similarity of tasks, difference of systems, and which listing type participants found easier to use, liked the best overall, and liked the least overall as dependent variables.

5. Results

5.1 Characteristics of the participants

The study participants included 15 women and 13 men. Their average age was 36.7 years (s.d.=14.47) and ranged from 21 to 62. All participants use a computer daily, and 27 have a great deal of experience in using a point-and-click interface (e.g., Macintosh OS, Microsoft Windows); one participant has slightly less experience. The participants averaged 10.5 years of experience in online searching (s.d.=2.31). They are frequent searchers. Twenty-four of the 28 participants conduct a search daily (3 weekly, 1 occasionally).

The participants varied in their familiarity with types of search systems (World Wide Web search services, computerized library catalogs, CD ROM systems, commercial online systems, and other systems; see Table 4). Participants are most familiar with World Wide Web search services, and 26 participants are also familiar with searching computerized library catalogs. Participants were less familiar with searching CD ROM systems, and were least familiar with commercial online systems. Some participants have experience with other general and specialized search systems, such as Google, Facebook, Wikipedia, commercial book catalogs, Intranet systems, Prophet (audio search system), and PIDI database, but 18 did not mention using an additional search system, and four did not specify the additional system used.

Table 4. Comparison of participant search system familiarity in pre-session questionnaire (number of participants giving each response)

How much experience have you had…	No experience		Some experience	A great deal of experience	
searching on computerized library catalogs either locally (e.g., your library) or remotely (e.g., Library of Congress)	1	1	9	7	10
searching on CD ROM systems (e.g., Encarta, Grolier, Infotrac)	4	6	13	3	2
searching on commercial online systems (e.g., BRS Afterdark, Dialog, Lexis-Nexis)	1	10	9	2	6
searching on world wide web search services (e.g., Alta Vista, Excite, Yahoo, HotBot, WebCrawler)	0	0	1	5	22
searching on other systems	4	0	1	1	4

Note: 18 participants did not specify that they searched on other systems.

All participants take or view pictures, but with varying frequency. Twenty-three take or view pictures on at least a monthly basis (nine monthly, eight weekly, five daily), and five do so occasionally. All participants but one actively search for images each month. Twenty-one search for images on an occasional or monthly basis (nine occasional, 12 monthly) and six search for images more frequently (four weekly, two daily). When they search for images, twenty-seven participants search online. Five also search in newspapers or magazines, five search in image archives or collections as well, and one searches in books. Twenty-four of 28 participants usually search for images by topic, and three search by title. One of those who search by title also searches by author, as do two other participants. One participant also searches by album.

Participants seek out images for particular needs. Their searches are targeted to fill informational, entertainment, and product-oriented requirements for personal, school, and work equities. Informational uses fall into a broad range, from finding out initial information (personal information or backgrounds, seeing what people or things—such as wildlife, buildings— look like, seeing referred pictures, maps, or diagrams, and product searches) to adding to information the seeker already possesses (clarifying information in a news article, adding to information about a subject or person, and understanding other people's interpretations of words—such as seeing what art pieces they might index with the term "yellow" on DeviantArt).

Participants entertain themselves in the course of image searches, finding images of friends or reunions, and looking at images they find "cute" or otherwise interesting. They add to the aesthetic qualities of their computer environment by using images they find as screen backgrounds or wallpaper, and as icons. They use images for personal communications and as information included in greetings to others.

Found images are used in the product lifecycle, both to inform the design concept and when assembling product content, as supplemental material. These products (coursework or essays, writing projects, educational material and lectures, artwork, illustrations, and presentations) often communicate information to others. Seven of 28 participants used images in presentation products (briefings, course lectures), and four used images in work-related publications, such as website biographies, letters to donors, and business proposals.

While participants were randomly assigned to groups (each group being exposed to the three representation types in a different order), an analysis was conducted to confirm that there were no differences in background characteristics between the three groups. Analysis of variance results for age and years of computer experience and

Fisher's exact test for the remaining variables indicated no statistically significant differences between the groups (p ranged from 0.060 to 1.000).

5.2 User performance

As described above, the time and scoring data reported in Table 5 were calculated across all 28 participants. The mean score and time is the average score and time achieved by *each person* on all the relevance judgments they performed.

Table 5. Summary of performance, by system

	Score		Time (in seconds)	
	Mean	*s.d.*	Mean	*s.d.*
Text	23.04	*5.774*	91	*228*
Image Preview	27.69	*3.506*	49	*25*
Text + Image Preview	27.69	*3.517*	53	*23*

Note: These data do not include data from the Training sets (Scenarios 1, 2, and 3).

5.2.1 Score

The "correctness" of the participants' relevance judgments was scored for each item in each results list, then summed across the items in each results list. The individual's scores were then averaged across all scenarios viewed in each of the three systems. The differences in score across systems were statistically significant (F= 41.350, p= 0.0000). Post Hoc Analysis indicated that both Image Preview and Text + Image Preview systems supported better relevance judgments Text system.

5.2.2 Time

The assessment of time it took each participant to review a results list and make relevance judgments was averaged across all the scenarios viewed in each of the three systems. The differences in time used to make relevance judgments across systems were statistically significant (F=3.346, p=0.0360). Post Hoc Analysis indicated participants took less time to perform judgments in the Image Preview system and the Text + Image Preview System than in the Text system.

5.3 User perceptions

Two measures of user perceptions were taken in relation to the three systems: perceived usefulness and perceived ease of use. After completing the four assigned scenarios (plus one training scenario) for each system, each participant completed the measures of usefulness (3 items) and ease of use (3 items). The results of these measures are shown in Table 6.

Table 6. User perceptions, based on post-system measures

	Perceived Ease of Use		Perceived Usefulness	
	Mean	*s.d.*	Mean	*s.d.*
Text	3.2	*1.287*	3.8	*1.427*
Image Preview	1.9	*0.924*	2.0	*0.964*
Text + Image Preview	1.7	*1.079*	1.9	*1.102*

Note: Lower scores indicate more positive attitudes.

Differences in perceived ease of use and perceived usefulness were both statistically significant (ease of use: F=15.361 p=0.00; usefulness: F=24.141, p=0.00). Post hoc analysis of perceived ease of use indicated that the Text + Image Preview and the Image Preview systems were viewed as easier to use than the Text system (p=0.00). Post hoc analysis of perceived usefulness indicated that participants perceived both Text + Image Preview and Image Preview systems as more useful than the Text system (p=0.00).

After working with all three systems, the participants were asked questions about the search tasks and systems, and were asked to make direct comparisons of the three systems; their responses are shown in Table 7 and Table 8. Most participants viewed systems as different from one another. The Text + Image Preview system was viewed as easier to use and liked the best overall. The Text system was liked least overall.

Table 7. Comparison of search tasks and systems in post-session questionnaire (number of participants giving each response)

	Not at all	Somewhat	Completely	No response
To what extent did you find these tasks similar to other searching tasks that you typically perform?	3	18	6	1
How different did you find the systems from one another?	2	19	6	1

Table 8. Comparison of systems in post-session questionnaire (number of participants giving each response)

	Text	Image Preview	Text + Image Preview	No difference
Easier to use	0	10	16	2
Liked the best overall	0	5	23	0
Liked the least overall	27	1	0	0

6. Discussion

The primary goal of the current study was to determine possible improvements to the results-browsing stage of the retrieval process, investigating user interaction, performance, and satisfaction with different types of image surrogates in the context of result sets. The study focused on result sets of search tasks that targeted images with specific conceptual and literal content. Result sets were represented in a familiar World Wide Web-style interface. Twenty-eight participants completed relevance judgment tasks for 15 search scenarios: a training scenario and four study scenarios for each type of surrogate. The surrogate types were: a Text surrogate, which contained author, copyright owner, copyright date, location (optional), title (optional, up to 50 characters), caption (optional, up to 50 characters), location in subject-category hierarchy, as entered in a commercial image database site, and resolution; an Image Preview surrogate, which contained a thumbnail image with associated filename (e.g., img099.jpg) of at least 1.92 effective pixel resolution; and a Text + Image Preview system which consisted of both Text and Image Preview system data. Participants made relevance judgments by marking a listing as Possibly Relevant, Relevant or Not Relevant (i.e., by not selecting it as Possibly Relevant or Relevant).

