Title
Search Log Analysis of the ARTstor Cultural Heritage Image Database

Author
Lowe, Heather Ann

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Search Log Analysis of the ARTstor Cultural Heritage Image Database

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Library and Information Science

by

Heather Ann Lowe

2013

ABSTRACT OF THE THESIS

Search Log Analysis of the ARTstor
Cultural Heritage Image Database

by

Heather Ann Lowe

Master of Library and Information Science
University of California, Los Angeles 2013
Professor Jonathan Furner, Chair

Search log studies are widely used to gain an understanding of how users interact with an information resource. However, of the studies that examine image databases, the information systems studied are image sites, local collections or general web engines. Few if any studies examine the way users search within a general art and cultural heritage image database. This thesis seeks to discover the types of queries users submit, whether these change over time, and if these differ depending on the user's institution type. Samples from the ARTstor search log were coded for uniqueness, category, and type of art historical information they contained. The results suggest that the way users query ARTstor has shifted over time but that they maintain a large reliance on art

historical information. In order to further research on image databases across cultural heritage institutions, this thesis makes recommendations for further research and transaction logging.

The thesis of Heather Ann Lowe is approved.

Johanna R. Drucker

Ellen J. Pearlstein

Jonathan Furner, Committee Chair

University of California, Los Angeles

2013

For my brothers Jeff and Jason,

my compass, my first and best friends.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

There are many to whom I am deeply indebted for the completion of this thesis. As my research assistants Brooke Sansosti and Regina Kammer, I owe a debt of gratitude particularly because they tackled a tedious classification project with such enthusiasm that it kept me energized. As a friend and motivator, Dustin Locke helped me keep focus and balance throughout the process.

Many thanks are owed to Susan Chun for her support and the introduction into transaction log analysis. My gratitude also extends to ARTstor for releasing the search log data set, to the Indianapolis Museum of Art for hosting the data, and to Dustin Wees at ARTstor for his curiosity in and support of the project.

I am also indebted to all of the wonderful faculty and staff of the Department of Information Studies at the University of California, Los Angeles. I offer sincere thanks to Susan Abler and Gregory Leazer for their help in navigating the administrative hurdles involved. Most of all, my complete gratitude goes to my committee, Johanna Drucker, Jonathan Furner, and Ellen Pearlstein for their guidance, support, and generosity.

1. Introduction

As professional workflows and educational experiences become increasingly mediated by digital interfaces, there is a greater need to find, use, and provide images from a range of sources. With this changing paradigm, museums, educational institutions, and other collecting bodies view the digitization of their collections as a necessity. Now that digital images are both desired by end-users and available to them, the greatest difficulty in providing true access to images is the problematic task of searching for visual material using a verbal language. Though research into improving image access spreads in numerous directions including content-based retrieval and social tagging, we still know so little about how users actually search for images. Few studies verify the types of terms searchers use, what affects the users' choices, or how users modify their search strategies when unsuccessful.

Learning users' search habits and preferred term types is essential to properly designing effective user interfaces. Using transaction logs, which track not only search terms used but other interactions with a web interface, we can pinpoint areas of frustration for users, places where they abandon their search and the most often used means of access. Because many educational databases, like ARTstor an academic cultural heritage image database, are accessed through an institutional subscription, we can examine the differences in search habits across different types of users. This might lead to creating specialized interfaces for different institutions, perhaps one aimed at museum professionals and another aimed at k12 teachers and students.

A better understanding of the ways in which users search might also inform other factors involved in the retrieval of images such as how we catalog. If users overwhelming search based on certain categories of terms, we might begin to design browsing structures to facilitate finding images along such categorical lines (e.g. artist name, color, etc.). Generating better browsing structures may prove important as many of the users of art resources note that browsing is an important strategy within their information-seeking toolkit.

The vast majority of studies to date focus on large general image web databases like Yahoo! Images or small, specialized museum image collections. Because these types of image databases are quite different in content to general academic cultural heritage image databases, little from these studies can inform how these types of databases are used. In 2009, however, ARTstor released the contents of its search log to the public for study. The data set provides the first opportunity to investigate how users interact with academic art image databases. The search log tracks little outside of the actual terms used to search rather than the full range of user interaction with the interface, so what we can learn from the search log is somewhat limited.

However, the ARTstor search log does provide an important opportunity to create a baseline study of how users search in investigating the most basic unit of access—the words used to search. Through study of these terms, we can determine the types of search terms used and compare searching trends across different types of institutions. Because the service is subscription-based and

accessible only through institutional proxy, searches can be associated with a type of institution, such as a library, museum, k12, or college.

The ARTstor interface and collections have also changed over time, so it is possible to discover and compare any corresponding changes in how users searched. If these are discovered, it might suggest that there are factors beyond information need that affect how a user structures her search.

The impetus for the specific aims of this project stem from the hypothesis that users search for art images in a manner different from general images. Largely this may stem from divergent needs that bring the user either to an art image database or a general image database. Given such a supposition, this research seeks to ask the question what is the relationship between the nature of the collection held within a database and the way in which users approach searching within that collection. Additionally, the research asks what is the relationship between the community from which the user comes and the way in which the users approaches searching. Using the ARTstor search log, this study seeks to do the following:

a. To discover the types of terms users employ when searching for general art images, so that we might better understand how to catalog images, structure browsable categories, and react to user information needs.
b. To discover if there are major differences among types of terms used based on the institution through which the user accesses ARTstor, so that we might understand how to better serve different

types of communities and pinpoint how to improve interfaces delivered to these communities.

c. To discover if the nature of searches have changed over a period of time and if this seems to mirror the changing make-up of the database itself, so that we might better understand the factors affecting users' search strategies.

d. To suggest ways in which image databases might improve transaction logs to better capture meaningful data about user interaction, so that we as a community of cultural heritage and information professionals might reach a consensus as to the best practices for researching user interaction with image databases.

The aims of this research project seek to create a baseline from which other researchers can begin to ask new questions. Beyond simply demystifying the ways in which users search for art images, the manual classification of search terms may also create an important data set for training machines to automatically categorize searches or image tags. With a better understanding of the basics of user art image search and what is still unknown, we can make better decisions about what types of interactions should be recorded in transaction logs in the future.

2. Literature: Image Description and access

2.1 Overview

Access to art and its visual surrogates has been a persistent priority for artists, art historians, and other cultural heritage professionals, one that predates the advent of digital images and online databases. It is not art historians alone who wish to use cultural heritage images but also a wide range of those from various backgrounds including general humanities scholars, museum professionals, k12 educators, artists, students, and many others. Not only do the types of people seeking images vary but also the reasons for which images are sought. Within a single academic image collection, users report a range of purposes for images from personal research and collection management to outreach and classroom education (Frost et al 2000, 293). Some informational needs, such as those of practicing artists, may not be articulated concisely but rather inhabit a more generalized need.[1]

Even when a user has a specific quantifiable information need, finding images can be problematic. Images contain no linguistically embedded subject clues to enable subject searching unlike text-based work completely comprised of searchable terms from which keywords may be drawn even without expert cataloging.[2] Complicating retrieval further, cultural heritage images are generally open to a multitude of interpretations some conflicting, and bibliographic

[1] In a survey of studio art students, Cobbledick found that students often engaged in a browsing that was not simply a means to obtain specific information but a method by which students simultaneously found the information they needed and realized what that need may have been (1999, 450-451). Therefore such browsing or searching may not have a namable information need for some users until after they find an object they wish to use.

