IMAGE MANIPULATION AND USER-SUPPLIED INDEX TERMS

Leah Schultz, B.A., M.L.I.S

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

May 2009

APPROVED:

Samantha H, Hastings, Major Professor
Philip M. Turner, Major Professor
Brian C. O'Connor, Committee Member
Jon I. Young, Committee Member
Linda Schamber, Associate Director of Information Science Ph.D Program
Herman L. Totten, Dean of the College of Information
Michael Monticino, Interim Dean of the Robert B. Toulouse School of Graduate Studies

Schultz, Leah. Image manipulation and user-supplied index terms. Doctor of Philosophy (Information Science), May 2009, 139 pp., 19 tables, 41 figures, references, 68 titles.

This study investigates the relationships between the use of a zoom tool, the terms they supply to describe the image, and the type of image being viewed. Participants were assigned to two groups, one with access to the tool and one without, and were asked to supply terms to describe forty images, divided into four categories: landscape, portrait, news, and cityscape. The terms provided by participants were categorized according to models proposed in earlier image studies.

Findings of the study suggest that there was not a significant difference in the number of terms supplied in relation to access to the tool, but a large variety in use of the tool was demonstrated by the participants. The study shows that there are differences in the level of meaning of the terms supplied in some of the models. The type of image being viewed was related to the number of zooms and relationships between the type of image and the number of terms supplied as well as their level of meaning in the various models from previous studies exist.

The results of this study provide further insight into how people think about images and how the manipulation of those images may affect the terms they assign to describe images. The inclusion of these tools in search and retrieval scenarios may affect the outcome of the process and the more collection managers know about how people interact with images will improve their ability to provide access to the growing amount of pictorial information.

ACKNOWLEDGEMENTS

This journey could not have been completed without the help of many people along the way. My heartfelt thanks to Dr. Samantha Hastings for her patience, her inspiration and her guidance through this long and arduous process. Additional thanks to Dr. Philip Turner, Dr. Brian O'Connor, and Dr. Young for their insight, suggestions, and hard work. Friends and colleagues who helped and encouraged along the way were invaluable but a special thanks goes to Tracy Holtman for her assistance in coding data and her many hours of work she dedicated to this project. Thank you to Dr. Jason Sharp whose moral support and conversations kept me sane through the process. Without the technical assistance, friendship, and support of Mark Adams, this project could not have been completed.

Finally, thank you to my family for their love and support throughout this and every other endeavor I've undertaken. It's impossible to express how important you've been by encouraging me, nagging me, and routing for me.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

In this age of information, improvements in technology allow for storage of immense amounts of data. In order to quantify the trends and amounts of data being stored, Lyman and Varian (2003) estimated that 5187.13 petabytes were stored on magnetic media and an additional 102.9 petabytes were stored on optical disks. They also determined that the amount of stored information increased 30% each year of their study. When looking at the world of digital images, the trends appear to be the same. In 2006, 105 million digital cameras were shipped to stores, with that number expected to increase to 122 million in by the end of 2007 (Graham, 2007). Add to that the increase in cellular phones with photo taking capabilities and the number of devices capable of producing digital images increases significantly. In 2003, Lyman and Varian estimated that the storage capacity of all digital cameras and camera phones sold could hold approximately 1.2 petabytes of digital images.

What happens to those images after they are taken and how do institutions such as museums and libraries accommodate digital access to large collections of images? These questions have garnered much study over the last 15 years. Centuries of study have been dedicated to the access and description of textual material with very little attention to images. With the sheer increase in the amount of pictorial information being produced, many researchers and practitioners are finding difficulties with imposing the existing systems of bibliographic control on graphic material.

With the proliferation of these databases in institutions and on the World Wide Web, new approaches to image access are growing as well. From formal stock

photography databases such as Getty Images® to community-based tools such as Flickr®, the amount of photographic material organized into databases on the World Wide Web is immense. Getty Images displays 3.2 billion thumbnails per month while Flickr has thousands of new photos uploaded every minute. A quick examination of the finding aids at these sites make it very clear that image retrieval has moved beyond the structured approaches of traditional bibliographic control.

Flickr allows creators to assign up to 75 tags to their images and allows for images to be grouped into sets which are subsets of groups. Images can also be tagged geographically with a cartographic depiction of where images were taken. Zooming in on a map application many times pinpoints the street location of the image. Other finding aids also allow visitors to search for images by camera type or by looking at the favorites of other members. Popular tags, most recent uploads, and the explore option that displays “interesting” photos all promote browsing behaviors as well.

At gettyimages.com, the approach is more traditionally based on keyword searching with the ability to narrow searches by options common to users of stock photography. Users can search for abstract ideas as well as concrete objects. Options to limit are tied closely to the use of the image, from rights management to image orientation. Alternate search options allow users to explore the collections in more creative ways. The Moodstream™ allows users to set levels of happiness, temporal, warmth, etc to begin exploring image sets. The catalyst search option allows users to start with one keyword and drag and drop related words from the database into a search box to mine through the image database.

These are two examples of the multitudes of databases containing digital images and represent two institutions'' unique approaches to providing access in image databases. Traditionally, the process of searching for images was mediated by a professional. Now the search many times is conducted by the person with the information need. The shifting in search responsibilities was noted when Jörgensen and Jörgensen (2005) noticed different results when analyzing search queries for images in a stock photography database. The researchers noticed differences in query formulation from previous studies and attributed those differences to the inclusion of a search mediator.

To accommodate the shift in responsibility and to provide access to the growing number of photographic resources, researchers initially took two approaches to indexing images for retrieval. One group embraced the traditions of library science and assigns text descriptors for the image, relying on traditional test-based information retrieval systems. As the field grew, best practices began to emerge and systems or rules such as the *Anglo-American Cataloging Rules, Revision 2*, *Art and Architecture Thesauraus,* and *Library of Congress Thesaurus of Graphic Materials* were developed to guide indexers in preparing images for retrieval (Jörgensen , 1999).This approach can be time consuming, expensive, and ineffective in many cases as demonstrated by studies of relevance in image retrieval (Shatford, 1984).

The second approach to image indexing focused on the use of computers to index images based on content. The computer can index images based on many of its physical characteristics such as shape, color, and texture. There have been many advances in this area and content-based retrieval has been particularly effective in

scientific fields such as medicine and astronomy (Goodrum, Rorvig, Jeong, and Suresh (2001). Many researchers are examining ways to find common links between the two fields and use the physical characteristics of an image to convey higher levels of meaning. Colombo, Del Bimbo, and Pala (1999) combine the fields of semiotics with the automatic detection of characteristics to convey meaning. The combination of colors and the resulting emotional states and the implementation of line slope by artists to denote meaning are being explored as possible advances in content based image indexing.

However, many argue that these tools are being tested against queries that are not realistic and that current computer algorithms cannot determine the subject or aboutness of an image to satisfy many queries (Enser, 2000). Greisdorf and O'Connor (2008) also discuss the lack of ability to search beyond the color and texture of an image, which leads to exclusion of important semiotic elements such as metaphor, hyperbole, and allegory among others. Eakins, Briggs, and Burford (2004) contend that many recent developments in image retrieval have been technology driven and not user driven. The authors believe that the application of the color based search at the Hermitage Museum is a prime example, as there is rarely discussion on why anyone would actually want to conduct this type of search. Studies that survey users on the different attributes of images continue to show that users prefer concept-based retrieval over content retrieval (Eakins, Briggs, & Burford, 2004) even though content-based research currently dominates the literature (Chu, 2001).

In 1995, the Louvre began to put images of their collections online and it was not uncommon for visitors to wait many minutes to view images that could be found in most

art history books. When image collections first began to appear on the web, download rates were slow and images many times were compressed to accommodate these slow connections. Due to the compression, the images were many times of lower quality and made it difficulty to examine details closely. Even though new networking technologies continue to improve and become affordable for many Internet users, the concept of zooming still has relevance in a high speed world. As Internet speeds increase and become more widely deployed, new tools for creating larger, higher detail images are becoming more affordable. Camera backs for large format cameras are capable of creating 125 megapixel digital cameras that produce images that could still take a significant amount of time to deploy via the Internet. In opposition to the increases in image size, the decrease in display size on many mobile devices hinders the display of detailed, high resolution images. The ability to download an image to a mobile device and then inspect it closer with a zoom tool could also be useful.

In addition to the issues of access, managers of image databases also have access to an increasing variety of tools to assist image searchers in viewing and selecting images. Tools that allow users to zoom, rotate, crop, and compare multiple images are becoming easier to incorporate into image database interfaces. Users are interacting with digital images in digital ways, ways that are different than traditional viewing of analog photographs.

Statement of the Problem

Even though image data has existed for thousands of years, the focus in information retrieval has traditionally been text based. Images are difficult. They do not have convenient title pages or publication information attached that most books and

periodicals have. The nature of the image is not textual and extracting meaning is difficult, if not impossible. But the reality of image databases remains and something must be done to allow users to locate images in the vast body of digital images available. Understanding how users interact, react, and assign meaning to images is invaluable in improving access to digital images and has been studied by many researchers.

In addition to understanding access to the image, researchers also need to understand how users interact with images and how this can potentially affect the way searchers assign meaning to images. With new digital tools available, will users change the way they think about images due to the tools available? For example, some researchers (Hastings, 1999; Green, Marchionini, Plaisant, & Schneidermann, 2000) recommend the inclusion of a tool that allows comparison of images in the search process. With at least two images being considered together, does the content and meaning of one image affect the user's perception of meaning in the other? Will a user viewing an image of a an anti-nuclear rally with an image showing the devastation of the Hiroshima bombings attach different meaning to the first image if it were paired with an image of a nuclear power plant?

Much of the research in image retrieval focuses on the queries used or the text descriptors assigned for retrieval purposes but little research has been done to determine how the interface design of the system may affect these processes. The transition from analog to digital imagery has provided many new tools that allow the searcher to interact with images in different ways. Many of these tools are inexpensive and easy to implement in databases. Understanding what role these tools may play in

the retrieval process and their effect on the way users determine meaning in an image needs further study to assist image collection managers in providing access to their digital collections.

Purpose of the Study

The purpose of this research is to determine how the implementation of one of these interface tools that allows users to zoom in and out on a digital image affects the numbers and meaning of the terms supplied by participants to describe the images in different categories. The use of the tool allowed users to zoom in and examine details in the digital image and allowed users to pan across the image at various levels of zoom. Studying the terms submitted by students provided insight into the effects of the use of this tool, which is becoming widely distributed through its recent inclusion in Adobe® Photoshop® software, the popular photo editing package. Providing users with the ability to zoom in and examine details of an image may change the way they describe and ultimately retrieve images from a database.

Research Questions

In order to examine the relationships between the use of an interface tool and the terms submitted by participants, the following research questions were considered. The main question this research strives to answer is: How does the use of a zoom tool affect the way participants view and describe images and what role does the type of image play in the process of manipulation and terms supplied? The hypotheses to be tested include:

- Hypothesis 1: There is no difference in the number of terms supplied by participants to describe an image by those supplied a zoom tool and those who are not.
- Hypothesis 2: Access to the zoom tool affects the level of meaning of the terms supplied in existing models of image meaning;
    - Sub-hypothesis 2-1: A relationship exists between the level of meaning of terms in Shatford Layne's model and access to the zoom tool.
    - Sub-hypothesis 2-2: A relationship exists between the level of meaning of terms supplied in Enser's model and access to the zoom tool.
    - Sub-hypothesis 2-3: A relationship exists between the level of meaning of terms supplied in Jörgensen's model and access to the zoom tool.
- Hypothesis 3: A relationship exists between the amount of use of the zoom tool and the type of image being viewed (portrait, landscape, cityscape, and news).
- Hypothesis 4: The type of image being viewed affects the level of meaning of the terms supplied in existing models of image meaning.
    - Sub-hypothesis 4-1: A relationship exists between the level of meaning of terms in Shatford Layne's model and viewing of the four categories of images.
    - Sub-hypothesis 4-2: A relationship exists between the level of meaning of terms supplied in Enser's model and viewing of the four categories of images.

    - Sub-hypothesis 4-3: A relationship exists between the level of meaning of terms supplied in Jörgensen ’s model and and viewing of the four categories of images.
- Hypothesis 5: A relationship exists between the number of terms supplied by users with the type of image being viewed.

Data gathered from this study serves the general purpose of providing further insight into how viewers conceptualize images in the search process. With the introduction of the zoom tool, digital libraries can offer access to more details in large images and potentially provide access to information in these images that was not possible due to the need for image compression to increase download speeds. From a more general approach, this research also examined changes in the provision of a different interface and how it may affect the user’s interaction and definition of meaning of images. With findings from this study, digital library managers could determine if inclusion of image manipulation software in their online digital library would be a benefit to their patrons. Managers can use the results in indexing images if a relationship emerges between the use of the tool and the terms supplied. For certain types of images that are not zoomed often or generally receive lower numbers of indexing terms, managers may make generalizations about indexing practices for these images. For images that are either highly manipulated or receive terms that fall in the different levels of meaning in the models presented, managers may wish to review the indexing guidelines and make adjustments for depth of indexing if necessary.

CHAPTER 2

REVIEW OF THE LITERATURE

In studying the use of a tool for manipulation of images and its effects on the user's approach to supplying potential index terms, a review of the literature in many fields must be explored. The field of cognition sheds light on how users perceive images at a basic level. Studies of previous queries to image collections explore how users translate this basic cognition to meaning and aboutness in an image. From these studies, several scholars have developed models of how meaning is expressed in images and how these models can be applied to a user's search behavior and retrieval in image collections. One approach that was utilized in this study to increase understanding of users' search behavior was the assignment of potential search words and phrases that the user believed would make good indexing terms.

Information Retrieval

The role of language in the retrieval process is of concern to many in the field of information retrieval and is of particular concern in image retrieval. O'Connor, O'Connor, and Abbas (1999) discuss the potential for problems in using words to describe meaning of images. Because we cannot hear or experience the viewer's thoughts, researchers many times have to fall back on language as a way to try and understand how people assign meaning to images. Blair (2003) discusses the language problem in information retrieval in depth. Blair posits that searchers use language not only to describe their information needs but also to discriminate what they want from what they do not want. Blair also notes that this is the area where the complexities of language many times complicates the process. Users may not be able to describe their

information needs in terms of the controlled vocabulary or choose the terms that an indexer has assigned to a document. As for discrimination, Blair believes this a problem exacerbated by online collections of documents. Because weeding is less frequent or non-existent, index terms will become less discriminate as the collection grows and will result in greater recall and less precision.

Other problems associated with using text descriptors to retrieve images is the ambiguity of language. Bjarnestam (1998) worked with an image database to create a tool to limit the ambiguity in retrieval of stock photography. Problems with language can be demonstrated in a keyword such as “orange.” A search in the stock photography database may return images that are comprised mainly of the hue orange or can return images of the fruit. Large result sets with both included could frustrate users and lead to less precision in searching.