The overall finding of this study is that both Image Preview and Text + Image Preview systems were superior to the Text system for all measurements of performance (speed and score) and user perceptions (ease of use, usefulness). In addition to measurable superiority, examination of affective response, or user preferences ("liking"), of systems ranked the Text system as least "liked" of the three systems. This finding is in accordance with Davis' (1989) technology acceptance model, where perceived usefulness and ease of use are posited as valid predictors of system acceptance and use. Additionally, answers to the post-session questionnaire indicated that users prefer the Text + Image Preview surrogate over the Image Preview surrogate, as more people liked it best overall and more people considered it easier to use. In consideration that scoring and time differences between Text + Image Preview systems and Image Preview were negligible this preference is notable.

A difference emerged between surrogate types in time and scoring: both the Image Preview and Text + Image surrogates consistently outperformed the Text surrogate on these measures. From these results, we can infer the advantages of adding an image preview element to the surrogate. Participants noted in questionnaires that the image preview allowed them to assess the relevance listing without or before reading text elements, and to quickly discard non-relevant listings from consideration. Thus, the overall speed and "correctness" of relevance judgments improved from the availability of the image preview. Multiple participants reiterated, "A picture is worth a thousand words," and that the picture conveyed selection criteria with greater immediacy and less thinking required for relevance judgments. They stated that seeing only text information did not provide enough information to make relevance judgments for images and make comparisons between candidates. With images, they could see what they were getting, noting that words "can't convey images very accurately sometimes" and "it's difficult if not impossible to compare things as subjective as pictures through text alone." As experienced online searchers, participants have likely come to expect the inclusion of image previews in image search result sets (such as with Google Images searches), and that expectation may contribute to these findings. However, their explanations of how they used the image preview and how its presence affected their relevance judgments support the conclusion that this finding reflects more than habituation.

User perceptions of the usefulness and ease of use of a system affect their acceptance and use of a system (Davis, 1989). Perceived usefulness, particularly, is considered a consistent and valid predictor of system adoption (Ma & Liu, 2004). Study participants perceived the Text + Image Preview system and the Image Preview system as more useful and easier to use than the Text system. This perception held true throughout the study: in post-system questionnaires assessing ease of use and usefulness, and in a post-session questionnaire, where participants reported that the Text + Image Preview system was the easiest to use and most useful, followed by the Image Preview system, and then the Text system. This preference was also reflected in their affective responses during direct comparison of systems, as more participants selected the Text + Image Preview system as the one they liked best, followed by the Image Preview system. None liked the Text system best.

Open-ended comments on both the final and post-system questionnaires offer some insight into how participants formulated their assessments of system usefulness and affective judgments. Many participants thought the presence of an image preview element within the surrogate was essential to relevance judgments, and were pleased when it appeared in a surrogate system and displeased when it was absent. For them, looking at the image preview enabled quick inclusion or exclusion judgments. They also noted that the addition of text elements assisted in sorting and discrimination, especially if one wasn't sure what one was looking at, one needed to clarify context or content, or one needed information not embedded in the image (such as resolution, author, and copyright). They noted that, on the first look, they determined likely candidates from the image previews, and then used the text to further discriminate and sort – eliminating candidates and confirming or comparing content and other information pertinent to final selection decisions. When an image element was not available, participants did note that using the text element helped with elimination and comparison—but not as well as the image element. Inclusion of the elements in both Text and Image Preview surrogates added contextual and content information that made a difference to user perceptions—they commented that seeing the content, form, and color of an image (image preview) and the context and associated details (text) was the most helpful combination of access points. These comments are very consistent with the data from a recent eye-tracking study (Hughes et al., 2003) that examined use of both textual and image surrogates; the eye-tracking data confirm that people use the images to make initial judgments and confirm those judgments by consulting the textual data. Considering that participants viewed the Text system as the least useful, least easy to use and least liked, but the combination of Text + Image Preview as the most useful and easy to use and liked, we can conclude that, while the image preview part of a surrogate is a key element, the addition of some information absent in images and present in text descriptions adds to ease of use and usefulness of a system and, thus, its likely adoption.

Users expressed several other preferences in the post-system and post-session questionnaire, for both surrogate access points and user-interface design features. What access points the surrogate should include and what constraints or considerations should affect access point values were addressed in open-ended comments.

What to include in or exclude from image surrogates varied across users, but all agreed that the image preview element was essential. Size of images was a factor, not particularly in terms of screen space, but in how well the thumbnail expressed clarity and detail of content—aspects which can vary across images during the thumbnail-generation process, for instance, in a focused close-up versus a landscape. The addition of Zoom functionality or linkage to a larger image that does not require abandoning the current screen is a possible solution to this concern.

For some users, resolution (a measure of quality), source (website), and copyright information were important and served as exclusion criteria. For instance, when asked to choose the best picture or when considering its use (e.g., in a publication), they would factor in resolution and copyright when making their choices and eliminate candidates based on that access point value. Participants also commented that the inclusion of a title for the image helped in selection, as did the category (e.g., Top Downloads > Animals). Descriptions were viewed as helpful in some cases, but not always useful, substantive, objective, or accurate. Users preferred shorter descriptions (such as titles) that reflected content of images, and commented that the quality and trustworthiness of descriptions varied both in this study and on the Internet as a whole. They considered non-descriptive image filenames visual clutter, and extended that assessment to parts of the text surrogate that did not address their needs, considering the excess information distracting. A possible approach for enabling inclusion of helpful access points and exclusion of "distracting" additional information is a search interface that allows either individual access point selection for inclusion or exclusion or selection of a surrogate profile, each profile specifying a set of commonly selected access points. Thus, it is recommended that image retrieval system designers should include textual metadata, but should focus the content of descriptive elements on substantive and objective commentary about image content. Additionally, they should limit the length of in-depth descriptions.

Participants additionally expressed preferences for user interface design features. They commented that it would be useful at the initial result set return to see results presented in logical groupings, where listings sharing similarities would be arranged in proximity to one another. Users also expressed a preference for uniformity of presentation and clear organization of result listings within result sets.

Participants further mentioned they would prefer not to have to scroll up and down pages, to have a scrollable page of listings on one page with a Next button linking to additional pages of listings if necessary. This preference may require design tradeoffs with surrogate elements, linkages to enhanced views (Zoom, pop-up windows, or rollover functionality for additional text or image detail), or tabbed groupings of result sets by similarity or category to enable convenient comparisons without requiring excessive scrolling or screen navigation.

Finally, participants related a desire for the ability to rearrange or group result sets by user action, for instance, to place Relevant and Possibly Relevant listings together for further comparisons, eliminating listings judged to be non-relevant.

As with any empirical study, the methods used in this study had weaknesses. In particular, the scoring method used in this study likely had an effect on performance measurements, and a different scoring algorithm may have returned somewhat different findings. This scoring algorithm was weighted in favor of finding a Relevant result, and penalized participants for selecting non-relevant results as Relevant or Possibly Relevant. While this scoring algorithm was developed to mimic the penalties associated with making incorrect relevance judgments in realistic image retrieval situations, other researchers may have chosen a different tactic.

7. Conclusion

User interaction with a result set is a pivotal stage of the search process. The information users receive in this step influences not only their satisfaction with the system but how effectively they achieve their search goals. Providing information in such a way that it assists a user's ability to review results and make relevance judgments quickly and accurately is a prime opportunity to optimize the search experience. Because images are costly (in terms of time and screen space) to display in the result set, and do not always provide all the information a user needs to make a relevance judgment for a search scenario, optimizing surrogates and the access points they provide would make result browsing more effective and satisfying for users. The type of surrogate used in search scenario result sets was manipulated in the current study to investigate their effects on user performance and perceptions. Study results indicated that the Image Preview and Text + Image Preview surrogates outperformed the Text surrogate on all measures: of time, relevance judgment scoring, and perceived ease of use and usefulness. Results also showed that, while relevance judgment scoring was identical between Image Preview and Text + Image Preview surrogates, answers to the post-session questionnaire indicate that users prefer the Text + Image Preview surrogate, as more people liked it best overall and more people considered it easier to use.