[2] Image files can, of course, have embedded metadata, but this requires cataloging to input such information into the image file.

cataloging no matter how thorough is unable to supply all possible descriptions.

Some scholars suggest that Content-Based Image Retrieval (CBIR) may offer a significant supplement to bibliographic metadata. In fact, facial recognition and object recognition have made their way into common consumer products, particularly in mobile contexts like iphoto and Google goggles.

Despite great progress in the field, the capabilities of such retrieval are unlikely to fully replace human interpretative cataloging in the near future (Enser et al 2007, 468-469). Were Content-Based Image Retrieval perfected, it would still face serious challenges to meet many search requests. Among these challenges are determining the creator of an image, a specific type of object, art historical interpretation, and cultural significance (Jörgensen 2003, 127). Enser also identifies several "non-visual features" that would require complex determinations frequently appearing in representations: time, space, and events or significance (e.g. costume representing a specific festival) (Enser et al 2007, 471-473). Therefore, we cannot view content-based image retrieval as the sole solution to improved access. Rather both content and subject based retrieval should be improved as different but complementary methods of improving overall efficiency and recall within image collections.

2.2 Describing images

Accurately cataloging visual images has long been an area of discussion and research (A small sample: Shatford Layne 1986, Svenonius 1994, Shatford Layne 1994, Krausse 1998, Chen and Rasmussen 1999). Scholars and librarians

have yet to reach complete agreement on best practices for image cataloging and how it affects retrieval. Disagreements persist about the types of descriptions that might be contributed to a visual image. Some scholars (Shatford Layne, 1986) use Panofsky's (1939) distinction among the pre-iconographic, the iconographic and the iconological. Each level requires a greater knowledge about the world and the context within which a work was created. The pre-iconographic consists of the "factual ('ofness') or expressional ('aboutness')" or those things that might be readily observable without expertise (Chen and Rasmussen 1999, 293). Iconographical description requires an immersion in the symbols and themes within the culture from which the image originates. For example, the pre-iconographical description of a painting might include the term "lily" to describe a flower; whereas the iconographical description would name the flower as a representation of the Virgin Mary. While this level might take expert cataloguers and is therefore more difficult and costly to apply, it is often important to subject specialists such as art historians or curators. The iconological description relies on the ability to interpret not only the iconographic material within an image but also in-depth knowledge of cultural significances, artistic media, and context of a work.

While it might make sense to try distinguishing all three levels of interpretation within some images, for example the Renaissance works that served as Panofsky's model, more contemporary works do not rely on the same broad conventions and lack such distinctly nameable iconography. Rather than try to fit contemporary works into an anachronistic model, Shatford Layne

suggests using a type of faceted and investigative strategy for the description of images. She suggests the “general of,” “specific of” and “about” using the questions who, what, when, and where for each category (1999). Here the “general of” aligns somewhat with the pre-iconographic in that it is the group of identifiable objects within a work. The “specific of” comprises all the objects within a work that might be identified as a specific individual within a group. Finally the “about” category of Shatford Layne’s descriptive classes are those interpretive meanings which are subjective. While this mental model of cataloging may not always map neatly into a museum catalog or image database, they may provide the cataloger a good mental model for recording subject metadata. It may also be helpful to consider these strategies of bibliographic description when examining search queries, information needs, and behaviors.

One successful effort to better encompass user queries and still retain high levels of accuracy is the use of controlled vocabularies. Baca offers a concise definition of a controlled vocabulary as an “organized collection of words, phrases, and/or names, structured to show the relationship between the terms” (2010, 1277). Controlled vocabularies gather variant names together into one authority record that designates a preferred term. This method allows a catalog to display the desirable term or name in a record while still allowing for retrieval by variant terms or names. When compiled into a thesaurus, these terms can be arranged hierarchically, allowing objects to be narrowly cataloged but retrieved by a range of broader terms. For example, Giacomo Balla’s *Dynamism of a Dog on a Leash* could be catalogued with a subject term “dachshund” but could still

be retrieved by the broader term, “dog.” However, were a user to want an image of a dachshund, she would only find images containing that specific breed. A few of the widely used controlled vocabularies are the Library of Congress Subject Headings (LCSH), *Art & Architecture Thesaurus (AAT)*, *Thesaurus of Geographic Names (TGN)*, *Union List of Artist Names (ULAN)*, *Thesaurus of Graphic Materials I and II.* Though these tools have the potential to greatly increase recall and precision, they may still leave gaps between the language of the searcher and that of the cataloguer and the controlled vocabularies. For example, if a five-year old searched using the term “wienie dog,” she may not find a single image of a dachshund if the term was deemed unworthy of being an alternate name within the controlled vocabulary.

On the opposite pole of image retrieval lie the linguistic and behavioral strategies involved when someone seeks an image. In decades past, the image-based informational need of a particular art historian, artist or layman might be met within the confines of a slide library. Generally, a librarian with an extensive knowledge of the collection contents would oversee such collections. Were the patron not able to immediately locate the object she seeks, the librarian would be able to assist her through a conversation that relies on the librarian’s knowledge and leading questions. However, today interactions with image collections rarely occur within the confines of a room containing a librarian and a physical slide collection but rather occur from a user’s computer through a mediated interface. Therefore, the user must rely much more heavily on her ability to describe the object sought. Unlike the response a patron might get from a slide librarian, the

potential for aid in the digital library is often fixed within the system (e.g. a list of frequently asked questions or algorithms to suggest alternate spellings) or non-existent. Therefore, aid provided to a user must be done so in anticipation of a need. Due to this necessary application of forethought, it is imperative to make all efforts to bring the language of cataloguers closer to the language of searchers. One way of achieving this feat is a greater understanding of how users search.

3. Literature Review: Transaction Log Research

3.1 Transaction Log Analysis Overview

Transaction log analysis is an important way scholars gain a window into the world of the user. Though transaction logs vary greatly across information systems, a simple definition of a transaction log would be a record of interactions between some information system, such as an Internet search engine or library catalog, and a user. The *transactions* recorded in a log might be as simple as a collection of search terms entered into a web page or nearly infinitely complex such as mobile web app transaction logs that record dozens of data points with each interaction. The information recorded in transaction logs is then used as a window into real-world user behavior. Though it might appear that transaction logs are a relatively new phenomenon, interest in transaction log analysis (TLA) as a methodology has been growing since the 1960s (Meister and Sullivan, 1967). TLA has a wide range of uses from the analysis of large, general search engines such as Google to simply looking at the searching mechanisms within a specific web site or resource such as a library catalog. The transaction log records pre-determined ranges of interactions between the user and the system (Jansen 2006, 408). Users may be defined as software or humans, and extracting human actions from the mechanized searching is crucial for obtaining accurate results.