The problem with language worsens with image retrieval in a traditional cataloging scheme. Svenonius (1994) tackles the issue of using lexical terms to describe visual data and the shortcomings of such an indexing scheme. She argues that traditional subject indexing does not meet the needs of users because the concept of subject in library cataloging refers to the “aboutness” in a text based tradition. She asserts that assuming text based retrieval that is successful for some media may not necessarily be the best retrieval scheme for all media. Instead she champions the works of others such as Shatford Layne (1994) who discusses various attributes of various media as indexing terms.

In their discussion of the snapshot as a visualization process, O’Connor and Greisdorf (2004) also discuss the context or the purpose of an image and its to the

meaning of the image. A snapshot, the kind that is commonly taken by tourists or participants in an event, is many times created to serve as a record of being present at a site or an event and not solely “about” the objects captured in the image. This reliance on named objects does not capture the additional meaning of the circumstances that surround the creation of the image.

O’Connor and Wyatt (pp.108-100) also discuss the idea of “engagement” with the photograph. The discussion centers on the viewer’s relationship to the image. The example that the authors provide is an image of a woman on horseback. Viewers without a relationship to the woman might name the objects and actions in the photo whereas people who are either familiar with the individual or of horse riding might provide different terms due to their relationship or “engagement” with the photograph. Traditional retrieval methods rarely incorporate other engagements of the image, leaving much of the “meaning” untouched.

Roberts (2001) discusses a similar problem in her studies of art indexing. She argues that divorcing the image from the text can also be destructive to the meaning of an image and its ultimate retrieval. Roberts states that art historians are primarily concerned with constructing a verbal discourse of art and are dependent on words and text, although their main occupation is the study of visual material. The image itself has meaning beyond the content of the visual work. Information about the artist, the motivation for creating the work, the environment in which it was created, the materials used to create it, the time period in which it was created, the history of the work, and it’s damage and conservation all add to the story of a work of art. Roberts argues that a system that relies on indexing features of the image itself does not tell the entire story

and will make retrieval under these circumstances difficult and unusable for art historians.

Cognitive Viewpoints on Visual Thinking

Any study that seeks to understand a user's needs for image retrieval and the ways in which users conceptualize and utilize images must begin at the basic level of cognition. Determining how users assign meaning to an image is a daunting task since we do not have the ability to read one's thoughts as one looks at an image. These building blocks of cognition assist researchers in understanding the processes of how viewers receive visual information and ultimately interpret that information. With understanding of these processes, researchers can begin to develop strategies for helping users retrieve images that are congruent with the cognitive processes taking place.

How a user's eyes see and how a user's brain interprets images has been studied extensively in the fields of cognitive psychology and in the visual arts. In an effort to understand some of the basic cognitive approaches to imagery, Jörgensen (1996) discusses the importance of the Gestalt principles of proximity, similarity, good continuation, closure, common fate, area, and symmetry. She states that these are important to image retrieval, but the way they interact and contribute meaning are difficult for an image retrieval system to accommodate. In addition to these cognitive basics, Jörgensen discusses the importance of prior knowledge and expectations in image cognition. To understand these implications, Jörgensen conducted experiments to see which attributes of an image were named the most. She found that the highest percentage of responses were nouns for concrete objects such as dog, tree, etc. The

second most commonly mentioned attribute was color, although it was named much less frequently than objects.

These results can be explained by the theories of Biederman, particularly his theory of Human Image Understanding. Biederman (1987) believes human image understanding is based on phonemes, or a basic number of concepts that are mapped to represent words and concepts. Because a person recognizes and builds meaning in an image from these basic concepts, the studies by Jörgensen (1996) appear to be based in the perception of the image from a cognitive standpoint.

Heidorn (1999) expands on cognitive psychology theory and applies it to image retrieval in his exploration of mental models and their use in image retrieval. Hiedorn discusses search strategy in terms of a preconceived mental model that changes as the search progresses and says that users match retrieved images against this mental model. Because the indexer must share many cultural and educational background elements with the searcher, text alone is often ineffective as a retrieval mechanism. Instead, Heidorn champions a marriage between concept-based indexing with context-based indexing as the best chance of matching preconceived mental maps.

Mackworth and Morandi (1967) studied the visual fixations of the eye on images to determine what details capture the eye and attention of the viewer. The study showed that eye fixations landed more often on dominant areas of images that had unpredictable or unusual details. Over half of the image received no attention at all and areas of different texture were rarely registered on the fovea, particularly when they were smooth and predictable. Although this information can be useful to researchers in content analysis, it is should also of interest to researchers in concept based research

to determine which areas of an image are more likely to contain information that needs to be considered when providing access to the image.

In a study of user reactions to images, Greisdorf and O'Connor (2001) turned again to cognitive psychology to explain what users see when they look at an image. One aspect of their study explored the effect of prototypes and exemplars in image cognition and their potential effects on image retrieval systems. In this case, the researchers discovered that temporal prototypical displacement occurs, causing more information to be reported than appears in the image, and recommend it be considered in the retrieval process. They also discuss the use of exemplars, specific properties that play a major role in identification or classification, as possible cognitive agents in the retrieval process.

Greisdorf and O'Connor (2008, pp.57-59) also discuss the problems of "crippled viewer syndrome" which is the cognitive disconnect that a view experiences when viewing an image. The authors argue that the question "What do you see?" is a very different question than "What does it mean?" The identification of objects in an image are just signs that lead a viewer to construct meaning and if the viewer does not understand the signs, then the cognitive disconnect occurs. Combine with that the idea of symbolism and codes from the field of semiotics that provide different levels of meaning that can lead to disconnect. A lack of understanding of social, cultural, and historical systems of code can lead to the crippled viewer syndrome when determining what an image means.

In his book *Visual Thinking*, Arnheim (1969) discusses the process of visual thinking from perception to abstraction. He examines the effects a viewer's memories

has on the process, and integrates the process of seeing and its effect on visual thinking. For example, in nature, the human eye has a wide scope of data to assimilate and people automatically focus on certain areas or details while leaving other areas of the field of vision to be accommodated by shaper perception. He then demonstrates the effect of this natural perceptive tool and its effect on art. Arnheim argues that an artist uses these visual tendencies when creating works of art and therefore focuses on some items in a work and leave other details less well defined or of less prominence in a work. On the subject of categorization and its relationship to abstraction, Arnheim argues that our intellectual culture puts too much emphasis on categorization and warns against it in its application to visual objects. In areas such as biology, categories have evolved as similarities in functionality and structure. Arnheim warns about abstracting visual information in “correct” categorizations because the various cognitive abilities such as perception and memory affect the way an individual categorizes things. Within areas of academic study, there may be acceptable ways of categorizing items but this does not necessarily transfer to an individual who adopts schemes of organization based on their own observations and experiences.

In their book studying perception, knowledge, and the computer, Fishcler and Firschein (1987) discuss many aspects of perception that make classification of images difficult to do manually and in some cases seemingly impossible for computers to do. As to representation of the world and in particular, the use of language to describe it, the researchers point out that these things are heavily influenced by culture and the environment in which a person lives. The example the authors refer to is the number of words Eskimos have for snow. Because it is important for them to speak of and interact

with their environment, the prototype of what snow is does not necessarily coincide with the mental models of people living in a different environment.

Other factors that differ between users may affect the way images are viewed and understood. Beyond the obvious differences in visual abilities at the perceptual level such as loss of vision, color blindness, and peripheral vision issues, there are also many other cognitive properties that need to be understood in order to improve image retrieval and the understanding of the user and their interpretation of the image. Motivation in searching, spatial reasoning abilities, and other psychological factors need to be studied for further understanding of the user and how their cognitive abilities interact in the search process. Recent research by Neal (2006) incorporated the concept of locus of control and how it affected the image retrieval behaviors of journalists. Neal's study was based on the work of psychologist Julian Rotter and his concept of locus of control which states that people believe that the results of their behavior are either under their own control or under the influence of outside sources. In the study, Neal examined search behaviors of journalists and their locus of control. In her findings, Neal states that journalists preferred text based searching due to locus of control. Having more familiarity with text-based searching than other paradigms such as query by example, the journalists felt they had more control over the search process. The hidden nature of the computer algorithms in content-based systems that matched on characteristics such as color or line did not provide enough transparency for the participants and lead them to feel less in control during the search process.

In addition to a discussion of language, Fischler and Firshein (1987) also discuss some of the characteristics of reasoning that are currently a large hurdle in the area of

content-based classification. They discuss an image of what appears to be a woman pushing a man which will result in the man falling in the well. A person with an understanding of how the world works analyzes the relationships between the items and applies understood concepts of the world such as gravity to tell the story of the image. Using clues and relationships of objects in the image, a person can surmise that the woman has pushed the man and because most of his weight appears to be off balance, the ultimate result will be a fall. Because he is positioned over the well and knowing the characteristics of a well, the viewer can easily predict that the man will fall inside the well . Fischler and Firshein believe this kind of reasoning is difficult if not impossible for a computer.

These studies demonstrate that at the basic level of cognition, the brain is assigning meaning to an image upon viewing. The use of language to express these cognitive processes is problematic and requires refinement in how we discuss the ideas of meaning when applied to visual content. Because many different variables affect the process of perception and cognitive performance, the ability of a researcher to predict the meaning assigned to an image by potential searchers is daunting.

Models of Meaning in Images

The study of cognitive attributes of images have led several researchers to develop models for the purpose of assigning meaning to an image. Indexers have used these models in developing guidelines for indexing and access. Many models cited in current information science literature are based on the work of art historian Edwin Panofsky, who described three categories for the meaning of an image in his book *Meaning in the Visual Arts* (1955).The categories were labeled pre-iconographical

description, iconographical analysis, and iconological interpretation. Each layer of interpretation in Panofsky's model requires a deeper understanding of the culture in which the item was created. In pre-iconographical description, the object and actions, without cultural meaning, are listed. In iconographical analysis, stories, ideologies, and allegories from the culture bring a second layer to interpretation. Iconological interpretation requires a thorough knowledge of the culture in which the item was created and is very abstract and subjective in its interpretation as compared to the other layers.

Shatford Layne (1986) expanded on Panofsky's work for the purpose of developing a model to enhance indexing of images. She loosely adapts Panofsky's three levels in a model that includes, *generic of* ,*specific of,* and *About*. These three levels of interpretation are then applied to the facets of who, what, where and when (Table 1).

Table 1

Shatford Layne's Model of Meaning

Level	Definition	Example
Generic of	Maps to Panofsky's pre-iconographical level that requires no special cultural knowledge for interpretation. Generally relies on naming objects and physical descriptions of objects such as color	Baby
Specific of	Maps to Panofsky's icongraphical level that requires some interpretation or identification. Generally names a specific instance of an object, person, place, etc.	Jesus
About	Maps to Panofsky's iconological level that requires subjective interpretation and often understanding of the culture in which the item or object was created.	Salvation

Note. Adapted from "Analyzing the Subject of a Picture: a Theoretical Approach," by S. Shatford, 1986, *Cataloging and Classification Quarterly*, 6, p.49.

After studying requests submitted to the Hulton Deutsch Collection, Enser (1993) proposed his own model based on actual queries. Enser notes that generic requests were rarely encountered and that requests generally fell into two categories: specific

and nonspecific. He also notes that requests were generally qualified with information such as place or time. As a result, Enser proposed a model based on these characteristics. (Table 2).

Table 2

Enser's Categories

Category	Explanation	Example
Unique	A specific stated need of an individual, place, item, etc.	Richard Nixon
Unique with refiners	A specific item refined with relation to time, location, action, event, or technical detail	Richard Nixon in Air Force One
Non-unique	A generic object request	Female movie star
Non-unique with refiners	A generic object refined by time, location, action, event, or technical detail	Female movie star of the 1950s in Hollywood

Note: Adapted from "Query Analysis in a Visual Information Retrieval Context," by P. Enser, 1993*, Journal of Document and Text Management*, 1, pp.28-29.

Jörgensen (1996) proposed her own model of meaning for images after her studies of image attributes. Her model divides meaning into three classes: perceptual, interpretive, and reactive. Perceptual is comparable to the pre-iconographic level of Panofsky in that it includes perceptual cues without interpretation. Interpretive includes

information found in Panofsky's iconographical level, and reactive includes emotional and personal reactions to images. Jörgensen also includes subclasses to assist in qualifying meaning that range from visual elements content/story.

Table 3

Jörgensen 's Attributes

Class	Attribute	Explanation	Example
	Literal Object	Named object	Car, bridge, building, tree
	People	Named or unnamed person	Woman, toddler, George Washington
Perceptual Attributes	Color	Specifically named colors	Red, white, blue
	Visual Elements	Composition elements such as focal point, shape, texture	Close-up, blurry, lens flare
	Location	General and specific locations	Paris, Europe, underwater
	Description	Adjectives	Metal, shiny, giant

(table continues)

Table 3 (*continued*).

Class	Attribute	Explanation	Example
Interpretive Attributes	People-related	Concepts relating to people such as social class or emotions	Middle class
	Art Historical	Elements of the production of the image	Black and white photograph, watercolor
	Abstract Concepts	Atmosphere, thematical, or symbolic elements	Gloomy, resurrection, peace
Interpretive Attributes	Content/story	Specific instance or event	Battle of the Alamo, Washington's Inauguration
	External Relationships	Similar to other attributes	Looks like Degas' Dancers
Reactive Attributes	Viewer response	Viewer's personal feelings about the image	Makes me happy, reminds me of my tenth birthday

Note: Adapted from Image Retrieval: Theory and Research by C. Jörgensen, 2003, pp.235-236.

Focusing on the nature of the retrieval task, Fidel (1997) has proposed a general framework for meaning in images and their ultimate use. The model proposed by Fidel defines two poles, the data pole and the object pole, based on the motivation for the

search. Even though Fidel uses the word pole, he also recognizes that many images fall somewhere between these two poles. The data pole revolves around a system of image retrieval where the images are used as sources of information such as medical slides or maps. The object pole represents images that will eventually be used as an object. For example, images that will be used in advertising or illustration will many times be concrete, specific searches for ideas or objects. Many users retrieve images that fall in between these two categories. Fidel gives the example of the graphic artist who retrieves images of trees for different purposes. One image may be retrieved to include in a design another may inform the designer on types of trees and shapes of trees that exist. Fidel's model is tied very closely to the information retrieval field and is concerned with levels of precision and recall.

Buford, Briggs, and Eakins (2003) have developed another model of meaning to be extracted from image data that they hope can be mapped onto the interface of image retrieval systems. The basis of their model is a thorough review of literature on the image in different disciplines. The researchers contend that all previous models of meaning are system-centered in that they approach the idea of meaning based on the system and the classifications within that system of image retrieval. The researchers in this study intend to present image meaning in the context of the user and the differences between users. In this model, the researchers believe it is not uncommon for there to be overlap within categories depending on the user because of their background and their motivation for using certain terms. The taxonomy developed has nine classes of meaning presented in the table below.