Future studies should focus on three things: surrogate composition and presentation, optimization of surrogates for specific search scenario types, and contextualization or grouping of surrogates. Studying surrogate composition and presentation would assist in determining high and low thresholds of indexing requirements, and isolating the most effective spatial representation of surrogates and their composite access points for information extraction. Determining optimal surrogates for specific search scenario types (high or low specificity, open-ended or close-ended, specific prospective uses of items sought) would enable users to specify their search type and ensure that necessary access points were included and superfluous access points were omitted from surrogates, speeding processing time. Examining the effect of grouping results according to either commonalities of access points or contextual pointers (i.e., results from each source category are grouped together) could enable user selection of access point weighting and use of spatial arrangement to enhance relevance decision-making.

8. References

Aslandogan, Yuksel Alp and Clement T. Yu. (2000)“Evaluating Strategies and Systems for Content Based Indexing of Person Images on the Web.” Los Angeles, *ACM Multimedia 2000*, 313-321.

Borlund, P. (2003) “The IIR evaluation model: A framework for evaluation of interactive information retrieval systems.” *Information Research, 8*(3), paper no. 152. [Available at: http://informationr.net/ir/8-3/paper152.html]

Chang, Shih-Fu , John R. Smith, Mandis Beigi, and Ana Benitez. (1997) “Visual Information Retrieval from Large Distributed Online Repositories.” *COMMUNICATIONS OF THE ACM 40*(12), 63-71.

Chen, Hsin-Liang. (2001) "An Analysis of Image Retrieval Tasks in the Field of Art History." *Information Processing & Managment 37* (5) , 701-20.

Chen, Hsin-Liang, and EM Rasmussen. (1999) "Intellectual Access to Images." *Library Trends 48* (2), 291-302.

Cole, J.I., Suman M., Schramm, P., Lunn, R., Aquino, J-S and H. Lebo. (2003) “The UCLA Internet Report: Surveying the Digital Future, Year Three.” (UCLA) UCLA Center for Communication Policy. [Available at: http://ccp.ucla.edu/pdf/UCLA-Internet-Report-Year-Three.pdf]

Davis, Fred D. (1989) “Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology.” *MIS Quarterly*, *13*(3), 318-340.

Dumais, Susan , Edward Cutrell, and Hao Chen. (2000) “Optimizing Search by Showing Results In Context.” *CHI 2000*, *3*(1), 277-284.

Dziadosz, Susan and Raman Chandrasekar. (2002) “Do Thumbnail Previews Help Users Make Better Relevance Decisions about Web Search Results?” *SIGIR’02*, August 11-15, 2002, Tampere, Finland. ACM 2002, 365-366.

Efthimiadis, Efthimis, and Raya Fidel. (2000) "The Effect of Query Type on Subject Searching Behavior of Image Databases: an Exploratory Study." *SIGIR 2000: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Eds. Nicholas J. Belkin, Peter Ingwersen, and Mun-Kew Leong, 328-30. New York: ACM, 2000.

Enser, P.G.B., and C.G. McGregor. (1993) "Analysis of visual information retrieval queries." (British Library R&D Report No. 6104), London, England: British Library Board. Enser, P.G.B. (1995) "Pictorial information retrieval." *Journal of Documentation*. *51*(2), 126-170.

Fidel, Raya. (1997) "Image Retrieval task: Implications for the Design and Evaluation of Image Databases." *The New Review of Hypermedia and Multimedia, 3, 181-199.*

Frost, C. Olivia, Bradley Taylor, Anna Noakes, Stephen Markel, Deborah Torres, Karen M. Drabenstott. (2000) "Browse and Search Patterns in a Digital Image Database." *Information Retrieval*, *1*(4), 287–313.

Garber, Sharon R., and Mitch B. Grunes. (1992) "The Art of Search: a Study of Art Directors." *CHI '92*, 157-63. New York: ACM, 1992.

Hastings, Samantha K. (1999) "Evaluation of Image Retrieval Systems: Role of User Feedback." *Library Trends 48*(2), 438-52.

Hastings, Samantha K. (1995) "Query Categories in a Study of Intellectual Access to Digitized Art Images." In T. Kinney (Ed.), *ASIS '95* (Proceedings of the 58th annual meeting of the American Society for Information Science, October 9-12, 1995, Chicago, IL.). Medford, NJ: American Society for Information Science, 3-8.

Heidorn, P. Bryan. (1999) "Image Retrieval as Linguistic and Nonlinguistic Visual Model Matching." *Library Trends 48*(2), 303-24.

Hirata, Kyoji, Sougata Mukherjea, Wen-Syan Li, and Yoshinori Hara. (2000) "Integration of Image Matching and Classification for Multimedia Navigation." *Multimedia Tools and Applications*, *11*(3), 295–309.

Hughes, A., Wilkens, T., Wildemuth, B., & Marchionini, G. (2003) "Text or pictures? An eyetracking study of how people view digital video surrogates." *Proceedings of CIVR 2003*, 271-280. [Available at http://www.open-video.org/papers/hughes_civr_2003.pdf.]

Johnson, J. (2000) *GUI Bloopers: Don'ts and Do's for Software Developers and Web Designers.* San Francisco: Morgan Kaufmann.

Jorgensen, Corinne. (1998) "Attributes of Images in Describing Tasks." *Information Processing & Management 34*(2/3), 161-174.

Jorgensen, Corinne. (1995) "Image attributes: a investigation." Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.

Layne, S.S. (1994) "Some issues in the indexing of images." Journal of the American Society for Information Science, *45*(8), 583-588.

Ma, Q., and Liu, L. (2004) "The technology acceptance model: A meta-analysis of empirical findings." *Journal of Organizational and End-User Computing, 16*(1), 59-72.

Marchionini,G. (1995) *Information Seeking in Electronic Environments*. Cambridge University Press.

Markkula, Marjo, and Eero Sormunen. (2000) "End-User Searching Challenges Indexing Practices in the Digital Newspaper Photo Archive." *Information Retrieval 1*(4), 259–285.

Mukherjea, Sougata, Kyoji Hirata and Yoshinori Hara. (1999) "AMORE: A World Wide Web image retrieval engine." *World Wide Web 2*(3), 115–132.

"NUA Internet Surveys: Photo-sharing sites prove popular." (2003) (NUA)[Available at: http://www.nua.ie/surveys/index.cgi?f=VS&art_id=905358710&rel=true]

Panofsky, Erwin. (1939) *Studies in iconology: Humanistic themes in the art of the Renaissance.* New York: Oxford University Press.

Panofsky, Erwin. (1955) *Meaning in the visual arts: papers in and on art history.* Garden City, N.Y.: Doubleday Anchor Books.

Rainie, Lee and Jeremy Shermak. (2005) "Big Jump in Search Engine Use." Pew Internet & American Life Project Memo, 1-9.
[Available at http://www.pewinternet.org/pdfs/PIP_SearchData_1105.pdf]

Rodden, Kerry, Wojciech Basalaj, David Sinclair, and Kenneth Wood. (2001) "Does Organisation by Similarity Assist Image Browsing?" *CHI 2001 3*(1), 190-197.

Rui, Yong, Michael Ortega, Thomas S. Huang, and Sharad Mehrotra .(1999) "Information Retrieval Beyond the Text Document." *Library Trends 48*(2), 455-74.

Shatford, S. (1986) "Analyzing the subject of a picture: A theoretical approach." *Cataloging & Classification Quarterly*. *6*(3), 39-62.