The level of information collected in transaction logs varies greatly depending upon the context. For example, some systems' transaction logs might record all interactions within the system such as log-in and log-off times, queries,

results, clicked results, tools used and other details of actions while simple transaction logs may be limited to only a time stamp, IP address and query string. Penniman and Dominick (1980, 23) classify the granularity with which data can be collected as follows:

- General session variables: the least amount of data collected limited to broad actions such as databases used, query string, number of results and other simple data.
- Function/state traces: the transaction log uses predetermined categories to classify the data as it is being recorded.
- Complete protocol: the highest level of granularity recorded includes all user interactions with a system. In addition to query data, actions captured might be the types of tools used, results viewed, and a time stamp for all of these.

The types of transactions recorded generally reflect the character of the information system. Closed systems such as a library catalog that requires users to log in may record every action using the complete protocol. Conversely, the transaction log for an open web search engine may be more limited due to several factors such as the volume of users or the fact that users will navigate away from the point of origin and may or may not return. Privacy concerns may also limit what types of information are recorded.

Unsurprisingly the fields which are recorded determine what types of analysis may be conducted. While there are nearly as many combinations of

transaction variables as there are information systems, some terms have emerged as standard collected fields. Wang et al (2006) identify these fields as:

- Query: search statement entered.
- Time stamp: date and time of interaction's occurrence.
- User identification: IP address, a randomized number representing a unique IP address, or a cookie id.
- Click through: the order and time at which the user clicked on search results.

Though not a definitive list, these fields should be applicable to many of the digital information retrieval environments despite other important differences.

The methods used to analyze the data sets collected in transaction logs has been both the subject and process within various studies. The granularity of analysis consists of three levels:

- Session: an entire period of user interaction with the system. While extremely useful, determining the boundaries of an individual session within an information system can prove extremely difficult. Session analysis may seek to answer whether transactions are successful, how long users interact with the system, what tools or strategies users employ and how queries are reformulated over time.
- Query: an entire search string entered by a user. The complexity of the query, the length, and the type of terms used are often examined in query analysis.

- Term: the individual words or fragments of a query. These are generally determined using spaces or other delimiters. Term analysis is predominantly used to answer questions about how terms are related.

These levels of analysis may be applied to a variety of research motivations including identifying user goals, user strategies, identifying trends and identifying types of sessions.

3.2 Term analysis

Term analysis can aid researchers in determining linguistic patterns and associations among words. This can then aid designers in creating relational and ranking algorithms. Term frequencies, a more simplistic type of term analysis, reveal top query terms and have the potential to show variations in searcher's interests over time. Potentially more informative than term frequency is determining term co-occurrence (Jansen, Spink & Saracevic, 2000; Wolfram, 2000). Algorithms testing term co-occurrence can be particularly helpful in creating an automated suggestion tool for searchers (e.g. when the user types car, a selection box pops up with "car sales, car ratings, used car").

3.3 Query Analysis

Among the motivations for structural query investigations is also the investigation of the syntactical structure of queries. This includes investigations into whether searchers employ Boolean operators, special characters, and the

form and length queries take. Across many types of information retrieval systems, there is a relatively low use of operators with the exception of traditional informational retrieval environments such as library catalogs and other more carefully structured information systems. For example, Hölscher & Strube (2002) found fewer than 2% of web users employed operators and Jansen, et al (2000) found around %8 web users employed operators while Siegfried et al (1993) observed %37 of users employed operators in traditional IR environments. An extreme example of the use of operators with more traditional IR environments might be the Jones, et al. (1998) digital library study that found over a quarter of users employed Boolean operators. In contrast, the trend away from using operators is so consistently observed in web environments that in a literature review Markey questions whether it still retains its relevance (2007). However, in some earlier and later studies the appearance of Boolean operators within queries aides clustering of user or session types (Bendersky & Croft, 2009; Wen, Nie & Zhang, 2002; Wolfram, Wang, & Zhang, 2009).

Even without extensive click data, some studies point toward using query construction as an indication of the ease or difficulty a user may be having meeting their informational need. Bendersky and Croft (2009) use the MSN query log to compare query length with location of the result clicked. The longer the query, the farther down in the list the users tended to click and the higher the abandonment rate was. Therefore, Bendersky and Croft argue long searches (e.g. over five terms in length) are indicative of user struggle.

In line with examinations of query lengthy and complexity are studies of the use of natural language, such as those prompted by sites like AskJeeves. When Bendersky and Croft compared long queries that were formulated as natural language questions to other types of long queries, they found that there was a lower rate of user abandonment with the natural language queries. (2009, 11) Inquiry into the uses of natural language could enlighten the ways in which some search engines treat question indicators (e.g. when, who, where). Wen, Nie and Zhang suggest that rather than ignore these words, information retrieval systems could use them to direct users toward more effective matches than simple keyword retrieval might. For example, if a user included question indicators like *who* or *where*, the IR system might weight results that included people and places respectively (2002, 62).

Query analysis particularly when used alongside or in conjunction with other indicators from the transaction logs can be used to determine a wide range of information about the users. It may also yield important results that may aide programmers in more effectively designing algorithms for returning relevant results.

3.4 Session analysis

Modeling of general user behaviors can be achieved with quantitative methods. These models include forming generalized average characteristics of search sessions as in the Jansen and Spink (2006) study comparing nine different search engines. Baeza-Yates et al. (2005) use click data and query

submission within single sessions to model average session length, reformation rates, and click rates. Other studies try to cluster sessions into meaningful groups through examining similarities in traits. Wolfram et al (2009) pinpoint three distinct session types across three searching environments: an academic site (University of Tennessee, Knoxville website), a specific field resource (Healthlink), and a general search engine (Excite). Sessions could be identified as short queries in short sessions (generally using popular search terms), long queries in short sessions, and long queries in long sessions.

The vast majority of these session-identifying studies use term repetition across queries and statistical time limits to determine a search session's boundaries. Hollink et al. call this type of session identification *syntactic* and claim that it misses many important query reformations (2011). Rather than depend solely on the terms in a set of queries overlapping, the researchers sought to expand sessions toward a more comprehensive inclusion of general ideas. To do this, they used RDF triples.[3] This expands the ability to match query reformation through terms associated with the primary search term, therefore capturing a wider spectrum of searches within a session including those that do not contain any of the same terms (Hollink et al 2011, 709).