Table 4

Classes of Image Content Developed by Buford, Briggs, and Eakins

Class	Description
Perceptual Primitive	Very low-level perceptual cues such as color and texture
Geometric Primitive	Simple shapes such as lines, arcs, square, circles, etc.
Visual Extension	Visual meaning that relies on additional perceptual cues such as depth and perspective
Semantic Units	General and specific names
Contextual Abstraction	Interpretations that depend on environmental knowledge
Cultural Abstraction	Interpretation which requires specific cultural information or background
Technical Abstraction	Expert knowledge in a field that requires knowledge of detail and vocabulary
Emotional Abstraction	Emotional associations that many times rely on the viewer's own background and experiences

(table continues)

Table 4 (*continued*).

Class	Description
Metadata	Information not present in the image but can be useful to describe the image itself in terms of image format, size, technical details, etc.

Note. Adapted from "A Taxonomy of the Image: On the Classification of Content for Image Retrieval," by B. Burford, P. Briggs, and J.P. Eakins, 2003, Visual Communication 2(123), p.130.

Through the use of models such as these, indexers can begin to establish practices for concept-based indexes that allow people to search for images in the context in which they perceive and understand images. Because this understanding varies by individual, there are many discussions in the literature that debate the ability of indexers or catalogers to provide access to pictorial material as outlined in these models.

Problems with Concept Indexing

Even though researchers strive to create models to guide indexers in the task of assigning indexing terms, there are still problems with subjectivity when people manually assign indexing terms. Many authors note that the manual assignation of subject terms for visual material is time-consuming and expensive. The amount of resources increases as the level of indexing increases. In large collections that contain millions of images, manual indexing is impractical if not impossible.

In her study on image retrieval in a stock photography database, Bjarnestam (1998) estimated that the assignment of keywords takes an average of seven minutes per image. In their article suggesting the combination of content and concept-based

approaches to image access, Goodrum, Rorvig, Jeong, and Suresh (2001) succinctly summarize the pros and cons of manual assignation of index terms. They state that concept-based indexing has an advantage in that meaning can be represented at a generic or specific level, but that problems with this approach include, cost, time, and low levels of consistency between indexers. Chen and Rasmussen (1999) concur, stating that the tradeoff between consistency and cost is the ability to index higher-levels that cannot currently be obtained by content-based indexing systems.

In addition to problems with expense and time, other factors such as breadth of a collection or expertise needed to adequately describe images may be difficult for some institutions to implement. In his study of digitization of lantern slides, O'Connor (1992) describes the difficulty in assigning terms because of expertise that might be needed in describing antique subjects depicted in the slides. Instead, O'Connor posits a system that allows users to assign terms that results in descriptions that are not bound by normal rules or text representation. In their book, Greisdorf and O'Connor (2008, pp.122-123) discuss the implementation of a system like this. The authors discuss the concept of tagging on the photo site Flickr. The implementation of tags by users allows them to organize and provide access to their photographs by keyword or topic. The tags are then organized by Flickr into tag clouds which indicate the prevalence of some terms over others in the Flickr universe. The authors believe that the non-hierarchical nature of tagging allows for more freedom to represent the interaction between the user and the photograph.

The nature of the photograph adds to the difficulties in indexing visual collections. Graham (2001) points out in her examination of the current practices in image

cataloging and indexing, that the image is a complex document that can be used by users from different domains in ways that the creator of the work never intended. The process is made more difficult when compared to traditional cataloging of text based documents. It is very difficult to identify standardized information in an image to indicate its aboutness, unlike the text counterparts which contain title pages, tables of content, abstracts or summaries on the dustjacket of a book. Many times information such as publication date or author information is fixed as part of the text-based document. This is often untrue of photographs. The photographer of the image may be unknown and even the date and subject of photographs can be difficult to identify if there is no text or visual clues accompanying the photograph. Unlike traditional cataloging of text items, the process of providing access to images is not a process of extraction because, generally, there is no text to extract from the image object (O'Connor & Wyatt, pp. 105-107). Even when there are captions or titles accompanying the image, they may not originally have been intended to define meaning in the image (Enser, 1995).

Even though Cooper (1969) indicates that inter-indexer consistency is not necessarily a measurement of "good indexing," he does note that almost all indexing will have some level of inconsistency and that there are even inconsistencies when one indexer assigns terms at different times. In a later article, Cooper (1973) indicates that evaluating retrieval systems based on cost and benefits is many times carried out from different points of view from system managers to designers. Cooper argues however that the best method of evaluation should be from the user's point of view, even though it is highly subjective. Cooper's ideas about retrieval systems seem to apply equally to

the problems in indexing images and the difficulties in accommodating the wide range in differences of meaning applied to images.

Markey (1984) explores the subjective nature of subject analysis of images through her study of inter-indexer consistency tests. The results from this study demonstrate that consistency between indexers providing indexing terms is very low. These findings imply that a lot of time and effort are spent in assigning terms that are not very uniform and predictive of the object that will be retrieved. However, Turner (1995) found that indexing terms supplied by users had a high percentage of matches with terms supplied by professional indexers.

Throughout the literature, several authors mention the subjectivity of indexing, particularly at higher levels of meaning as described in the previous models. In his discussion of intellectual access to images, Krause (1998) discusses the tradeoffs between what he terms hard-indexing which relates to the lower levels of meaning and soft-indexing which relates to the more interpretive levels of meaning. Krause argues that images, unlike books, provoke emotion upon initial inspection. Krause argues that these feelings or emotions can determine whether a user selects an image to fulfill an information need. Krause also warns that these emotions affect the assignment of indexing terms at higher levels of meaning, possibly revealing indexer bias or giving the impression that the indexer is passing judgment on the subject or creator of the image.

A searcher's experiences which obviously vary greatly can have an effect on the interpretation of meaning in images. In their study to indentify visual impressions in images, Bianchi-Berthousze and Wakamatsu (2001) found that our experiences and state of mind change over time and can have an impact of the meaning assigned to

images. In addition, they believe that this also affects the lower level of perception in filtering what our eyes select and pay attention to in an image, ultimately influencing our understanding of the image. The authors also suggest that these forces can be very latent and difficult to externalize, thus making research in this area even more difficult.

Studies of User Behavior and User Queries in Image Databases

In order to better describe understanding of image meaning, many researchers have examined user behavior in searching image databases or in simulated image retrieval situations.

Many of the studies have focused on specialized collections or queries of experts in specific types of image database. Colombo, Del Bimbo and Pala (1999) found that it was very difficult to define meaning in situations dealing with general subject matter and that these models of meaning many times need to be applied only to specific users in specific fields. However, in studies where information needs are perhaps different than those of users of general databases, researchers found many similarities to the models outlined previously. Hastings (1995) found that queries provided by art historians fall into certain categories, from Level 1 of identification to Level 4 of complex queries dealing with interpretation of meaning. Choi & Rasmussen (2003) found that queries from scholars involved in American history research also fell mostly in Shatford Layne's categories of general/nameable needs and specific needs. Only rarely were abstract or subjective requests formulated in the queries. Collins (1998) reported similar results when analyzing queries for archival material. However, in analyzing image queries of journalists, Markulla and Sormunen (2000) found that queries analyzed using an indexing scheme based on Shatford Layne's model of meaning differed based on the

ultimate use of the image. Images to illustrate hard news stories were more likely to be retrieved at the lower levels of meaning, whereas photos retrieved at the higher levels were generally used in feature articles. Similarly, in a study of newspaper image queries, Ornager (1997) suggests an additional level of meaning or indexing point as that allows for retrieval of images based on how the photograph could potentially be used.

Even within subject areas, researchers have found that the level of expertise can add additional dimensions to queries. In their study of image information seeking with geography faculty, Borgman et al (2005) suggest that students who have less expertise in the subject area rely more on location searching than experts in the field who were more likely to search for primary resources by concept or theme. Jörgensen (2004) also discovered that term specificity was also dependent on expertise in a domain. For example, bird may be the term assigned by some where an expert may give a common bird name or a specific genus and species name. In a study of queries submitted by professional intermediaries to a stock photography database, Jörgensen and Jörgensen (2005) discovered that unique term searches were much less common than expected and that there was a higher occurrence of theme and descriptive queries than expected. The researchers also discovered that a large percent of queries were modified and that browsing strategies resulted in more image downloads that queries.

In order to provide background on the types of users that may be searching for visual information, Connis and Ashford developed a taxonomy with the intent to study the various uses for image data. These uses may reflect differences in retrieval tasks in

the various studies conducted so far. In their studies, the researchers found a great difference in searchers based on their ultimate use of the image.

Table 5

Seven Classes of Image Use from the VISOR project

Class	Explanation
Illustration	Used to illustrate accompanying media such as news stories
Information Processing	The data contained in the image is the primary concern such as an x-ray
Information Dissemination	For the purposes of transmission of information such as a mug shot
Learning	People gain knowledge from the image content such as art slides used in an educational setting
Generation of Ideas	Used to instigate creativity or provide inspiration such as source material for architects or interior designers
Aesthetic Value	Images are selected for their decorative value
Emotive	Images are selected for the emotions they evoke such as advertising images

Perhaps the most comprehensive study of users' queries to an image database was undertaken by Armitage and Enser (1997) in their study of requests for image retrieval from the Hulton Deutsch Collection, one of the largest image collections in Europe. Unlike the collections in many other studies, this collection is not subject-specific and serves the needs of the general public. Armitage and Enser studied written and phone queries from more than 1,000 users. From these data, they characterized

images into the Enser's previously discussed model of meaning. The two concepts they report as characterizing the largest percentage of queries was the refinement of terms to specific time but also more notable according the to the researches is the fact that almost 70% of the queries requested unique items. According to the authors, these results differed from user behavior previously witnessed in library OPACs built for retrieval of mostly text-based materials. Users requesting visual information were more specific in the beginning of a search compared to those using OPACs, who generally were more general when beginning their search. The authors noted that this could also be due to the nature of the process of submitting queries to the Hulton Deutsch collection: users may be more specific because they must either write down their requests or must speak to a researcher who writes down their requests for them and possibly asks for refinement from the user. Enser and Armitage also say, however, that users may have a clearer mental model of a photograph that would suit their information needs than do users searching textual material.

Also noteworthy is the Frost et al. (2000) study of browsing behavior in image databases. The authors observed browsing behavior among expert art historians and some generalists in an image database of fine art images. They found that a large percentage of the users surveyed preferred some level of browsing using thumbnail groupings that could be scanned quickly for suitability. However, the hybrid system that used traditional searching combined with browsing received more positive reactions than straight browsing or straight searching. It is interesting to note the high percentage of participants who said that image resolution was a very important aspect of online image databases.

Studies of User-Supplied Search Terms

A number of researchers have studied how people perceive images or form search strategies by asking participants to supply potential index terms or provide descriptions of the image. The researchers have utilized this strategy in order to classify these possible terms into models of meaning and models of the search process.

Through analysis of potential search terms, researchers can begin to understand how users conceptualize the image as a whole, determine which elements are important, and assign levels of meaning to images (Jörgensen, 1998). Users' search behavior is a complex phenomenon to observe; in order to gain some control, researchers often rely on contrived searches that do not represent real information needs. The technique of having users provide potential search terms for images does not reflect users' real search strategies as they search for images in a database. Nevertheless, a research design that requests users to emulate an aspect of the search task and provide potential retrieval terms can provide insight into how users begin their approach to image searching and guidance for indexers on how to modify existing indexing practices or create new, innovative approaches to image access.

In two studies, Jörgensen (1996, 1998) investigated the importance of attributes in image retrieval. She used a variety of methods to study users' perceptions of images. In these studies, she used three descriptive tasks: a basic narrative of the image, a simulated search description, and description from memory after a specified amount of time. The search task was intended to elicit responses as if the participant were searching for a known item. Interestingly, a comparison between the descriptive tasks

and the search tasks revealed a higher level of story attributes in the search and memory tasks than in the descriptive tasks.

O'Connor, O'Connor, and Abbas (1999) obtained user-supplied descriptions of images in order to study reactions to images. Their original survey instrument, which asked library students to describe the images and provide any reactions, resulted in terms that mirrored standard classification studies The researchers then revised the instrument to ask participants to write captions for images and to list words and phrases that described the images or how the images made them feel. Analysis of these images showed interesting trends in the data. The trend of antonyms describing the same image was not uncommon and that as the tasks continued, participants became progressively more comfortable with the process and narrative increased. Narrativity in which participants "told little stories" was very common in user responses and captions. Full image adjectives were more common in the response task compared to the caption task, with and average of 3 adjectives per response.

In an attempt to study how users viewed moving images, Turner (1995) utilized a similar methodology in asking users for possible retrieval terms for stock footage. Viewers were asked to supply terms that that would be useful in retrieving the image for themselves or others. In this study, Turner discovered trends in consistency with the most often supplied terms when compared to terms supplied by professional indexers and that pre-iconographical objects were listed more often than iconographical aboutness terms.

In 1996, Brown, Hidderley, Griffin, and Rollason investigated the possibility of building an image retrieval system around the indexing terms supplied by users of the

system. In 1997, some of the researchers applied the same process for the retrieval of moving pictures (Hidderly & Rafferty, 1997). In both projects, users were asked to supply terms to an image database that were indicative of good retrieval terms. Terms supplied by users were compiled into a public database but users who participated could also organize their personal search space with their own private indexes. In unrestricted description, the researchers discovered as many had before them that as the number of terms increased, the agreement between indexers decreased (Brown et al.1996). Complicating the process even further, Greisdorf and O'Connor discovered (2001), terms were sometimes supplied for objects that did not exist in the photograph. For example, an image that had a hint of a path resulted in a statistically significant provision of the term "walking" even though that action is not actually depicted in the image.

An additional study of browsing and the categorization of images for improvement in browsing by Rorissa and Hastings (2004) indicates that interpretive attributes in grouping images were better candidates than perceptual attributes. In image describing tasks, perceptual tasks were more prevalent in previous studies but sorting tasks resulted in different findings. Using cluster analysis on sorting categories supplied by participants, the researchers discovered that perceptual attributes such as color, texture, and shape were rarely if ever supplied by the users and interpretive terms were more prevalent.

Manipulation of Images in Image Databases

Very little research has been written on systems that allow users to manipulate the images in some way and in particular how this might effect the way images are indexed. Most studies being conducted with zoom software is in the area of interface design and improving browsing abilities in databases.

In their work on creating an image browser that allows for viewing of large images on mobile devices such as cell phones, Liu, Xie, Ma, and Zhang (2003) studied the best way to automatically browse the features of a large image on a small screen. Using an image attention model to determine the information structure of an image, the researchers created a tool that automates the scanning and zooming of an image based on the most salient features of the image. This in turn, allows the viewer to see the important parts of the image as determined by the researchers without having to manually manipulate the image on the smaller device.