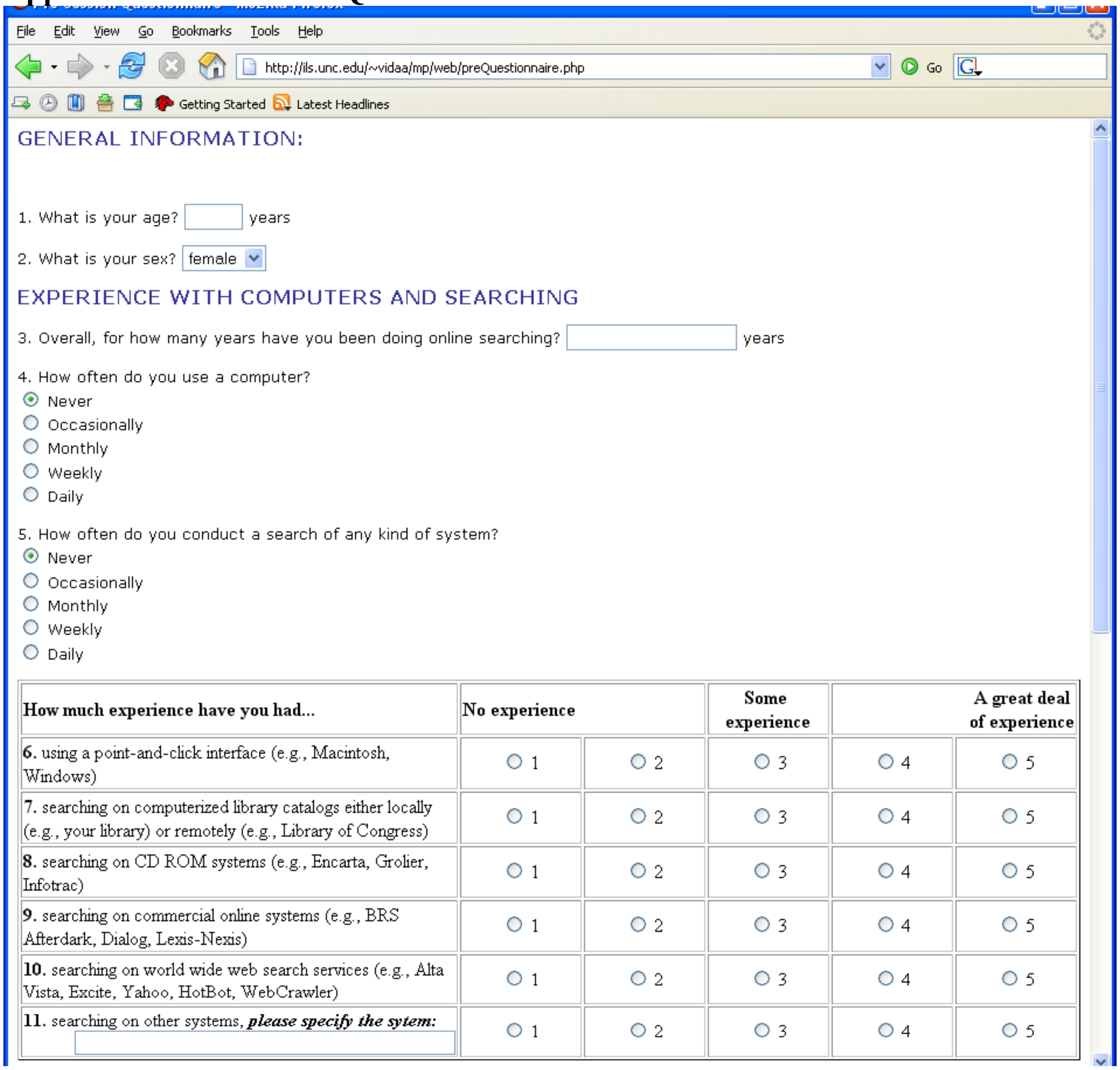
Woodruff, Allison, Andrew Faulring, Ruth Rosenholtz, Julie Morrison, Peter Pirolli. (2001) "Using Thumbnails to Search the Web." *CHI 2001 3*(1), 198-205.

Wu, Jian Kang and A. Desai Narasimhalu. (1998) "Fuzzy Content-Based Retrieval in Image Databases." *Information Processing & Management 34*(5), 513-534.

Yee, Ka-Ping, Kirsten Swearingen, Kevin Li, Marti Hearst. (2003) "Faceted Metadata for Image Search and Browsing." *CHI 2003 5*(1), 401-408.

Zhu, Lei , Aibing Rao and Aidong Zhang. (2002) “Theory of Keyblock-Based Image Retrieval.” *ACM Transactions on Information Systems 20*(2), 224–257.

Appendix A: Pre-Session Questionnaire



GENERAL INFORMATION:

1. What is your age? [ ] years

2. What is your sex? [female]

EXPERIENCE WITH COMPUTERS AND SEARCHING

3. Overall, for how many years have you been doing online searching? [ ] years

4. How often do you use a computer?
- ◉ Never
- ○ Occasionally
- ○ Monthly
- ○ Weekly
- ○ Daily

5. How often do you conduct a search of any kind of system?
- ◉ Never
- ○ Occasionally
- ○ Monthly
- ○ Weekly
- ○ Daily

How much experience have you had...	**No experience**		**Some experience**		**A great deal of experience**
6. using a point-and-click interface (e.g., Macintosh, Windows)	○ 1	○ 2	○ 3	○ 4	○ 5
7. searching on computerized library catalogs either locally (e.g., your library) or remotely (e.g., Library of Congress)	○ 1	○ 2	○ 3	○ 4	○ 5
8. searching on CD ROM systems (e.g., Encarta, Grolier, Infotrac)	○ 1	○ 2	○ 3	○ 4	○ 5
9. searching on commercial online systems (e.g., BRS Afterdark, Dialog, Lexis-Nexis)	○ 1	○ 2	○ 3	○ 4	○ 5
10. searching on world wide web search services (e.g., Alta Vista, Excite, Yahoo, HotBot, WebCrawler)	○ 1	○ 2	○ 3	○ 4	○ 5
11. searching on other systems, ***please specify the sytem:*** [ ]	○ 1	○ 2	○ 3	○ 4	○ 5

(continued on next page)

IMAGE USE:

12. How often do you take or view pictures?
- (•) Never
- ( ) Occasionally
- ( ) Monthly
- ( ) Weekly
- ( ) Daily

13. How often do you search for images?
- (•) Never
- ( ) Occasionally
- ( ) Monthly
- ( ) Weekly
- ( ) Daily

14. When you search for images, where do you go? (Check all that apply)
- [ ] Online
- [ ] Newspaper or magazine
- [ ] Image archives or collections
- [ ] Other:

15. How do you usually search for images? (Check all that apply)
- [ ] By title
- [ ] By author
- [ ] By topic
- [ ] By preview (thumbnail)
- [ ] Other:

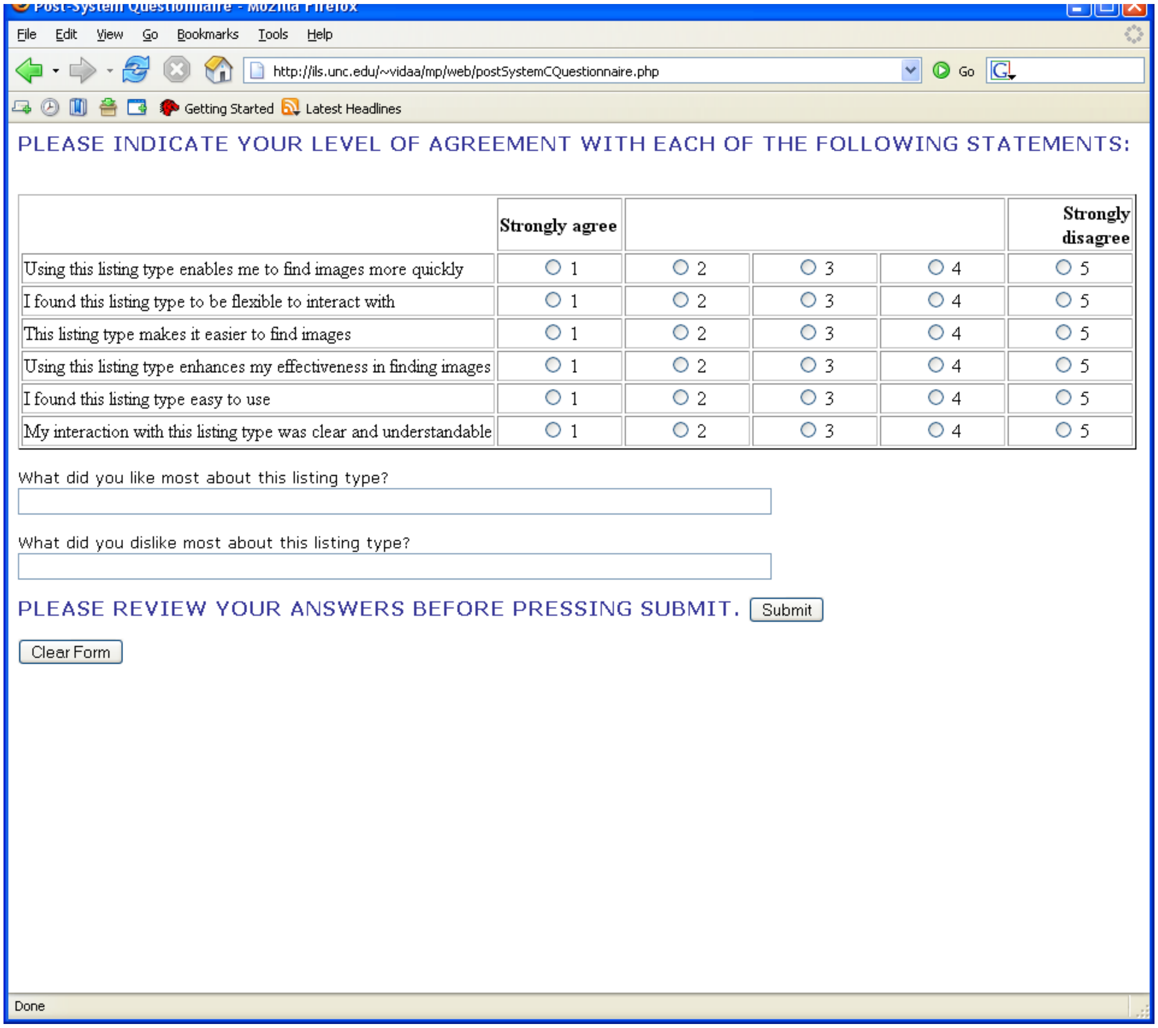
16. For what purposes do you usually search for images?