Stemming from studies focusing on session characteristics, some studies make an effort to identify user goals, Quite often these goals draw from or at

[3] RDF stands for Resource Descriptive Framework. These triple sets include the subject term, a predicate and an object that are predicated by subject. For example, a sample RDF triple with 'David Beckham' as the subject might also have 'married to' as a predicate and 'Victoria Beckham' as an object (e.g. David Beckham (married to, Victoria Beckham). Likewise another RDF triplet with David Beckham as the subject might have 'plays' as the predicate and 'football/soccer' as the object. Therefore, a search for David Beckham might be linked to a search for Landon Donovan because they both are part of a triple with a predicate and object 'plays' 'football/soccer' (Hollink, et al 2011, 692).

least mention the taxonomy of web search laid out by Broder (2002): *navigational*, search to find a site URL; *informational*, search to seek some kind of information; and *transactional*, search to perform an interaction within the space of the internet (e.g. email or shopping). Studies like Jansen, Booth & Spink (2008) and Rose & Levinson (2004) use manual classification to evaluate user goals based on the query itself and the corresponding click data.

Some studies try to form a method of automatically identifying user goals; Lee, Liu & Cho (2005) base their algorithm on where the query terms were anchored within a site and how many clicks occur within the results. If the anchor link is a homepage, the query is determined to be navigational, if on a content page, informational. From click-through data, goals were determined to be navigational when the user clicked on just one result. For simplification, the goals were limited to navigational and informational only. Likewise, Baeza-Yates, Calderón-Benavides & González-Caro (2006) focus on automatically identifying whether searches are informational, non-informational or ambiguous through supervised and unsupervised machine learning.

Strohmaier and Kröll (2011) also seek to automatically determine user goals. However, they move away from the Broder's (2002) broad classifications for user goals and rather focus on individual goals. Strohmaier and Kröll use a knowledge base called ConceptNet that not only contains a large number of goals (e.g. *how to change a tire*) but also relationships between these goals "*MotivatedByGoal, UsedFor* and *CapableOf*" (2011, 12).

One method of identifying user goals is external to the searches themselves. Clustering techniques are often used to determine the nature of different types of search sessions, finding methods of clustering search terms around a particular topic or group of documents can aid search engines in retrieving materials that might not contain the exact phrasing of the user query. One such study, Wen, Nie & Zhang (2002) designs a method of clustering queries around the documents clicked within results lists deeming that if two different queries were answered by the same document, it is likely those queries are related. Other studies focus on the clustering of search results. Wang & Zhai (2007) use OKAPI similarity testing to group search results from a query in a method called star-clustering that breaks the results lists into more narrow topical sections.

Over the last few decades, transaction log analysis has blossomed into a means to discover a wide variety about the way in which users interact with information systems, the goals they have, and the ways they alter their behavior to reap better results.

4. Literature review: Image searching

One drawback to query log analysis is that all of the data is collected on the end of the server. Transaction logs cannot capture actions like copying text, using multiple resources, or the mental state of the user. These logs only allow an examination of the traces of human behavior. In order to put the information gleaned from any kind of transaction log within the larger context of user searching, we need methods for creating and investing cognitive models, informational needs and information goals.

4.1 Cognitive models

The subsets of information, including cultural, technical, and scholarly knowledge, that users bring to a search will impact the way in which a user searches. Heidorn (1999) examines how indexers and searchers create mental models of images in information retrieval environments. Heidorn argues that the image retrieval process is essentially an act of communication between the indexer and the searcher, so such cognitive models play an important role in achieving successful search results (1999, 306). Image retrieval then relies upon shared social constructs and knowledge that structure both groups' mental models of an image. Alignment of these models must occur in two different manners:

(1) *Mental model-to-index*: how the indexing terms match the cognitive models of the searcher, and

(2) *Cognitive-model-to-interface*: how well the searcher can express these mental models within the information retrieval setting (Heidorn 1999, 309).

This means that in order for a user to find an image, their terms must match the types of terms used to catalog the image *and* the system must allow (or even help) the user input these terms.

As a searcher seeks an image, his or her mental model shifts. Searching for an image relies on a combination of both long-term and short-term memory. The searcher utilizes various perceptual and linguistic categories from long-term memory to change or complete a dynamic mental image held in short-term memory. The model in short-term memory can also be shaped by inputs from the information retrieval system (Heidorn, 1999). This suggests that more responsive user interfaces with expanded choices for browsing, suggested terms, and displays of similar items may increase the effectiveness of image retrieval systems through allowing more potential methods of altering the dynamic mental model in short-term memory and triggering more terms and strategies from long-term memory

Beyond having a cognitive model to explain how one might understand an image, Matusiak suggests we also need a model for how the user interacts with an image retrieval system (2006, 481). Using a historical image collection, “Milwaukee Neighborhoods: Photos and Maps, 1885-1992” created by the University of Wisconsin-Milwaukee Libraries, Matusiak studied both university students and community members interaction with the system through direct observation, self-reported logs and follow-up interviews (2006, 482). Users predominantly relied upon either browsing or keyword searching initially but then most incorporated both strategies as they continued their interactions with the

system. Experience with web-based searching and confidence in computer skills seemed to determine the type of mental model of the information system the user formed. Students, who regularly employ web search in their daily activities, tended to have a mental model of a text-based website. In contrast, the community members less familiar with web searching tended to create an exhibition-based cognitive model (Matusiak 2006, 485-486). These findings reaffirm that the design of image retrieval systems should allow for multiple pathways to the same image as well as multiple types of experiences (e.g. enjoyable browsing vs. targeted searching) to be most effective for the broadest possible audiences.

4.2 Behavioral Studies:

While information retrieval studies that focus on cognitive models help us understand how information-seekers conceive of information, behavioral studies seek to understand motivations for and actions to fulfill information needs. Information-seeking behavior studies relevant to image retrieval generally fall into two categories: broad information need studies of particular groups who may seek images to meet some needs (e.g. art historians, cultural heritage experts, etc.) and transaction log studies of image collections. The act of searching for images is generally contextualized by a particular informational need that can be further constrained by an occupation or personal interest. Therefore, many clues to the goals and behaviors that drive image searching are embedded in larger studies of information-seeking behaviors within specific user groups such as art

historians, cultural heritage professionals or journalists. Understanding the larger context within which image-searching illuminates how and why users turn to specific digital image collections in the first place.

Frost et al (2000) set out to determine the differences in search strategies between experts and non-experts within an art image database. Again it was shown that experts tend to rely on keyword searching while generalists were more likely to use hybrid or browsing techniques (294). Part of this may result from the experts' perception that browsing techniques are much slower than keyword searching while generalists found browsing to be a more forgiving discovery method (298). Artist name and subject categories were used in similar proportions across the expert and generalists samples (297).