In her study of evaluation of image retrieval systems, Hastings (1999) supports the idea that improvements need to be offered to users in the form of tools that enhance browsing, allow manipulation and comparison of images in the database. In an article by Greene, Marchionini, Plaisant, and Shneidermann (2000) exploring the interface design of digital libraries to support visual seeking, the authors suggest an increase in the user's control over the view of the data retrieved. The authors recommend improvements in usability factors that include multiple window comparisons, zooming, and collages for browsability of surrogates.

To aid retrieval and browsing of images, Joon (2006) suggests an interface that provides connection between images based on connotative messages derived from

denotative signs. The framework developed by Joon predicts greater user satisfaction by not only providing links between images that have similar denotative messages but also groupings of connotatively linked messages. Using records from Artefacts: Canada and terms from the *Art & Architecture Thesaurus*, Joon's research demonstrates a relationship between connotative and denotative elements that can lead to a connotative message being derived from the denotative. This relationship was shown to improve image retrieval and browsing.

In addition to studying terms and approaches to searching for primary geographical image data, Borgman et al. (2005) also studied how faculty members manipulated data they discovered. In addition to computer tools that allowed them to enlarge and shrink images, make selections within the image, and add annotations, faculty members and researchers also used many manual tools such as photocopiers and overhead projector slides and markers. The researchers found these to be important additions to the interface that could eventually add more capability of the image database through use of shared spaces and personal digital libraries.

Researchers in information visualization have studied zoomable interfaces as an aspect of information retrieval. The focus of these studies did not examine how zooming affected a participant's understanding of an image. Instead, many focused on the addition of a zoom tool as a finding aid in image databases, particularly in browsing tasks. Combs & Bederson (1999) reported that a zoomable interface decreased the number of incorrect responses in their study as well as gaining favor with participants as a useful tool. Hornabaek, Bederson, and Plaisant (2002) studied the effect of providing an overview window in a browsing task with a zoomable interface and providing the

interface without the overview. The overview window displayed in the corner showed the entire original image while participants zoomed and panned across the images. Overall, subjects preferred the interface with the overview over the non-overview image although there were problems in recall tasks and timed tasks when combining the two.

Because tools such as zoomable interfaces are being investigated as potential interface enhancements to image browsing, it is imperative to understand how the manipulation of images affects the viewer's conception of the image in a retrieval environment, be it browsing or searching.

Summary

Overall, these aspects combined provide a background to support the undertaking of this research. Information from cognitive science allows for examination of mental models and other cognitive processes which helps determine how people conceptualize images and, in this study, the effect the zoom tool may or may not have on those processes.

Studies of user behavior and user-supplied queries have allowed researchers to prepare models that demonstrate how people assign meaning to images and how they formulate search queries. These models are used to evaluate the terms supplied by participants in this study to determine the levels of meaning for the terms supplied. The terms provided by participants also shed additional light on how useful these models might be in assigning terms for indexing. In addition to the stated goals of this research, the knowledge gained also supplements the growing understanding of how people think about, interact with, and search for the growing number of digital images in databases.

CHAPTER 3

METHODOLOGY

This is a descriptive study of terms generated by participants in response to a digital image reviewing task, involving the use of an image zoom tool. This chapter discusses the participants in the study, the tools used to collect the data, the procedures used to collect the data, the analysis of data, and finally, the limitations of the study. In order to answer the research questions this study provided the ability to manipulate images to half of the participants. The terms supplied and their use patterns of the zoom tool were analyzed for relationships. In addition to the relationship between terms and the zoom tool, the data were also analyzed for relationships between the use of the zoom tool and particular types of images and the types of terms provided for certain types of images.

Definitions

- *Image manipulation* – Although manipulation is a general term, for the purposes of this study the use of the word manipulation refers to the functionality included in the tool used in the data collection tool. This particular tool allows the user to zoom in and out of an image as well as re-center the focus of the image. While zoomed in, users can pan across the image by dragging with the mouse.
- *Landscape photographs* – Defined according to the *Library of Congress' Thesaurus for Graphic Material* (*TGM*), landscape photographs include "general or broad views of natural scenery, including inland bodies of water; may also include figures or man-made objects, but these are of secondary importance to the composition."

- *Portrait photographs* - Defined according to the *TGM*, portrait photographs are "formal, posed studio portraits, as well as informal, candid photographs of persons in natural or spontaneous situations."
- *News photographs* – Taken from the *TGM* entry for 'Peridocial Illustrations', news photographs are intended for publication or actually are published as illustrations in periodicals.
- *Cityscape photographs* – Photographs in this category are "general or broad views of cities and towns or sections of them," according to the *TGM*.

Participants in the Study

In this study, a convenience sample of 64 undergraduate students at Tarleton State University participated in the data collection phase of the study. Students were not required to have any specific subject knowledge to participate and students were recruited from all backgrounds. With the assistance of faculty members in the Department of Computer Information Systems, students were recruited from an introductory computer concepts class that is a requirement of many different majors at Tarleton State University. In the first round of data collection sessions, attendance was much lower than the minimum number of students desired. To increase the number of participants, additional volunteers were recruited from upper-level computer information systems classes and technical writing classes in the Department of English and Languages. Students were recruited by faculty teaching the classes and most participants were given extra credit for attending the data collection sessions. In addition to extra credit incentives, all students who attended a data collection session were eligible to win a gift certificate from a local store. Participation was not required for any

of the participants and the data collection activity was not directly related to any classroom activity beyond the extra credit.

As part of the data collection process, students provided demographic information on their age, race, gender, and academic characteristics. Students were not required to provide any information that could identify them as individuals. Students were allowed to discontinue the study at any time for any reason and still receive credit for attending the data collection session.

Research Tools

Images used in study.

Images were selected for this study from a variety of sources encompassing government collections, news agency images from the *Dallas Morning News*, stock photography resources, and images from my personal collection. The images do not focus on one specific subject area and do not require expert subject knowledge in a field such as art history. The original images are all in color and were taken with a variety of different cameras, and in many cases, this information is unknown because images were procured from various sources. The original images were in color and 3072 pixels by 2048 pixels. When viewed in the application, the initial view of the image and the non-zoomable images are presented to the participant at 307 by 237 pixels, approximately 10 times smaller than the original image.

Images were selected to represent different categories based on common genre terms selected from the Library of Congress' Thesaurus for Graphics Materials II: Genre and Physical Characteristic Terms (TGMII). The categories used for this study include landscape photography, portrait photography, cityscapes photographs, and news

photography. The selection of these categories represents two general types of images, detailed and non-detailed images. Generally speaking, the portrait and landscape images have fewer details included as compared to the news photography and the city scenes.

In order to ensure agreement on the categorizations of the images used in the study, a set of printed images was sorted into the four categories by myself and two librarians with experience in cataloging and classification. Librarians were selected for this task for their familiarity with the categorizations outlined by the TGM. From the images that had consensus in categorization, 40 images were selected for the study, with 10 images from each category. It was decided to keep the number at 40 images because of the amount of time required to complete the task on the part of the volunteer and to try to lessen the effect of fatigue with the data collection process. All images used in the data collection process are included in Appendix C.

Data collection instrument.

After the images were selected, they were processed using the Zoomify™ software produced by Zoomify, Inc., now an included tool in Adobe® Photoshop® software. The application dissects the original images into smaller, easily downloadable parts and reassembles them as the viewer zooms in and out of the image. In addition to changing the level of the zoom, the tool also allows the user to change the central point of the image at any point on the image. The tool is similar to popular satellite imagery and mapping applications that allow the user to set the focus and zoom level of the application. The user can pan across an image or can reset the center point by clicking another area on the image.

Once the images were processed with the software application, the images were placed into a custom programmed online application that presented the images one at a time, in a predetermined order. This was done to allow for comparisons across participants while lessening the effects of fatigue with the task. If images were presented randomly, and all of the images in one category were presented towards the end of the task, this could have a potential effect on the data. The application was programmed to capture the terms supplied by the participants as well as any zoom movements of re-centering of the image that the participant executed. The application gave the participant 10 blanks to provide terms for each image. If the participant supplied a 10th term, the application automatically added new blanks in sets of four until the participant clicked the button labeled “Done.” The application was also programmed to collect demographic data supplied by participants (see Appendix A).

Figure 1. Example of screenshot from online application.

The application was programmed in Macromedia® Flash® software with the assistance from the web programmer at Tarleton State University and the application resided on the Tarleton State University web server. Microsoft SQL Server® database software was used to collect the data submitted via the application. All data were assigned sequential numbers in the database and were linked by this number, thus eliminating the possibility of linking data to individuals. After the collection period was complete, all data were exported from the Microsoft SQL database into an Microsoft Access® database for the purposes of creating the forms used for coding purposes. Data were eventually exported from the database software into Microsoft Excel® spreadsheet software and SPSS ® statistics software package for analysis.

Data Collection

Data collection sessions.

Data collection sessions occurred in four sessions over a three week period in the spring semester of 2006. All sessions took place in the afternoon on different days to accommodate as many student schedules as possible. Sessions were held in a central campus location in a teaching lab in the campus library. All of the computers in the lab were identical and all had 17” flat panel monitors with resolutions of 1024x768.

When students began the exercise, they read and agreed to the consent form or chose to end their participation. Only one student chose not to continue participation. For the students who chose to continue, the application checked to see if the previous participant who accessed the application was assigned to the control or treatment group and assigned the next participant to the alternating group. Because one student chose

not to participate and a couple of students accidentally restarted the application, there was ultimately a small inequality in the number of participants assigned to the two groups. Alternating assignments to the control and treatment group helped to eliminate bias in inclusion of subjects in the groups.

Participants assigned to the control group did not have access to the zoom tool and were provided with the following written instructions:

> In the following exercise, you will be shown images and will be asked to provide terms to describe the images. You can provide as many terms as you feel necessary to describe the image.
>
> Once you have provided your terms for an image, click the Done button to continue until all images have been completed. Thank you for your time.

The treatment group had access to the zoom tool for the duration of the exercise. Once they have completed the demographic information, the participants in the treatment group were shown a short animation demonstrating the use of the tool. The treatment group were then provided with a practice image and the following written instructions :

> You can zoom in and out on the picture using the tool bar at the bottom of the image below. Practice with the toolbar on the picture below until you feel comfortable continuing.
>
> In the following exercise, you will be shown images and will be asked to provide terms to describe the images. You can provide as many terms as you feel necessary to describe the image.
>
> Once you have provided your terms for an image, click the next button to continue, until all images have been completed.

Both groups were presented with the 40 images, one at a time, and participants provided terms for each image before moving on to the next image. The task took students 30 minutes to 1 hour to complete the entire task. At the end of the task, students were provided with a slip of paper to return to their instructor for extra credit purposes. At no time were students identified and the data collected was never tied specifically to identifiable students.

Data Analysis

After the data collection period was complete, the data were exported from the Microsoft SQL Server database software into a Microsoft Access database to allow for printing of codebooks. The terms were printed by participant and every term provided was entered on one line. On each line, there were columns that corresponded to the different models of image meaning developed by Shatford Layne, Enser, and Jörgensen.

A librarian from the university volunteered to act as a second coder to increase the validity of the coding decisions. The second coder read the articles published about the models under study and participated in discussions about the different models. Using data from a pilot study similar to the research presented here, the two coders discussed a set of data and decided on the best coding schemes for the data presented in all three models of meaning. With another set of data, the coders worked independently and assigned codes to the practice data and compared results. All of the discrepancies between the coding on the practice data were discussed. This process was repeated with practice data twice more as the second coder became more familiar

with the systems and felt comfortable assigning terms to the data collected for this study.

Both coders worked independently on coding the terms supplied in all three models. Coding was done by hand in the coding books and the codes were re-entered into the Access database to make analysis easier. Coding took close to two months to complete because of scheduling and the amount of data to be coded. Both coders worked in short sessions in coding the data to avoid fatigue that could affect the quality of the coding. After the initial coding was completed by the second coder, it was reviewed and entered into the database. The codebook was returned to the second coder to make some adjustments. These adjustments included codes that were hard to decipher and codes that had been skipped. Once the coding was complete, the additions were entered into the database for analysis.

After all codes were entered into the database, the data set was organized for export into SPSS for statistical analysis. After initial analysis, it was determined that the large number of categories utilized in the Jörgensen model resulted in ambiguous results. To be able to make meaning of the significant results of the analysis, the data in the Jörgensen model were regrouped into the three broader categories of perceptual, interpretive, and reactive as defined by Jörgensen, instead of the more specific 12 attributes in her model. (2003).

Research Hypotheses

In order to test the hypotheses, the following data was analyzed to determine if the hypotheses could be accepted or rejected. The following approach was applied to test each hypotheses.

- To test hypothesis 1, which states that a relationship exists between the number of terms supplied by a participant and access to the zoom tool, the mean number of terms supplied by participants who had access to the zoom tool was compared to the mean number of terms supplied by those without access to the zoom tool.
- In order to evaluate hypothesis 2 and the associated sub-hypotheses, the number of terms coded into the categories of the various models was evaluated for relationships between participants who had access to the zoom tool and those that did not.
- To determine if a relationship existed between the use of the zoom tool and the category of image being viewed, the means of the number of zooms per image type was compared to the other categories.
- To test hypothesis four and the associated sub-hypotheses, the data coded into the levels of meaning in the three models selected was analyzed for relationships between the levels of meaning and the type of image being viewed.
- The means of the numbers of terms supplied for each type of image were compared to determine if a pattern exists for hypothesis 5.

Methodological Issues

Limitations.

Potential limitations exist in the selection of the population. Many previous studies have focused on the search behaviors of specialists in a certain field and their retrieval of images. In this case, however, a general population was desired without specific knowledge of the subject matter contained in the images. For this reason, undergraduate students were selected over graduate students in a related field. Because these students are drawn from a university that differs slightly in the educational, racial, and age makeup of our society at large, generalization at that level would require further testing. However, students at the university do represent various socio-economic levels, and students selected represent many areas of study at the university and different levels of computer experience. For these reasons, the results of the study can be generalized to represent a population of undergraduate students of various disciplines. Future research may expand on this by requesting volunteers from a wider population.

Another potential bias in the population is the dependency on volunteers. Research has shown that volunteering can potentially bias a study because the characteristics of volunteers may differ greatly depending on the task (Gall, Gall, & Borg, 2003). This task does not introduce tasks that would discriminate against one population over the other in requiring any economic action on the part of the participant. Other biases such as a volunteer's desire to change the outcome of research to support a personal or political view is unlikely in this case. When volunteers were requested, students were told the general nature of the experiment and some were offered

incentives at the discretion of the instructor, such as extra credit. All students who participated were elgible for the gift cards awarded at each data collection session.

Validity.

Krippendorf (2004) outlines the various constructs of validity for content analysis. Content analysis many times relies on face validity or the common sense behind data. Krippendorf also believes that content analysis can also exhibit social validity as well as empirical validity. In the case of this study, in addition to face validity, it is also possible to demonstrate other types of validity. Sampling validity is not an issue because all of the data collected was analyzed and therefore does not have sampling validity issues in the content analysis. Of course concerns about sampling of participants as discussed above need to be taken into account about sampling, but not in the validity of the analysis itself. Functional validity can also be demonstrated because the categories for coding have been used in studies previous to this one. Enser (Enser 1993; Enser 2000; Armitgae & Enser, 1997) and Jörgensen (Jörgensen , 1996, 1998) have both used their models for analysis in multiple inquiries as have others such as Chen's (1999) analysis of art queries.However, there also may be limitations in the use of previous models in that the categorization of the data into existing categories may disguise other information to be gained from the data.