PLEASE REVIEW YOUR ANSWERS BEFORE PRESSING SUBMIT. Submit

Clear Form

Done

Appendix B : Post-System Questionnaire



PLEASE INDICATE YOUR LEVEL OF AGREEMENT WITH EACH OF THE FOLLOWING STATEMENTS:

	Strongly agree				Strongly disagree
Using this listing type enables me to find images more quickly	1	2	3	4	5
I found this listing type to be flexible to interact with	1	2	3	4	5
This listing type makes it easier to find images	1	2	3	4	5
Using this listing type enhances my effectiveness in finding images	1	2	3	4	5
I found this listing type easy to use	1	2	3	4	5
My interaction with this listing type was clear and understandable	1	2	3	4	5

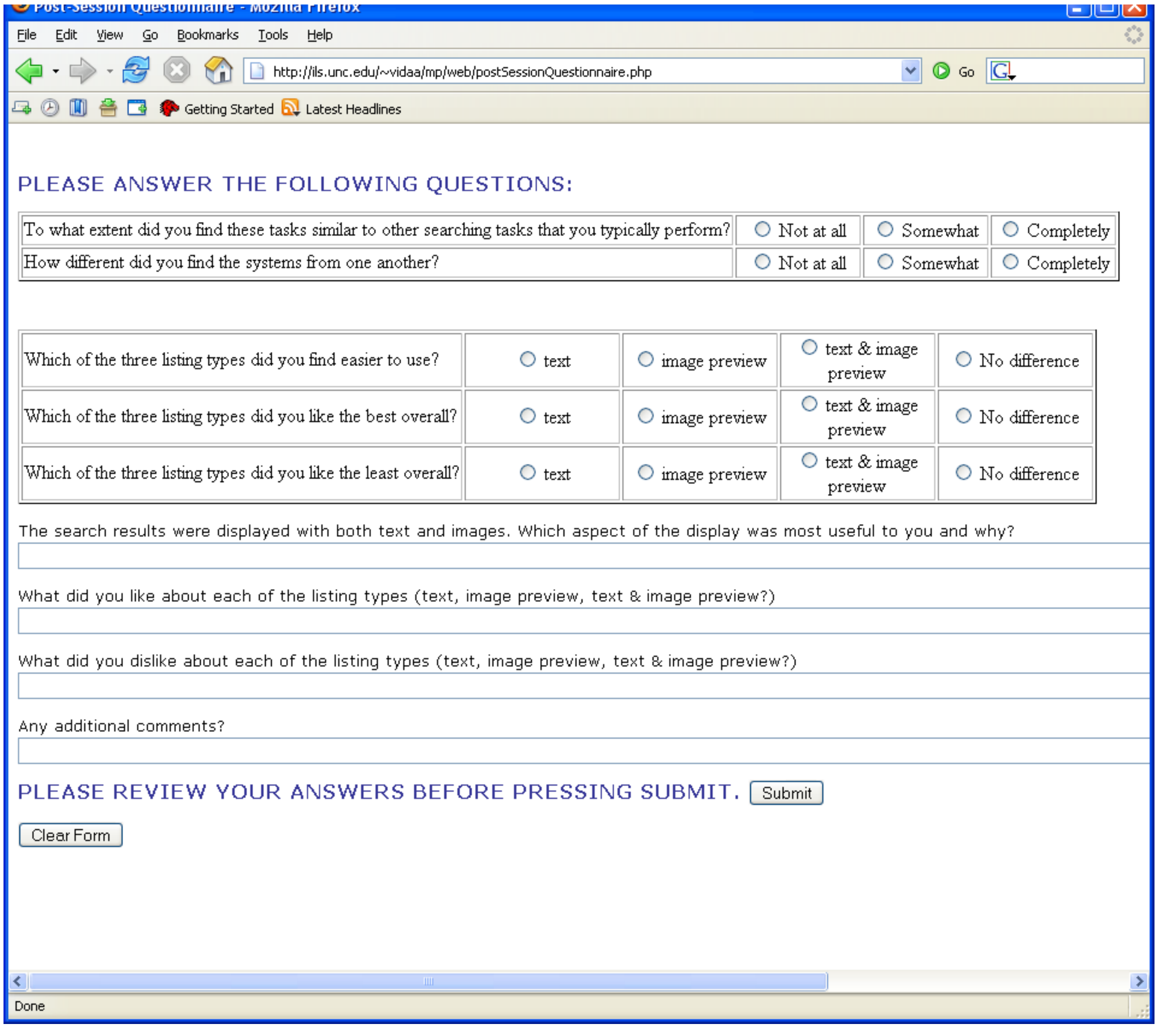
What did you like most about this listing type?

What did you dislike most about this listing type?

PLEASE REVIEW YOUR ANSWERS BEFORE PRESSING SUBMIT. Submit

Clear Form

Appendix C: Post-session Questionnaire



Post-Session Questionnaire - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://ils.unc.edu/~vidaa/mp/web/postSessionQuestionnaire.php

Getting Started Latest Headlines

PLEASE ANSWER THE FOLLOWING QUESTIONS:

To what extent did you find these tasks similar to other searching tasks that you typically perform?	○ Not at all	○ Somewhat	○ Completely
How different did you find the systems from one another?	○ Not at all	○ Somewhat	○ Completely

Which of the three listing types did you find easier to use?	○ text	○ image preview	○ text & image preview	○ No difference
Which of the three listing types did you like the best overall?	○ text	○ image preview	○ text & image preview	○ No difference
Which of the three listing types did you like the least overall?	○ text	○ image preview	○ text & image preview	○ No difference

The search results were displayed with both text and images. Which aspect of the display was most useful to you and why?

What did you like about each of the listing types (text, image preview, text & image preview?)

What did you dislike about each of the listing types (text, image preview, text & image preview?)

Any additional comments?