Three studies in particular examine how different demographic groups of users interact and perceive image retrieval systems. These studies make efforts to categorize when images satisfy information needs, characteristics of effective image retrieval systems, and the variety of services rendered by image systems: the Visual Information Seeking Oriented Research project (VISOR I), the Museum Educational Site Licensing project (MESL), and the Pennsylvania State Visual Image User Study (VIUS) (respectively: Conniss et al 2000; Stephenson and McClung 1998; Pisciotta et al 2005). The VISOR I study took a broad approach to examining image uses across disciplines in organizations as disparate as radiology centers, a police departments, and regional museums. Focusing on cultural heritage images alone, the MESL study supplied museum images in a digital archive format to several universities surveyed students and

faculty beforehand and afterward to reveal their insights on image-retrieval (Conniss et al 2000; Stephenson and McClung 1998). The VIUS project conducted from 2001-2005 was modeled after both these earlier studies (Pisciotta et al 2005, 35). Some similarities were discovered between MESL and VIUS, namely that students tended to value rich metadata more than faculty, that overall image resources will be perceived as valuable only if the content matches the immediate need of the user, and that database usage increases in correlation to image-driven assignments (Pisciotta et al 2005, 37, 43, 49).

Though these studies reveal that image need, attitude toward image resources and search strategy can vary depending on the context and individual. However, more contemporary studies may be necessary to investigate whether usage patterns have shifted over the last 8 to 12 years when the VIUS project was conducted. During the faculty interviews, Pisciotta et al found that among Arts and Architecture faculty the usage of analog images per semester was greater than the use of digital images (2005, 39-40). Additionally, over one-quarter of the faculty reported having some level of apprehension when using technology in the classroom (43). With the quality and ease of transfer and storage of digital images improving drastically over the last decade, it is very possible that some of these attitudes, practices, and search techniques have begun to change among faculty and students.

Chen (2007) focuses directly on museum practitioners, examining their motivations, needs, expectations and involvement with image databases and the digital museum. Chen points to some factors that impair the effectiveness of

digital museum collections: lack of understanding of online user demographics, limited understanding of technology, imprecise goals for collections, and lack of reward structure are particularly problematic and work as impediments for progress in digital collections (2007, 24). In order to provide a framework to address this issue, Chen investigates museum practitioner's own image practices. Chen shows that the motivation for seeking out images ranges widely from locating an object, patron requests, inventory, research and educational programs (2007, 27-29). Similarly, image-seeking practices still largely relied on using slide libraries or other analog collections (30-31). The responses from participants suggest that the apparent idiosyncrasy within image-seeking may grow out of the idiosyncratic nature of cataloging such images and objects that arises out of the particular nature of a museum collection. Thus, the results suggest that both the context of the collection and the system itself (including cataloging practices) effect the information-seeking behaviors of the user.

Amin et al (2008) examine cultural heritage experts' information-seeking tasks across all informational needs. Largely, their findings confirm the motivations suggested by Chen (2007) (e.g. research, collection management, exhibition planning, etc). The Amin et al (2008) study helps to streamline further study of cultural heritage professionals by creating a taxonomy of information seeking tasks specific to this discipline: fact finding, information gathering and keeping up-to-date (40). Information gathering comprised 63% of information seeking tasks and included the subsets of comparison, relationship search, topic search, exploration, and combining matches across types of information (43).

Importantly, the research team points to the need to continually rely upon the expert's own knowledge of the domain and available resources as well as the particular ineffectiveness of many types of information retrieval systems to meet the needs of comparison and relationship searching (44-45).

Moving toward transaction log studies of image collections, Hastings (1999) outlines a qualitative method by which the effectiveness of image retrieval systems might be evaluated. The investigation uses observed user behaviors and responses to surveys to investigate areas including (1) how queries might be satisfied by thumbnails, (2) relationships between query and image manipulation, (3) comparison of query to access points, (4) categories that might be identified to increase browse functionality, and (5) image manipulations that aid queries. Hastings suggests four categories of tasks that might be evaluated in an image retrieval system. *Known item query* might be evaluated by user feedback and measures of time and effort. *Unknown item query* might be evaluated by user supplied terms, survey, feedback, and measures of time and effort. *Style and image content queries* can be investigated by log analysis, screen capture, and survey. *Subject queries* might be evaluated by user observation and measure of user effort.

Transaction log studies of image databases still comprise a type of behavioral investigation even if it is limited to simply capturing the interactions that occur between the user's computer and the web server. While transaction log studies of general web use are plentiful as described in the previous section, fewer studies have turned an inquisitive eye toward the same kind of in-depth

analysis on image catalog transaction logs. Several reasons for this may exist. It may be that many users conceive of the image search environment in the same way as text-based dynamic search environments. Another contributing factor could be that the proprietary nature of images as well as their file size slowed their appearance on the web forcing interest in digital image searching to naturally follow some years behind that of text-based web resources. Furthermore the explosion of efforts to expand the abilities of content-based retrieval may have drawn attention away from investigating subject-based image retrieval.

In the few transaction log studies that have occurred already, there are significant signs of difference between the practices of text-based and multimedia searches. Goodrum and Spink (2001) found that queries for images tended to be longer than those submitted to general search engines and had a higher rate of query reformation. These findings confirm a previous study (Jansen, Goodrum and Spink, 2000) in which image and audio queries were also observed to be longer on average. As reported in other realms of web search study, searches of explicit sexual content were high and many of the most frequently submitted terms could be interpreted to qualify or correspond to these types of queries and information need (Goodrum and Spink, 2001).

The Goodrum, Bejune, & Siochi (2003) study confirmed what information seeking surveys suggested (Frank, 1999; Bates, 2001; Cobbledick, 1996) that there is a high rate of browsing used in seeking images. Jörgensen and Jörgensen (2005) investigate the methods used to search for images within a

photojournalism database. They noticed use of Boolean operators was higher than might be expected but that they were often misused (Jörgensen and Jörgensen 2005, 1353). Additionally, there were mixed results as to the success of searches as nearly half of the queries were reformulated but about a quarter of the sessions ended with a download that was considered a success by the research team (1354-1355).

4.3 Query Classification Studies

Because the vast majority of image collections still rely predominantly on semantic-based image retrieval, it is imperative that we understand the terms and types of terms searchers employ. Classifying queries of image databases informs concrete decision making in the technical programming and cataloging practices of image retrieval systems. Studies that suggest classification systems for image queries generally stratify queries by complexity or by object groups (e.g. animals, emotions, etc).

In the background section of their report on Content-based Image Retrieval, Eakins and Graham (1999) divide queries into classifications based on complexity. For their purposes, the researchers exclude any queries that would be satisfied by art historical metadata (e.g. creator, date creation, etc.) because they claim these queries are purely text-based queries. Their schema consists of the following three levels of query types:

> *Level 1: The Primitive.* Consisting of compositional and visual elements of an image including color, shape, placement, and texture.