Reliability.

Reliability in content analysis many times is established by increasing the likelihood that the process can be replicated by others. Attempts to make the data more reliable was established according to rules outlined by Krippendorf (2004) in that coders worked independently of each other and contain information on the selection of coders

as well as coding instructions. Data that coders did not agree on was not included in the data analysis. The high percentages of agreement on the coding of the data for Shatford Layne and Enser led to inclusion of most of the data set. However, the lower percentage of agreement between coders on the coding of the Jörgensen led to 16% of the data being excluded from the analyzed data which could potentially be a weakness of the analysis of Jörgensen's model.

CHAPTER 4

FINDINGS

This study collected data on participants' use of a zoom tool to manipulate images and the descriptions subsequently given by the participant to describe those images. The data collected from the participants' responses to the demographic data, the terms supplied, and the log information collected from the zoom tool itself were analyzed to answer the research questions posed. This chapter examines the characteristics of the participants in the study, the number of terms supplied in relation to the use of the tool, and an examination of meaning of these images as posed by previous models of meaning. In addition to the user's actions, the different categories of images are analyzed for patterns of usage with the zoom tool as well as levels of meaning for terms supplied for the different types of images.

Participants

In the study 61 students volunteered to participate. All of the students who participated were undergraduate students with an average age of 22.4 years of age (*SD*=4.9). Forty-eight of the students were Caucasian, five African American, four Hispanic, three Asian, and one participant did not disclose their ethnicity. The group was very closely distributed by gender with 29 females, 31 males, and one participant who declined to identify gender.

The students represented a variety of academic backgrounds with the largest number coming from the computer information systems department (*n*=19). Participants rated their computer familiarity with more than half rating themselves at least somewhat familiar with computers. Specifically, 25 students rated themselves as very comfortable,

14 somewhat comfortable, 7 neutral, 5 somewhat uncomfortable, and 10 rated themselves as very uncomfortable.

Overview of Number of Terms Supplied

Overall, participants submitted 10,429 terms for all 40 images. The mean of terms submitted by participants was 170.97 terms per student.There was a mean of 260.73 terms submitted per image viewed.

Table 6

Descriptive Statistics for Number of Terms Supplied Overall

Total Number of Terms Submitted	Range	Min.	Max.	Mean	Std. Deviation
10,429	291.00	40.00	331.00	170.97	57.65

Description of Zoom Tool Use

The total number of zooms and re-centering of the image view was 2,744 for all members of the experimental group. With 32 members in this group, that results in a mean number of 85.75 uses of the tool per participant. Broken down by image, that would be an average of 2.14 uses of the zoom tool per image on average. The variation in the use of the tool however was large as demonstrated in table 7.

Table 7

Descriptive Statistics on Use of Zoom Tool

Number of Uses of Zoom Tool	Range	Min.	Max.	Mean	Std. Deviation
2,744.00	592.00	1.00	593.00	85.75	122.99

Number of Terms for Zoom/Non-Zoom Participants

Of the 63 students that participated, 29 students were assigned to the control group and did not have access to the zoom tool. The other 32 students were assigned to the treatment group and had access to the zoom tool during the exercise. The groups were not evenly assigned due to a technical problem with the computer that assigned the students to a group as they began the exercise.

The control group without access to the zoom tool supplied a total of 5199 terms to describe the 40 images. The test group with access to the zoom tool submitted 5230 terms. Applying an independent t-test to both means demonstrates the difference between the means of the non-zoom group (M=163.44, SD=54.91) and the zoom group (M=179.28, SD=60.40) is not significant statistically t (59)=-1.07s, p>.05. Even though the average number of terms supplied for the non-zoom group is higher than the zoom-group, the lack of statistical significance requires that the hypothesis that predicts no difference between to means is accepted.

Use of Zoom Tool and Shatford Layne's Model of Meaning

All of the terms were categorized by both coders and reached agreement on 98.7% of the terms submitted. The data the coders agreed upon were examined to determine if a relationship between the use of the zoom tool and the categories in Shatford Layne's model of meaning existed. Participants with access to the zoom tool provided the majority of their terms in the generic category (4953 terms). The levels dropped off to 189 terms in the specific category and only 7 terms in the abstract category. Participants who did not have access to the zoom tool also provided most of

their terms in the generic category with 4971 terms. At the specific level, participants provided 145 terms and supplied 14 terms in the abstract category.

Table 8

Number of Terms Supplied by Groups Using Shatford Layne's Model

Group	Generic	Specific	Abstract
Access to zoom tool	4953	189	7
Without access to tool	4971	145	14

Applying a chi-square test to determine if a relationship exists between participants' access to the zoom tool and the categorization of the terms they supplied revealed that a relationship exists ($x^2(2, N = 10279) = 8.13, p < .05$) between the use of the tool and the level of meaning in Shatford Layne's model which allows us to accept the hypothesis predicting the relationship.

Examination of actual frequencies compared to the expected frequencies reveals that participants with access to the zoom tool supplied terms in the specific category at a higher rate than expected and at a lower rate in the other two categories. Those participants without access to the tool supplied more terms in the abstract and general category than expected, and lower than expected in the specific category.

Use of Zoom Tool and Enser's Categories

After categorization by both coders, the 97.1% of records that both coders agreed on were selected for analysis. The coders agreed on 10111 categorizations for Enser's categories and were entered into SPSS for analysis. The data were analyzed first to determine if there was a significant relationship between the use of the zoom tool

and the designation between unique and non-unique. Within those categories, the types of refiners were also studied for relationships with the manipulation of the image.

A chi-square test was used to determine if access to the manipulation tool had an effect on the terms supplied and their categorization into generic and non-generic categories defined by Enser. Participants without access to the zoom tool supplied 4879 terms in the non-unique category and 175 terms that were considered unique by both coders. Participants with access to the zoom tool supplied 4915 terms in the non-unique category and 142 in the unique category.

Table 9

Number of Terms by User in Enser's Model

Groups	Unique	Non-Unique
Access to tool	4915	142
Without access to tool	4879	175

Results of the chi-square test show that there was not a significant relationship between the variables ((x^2(1, N = 10111) = 3.57, p > .05) Because the level of significance was greater than the .05 threshold, there is no evidence to support the hypothesis that predicts the existence of a relationship between the use of the zoom tool and the levels of categories within Enser's definition of image meaning.

When examining whether the participants with access to the zoom tool were more likely to refine their terms according to Enser's model, the chi-square test also reveals that there is no relationship between the number of refined terms and access to the zoom tool ((((x^2(1, N = 10111) = .190, p > .05) However, looking at the terms the

most commonly used refiner was by action on 56 terms, followed by location on 23 terms, by time on 5 terms and 3 terms were refined by technical detail.

Use of the Zoom Tool and Jörgensen's Attributes

An initial chi-square test was applied to the individual classes in Jörgensen's model, but the large number of the columns in the chi-square test made it difficult to determine what those relationships were. Instead, the data were examined at the higher level of attributes for significant relationships. Because the coding scheme was more involved than the other models examined previously, agreement on all codes for terms submitted was lower at 81.1%. This model, unlike other models, allowed for multiple coding schemes. For example, red car could be coded as an object and a color. If an image was assigned two codes and only one code matched between the coders, the non-matching codes were discarded and the matching codes were retained in the data set.

A total of 9,741 codes were analyzed for relationships between the zoom and non-zoom participants and categories of perceptual, interpretive, and reactive. Participants without access to the zoom tool provided 2,638 terms in the perceptual category, 1,984 terms in the interpretive category, and 277 in the reactive category. Participants with access to the zoom tool provided 2,286 terms in the perceptual category, 2,271 terms in the interpretive category, and 285 in the reactive category.

Table 10

Number of Terms by User in Jörgensen's Model

Groups	Perceptual	Interpretive	Reactive
Access to Tool	2,286	2,271	285
Without Access to Tool	2,638	1,984	277

The chi-square test indicates a statistically significant relationship between the use of the zoom tool and the attribute categories in Jörgensen's model ((((x^2(2, N = 9741) = 44.303, p < .01) which allows us to accept the hypothesis predicting the relationship.

Comparing the findings to the expected frequencies, it appears that users with access to the zoom tool provided more terms at the interpretive and reactive levels than expected and the participants without access to the zoom tool provided more terms than expected in the perceptual class.

Image Type and Number of Terms Supplied

. Another aspect of interest to the study was the relationship of the type of image to the number of terms supplied by the participants. In order to determine if there was a difference in the means of terms supplied for the different image types, a one-way ANOVA was performed. The lowest mean number of terms supplied was for the images in the portrait category followed by images in the landscape category. The two highest means belonged to the news and the cityscape categories of images.

Table 11

Mean Number of Terms Provided by Image Category

Groups	Mean	Standard Deviation
Portrait	38.00	34.41
Landscape	55.90	28.68
News	74.00	31.44
Cityscape	111.20	57.29

The results of the ANOVA indicate that there is significant difference between the means of the categories being studied, $F(3,36)=6.24$, $p<.01$.

Using the Tukey HSD post-test to determine the nature of the relationships between the means of the groups, it was determined that the mean of the city scene images was significantly different than the two groups of portrait and landscape ($p<.05$) There was no significant difference between the means of the remaining groups.

Table 12

Results of Tukey HSD

Group	Mean Difference	Sig.
City		
Portrait	73.20	.001
Landscape	55.30	.018
News	37.20	.173
Portrait		
City	-73.20	.001
Landscape	-17.90	.744
News	-36.00	.195
Landscape		
City	-55.30	.018
Portrait	17.90	.744
News	-18.10	.738
News		
City	-37.20	.173
Portrait	36.00	.195
Landscape	18.10	.738

Use of Zoom Tool in Relation to Image Categories

In addition to studying the number of terms applied to each type of image, the manipulation of the different types of images was also analyzed.Again, a one-way ANOVA was used to determine if the means were significantly different and the Tukey HSD test was applied to determine what the differences between the different image types were.

Table 13

Mean Number of Zooms by Image Category

Group	Mean	Standard Deviation
Portrait	1.07	3.56
Landscape	1.59	.4.91
News	2.15	5.15
Cityscape	3.32	7.72

On average, the zoom tool was used the least on the portrait and landscape categories and more often on the news and cityscape categories. The results of the ANOVA show that there is a significant difference between the means, $(3,1348)=10.29$, $p<.01$. The Tukey HSD post-test shows that the mean number of manipulations on the cityscape images was significantly different than the means of the other three categories. The means between the remaining categories were not significantly different.

Table 14

Results of Tukey HSD Post-test on Mean of Manipulations by ImageType.

Group	Mean Difference	Std. Error	Sig.
City			
Portrait	2.25	.426	.000
Landscape	1.73	.426	.000
News	1.17	.426	.031
Portrait			
City	-2.25	.426	.000
Landscape	-.52	.426	.613
News	-1.08	.426	.054
Landscape			
City	-1.73	.426	.000
Portrait	.52	.426	.613
News	-.56	.426	.551
News			
City	-1.17	.426	.031
Portrait	1.08	.426	.054
Landscape	.56	.426	.551

Type of Image and Shatford Layne's Model of Meaning

In examining how the terms fit into various models of meaning in relation to the type of image, a chi-square test shows a relationship between the type of image being viewed and the level of meaning supplied in Shatford Layne's model ($x^2(6, N = 10279) = 224.89, p < .01$) Looking at the relationships, the cityscape images were the only category that showed a lower than expected number of generic terms. In the specific category, the cityscape category showed more than twice the expected count while landscape and portrait did not even receive half the number of expected specific terms. The news category received very close to the expected count in each category. Because there were so few abstract terms provided, it is difficult to determine if a relationship might exist in this category.

Table 15

Number of Terms and Expected Frequencies in Shatford Layne's Model

Group	City	Landscape	News	Portrait
Generic				
Expected	2394.3	2566.2	2513.1	2450.3
Actual	2286.0	2607.0	2520	2511
Specific				
Expected	80.6	86.4	84.6	82.5
Actual	101	46	75	22
Abstract				
Expected	5.1	5.4	5.3	5.2
Actual	3	5	8	5

Type of Image and Enser's Model

How the different types of images fit into Enser's model of meaning was examined to see if any image type was more likely to be uniquely identified than others. The chi-square test indicated that with Enser's model the distribution of terms was very similar to that of Shatford Layne's model. The cityscape category received a lower percentage of terms in the non-unique category than any of the other categories and more than twice the expected count in the unique category. The portrait category received less than half the expected counts in the specific category.

Table 16

Number of Terms and Expected Frequencies in Enser's Model

Group	City	Landscape	News	Portrait
Unique				
Expected	76.0	82.5	80.4	78.1
Actual	186.0	46.0	67	21
Non-Unique				
Expected	2347.0	2549.5	2485.6	2411.9
Actual	2240	2586	2499	2469

Type of Image and Jörgensen's Model

An analysis of image type in relation to Jörgensen's perceptual attributes shows that there is a relationship between the type of image being viewed and the level of meaning in the terms supplied by participants. The results of the chi-square test show that the relationship is significant ($x^2(6, N = 9741) = 1176.77$, $p < .01$) which supports the hypothesis predicting a relationship between the terms categorized in Jörgensen's model and the type of image being viewed.

Examining the categories it appears that the categories appear to group differently than in previous models. Cityscapes and landscapes both have fewer than expected in the interpretive categories and more in the perceptual classes. The opposite is true of the categories news and portrait. These categories showed higher than expected numbers of terms in the interpretive category and fewer than expected in

the perceptual classes. With the reactive category, city and landscapes also registered higher than expected,

Table 17

Number of Terms and Expected Frequencies in Jörgensen's Model

Category	City	Landscape	News	Portrait
Perceptual				
Expected	1223.8	1241.0	1276.4	1182.9
Actual	1548.0	1548.0	742.0	1086.0
Interpretive				
Expected	1057.5	1072.4	1103.0	1022.1
Actual	712.0	663.0	1729.0	1151.0
Reactive				
Expected	139.7	141.6	145.7	135.0
Actual	161.0	244.0	54.0	103.0

Conclusion

This chapter has outlined the analysis performed on the data collected in this project and presented the findings and the statistics to explain the findings. Although there were areas where the data did not establish significant relationships between the variables being studies, there was evidence of significant findings in relationship to the research questions. The next chapter discusses these results and their significance to the field of image retrieval.

CHAPTER 5

DISCUSSION AND CONCLUSIONS

The purpose of this study was to determine if the use a tool to manipulate an image in a retrieval task could ultimately affect the way a person interacted with that image and determined meaning in an image. The secondary purpose of this image was to determine if there were any relationships between the type of image being viewed and how the user would interact with the image and assign meaning. This chapter explores the results of this study and discusses them in conjunction with existing theory in image retrieval and related fields. This chapter examines how these findings may provide additional information to practitioners in the field of image database management and access control. Finally, recommendations for further study in this area are discussed.