PLEASE REVIEW YOUR ANSWERS BEFORE PRESSING SUBMIT. Submit

Clear Form

Done

Appendix D: Search Scenarios

#	Order appears in study		Scenario Name	Scenario Text	Open or Closed ended	Granularity
1	1	Training – Block 1	Washington Monument	You are developing a brochure for an upcoming school trip to Washington D.C. and need a picture of the Washington Monument to put on the front. After entering a search for "Washington Monument" you see listings for a number of U.S. Monuments. Pick the best picture of the Washington Monument for your brochure.	close-ended	high specificity
2	6	Training – Block 2	Eiffel Tower	You are developing a brochure for an upcoming school trip to Paris. and need a picture of the Eiffel Tower to put on the front. After entering a search for "Eiffel Tower" you see listings for a number of French Monuments. Pick the best picture of the Eiffel Tower for your brochure.	close-ended	high specificity
3	11	Training – Block 3	Sphynx and Pyramid	You are developing a brochure for an upcoming school trip to Egypt, and need a picture of a pyramid and the Sphinx, to put on the front. After entering a search for "pyramid" you see listings for a number of pyramid monuments across the globe. Pick the best picture of the Sphinx and pyramid for your brochure.	close-ended	high specificity
4	9	Study – Block 2	Cactus Sunset	You are designing some flyers for a weekend dude ranch in Arizona and need a picture of something Southwestern at sunset, like a cactus. After entering a search for "Cactus" you see listings for a number of results featuring cacti. Pick a few pictures of a cactus at sunset to review later for use in your flyers.	open-ended	high specificity
5	2	Study – Block 1	Forest Path	You are designing some flyers for a weekend hiking	open-ended	high specificity

#	Order appears in study		Scenario Name	Scenario Text	Open or Closed ended	Granularity
				trip and need a picture of a forest with a path running through it. After selecting a category called “Forests” you see listings for a number of results featuring forests. Pick a few pictures of a forest with a path running through it to review later for use in your flyers.		
6	14	Study – Block 3	Ocean at Dusk	You are designing some flyers for a trip to the beach over Valentine's Day weekend and need a picture of the ocean at dusk to set a romantic mood. After selecting a category called “Oceans” you see listings for a number of results featuring oceans. Pick a few pictures of the ocean when night is falling to review later for use in your flyers.	open-ended	high specificity
7	7	Study – Block 2	Group of Horses	You are looking for a picture to use as a screensaver, and decide that you want to use an image of a group of horses. After entering a search for “Horse” you see listings for a number of results featuring horses. Pick the best picture of a group of horses for your screensaver.	close-ended	high specificity
8	12	Study – Block 3	Fall Path	You are looking for a picture to use as a screensaver, and decide that you want to use an image of the forest in autumn with a path running through it. After selecting a category called “Fall colors” you see listings for a number of results featuring autumn leaves. Pick the best picture of an autumn forest with a path for your screensaver.	close-ended	high specificity
9	4	Study – Block 1	Mountain Lake	You are looking for a picture to use as a screensaver, and decide that you want to use an image of a mountain and a lake . After selecting a category called “Mountains” you see listings for a number of results featuring	close-ended	high specificity

#	Order appears in study		Scenario Name	Scenario Text	Open or Closed ended	Granularity
				mountains. Pick the best picture of a mountain with a lake for your screensaver.		
10	5	Study – Block 1	Top Download Wolf	You are looking for some good-quality pictures of a wolf to use in a report on environmental issues. You decide to look at the “Top Downloads” for the “Animal” category, since those are probably fairly good pictures. After selecting “Top Downloads”, you see listings for a number of results featuring animals. Pick a few pictures of a wolf to review later for use in your report.	open-ended	low specificity
11	13	Study – Block 3	Top Download Lion	You are looking for some good-quality pictures of a lion to use in a report on environmental issues. You decide to look at the “Top Downloads” for the “Animal” category, since those are probably fairly good pictures. After selecting “Top Downloads”, you see listings for a number of results featuring animals. Pick a few pictures of a lion to review later for use in your report.	open-ended	low specificity
12	10	Study – Block 2	Top Download Tiger	You are looking for some good-quality pictures of a tiger to use in a report on environmental issues. You decide to look at the “Top Downloads” for the “Animal” category, since those are probably fairly good pictures. After selecting “Top Downloads”, you see listings for a number of results featuring animals. Pick a few pictures of a tiger to review later for use in your report.	open-ended	low specificity
13	8	Study – Block 2	Top Downloads Sailing	You are looking for some good-quality pictures of sailing to use on a flyer advertising a summer resort. You decide to look at the	close-ended	low specificity

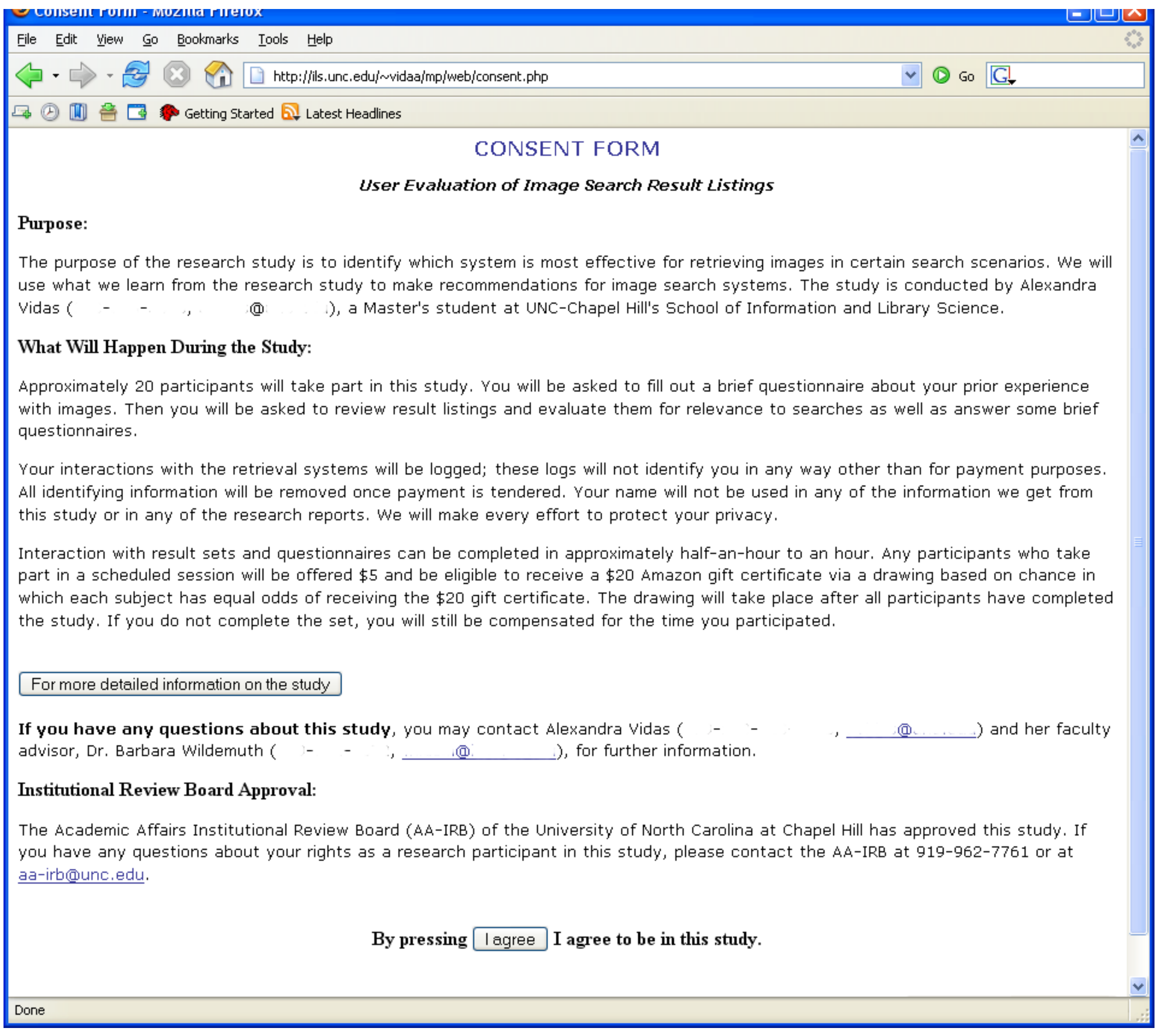
#	Order appears in study		Scenario Name	Scenario Text	Open or Closed ended	Granularity
				"Top Downloads" for the "Sports" category, since those are probably fairly good pictures. After selecting "Top Downloads", you see listings for a number of results featuring sports. Pick the best picture of sailing to use on your flyer.		
14	3	Study – Block 1	Top Downloads Surfing	You are looking for some good-quality pictures of surfing to use on a flyer advertising a summer resort. You decide to look at the "Top Downloads" for the "Sports" category, since those are probably fairly good pictures. After selecting "Top Downloads", you see listings for a number of results featuring sports. Pick the best picture of surfing to use on your flyer.	close-ended	low specificity
15	15	Study – Block 3	Sports Mix baseball	You are looking for some good-quality pictures of baseball to use on a flyer advertising a summer camp. After selecting the "Sports Mix" category, you see listings for a number of results featuring sports. Pick the best picture featuring baseball to use on your flyer.	close-ended	low specificity