> *Level 2: The Derived/Logical.* Consisting of some kind of named object either of a general type (e.g. dachshund) or a specific individual (e.g. Huckleberry Hound). These types of queries generally rely on objective descriptions of an image.
>
> *Level 3: The Abstract*. Consisting of more complex queries that name emotions, religious symbols, events and activities. These types of queries require either subjective judgment or sophisticated reasoning.
>
> (Eakins and Graham 1999, 7-8)

Eakins and Graham suggest that this classification serves to distinguish *semantic retrieval* (levels 2 and 3) from *content-based retrieval* (level 1) with the *semantic gap* falling between level 1 and level 2. Eakins and Graham's classification system is similar to Panofsky's pre-iconographical, iconographical, and iconological in that they are based around the intellectual requirements needed for each level description.

Breaking from the reliance of hierarchical levels of query classification, other studies focus on complexity from a more semantic perspective. By Eakins and Graham's standards, these classifications focus on queries that would fit into their levels 2 and 3. One such study is that of Enser and McGregor who stratify queries based on uniqueness and refinement or lack thereof (1992). The groups consist of: *unique* (e.g. specific person, Barack Obama), *unique with refiners* (e.g. unique place with dates, Trafalgar Square, 1941-1945), *nonunique* (e.g. dog), and *nonunique with refiners* (e.g. motorcycles in World War I).

Table 1. Enser-McGregor classifications

Classification	Description	Example
Unique	Specifically identifiable entity/object/location	Rosa Parks
Unique with refiners	Specifically identifiable entity/object/location with a restriction	Rosa Parks on a bus
Nonunique	General type of entity/object/location	Kitchen
Nonunique with refiners	General type of entity/object/location with a restriction	Kitchen from the 1950s

Armitage and Enser (1997) apply similar classifications that rely on ideas of uniqueness/non-uniqueness to art image collections. Their classification include a more bibliographic perspective of image seeking by including two types of known item searching, and the resulting classes are: by artist name, known item, unique, and non-unique (Armitage and Enser 1997, 287-289). To supplement this schema, Armitage and Enser also classify the queries according to the Panofsky/Shatford Matrix shown below:

Table 2. Panofsky-Shatford mode/facet matrix (Armitage and Enser 1997, 290)

	Iconography (Specifics)	Pre-iconography (Generics)	Iconology (Abstract)
Who?	Individually named person, place or thing	Kind of person or thing	Mythical or fictitious being
What?	Individually named event, action	Kind of event, action, condition	Emotion or abstraction
Where?	Individually named geographical location	Kind of place: geographical, architectural	Place symbolized
When?	Linear time: date or period	Cyclical time: season, time of day	Emotion, abstraction symbolized by time

Armitage and Enser note that though this schema works generally well to classify requests for images, it importantly leaves out queries concerning the media of a work as well as specific object attributes (1997, 294).

In order to develop a classification schema that would be consistent across describing tasks, Jörgensen (1998) had a group of participants describe an image from sight, during image retrieval and from memory. Drawing on the responses, Jörgensen developed a classification system to encompass all of the described attributes. Drawing on previous research and conceptual relationships among the attributes, Jörgensen developed the following schema to classify descriptions of images (1998, 174):

Table 3. Jörgensen's 12 attributes of image description

Attributes	Description
Perceptual	
• Objects	• Nameable, visually perceived objects (e.g. text, body part, clothing)
• People	• Human figures
• Color	• Color names and description of color values (e.g. bright)
• Visual Elements	• Composition, focal point, motion, orientation and similar attributes
• Location	• Both general and specific
• Description	• Descriptions of the work itself or objects within the work
Interpretative	
• People-related	• Abstract concepts related to human relationships: social status, emotion
• Art historical information	• Artist, format, media, technique, style, time reference, other related
• Abstract concepts	• State, abstract, atmosphere, symbolism, theme
• Content/story	• Activity, category, event, setting, time
• External Relation	• Similarities, references, comparisons
Reactive	
• Viewer response	• Personal reaction

Jörgensen notes that these attributes demonstrate the breadth of description necessary for image retrieval tasks, and that there is a discrepancy between the attributes used by searchers and the typical catalog records for visual images (1998, 173).

Though such classification schemas may at first appear trivial, having a robust schema of descriptive attributes used in the various image retrieval tasks enables the codification of such descriptions to be used in image cataloging, searches, and recall. In her 2001 study of the queries of art history students, Chen draws on the classifications schemas of Jörgensen (1998), Enser and McGregor (1992), and Fidel (1997). In order to verify the reproducibility of results, Chen recorded the level of agreement across three researchers using each of the schema. The results were striking, particularly with the Jörgensen and Enser-McGregor schema with agreement levels at 70% and 73% respectively (2001, 266). Such agreement suggests that these categorizations would be useful in future studies.

5. Methodology

5.1 Materials used: ARTstor sample

The data used for this study comes from the search log of ARTstor, a subscription-based cultural heritage image database. The collection of search terms was recorded over a more than 5 year period from January 2004 to October 2009. ARTstor released the dataset to a group of independent researchers from cultural heritage and educational institutions for analysis (Searching Museum Collections, 2009). The data was normalized through an automated process that removed punctuation, normalized accents, and removed white space. Erroneous, blank or corrupted data points were also removed to aide analysis. After the initial cleaning of the data, the records were stripped of identifiable information such as institution name. The total number of queries is 10.75 million (Searching Museum Collections, 2009).

For each query, information was collected both about the query and the institution. The fields recorded for a query on the ARTstor site are listed in the table below:

Table 4. Fields available in ARTstor dataset.

Search ID	A unique ID number given to each search
Institution ID	An identification key used to match search records to a particular anonymized institution.
Time Stamp	The recorded date and time of the search.
Session ID	An ID given to a set of searches automatically determined to be part of an individual's search session
Search term	The text string entered in a query.
Collections	Indicates whether a particular collection or all collections were searched.
Institution Type	ARTstor classification for the size of an institution within its class
Institution Class	ARTstor classification for the type of institution from which the search originated.

The institution ID recorded serves as an identifier to connect individual queries to information about each institution. The information available in the search log's institution records is listed below:

Table 5. Fields available in ARTstor dataset on institutions

ARTstor Type	A sub classification of Institution Class. Most of the designations here are determined by things such as FTE and/or expenditure per student. In general, this is field indicates the relative size of the institution.
Institution Class	ARTstor classification for the type of institution from which the search originated.
Access date	Date at which the institution subscription began.
Usage profile	Averages of the level of use and the types of features used on average.

5.2. Data preparation

In order to make the samples more meaningful, some restrictions have been placed upon which samples are to be selected. The samples will start at 2005 to avoid any potential abnormalities that might have occurred within the first year ARTstor was functional after its beta testing period. Additionally, although the ARTstor classification of universities and colleges arranges them by size, the sample will be drawn from the universities that fall into the "large" category based on Carnegie Classifications that use full-time enrollment. These schools average full-time enrollment of around 20,000 students and are likely to have the full range of academic programs, rigorous faculty research requirements, and graduate students giving a broader picture of typical academic requirements of tools such as ARTstor. The public library category will also be ignored due to its relatively low representation among subscribing institutions. Out of 1,332 institutions worldwide that subscribe to ARTstor, only 7 of these are public

libraries and include unique institutions such as the Library of Congress, the New York Public Library, and the Danish Royal Library.