Use of the Zoom Tool

The zoom tool in the study and its inclusion in the interface of the application drove the purpose of this study. How participants would use it, how that use would compare to participants without access to the zoom tool, what kinds of terms would be provided were the main points of this study and surprisingly, there was little use of the zoom tool overall. With an average number of uses per image slightly above two, it appears that the zoom tool was not preferred as a helper tool by many of the participants. However, there was large variation among users as well as the image that was being viewed. With an overall average of less than 90 zooms per session, the variation in the number of uses is great with the most uses of the tool over the 40

picture set at 593 compared to one student who only used the zoom tool once. Overall, because the use of the tool was so varied, it would be difficult to predict how the tool would be used in a real-life retrieval situation. Patterns of use within the group were hard to determine as well. Both of the extremes above were from male participants aged 19 and 20 respectively. Female participants also have wide variations from a low of 7 uses of the tool during the exercise to 300 used by another female participant. Because the demographic groups within each group are so small, it is very difficult to determine if any patterns of use are statistically significant and cursory examinations do not even suggest any patterns of use unique to one group.

With a larger group of participants, a study may be able to determine if there are relationships between the gender or age or race and the use of the zoom tool, but from a practical standpoint, this information may not be useful in the implementation of the tool. Because the tool was used so infrequently, managers of image databases that are designing interfaces may decide that the inclusion of the tool would not be an asset to their database. However, the large variation in use indicates that the average number uses of the tool does not tell the whole picture. Many of the participants obviously used the tool frequently to complete the task and not including an image manipulation tool such as the zoom tool examined here may be a hindrance to the many users searching for an image in a true retrieval situation.

When comparing the number of terms supplied by both groups, the fact that there is no significant relationship between the use of the tool and the number of terms supplied indicates that the tool did not lead to numerous descriptions of the image. The fact that the tool would allow participants to zoom in on details that would allow them to

identify more items in the image or to retrieve clues about the subject of the image did not seem to have an impact.

Types of Images and Use Patterns with the Zoom Tool

The patterns of use of the zoom tool on the types of images being viewed was not surprising in that the images with more detail and more objects were the ones zoomed most often. The images that fell in the news and city categories received many more manipulations than the other two categories of landscape and portrait. The average user was three times more likely to zoom a cityscape than a portrait. This is not surprising given the nature of these two types of photographs. The city images contain many different items that may be unique to an image and warrant closer inspection. A street sign, a building, a car, or a person in a cityscape may be of particular interest to different users or may be used to determine the location in the image. A portrait holds very few items that are unique to the image type. In other words, a portrait of a person contains the same objects most of the time and a person's familiarity with the components of the human body would rarely warrant a closer look. Different clothing or some background elements may receive cursory inspection or manipulation with the tool, but this was rare in this study.

Information Retrieval

In light of Svenious's (1994) argument that text may not be the best approach to retrieving images and O'Connor and Greisdorf's discussion of the context of the image (2004) is an interesting facet in this study. Overall when supplying terms students with access to the tool and those without access to the zoom approached the task in similar ways. Most of the terms supplied were terms that named objects or actions in the

image. The higher levels of meaning or the more abstract categories in some of the models discussed were rarely explored participants in the study. Svenious' argument that text hinders the indexing of the aboutness of the image is demonstrated clearly. Participants who named objects in the image were not necessarily describing what the picture was about. The problem lies in the concept of 'aboutness' as many have discussed before. When indexing text documents, there are text clues available to determine the overall aboutness of a document. Many argue that this is a difficult concept when discussing text but when indexing images, this concept of retrieval becomes even more evasive.

The term 'about' is almost incongruent with meaning in an image. If asked what a book is about, most people will give a general summary of the main concepts presented in the book or possibly the plot of the book if it is a work of fiction. Rarely, would one use the term 'about' to inquire about the meaning of an image. The question. "What is the image about?" would more likely be asked, "What is the picture of?" This linguistic choice perhaps influences the way people approach image retrieval and how indexers provide access to images. When providing access to a text document, a wholistic approach is generally taken and rarely would a cataloger take the time or effort to include access to the parts such as chapters, figures, etc. This reflects the concept of what the text document is 'about' versus what an image is 'of'. The opposite seems to be the case when participants in both groups described what images were about. With most of the terms naming objects and actions in the images, participants were providing terms that told what the image was 'of'.

Cognitive Theory

Many of the concepts presented in the literature review regarding cognitive theory is supported by the findings of this research. One of the more commonly demonstrated concepts is the idea of phonemes or mental models in assigning terms to an image. Regardless of whether the participant had access to the zoom tool or not, terms supplied by participants in the study shared many similarities to previous studies of researchers. When users described many of the images, there were indications that users relied on mental models for identification of subjects. Instead of being cautious and assigning terms on concrete aspects of the images, participants many times supplied terms based on assumptions they made from items in the picture that appear to rely heavily on these models. For example, an image that shows a palm tree on a white sand beach was identified as Hawaii and Cancun without any indication of where the actual geographic location was. Participants' ideas on what represented these geographical locations mapped to the appearance of the image and the terms were assigned. This happened frequently, particularly in landscape and cityscape images. Mountains were often associated with Colorado and lighthouses were associated with New England.

Even access to the tool that could disprove incorrect terms was not strong enough to overcome these mental models. When viewing the image taken from the street in New York's China Town, four of the participants incorrectly identified the scene as one from China. Because the image depicts s busy street scene with Asian writing, it is understandable that this misidentification would occur. What is interesting to note is two of the participants had access to the zoom tool. Use of the zoom tool would reveal

street signs in English and a license plate on a car from Pennsylvania. Of the two, only one zoomed on the image once and missed this information that would have probably caused them to reconsider the term China.

These results with geographic terms are similar to those witnessed by O'Connor, O'Connor, and Abbas (1999). An image taken in Maine was identified as Oklahoma Panhandle. The authors discuss the idea that a "wrong" image could actually fit the information needs of a searcher. If the details are similar enough and nothing in the image distinctly identifies the actual location, one could argue that the image could be used by the searcher, even though the image is not actually "about" the term supplied.

Arnheim (1969) and Fischler and Firschein (1987) discuss the interaction of perception with other cognitive abilities such as language, memory, as well as culture when discerning meaning in an image. These influences also appeared in the responses to the tasks. In the examples above, these mental models could have been formed through previous exposure to the places named and caused the student to relate that memory to the image before them. This is seems to be the case to when terms were assigned that seemed to have no apparent reason for its appearance. A student who assigned 'beer bottle' to the image of the beach and one who assigned the term 'viagra commercial' to an image of an older couple hugging demonstrate the affect of previous memories and culture. Because these elements are based on the experiences and memories of the viewer, access to the zoom tool would not seem to have an affect on this aspect of assignment of terms.

Culture and beliefs can also be revealed in assignment of terms to an image that can be very difficult to accommodate in a traditional information retrieval system. For

example, race and the participant's reaction to it were interesting to note. Two of the images that contain African-American women garnered the term 'uneducated' by one participant. Politically sensitive terms such as 'colored people' was submitted by one student and 'oriental' was submitted by more than one student. The use of these terms is one example of how culture and beliefs about race could potentially cause problems for the image indexer. Would it be appropriate for an indexer to exclude these terms if there is evidence to suggest that users of a particular database were likely to use these terms in their search? Would it be appropriate for them to include these terms that may be offensive to some people? These questions increase the difficulties encountered when trying to accommodate the background of searchers in an image database.

In relation to the concept of locus of control, there was not an apparent relationship between the participants'' access to the zoom tool and the way they assigned terms. Users who used the zoom tool would arguably have more control over the task. If they were unsure of smaller details or wanted to inspect something closer before assigning a term, then they would be able to manipulate or control the task to a greater extent. As Neal (2006) found in her study, searchers preferred the more transparent approach to the search because of their perceived control over the process. In the interface, however, there seems to be little effect on the number of terms assigned.

Models of Meaning

When studying the complex process of image retrieval, a model that could easily and correctly explain how users assign meaning to images would be a helpful tool in designing systems and programs to assist in image retrieval. As this study demonstrates, that is a very difficult task. The three models discussed previously, approach the task in different ways and have proposed systems with some similarities and many differences. In studying the use of the zoom tool and using existing models to characterize the level of meaning in the terms supplied, it is apparent how difficult this task really is. Shatford Layne(1986) proposed a model based on how much of a cultural understanding is needed to identify a photograph. Her model is simplistic and when terms from both groups are compared using this model, it appears that users with access to the tool supplied more terms than expected at the specific level and fewer at the generic and abstract. On the surface, this seems to be easily explained. Users with access to the tool were able to zoom in and see more detail and name specific objects in the image than those without access. Users without, viewed the image as presented and possibly thought of the picture as a whole entity and not a collection of the objects in the image. However, the strongest thing the data reveals regardless of the presence of the tool is that a very large majority of the terms for both users were supplied at the generic level.

This model shares many similarities with Jörgensen in that they both move up a perceptual to interpretive hierarchy. Surprisingly though, the relationship between the use of the tool and its absence seems reversed in the Jörgensen model. Participants with access to the tool provided more terms than expected in the interpretive and

reactive classes than in the perceptual class. A study of the same terms in two different models that present seemingly different results indicates a great difference somewhere in the model. In this case, it seems that the inclusion of activity in Jörgensen's model was the biggest difference between the two results. In Shatford Layne's model, simple emotions were included in the generic category in addition to generic activities. Shatford Layne's level of generic is presented as a term that does not require any cultural understanding. In her model, basic emotions and activities would be understood by anyone from any culture. In Jörgensen's model, these two types of terms would appear in the interpretive class.

This difference in the models demonstrates the difficulty in coming to consensus when discussing how people determine meaning in an image. One could argue that human emotions are interpretive in nature and could be considered cultural. Examining many of the terms given that named emotions, it seems that Shatford Layne's belief in international understanding may be flawed. Even though it can be argued that simple emotions such as anger and despair cross cultural boundaries and can be identified regardless of culture, it does not follow that these emotions are necessarily easy to identify in images. In support of Jörgensen's model, it may be additional information in the image or previous experience that leads to interpretation of emotions. In this study, many images with people showing emotion were many times interpreted differently. An image of Cindy Sheehan protesting her son's death in Iraq was described with terms ranging from despair to fear. The image of a police officer talking with a woman in the background resulted in terms from bored to curious.

On images where there was more uniformity in emotions there were still small enough differences to question whether the idea of generic emotions can carry weight. Many students used the terms love or happiness to describe the image of the older woman and younger woman smiling at each other. Because of the frequency of these terms, you could argue that this would be a case of a universally recognized emotions but less frequent terms such as compassion and admiration also appeared. One could argue that happiness and love are more recognizable than the other two emotions named and that the other two are a result of interpretation. Making these distinctions creates difficulties in applying Shatford Layne's model in this case.

Another difference in the model that led to the differences in the results is Jörgensen's inclusion of all activities in the interpretive class. Again, this seems like an oversimplification of a term that could cover a large set of terms. An image that depicts ice skating in central park may not be easily identified by someone in equatorial Africa that has never seen an ice skating rink and this would support the idea that activities are interpretive. However, there are many actions that would not require any more interpretation than the objects appearing in the image. Relaxing, smiling, listening are all action words provided in this study that represent activities that require little interpretation.

This problem is exacerbated in Jörgensen's model when discussing specific people, objects or locations. Jörgensen includes both generic and specific locations, objects, and people in her perceptual class. The implication that all of these things can be identified with little interpretation seems contrary to the subclasses she includes in the interpretive categories. Again, referring to incorrectly identified locations in this study

indicate that users do interpret clues in the image as well as rely on their own knowledge and experiences to identify specific locations. This could be extended to people as well as statues. Indentifying a specific sculpture or stating the identity of a celebrity in an image is hard to reconcile to a perceptual level instead of an interpretive one.

Enser's model was difficult to examine in this study. Enser's model is the result of studies of actual search queries submitted to an image collection and this study did not collect data in the form of queries. Enser's work focuses more on the structure of the user's query and not necessarily the meaning associated with the image. The main categories in Enser's study focus on the specificity or generic nature of the subject requested and whether the request is further refined by various attributes. The results in the study do show some parallels to Enser's results as well as the other two models. A large percentage of the terms supplied were generic and there was no significant difference between groups on how they described images and their use of the tools. From a practical standpoint however, this reveals very little on how people assign meaning to the image they are viewing.

Problems with Concept Indexing and User Supplied Terms

The results of this research reinforce the problems previously stated concerning the process of indexing the meaning of photographs. The variety and number and variety of terms assigned by the group of participants demonstrates that indexing an image for access is fraught with problems.

Many argue that time and cost are the prohibitive factors when assigning terms to images. This can be easily demonstrated by examining one of the images and the

assigned terms. In the picture of the man in the green jacket in the portrait group, there were 233 terms assigned to the image. Of these terms, well over 100 were unique terms. If an indexer or cataloger were to assign 100 terms per image, the cost and time would be problematic and add to that problem would be the fact that 64 different participants may have many more terms that were not given in this study. As Brown, Hidderley, Griffin, and Rollason(1996) discovered when investigating user-supplied terms as possible indexing terms, the number of terms can be great and as it increases, the agreement between terms decreases.

When examining the terms from this study, there is also evidence of more of the problems discussed by O'Connor, O'Connor, and Abbas (1999). Terms for some images appeared to be contradictory in nature and when considered as indexing terms, simply do not make sense. Terms supplied for the image of the soldier who has lost an arm seemed to be polar opposites. How could an indexer assign contradictory terms such as grief and joy? Theoretically, to meet the needs of these participants, both terms would need to be assigned.

In addition to the contradiction between terms, there were also quite a few terms submitted of things that did not exist in the image, similar to Greisdorf and O'Connor's (2001) findings. As discussed previously, a beach image was assigned the term 'beer bottle' when there wasn't anything in the image that resembled a beer bottle. In fact, there is nothing man made in the image at all. Images of mountains resulted in terms such as 'skiing' when there was no evidence of that sport and the image of a woman being arrested resulted in the term 'drugs'. With some of these, one can understand the connections made. This story-telling or word association occurrence is very difficult to

predict and would be very difficult to accommodate in a traditional indexing situation. In the example above, many would agree that the link from snowy mountains to skiing is easy to make and should possibly be included but including an exhaustive list of similar terms such as avalanche or snowboarding is implausible and very likely impossible.

User Behavior and User Queries

As discovered by Colombo, Del Bimbo and Pala (1999), defining meaning with general subject matter is very difficult. Participants in the study were not experts in a certain field and the subject matter of all of the images were general in nature. As a result there was little agreement on meaning in the images from both the control and the treatment groups. The results in this study did however resemble those found in previous studies. In all three models of meaning, the terms provided by both groups fell predominantly in the lower levels where named objects comprised the majority of terms supplied.