Appendix E: Participants and their surrogate orderings

Participant #	Surrogate Type –Block 1 (5, 14, 9, 10) (scenarios randomly drawn from pool to fit specificity and open or close-ended criteria)	Surrogate Type – Block 2 (7, 13, 4, 12) (scenarios randomly drawn from pool to fit specificity and open or close-ended criteria)	Surrogate Type - Block 3 (8, 11, 6, 15) (scenarios randomly drawn from pool to fit specificity and open or close-ended criteria)
Note: All blocks alternate high specificity and low specificity as follows: High/Low/High/Low. Ordering of close-ended and open-ended scenarios varies per block as follows: Block 1: Open/Closed/Open/Closed; Block 2: Closed/Closed/Open/Open; Block 3: Closed/Open/Open/Closed			
1	Text	Image Preview	Text & Image Preview
2	Text	Text & Image Preview	Image Preview
3	Image Preview	Text	Text & Image Preview
4	Image Preview	Text & Image Preview	Text
5	Text & Image Preview	Text	Image Preview
6	Text & Image Preview	Image Preview	Text
7	Text	Image Preview	Text & Image Preview
8	Text	Text & Image Preview	Image Preview
9	Image Preview	Text	Text & Image Preview
10	Image Preview	Text & Image Preview	Text
11	Text & Image Preview	Text	Image Preview
12	Text & Image Preview	Image Preview	Text
13	Text	Image Preview	Text & Image Preview
14	Text	Text & Image Preview	Image Preview
15	Image Preview	Text	Text & Image Preview
16	Image Preview	Text & Image Preview	Text
17	Text & Image Preview	Text	Image Preview
18	Text & Image Preview	Image Preview	Text
19	Text	Image Preview	Text & Image Preview
20	Text	Text & Image Preview	Image Preview
21	Image Preview	Text	Text & Image Preview
22	Image Preview	Text & Image Preview	Text
23	Text & Image Preview	Text	Image Preview
24	Text & Image Preview	Image Preview	Text
25	Text	Text & Image Preview	Image Preview
26	Text	Text & Image Preview	Image Preview
27	Image Preview	Text	Text & Image Preview
28	Text & Image Preview	Text	Image Preview

Note: Participants 1-24 were counter-balanced for surrogate ordering. Participants 25-28 were additional subjects.

Appendix F: Consent Form



CONSENT FORM

User Evaluation of Image Search Result Listings

Purpose:

The purpose of the research study is to identify which system is most effective for retrieving images in certain search scenarios. We will use what we learn from the research study to make recommendations for image search systems. The study is conducted by Alexandra Vidas (- - , @), a Master's student at UNC-Chapel Hill's School of Information and Library Science.

What Will Happen During the Study:

Approximately 20 participants will take part in this study. You will be asked to fill out a brief questionnaire about your prior experience with images. Then you will be asked to review result listings and evaluate them for relevance to searches as well as answer some brief questionnaires.

Your interactions with the retrieval systems will be logged; these logs will not identify you in any way other than for payment purposes. All identifying information will be removed once payment is tendered. Your name will not be used in any of the information we get from this study or in any of the research reports. We will make every effort to protect your privacy.

Interaction with result sets and questionnaires can be completed in approximately half-an-hour to an hour. Any participants who take part in a scheduled session will be offered $5 and be eligible to receive a $20 Amazon gift certificate via a drawing based on chance in which each subject has equal odds of receiving the $20 gift certificate. The drawing will take place after all participants have completed the study. If you do not complete the set, you will still be compensated for the time you participated.

[For more detailed information on the study]

If you have any questions about this study, you may contact Alexandra Vidas (- - , @) and her faculty advisor, Dr. Barbara Wildemuth (- - , @), for further information.

Institutional Review Board Approval:

The Academic Affairs Institutional Review Board (AA-IRB) of the University of North Carolina at Chapel Hill has approved this study. If you have any questions about your rights as a research participant in this study, please contact the AA-IRB at 919-962-7761 or at aa-irb@unc.edu.

By pressing [I agree] I agree to be in this study.

Appendix G: More information

ADDITIONAL INFORMATION

User Evaluation of Image Search Result Listings

Introduction to the Study:

We are inviting you to be involved in a research study of three alternative image result listing types. Your evaluation of them will help to improve their usability for others and possibly improve searching in image databases. The study is conducted by a Master's student at UNC-Chapel Hill's School of Information and Library Science, Alexandra Vidas (- - , @).

Purpose:

The purpose of the research study is to identify which system is most effective for retrieving images in certain search scenarios. We will use what we learn from the research study to make recommendations for image search systems.

What Will Happen During the Study:

Approximately 20 participants will take part in this study. You will be asked to fill out a brief questionnaire about your prior experience with images. Then you will be asked to complete a training trial, 4 study trials, and a brief evaluation for each listing type (3 in total). The trials consist of webpages displaying 15 result listings each. Listing types include text, image preview and a combination of text and image preview. Search scenarios will be given to you before viewing the search result set. You will be asked to rank the results and to select the order in which you would view them. After completing these steps for all 3 listing types, you will be asked to fill out a brief questionnaire.

Your interactions with the retrieval systems will be logged; these logs will not identify you in any way other than for enabling payment. Interaction with result sets and questionnaires can be completed in approximately half-an-hour to an hour. Any participants who take part in a scheduled session will be offered $5 and be eligible to receive a $20 Amazon gift certificate via a drawing based on chance in which each subject has equal odds of receiving the $20 gift certificate. The drawing will take place after all participants have completed the study. If you do not complete the study, you will still be compensated for the time you participated.

Your Privacy is Important:

- We will make every effort to protect your privacy.
- We will not use your name in any of the information we get from this study or in any of the research reports.
- The logs of your transactions will not contain any information that can identify you other than that necessary for enabling payment; after payment and the drawing for the Amazon gift certificate, that information will be erased.

Risks and Discomforts:

We do not know of any personal risk or discomfort you will have from being in this study.

Your Rights:

- You decide on your own whether or not you want to be in this study.
- If you decide to be in the study, you will have the right to stop being in the study at any time.
- You have the right to skip any question you choose not to answer.

If you have any questions about this study, you may contact Alexandra Vidas (- - , @) and her faculty advisor, Dr. Barbara Wildemuth (- - , wildem@ils.unc.edu), for further information.