The data sets for this study were downloaded from the Searching Museum Collections mySQL database hosted by the Indianapolis Museum of Art with the permission of the project lead, Susan Chun. The downloaded data was migrated to a local database and organized by institution type and date. This study follows the model Cochran's equation to estimate sample size for large populations (1963, 75, fig. 4.2):

$$ss = \frac{Z^2\, pq}{e^2} \qquad 384.16 = \frac{(1.96^2)(.5)(.5)}{.05^2}$$

Figure 1. Equation used to determine sample size

Here, the Z value corresponds to a 95% confidence interval. P represents the presence of a particular trait within a population while q represents the lack of that trait. Because the proportion of traits is unknown, Cochran suggests setting the p value to .5 as a "conservative estimate" because it assumes the most variation (1963, 75). Finally, e represents the error rate is .05 for 5%. Therefore, in order to reach the minimum requirement the sample size is rounded up, and each sample will be comprised of 385 individual queries, with a 95% confidence value that results fall within the plus or minus 5% range.

Samples were selected out of randomized groups of date and institution ranges. For each type of institution to be tested (k12, independent art schools, university, museum), samples for three date ranges were selected from the years 2005, 2007, and 2009. Therefore, there are 12 total sets of 385 sample queries with three time points across institution types. The total number of queries

selected was 4,620, a fairly demanding overall sample size to be manually coded. To raise the confidence value or lower the error rate even by a small amount would significantly increase the sample sizes without greatly increasing the value of the insights gained from the study.

5.3 Research protocol

In order to provide the reproducible results, the classifications will be drawn from previously devised and tested classifications. Therefore, the study will use two of the tested classifications used in Chen (2001). These are the Jorgensen (see Table 2) and Enser-McGregor models. These have been chosen because they had a high level of agreement across researchers during classification and because they represent a subject level classification and a complexity classification. Other schema such as Fidel poles have been rejected because they lack the ability to shed direct light onto methods of information system design or cataloging.

Because the samples will be classified subjectively, there must be a means of verifying and cross-checking the various judgments researchers make. In light of the results of Chen (2001) the reproducibility of the classifications need not come under study. Therefore, the Delphi method will be used in order to produce the classifications of queries. In this method, researchers independently categorize the data, and then they regroup to discuss and agree upon the categorizations they made. A similar method was also used in the Jansen et al (2007) study of user goals.

One reason for pursuing the Delphi method of agreement for the classifications is that it will streamline analysis. Rather than try to represent the variance across the cataloguers making the classification distinctions, the Delphi method removes the influence from individual cataloging idiosyncrasies and helps to arrive at a consensus for each query. This way, more nuanced judgments of each query may be recorded without over complicating the analysis. For example, the cataloguers assigned up to two Jörgensen attributes to each query. One drawback of the Jörgensen attribute classes is the inability to classify some complex queries. For example, the query, “Mississippi Architecture ” would not be clearly classified as either as a location or an art historical term alone. Using the Delphi method allowed the cataloguers to agree on joint classifications, so the query would be categorized as a query with both a location and an art historical attribute. Allowing for dual attribute assignment translated into an additional clue to the complexity of the search terms within a query.

5.4 Measurements and Analysis

The majority of the analysis conducted through this study is qualitative. While some general quantitative analysis of the search queries will be conducted, such as terms-per-search, the primary purpose of the study is focused on the manual classification of search queries and is therefore qualitative. More generalized quantitative data calculated is used to compare image searching to other types of information retrieval system searching and to compare the searches across time and institution type.

The frequency of occurrence of query types according to the Enser-McGregor and Jörgensen classifications will be used to answer the three primary research questions about how users search and as a guide to creating suggestions for future transaction log recording. Analysis will focus on the changes between sets from the same institution over time and among the different institution types overall. The hope is to illuminate whether or not there are meaningful differences in the ways that such communities of users search. The researcher's hypothesis is that the searches originating from k12 institutions will be much more distinct from those originating from the three other institutions of education or culture. However, there have been no studies to date which seek to determine whether there are differences among the way in which the different primary practices of the institutions influence what types of queries users submit in the same way that age-appropriate requirements might. This information may suggest or deny the need to tailor browsing or searching interfaces depending upon the subscribing institution.

6. Results

6.1 Word Count

Query word count was determined by measuring the character strings divided by a space. The average number of terms per query across all twelve data sets is 1.88 terms with a standard deviation of 1.28 terms. The maximum length was 15 terms. The majority of queries were only one or two terms long across all sample sets. Nearly half of the queries (49.70%) are only one word in length. Queries of two-term length comprise another 30.07%, while three word queries comprise 11.36%. Across all samples, 95.5% of queries are four words long or less.

Table 6. Query length in ARTstor data sets

	Query Length				
	Art school	K12	Museum	University	Year averages
2005	2.01	2.06	1.89	2.11	2.02
2007	2.43	2.30	2.18	2.47	2.35
2009	1.31	1.26	1.28	1.28	1.28
Institution averages	1.92	1.88	1.78	1.95	1.88

6.2 Rater agreement

Three researchers classified a total of 4,620 individual queries from 12 sets of 385 queries each. The main researcher selected two visual resources professionals from the Visual Resources Association to classify each of these terms according to the Enser-McGregor uniqueness/complexity scheme (see

Table 1) and the Jörgensen image attribute scheme (see Table 3). All three researchers are familiar with image cataloging and description practices, and they have an educational background in both the arts and information studies. Due to their experience only preliminary training was provided to the researchers so as to avoid skewing interpretations of classification the main researcher's interpretation. The Enser-McGregor and Jörgensen attributes were explained using the respective articles, and charts and examples explaining each classification were provided.

For Enser-McGregor classification, the researchers showed a similar but slightly higher level of agreement across the samples compared to the Chen (2001) study. All three researchers agreed on the classification for 63.66% of the queries (2,941 out of 4,620) compared to 51.06% in the Chen study (2001, 264). Because cataloging and classification can be fraught with inconsistency, Chen created a metric to evaluate general cataloguer consensus. This metric is called *effective judgment* and denotes that at least two of the three researchers agreed on a classification for a query. The researchers achieved an extremely high number of effective judgments with 98.51% of the queries receiving an effective judgment on the Enser-McGregor classification. When comparing each rater side by side, across samples the average rate of agreement between any two researchers was 75.25%. The high level of consistency among raters reaffirms Chen's findings that the Enser-McGregor scheme is both easily understood and applied by researchers.

Due to the nature of the Enser-McGregor scheme, only one assignment per query was permitted. The classifications are mutually exclusive when assigned to queries as well as somewhat objective, and thus, the Enser-McGregor classifications are more likely to be consistently applied. Conversely, the Jörgensen attributes cover more subjective qualities of queries and are not necessarily mutually exclusive when applied. Therefore, researchers could assign up to two different Jörgensen attribute types to each query. Thus, exact agreement is slightly more difficult to replicate across three researchers.