Many times the results were surprising with images where one might expect a reversal of this trend because of the easily identifiable content. In an image of George W. Bush making a speech, participants were still more likely to identify the image with the generic term president than specifically naming George Bush. Of the 244 terms for that image, only 24 named the president by name whereas 36 used the more generic terms of president or leader. In an image of New York City's Times Square which is a highly recognizable landmark, similar results were observed. Of the 301 terms provided, 36 named specifically either New York City or Times Square. However the relationship between these two tended to favor the more generic term with 24 terms naming New York City and only 9 terms naming the exact location of Times Square. Even though

access to the zoom tool provided more access to details in an image that could have led to more specific term, this was rarely observed. With access to the zoom tool, participants could have note the location of West 46th street in the shot of Times Square and New York or used the tool to determine the location from the subway map, but there were the only two occurrences of that phenomenon in the entire data set.

Although the data in this study provides additional information on how users assign descriptions in an image, translating this information can be difficult when considering search strategies. As Markulla and Sormunen suggest, the selection of search terms is also influenced by the ultimate use of the image. Students were asked to assign descriptive terms to images without a context for the activity. The potential exists for a similar difference in terms supplied as seen in a search arena. A quick, small sample of terms assigned to images in two different Internet databases were studied to see if a similar phenomena may exist. A quick, non-scientific review of images in popular Internet image databases shows possible support of this idea. At the website Flickr®, where website users upload their photos to share with other users on the site, the creator of the content is asked to assign the terms for their images. Selecting 40 images at random using the most recent photos page and studying the number of terms assigned by the user shows that the average number of terms provided by the creator of the photograph was 8.3. Doing a similar sampling from an online stock photography site where users also supply the search terms there are very different results. The purpose of this website is to sell photography through a community website and the easier it is to find a photograph, the more likely it is to be purchased. As might be expected, the average was much higher at 33.6 terms on average assigned to

each photograph. This difference in purpose of the database demonstrates that the lack of context of this study may also have an effect on the number of terms being supplied by participants.

Manipulation of Images and the User Interface

Previous research (Hastings 1995, Greene, Marchionini, Plaisant, and Shneidermann, 2000) suggests that helper tools such as the zoom tool used in this study would be welcome additions to many databases and could facilitate browsing. The tool seems to have limited influence on how the users supply terms to describe the image but the potential for the tool should not be dismissed as a useful addition to an image retrieval interface. Instead, focus may need to be shifted to determine the usefulness or effectiveness in helping users select images to meet their information needs. Because there is some indication that meaning is based on mental models (Heidorn, 1999) or prototypes (Greisdorf and O'Connor, 2001) the tool may possibly be more useful in assisting the user in discriminating between retrieved images and making final selections for the intended use of the searcher.

From the results of this study, the conclusion that the type of image being examined can also influence the use of the zoom tool and the way the user interacts with the image. For images that have many details like cityscapes, the zoom tool was used significantly more than the other categories. In databases that contain very detailed images, the zoom tool may be more useful and appropriate than in a database of other types of images such as landscapes or portraits. Again, the ultimate use and user of the image may have an impact on whether the tool should be included as brush strokes or techniques may warrant a closer look in an art history database containing

portraiture. Additional study may need to be conducted in specialized collections or with expert users to determine the usefulness of the manipulation tools.

Overall, in a generic set of images being examined by non-experts, the use of the tool overall seemed closely tied to personal preference. Even though there was no significant relationship between the number of terms supplied in relation to the use of the tool, it was very apparent from the results that there were great variances in the preferences of the users with access to the tool. Given the same task and looking at the same 40 images, the range of 592 manipulations in the data set is worth additional study as well. Were there similarities between the users beyond the data collected that could account for the differences in use of the tool. If so, this information could be useful to interface designers when creating interfaces for different applications in image retrieval.

Further Research

This study is just one aspect of the larger picture of image retrieval. Even though the results were not conclusive in many areas that were studied, there were some results that prove interesting. Because this study purposely did not center on one subject or analyze the keywords of experts in a certain area, there may be different results if studied in a different domain. If the use of image manipulation tools by art historians or medical personnel were studied, would one see different patterns of usage and different levels of meaning in the terms provided. .

In addition to the domain differences, further study needs to be conducted on the user and how they assign meaning to images. What motivated the users representing extremes on each end of the number of zooms or the number of terms assigned to

exhibit the behaviors they did. For example, Neal (2006) examined the concept of locus of control and its relationship with image retrieval. Research should be conducted to see if there are other cognitive differences or demographic differences between users that may affect the way meaning is determined. Further understanding of the user and differences that exist may lead to more effective retrieval tools and models of meaning and guidelines in concept indexing.

This study also focused on a very small subset of image manipulation. Although zooming and re-centering of the image are potentially useful tools in the image database interface, there are many other tools available that allow users to add notation to areas of an image, allow for cropping of images, comparisons of multiple images, among many others. Expanding the scope of this study to see if any of these tools may have an effect on the way users assign descriptions to images.

Conclusions

This study has examined the use of a tool that allowed users to manipulate an image by zooming and studied the effect the tool had on the terms supplied by the participants. In addition to the terms provided by the users, the types of images zoomed and also the number of assigned to different types of images were studied to provide further insight on the use of the tool and how meaning is assigned to images. In the field of image retrieval, the phrase “an image is worth a thousand words” has become something of a cliché that tries to illustrate the complexity of providing access to image databases.

Hundreds of years of research has been dedicated to the control of access to text based information and only recently have concerted efforts been directed towards

the retrieval of images. The different paths of concept versus content based research both strive to search for practical solutions of reducing the more than “thousand” possibilities to an effective and efficient procedure to control the growing number of computerized image databases. Even though we may never be able to pinpoint the meaning of an image since it is many times tied so closely to the user (O’Connor, 1996), better understanding of how the user thinks and the ways in which they interact with retrieval systems will ultimately allow storage and access to visual information that is growing exponentially in a digital age.

APPENDIX A

CONSENT FORM

Before agreeing to participate in this research study, it is important that you read and understand the following explanation of the purpose and benefits of the study and how it will be conducted.

Title of Study: Image Manipulation and User-supplied index terms.

Principal Investigator: Leah Schultz, a graduate student in the University of North Texas (UNT) Department of Library and Information Science

Purpose of the Study:
You are being asked to participate in a research study which will investigate how people think about retrieving images from a database and how certain online tools may affect the way people interact and think about images.

Study Procedures:
You will be asked to view 40 images online and provide terms that you think might be useful for categorizing the images. The task will take approximately one hour of your time.

Foreseeable Risks:
There are no foreseeable risks in participating in this exercise.

Benefits to the Subjects or Others:
This study is not expected to be of any direct benefit to you but will help researchers gain insight into how people search databases for images and ultimately improve the ability of users like you to find images in databases.

Procedures for Maintaining Confidentiality of Research Records:
Information that could identify you as individual will not be collected for the purpose of this study. We only ask that you provide us with basic demographic information in addition to the terms you supply for the images.

Questions about the Study
If you have any questions about the study, you may contact Leah Schultz at telephone number XXXXXX or Dr. Samantha Hastings, UNT Department of Library and Information Science, at telephone number XXXXXX.

Review for the Protection of Participants:
This research study has been reviewed and approved by the UNT Institutional Review Board (IRB). The UNT IRB can be contacted at (940) 565-3940 with any questions regarding the rights of research subjects.

Research Participants' Rights:
Clicking "I agree and wish to participate" below indicates that you have read or have had read to you all of the above and that you confirm all of the following:

- Leah Schultz has explained the study to you and answered all of your questions. You have been told the possible benefits and the potential risks and/or discomforts of the study.
- You understand that you do not have to take part in this study, and your refusal to participate or your decision to withdraw will involve no penalty or loss of rights or benefits. The study personnel may choose to stop your participation at any time.
- You understand why the study is being conducted and how it will be performed.
- You understand your rights as a research participant and you voluntarily consent to participate in this study.
- You have been told you will receive a copy of this form.

I agree and wish to participate I do not wish to participate

(Participants will select a link implying consent to participate)

APPENDIX B

SURVEY

Demographic Information

1. Gender: Male ☐ Female ☐

2. Major: _____

3. Classification: ☐ Undergraduate Student ☐ Graduate Student

4. Age: _____

5. Ethnicity: ☐ African American ☐ Asian ☐ Caucasian ☐ Hispanic ☐ Other (please specify) _____

6. Please indicate your comfort level with computers:
☐ Very Uncomfortable ☐ Somewhat Uncomfortable ☐ Neutral ☐ Somewhat Comfortable ☐ Very Comfortable

Instructions

In the following exercise, you will be shown images and will be asked to provide terms to describe the images. You can provide as many terms as you feel necessary to describe the image.

Once you have provided your terms for an image, click the next button to continue, until all images have been completed.

Thank you for your time.

(TREATMENT GROUP INSTRUCTIONS) Below is a short animation showing the use of the zoom tool. (short animated video inserted showing the use of the zoom tool)

You can zoom in and out on the picture using the tool bar at the bottom of the image below. Practice with the toolbar on the picture below until you feel comfortable continuing.

(Users will then be presented with the 40 images on page similar to the one below.

APPENDIX C

IMAGES USED IN STUDY


Kodak


MOULIN ROUGE
Féerie

Note. Copyright by istockphoto.com. Reprinted with permission

Note. Copyright by istockphoto.com. Reprinted with permission

Note. Copyright by istockphoto.com. Reprinted with permission.

Note. Copyright by istockphoto.com. Reprinted with permission.

Note. Copyright by istockphoto.com. Reprinted with permission.

Note. Copyright by istockphoto.com. Reprinted with permission.

Note. Copyright by istockphoto.com. Reprinted with permission.

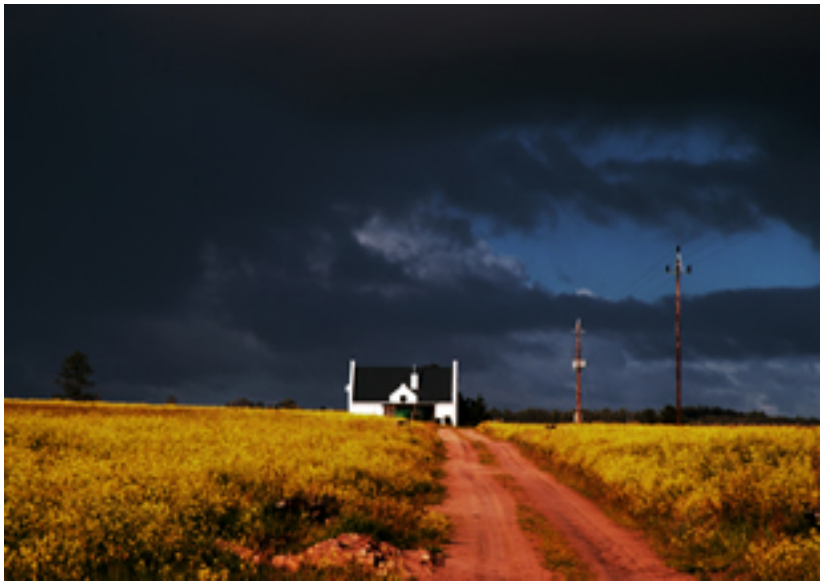
Note. Copyright by istockphoto.com. Reprinted with permission.

Note. Copyright by istockphoto.com. Reprinted with permission.

Note. Copyright by istockphoto.com. Reprinted with permission.

Note. Copyright by istockphoto.com. Reprinted with permission.

Note. Copyright by istockphoto.com. Reprinted with permission.



Note. Copyright by istockphoto.com. Reprinted with permission.

Note. Copyright by istockphoto.com. Reprinted with permission.

Note. Copyright by istockphoto.com. Reprinted with permission.

Note. Copyright by istockphoto.com. Reprinted with permission.

Note. Copyright by istockphoto.com. Reprinted with permission.

Note. Copyright by *Dallas Morning News*. Reprinted with permission.

Note. Copyright by *Dallas Morning News*. Reprinted with permission.

Note. Copyright by *Dallas Morning News*. Reprinted with permission.

Note. Copyright by *Dallas Morning News*. Reprinted with permission.

Note. Copyright by *Dallas Morning News*. Reprinted with permission.

Note. Copyright by *Dallas Morning News*. Reprinted with permission.

Note. Copyright by *Dallas Morning News*. Reprinted with permission.

Note. Copyright by *Dallas Morning News*. Reprinted with permission.

Note. Copyright by *Dallas Morning News*. Reprinted with permission.

Note. Copyright by *Dallas Morning News*. Reprinted with permission.

APPENDIX D

SAMPLE DATA

Sample data from two participants, one with access to the zoom tool and the other without access.

Table D.1

Terms supplied by participant without access data to the image tool..

Image	Search Term
Image 1	bright beautiful calming engaging
Image 2	funny happy silly imbarrised
Image 3	brilliant holy secret orange
Image 4	messy destruction construction
Image 5	Light city

Image	Search Term
	exciting
	busy
Image 6	innocent
	quiet
	happy
	young
Image 7	peaceful
	distant
Image 8	Sad
	distress
	content
Image 9	gloom
	noisy
	grey
Image 10	joyful
	happy
	innocent
	young
	nyeve

Image	Search Term
Image 11	Blue
	warm
	cool
	calm
	relaxing
	beautiful
Image 12	irritated
	obligated
	annoyed
	tired
Image 13	Cold
	busy
	fun
	social
Image 14	family
	happy
	joyful
Image 15	beautiful
	warm
	green
	autum

Image	Search Term
	colorful
Image 16	proud
	relief
	sorrow
	pride
Image 17	busy
	crowded
	loud
Image 18	family
	happy
Image 19	calming
	breezy
	cool
	green
Image 20	turmoil
	anger
	distress
	confusion
Image 21	classic
	rustic

Image	Search Term
	culture
Image 22	innocent
	beautiful
	simple
	elogant
Image 23	Rich
	content
	peacful
Image 24	Sad
	hurt
	pain
Image 25	gloomy
	classy
	classic
	vintage
	narrow
Image 26	laughter
	silly
	imbarrased
Image 27	pretty
	rich

Image	Search Term
	calm
	alive
Image 28	stern
	assurance
	hesitance
	confusion
	adimincy
Image 29	crowded
	busy
	narrow
	long
Image 30	Love
	life
	laughter
	happy
	joyful
Image 31	georgous
	cool
	calm
	thoughtful
	pure

Image	Search Term
Image 32	distress
	distraght
	upset
Image 32	Sad
Image 33	model
	same
Image 34	colorful
	content
	satisfied
	shopping
	radience
Image 35	narrow
	path
	green
	trip
Image 36	president
	authority
	promises
	asurance
Image 37	medal
	busy

Image	Search Term
	alive
	exciting
	extravagant
Image 38	cool
	satisfied
	attractive
	calming
	tired
	content
Image 39	Life
	alive
	green
	rich
	new
Image 40	Tired
	sad
	disappointed
	distressed
	upset

Table D.2

Terms supplied by participant with access data to the image tool.