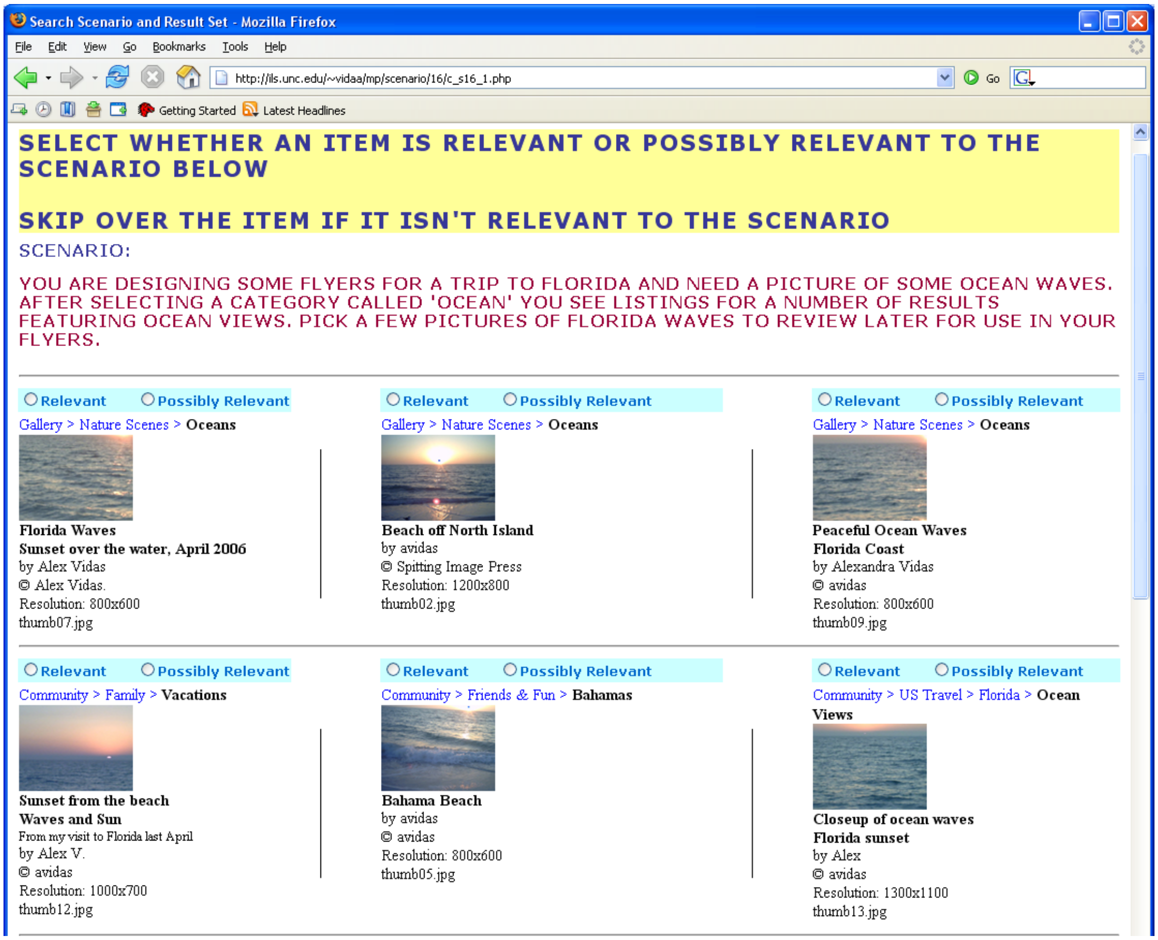
Institutional Review Board Approval:

The Academic Affairs Institutional Review Board (AA-IRB) of the University of North Carolina at Chapel Hill has approved this study. If you have any questions about your rights as a research participant in this study, please contact the AA-IRB at 919-962-7761 or at aa-irb@unc.edu.

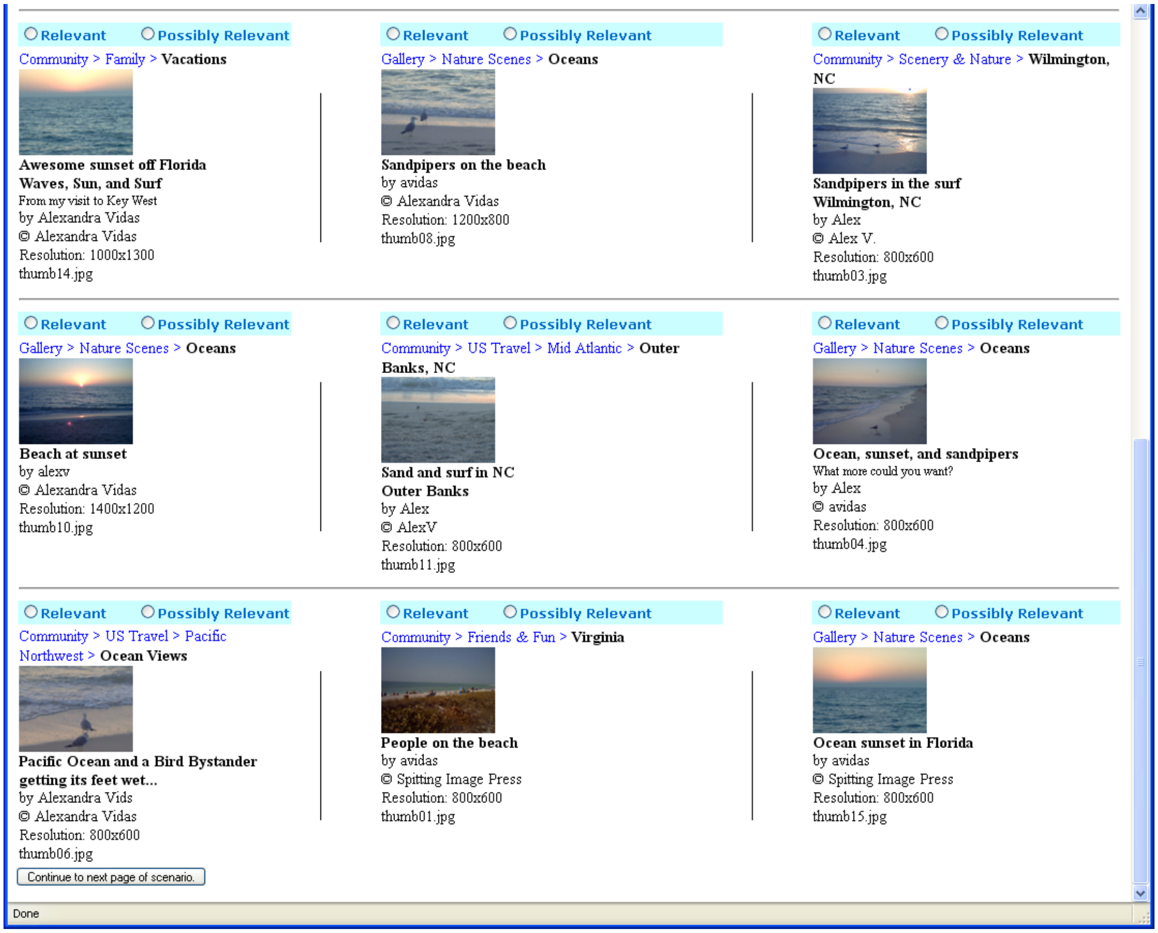
By pressing [I agree] **I agree to be in this study.**

[I do not want to participate]

Appendix H: Relevance Judgment Interface



(continued on next page)


Relevant
Possibly Relevant
Community > Family > Vacations
Awesome sunset off Florida
Waves, Sun, and Surf
From my visit to Key West
by Alexandra Vidas
© Alexandra Vidas
Resolution: 1000x1300
thumb14.jpg
Relevant
Possibly Relevant
Gallery > Nature Scenes > Oceans
Sandpipers on the beach
by avidas
© Alexandra Vidas
Resolution: 1200x800
thumb08.jpg
Relevant
Possibly Relevant
Community > Scenery & Nature > Wilmington, NC
Sandpipers in the surf
Wilmington, NC
by Alex
© Alex V.
Resolution: 800x600
thumb03.jpg
Relevant
Possibly Relevant
Gallery > Nature Scenes > Oceans
Beach at sunset
by alexv
© Alexandra Vidas
Resolution: 1400x1200
thumb10.jpg
Relevant
Possibly Relevant
Community > US Travel > Mid Atlantic > Outer Banks, NC
Sand and surf in NC
Outer Banks
by Alex
© AlexV
Resolution: 800x600
thumb11.jpg
Relevant
Possibly Relevant
Gallery > Nature Scenes > Oceans
Ocean, sunset, and sandpipers
What more could you want?
by Alex
© avidas
Resolution: 800x600
thumb04.jpg
Relevant
Possibly Relevant
Community > US Travel > Pacific Northwest > Ocean Views
Pacific Ocean and a Bird Bystander getting its feet wet...
by Alexandra Vids
© Alexandra Vidas
Resolution: 800x600
thumb06.jpg
Relevant
Possibly Relevant
Community > Friends & Fun > Virginia
People on the beach
by avidas
© Spitting Image Press
Resolution: 800x600
thumb01.jpg
Relevant
Possibly Relevant
Gallery > Nature Scenes > Oceans
Ocean sunset in Florida
by avidas
© Spitting Image Press
Resolution: 800x600
thumb15.jpg
Continue to next page of scenario.
Done