Image	SearchTerm
image1	Lights Night time Dallas Watar Beautiful Buildings
image10	school colorful fun computers kids
image11	big motor sand water palm trees boat georgeous secluded beach

Image	SearchTerm
image12	police unhappy people crazy hat
image13	winter ice skating city scape park people
image14	grand daughter grandma picket fence braces family
image15	mountains colorful
image15	snow trees

Image	SearchTerm
image16	father reunion proud welcome home american marines missing hand son sad/happy
image17	cargo vans trading company asian mural work truck busy street
image18	family picture golf many colors trees
image19	green grass american flag

Image	SearchTerm
	water light house cliff
image2	runner rainy asian happy cool outside green
image20	boxer saddam husaine another flag killing protest american flag
image21	red down town fun place stores cargo truck wind mill

Image	SearchTerm
image22	green eyes half face brown hair little girl green plants childish
image23	rock wall trees rocks ocean mountains
image24	blood protective clothing ambulance medical workers backboard crime scene sign
image25	old town red building steepel

Image	SearchTerm
	dark sky
image26	brooklyn weight loss card new york cold happy map subway ride city
image27	rocks bright light trees brook plants stream
image28	conference microphone city meting dallas sheriff press

Image	SearchTerm
image29	line of cars on street bridge bell tower congested city
image3	rocks orange shining light ridgid cave
image30	marriage loving happy older couple blue coat brown cout fall tree color
image31	pine trees colorful sky snow cold beautiful sky

Image	SearchTerm
image32	handcuffs accident busy road pregnant woman police yellow hat
image33	warehouse building pale colors cottage houses old buildings
image34	hot chick colorful bags fun times bras panties department store jean jacket towles reflection of a woman in the glass shopping

Image	SearchTerm
image35	tunnel orchard light at the end of the tunel leaves trail trees
image36	US seal meeting podium american flag pin funny hats president bush
image37	city ferry boat castel weird building designs glass
image38	surfer people wet suit sand

Image	SearchTerm
	beach surf board ocean
image39	yellow flowers house people orchard garden
image4	trash tragedy barrel workers storm pool broken fence
image40	lots of people sick woman war protester water hot

Image	SearchTerm
	unhappy big hat
image5	bright colors down town big city lots of people sky scrapers
image6	young girl woode barn happy dress blonde hair cute picture swinging doors
image7	dirt road coastel field stormy night deserted country

Image	SearchTerm
	farm house gloomy
image8	candels
image8	stuffed animals
image8	flowers
image8	friends
image8	balloons
image8	sadness
image8	morning
image8	death
image8	family
image9	busy town
image9	antiqueish building
image9	buses
image9	trains

APPENDIX E

INSTITUTIONAL REVIEW APPROVAL FORM

UNIVERSITY of NORTH TEXAS

Office of Research Services

January 20, 2006

Leah Schultz
School of Library and Information Sciences
University of North Texas

Re: Human Subjects Application No. 05-328

Dear Ms. Schultz

As permitted by federal law and regulations governing the use of human subjects in research projects (45 CFR 46), the UNT Institutional Review Board has reviewed your proposed project titled "Image Manipulation and User-supplied Indexing Terms." The risks inherent in this research are minimal, and the potential benefits to the subject outweigh those risks. The submitted protocol and consent form are hereby approved for the use of human subjects in this study **Federal Policy 45 CFR 46.109(e) stipulates that IRB approval is for one year only**.

Enclosed is the consent document with stamped IRB approval. Please copy and **use this form only** for your study subjects.

It is your responsibility according to U.S. Department of Health and Human Services regulations to submit annual and terminal progress reports to the IRB for this project. Please mark your calendar accordingly. The IRB must also review this project prior to any modifications.

Please contact Shelia Bourns, Research Compliance Administrator, or Boyd Herndon, Director of Research Compliance, at extension 3940, if you wish to make changes or need additional information.

Sincerely,

Scott Simpkins, Ph.D.
Chair
Institutional Review Board

P.O. Box 305250 ♦ Denton, Texas 76203-5250 ♦ (940) 565-3940
Fax (940) 565-4277 ♦ TTY (800) RELAY TX ♦ www.unt.edu

REFERENCES

Armitage, L.H. & Enser, P.G.B (1997). Analysis of user need in image archives. *Journal of Information Science*, *23*(4), 287-299.

Arnheim, R. *Visual thinking*. Univeristy of California Press, Berkely, 1969.

Bianchi-Berthouze, N. & Wakamatus, A. (2001). Kansei-Mining: Identifying visual impressions as patterns in images. IFSA World Congress and 20th NAFIPS International Conference, Vancouver, Canada, *4*, 2183-2188.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115-147.

Bjarnestam, A. (1998). Text-based hierarchical image classification and retrieval of stock photography in the challenge of image retreival 1998, Newcastle, United Kingdom, retrieved online November 18, 2006 from http://www.bcs.org/server.php?show=ConWebDoc.4432

Blair, D.C. (2003). Information retrieval and the philosophy of language. *Annual Review of Information Science and Technology*, *37*, 3-50.

Borgman, C. L., Smart, L.J., Millwood, K.A., Finley, J.R., Champeny, L. Gilliand, A.J., Leazer, G.H. (2005). Comparing faculty information seeking in teaching and research: Implications for the design of digital libraries, *Journal of the American Society for Information Science and Technology, 56*(6), 636-657.

Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology, 54*(10), 913-925.

Bovey, J.D. (2003). Providing web access to a catalogue of British newspaper cartoons. *Program: Electronic Library and Information Systems*, *37*(1), 16-24.

Buckland, M. & Gey, F. (1994). The relationship between recall and precision. *Journal of the American Society for Information Science*, *45*(1),12-19.

Burford, B., Briggs, P. & Eakins, J.P. (2003). A taxonomy of the image: On the classification of content for image retrieval. *Visual Communication*, *123*(2), 123-161.

Chen, H. (1999). An analysis of image queries and image retrieval tasks in the field of art history. *Dissertation Abstracts International*, *61* (01), 14. (UMI No. 99577215)

Chen, H.(2001). An analysis of image queries in the field of art history. *Journal of the American Society for Information Science and Technology*, *52*(3), 260-273.

Choi, Y. & Rasmussen, E. (2003). Searching for images: The analysis of users' queries for image retrieval in American history. *Journal of the American Society for Information Science and Technology*, *54*(6), 498-511.

Chu, H. (2001). Research in image indexing and retrieval as reflected in the literature. *Journal of the American Society for Information Science and Technology*, *52*(12), 1011-1018.

Collins, K. (1998). Providing subject access to images: a study of user queries. *American Archivist*, *61*, 36-55.

Colombo, C., Del Bimbo, A. & Pala, P. (1999). Semantics in visual information retrieval. *IEEE Multimedia*, *6*(3), 38-53.

Combs, T.T.A., & Bederson, B.B. (1999). Does zooming improve image browsing? *Proceedings of the fourth ACM conference on Digital libraries,* (pp. 130-137), Berkely: ACM Press.

Cooper, W.S. (1969). Is inter-indexer consistency a deep hobgoblin? *American Documentation*, *20*(3), 100-110.

Cooper, W.S. (1973). On selecting a measure of retrieval effectiveness. *Journal of the American Society for Information Science*, *24*(2), 87-100.

Eakins, J.P., Briggs, P., & Buford, B. (2004).Image retrieval interfaces: A user perspective. Image and Video Retreival: Third International Conference, CIVR 2004, Dublin, Ireland.

Enser, P. (2000). Visual image retrieval: Seeking the alliance of concept-based and content-based paradigms. *Journal of Information Science*, *26*(4), 199-210.

Enser, P. (1995). Pictorial information retrieval. *Journal of Documentation*, *51*(2), 126-170.

Enser, P.(1993). Query analysis in a visual information retrieval context. *Journal of Document and Text Management*, *1*(1), 25-52.

Fidel, R. (1997) The image retrieval task: Implications for the design and evaluation of image databases. *New Review Hypermedia and Multimedia*, *3*, 181-198.

Frost, C.O., Taylor, B., Noakes, A., Markel, S., Torres, D., & Drabenstott, K.M. (2000). Browse and search patterns in a digital image database. *Information Retrieval*, *1*, 287-313.

Goodrum, A.A., Rorvig, M.E., Jeong, K., & Suresh, C. (2001). An open source agenda for research linking text and image content features. *Journal of the American Society for Information Science and Technology*, *52*(11), 948-953.

Graham, M.E. (2001). The cataloguing and indexing of images: Time for a new paradigm? *Art Libraries Journal*, *26*(1), 22-27.

Greene, S., Marchionini, G., Plaisnat, C., & Shneidermann, B. (2000). Previews and overviews in digital libraries: Designing surrogates to support visual information seeking. *Journal of the American Society for Information Science*, *51*(4), 380-393.

Greisdorf, H.F. & O'Connor, B.C.(2002). Modelling what users see when they look at images: A cognitive viewpoint. *Journal of Documentation*, *58*(1), 6-29.

Greisdorf, H.F. & O'Connor, B.C.(2008). *Structures of image collections: From Chauvet-Pont-d'Arc to Flickr*. Libraries Unlimited, Wesport, Connecticut.

Hastings, S.K. (1995). Query categories in a study of intellectual access to digitized art images. *Proceedings of the 58th Annual Meeting of the Society for Information Science, ASIS '95*, *32*, 3-8.

Hastings, S.K. (1999). Evaluation of image retrieval systems: Role of user feedback. *Library Trends*, *48*(2), 438-452.

Heidorn, B.P (1999). Image retrieval as linguistic and nonlinguistic visual model matching. *Library Trends*, *48*(2), 303-325.

Hidderley, R. & Rafferty, P.(1997). Democratic indexing: An approach to the retrieval of film. Library and information studies : research and professional practice : Proceedings of the 2nd British-Nordic Conference on Library and Information Studies, Queen Margaret College, Edinburgh, *2*, 90-102.

Hornbaek, K., Bederson, B.B., & Plaisant, C. (2002). Navigation patterns and usability of zoomable user interfaces with and without an overview**.** *ACM Transactions on Computer-Human Interaction*, *9*(4), 362-389.

Jefferson, G. (2007). Technology keeps people focused on new cameras. *USA Today*, 11/21/2007. Retrieved from Ebscohost on 8/13/2008.

Jörgensen , C. (1996). An investigation of pictorial image attributes in descriptive tasks. *Proceedings of the SPIE-the International Society for Optical Engineering*, *2657*, 241-251.

Jörgensen , C. (1998). Attributes of images in describing tasks. *Information Processing and Management*, *34*(2/3), 161-174.

Jörgensen , C. (1999). Access to pictorial material: A review of current research and future prospects. *Computers and the Humanities*, *33*, 293-318.

Jörgensen, C. (2003). *Image retrieval: Theory and research*. Scarecrow Press. Lanham, Maryland.

Jörgensen, C. (2004). The visual indexing vocabulary: Developing a thesaurus for indexing image across diverse domains. *Proceedings of the 67th ASIS&T Annual Meeting*, *41*, 287-293.

Jörgensen, C. & Joregensen, P. (2005). Image querying by image professionals. *Journal of the American Society for Information Science and Technology*, *56*(12), 1346-1359.

Krause, M. (1998). Intellectual problems of indexing picture collections. *Audiovisual Librarian*, *14*, 73-81.

Krippendorf, K. (2004*). Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage Publications.

Liu, H., Xie, X., Ma, W. & Zhang, H. (2003). Automatic browsing of large pictures on mobile devices*. Proceedings of the eleventh ACM international conference on Multimedia,* 148-155.

Lyman, P. & Varian, H.R. (2003) How much information 2003. Retrieved 9/2/2008 from http://www.sims.berkeley.edu/how-much-info-2003.

Macworth, N.H. & Morandi, A.J. (1967). The gaze selects informative details within pictures. *Perception and Psychophysics*, *2*(11), 547-552.

Markkula, M. & Sormunen, E. (2000). End-user searching challenges indexing practices in the digital newspaper photo archive. *Information Retrieval*, *1*, 259-285.

Neundorf, K.(2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage Publications.

O'Connor, B.C. (1992). Preservation and repacking of lantern slides within a desktop digital imaging environment. *Microcomputers for Information Management*, *9*(4), 209-224.

O'Connor, B.C. (1996) *Explorations of indexing and abstracting: pointing, virtue, and power.* Englewood, CO, Libraries Unlimited.

O'Connor, B. C. (2001) Classroom lecture, given during the spring 2001 semester at the University of North Texas.

O'Connor, B. and Greisdorf, H. (2004) Hey look at this! Illocutionary discourse and snapshots. *Proceedings of the 67th ASIS&T Annual Meeting, Vol. 41*.

O'Connor, B.C., O'Connor, M.K, & Abbas, J.M. (1999). User reactions as access mechanism: An exploration based on captions for images*. Journal of the American Society for Information Science, 50*(8), 681-697.

O'Connor, B.C. & Wyatt, R.B. *Photo provocations: Thinking in, with, and about photographs*. Scarecrow Press. Lanham, Maryland, 2004.

Organisation for Economic Co-operation and Development (2005). *OECD Broadband Statistics*, June 2005. Retrieved October 23, 2005 from http://www.oecd.org/document/16/0,2340,en_2825_495656_35526608_1_1_1_1,00.html

Ornager, S. (1997). Image retrieval: Theoretical analysis and empirical user studies on accessing information for images. *Proceedings of the 60th Annual Meeting of the Society for Information Science*, ASIS '97, *34*, 202-211.

Panofsky, E. (1955). *Meaning in the visual arts: Papers in and on art history*. Garden City, NY: Doubleday, Inc.

Rasmussen Neal, D. (2006). New photography image retrieval practices: Locus of control in two contexts. Unpublished doctoral dissertation, University of North Texas, Denton, TX

Roberts, H.E. (2001). A picture is a worth a thousand words: Art indexing in electronic databases. *Journal of the American Society for Information Science and Technology, 52*(11), 911-916.

Rorissa, A. & Hastings, S.K (2004). Free sorting of images: Attributes used for categorization. *Proceedings of the 67th ASIS$T Annual Meeting*, *41*, 360-366.

Shatford, S. (1984). Describing a picture: A thousand words are seldom cost effective. *Cataloging and Classification Quarterly*, *4*(4), 13-30.

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

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

Svenonius, E. (1994). Access to nonbook materials: The limits of subject indexing for visual and aural languages. *Journal of the American Society for Information Science*, *45*(8), 600-606.

Turner, J.M. (1995). Comparing user assigned terms with indexer assigned terms for storage and retrieval of moving images. *Proceedings of the 58th Annual Meeting of the Society for Information Science, ASIS '95*, *32*, 9-12.

Yoon, J. (2006). Improving recall of browsing sets in image retrieval from a semiotics perspective.(UMI No. AAT 3214504) Retrieved September 10, 2008 from *Dissertation and Theses database*.