Furthermore, the method for assigning Jörgensen terms varied across the individual researchers. The average number of terms assigned and the standard deviation of those assignments varied for each individual researcher's assignments and the final mutually agreed upon assignment. While *Researchers 1* and *3* had somewhat similar rates of attribute assignment (1.08 and 1.04 on average respectively), *Researcher 2* was much more likely to assign multiple attributes to a single query (1.68 on average). Additionally, the standard deviation for *Researcher 2* was higher than the other two researchers suggesting a more varied approach to classification assignments. Unsurprisingly, the lowest standard deviation of term assignment was for the final discussed and mutually agreed upon assessment. Likely this low standard deviation arises from the fact that viewing all three researchers' assignments and discussing these led to a much more systematic approach to attribute assignment.

Table 7. Variation among researcher's application of Jörgensen attributes.

	Jörgensen terms applied per query	
	Average # applied	Standard Deviation
Researcher 1	1.08	0.044
Researcher 2	1.61	0.108
Researcher 3	1.04	0.053
Final assessment	1.07	0.028

As might be expected, agreement across the Jörgensen attribute assignments was much lower than the Enser-McGregor classifications. Across all samples, all three researchers only had complete agreement (assigned same type and number of attributes) on 17.36% of queries. However, at least two researchers had exact agreement on 83.55% of queries. Again, *Researchers 1* and *3* showed much more similarity in the way they assigned attributes. They had exact matching classifications across 70.63% of queries, and at least one term assigned matched across 77.16% of the queries. Conversely, comparisons between *Researchers 2* and *1* and *Researchers 2* and *3* showed much more disparity with exact match percentages at 26.84% and 20.80% respectively. However, the rate for at least one matching term was 82.64% between *Researchers 2* and *1* and 77.32% between *Researchers 2* and *3*. Because of the large spike moving from exact matches to just one term across terms prescribed by the researchers, it is likely that *Researcher 2*'s propensity to assign more than one attribute to each query compared to the other researchers is the root cause of the discrepancy.

When the agreement data is broken down into types of agreement, this hypothesis seems to be supported. Exact matches with both raters assigning one term were quite high between *Researchers 3* and *1,* while they were much lower between *Researcher 2* and either of the other researchers. Conversely, one term matches in which one researcher assigned a single attribute and the other assigned two attributes, were quite low between *Researchers 3* and *1* while they comprised nearly half of all types of agreement/disagreement with *Reseacher 2.*

Table 8. Jörgensen rater agreement by type of agreement. Here the types of agreement or non-agreement are described by a number #-# and match or mismatch. The number indicates the terms assigned by each researcher to the same query, the match or mismatch indicates the level of agreement across those terms. For example a 1-1 match would mean both researchers assigned 1 attribute to the query and it was the same. A 1-2 match indicates one researcher prescribed 1 attribute and the other prescribed 2 attributes.

Rater agreement by type of agreement						
	R1-R2		R2-R3		R3-R1	
	#	%	#	%	#	&
1-1 match	1171	25.35%	940	20.35%	3237	70.06%
1-1 mismatch	592	12.81%	845	18.29%	932	20.17%
1-2 match	2369	51.28%	2512	54.37%	265	5.74%
1-2 mismatch	188	4.07%	192	4.16%	114	2.47%
2-2, 1 match	209	4.52%	99	2.14%	37	0.80%
2-2 match	69	1.49%	21	0.45%	26	0.56%
2-2 mismatch	22	0.48%	11	0.24%	9	0.19%

6.3. Types of terms used

The majority of queries were classified as *unique* on the Enser-McGregor scale (54.89%, with a 5% error rate). The next largest category was *nonunique* queries at 21.39%. Queries with refiners made up the last almost quarter of the queries at 23.73% combined. The individual samples varied somewhat, particularly across the *unique with refiners, nonunique,* and *nonunique with refiners* categories that all saw large standard deviations compared to the average distribution (see table 1 for an explanation of the Enser-McGregor classifications).

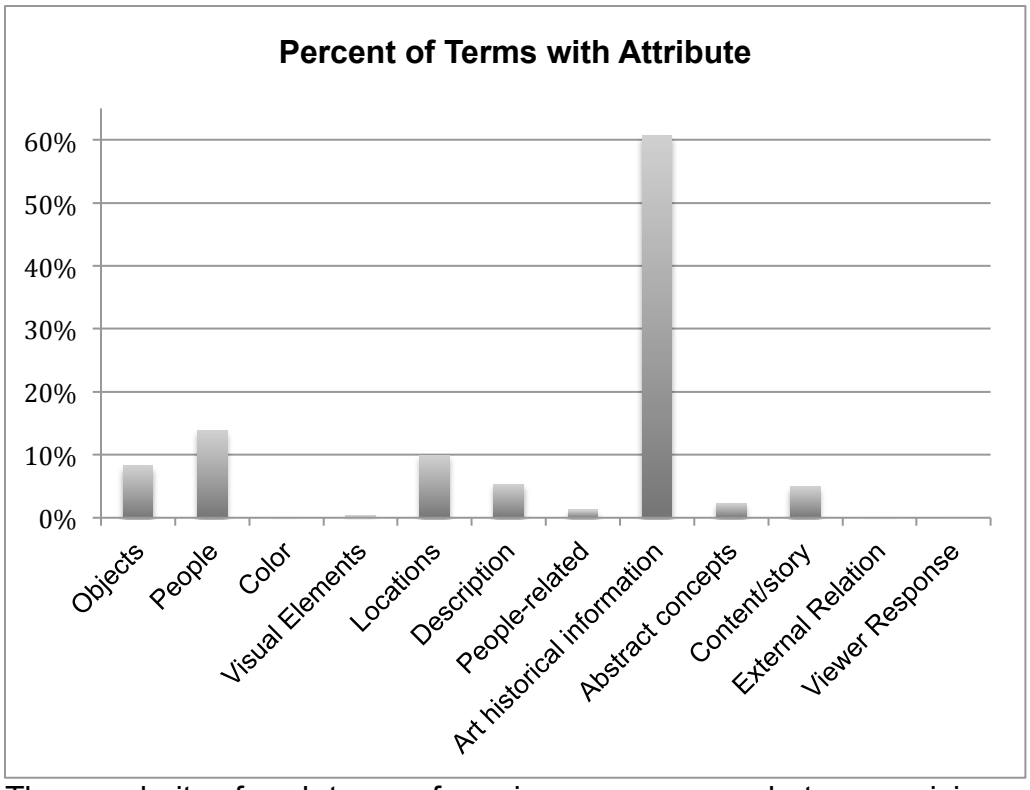
Table 9. Enser-McGregor classifications across samples

Enser-McGregor Classifications Across Samples				
	Total	Average per sample	Std. Dev.	Percentage
unique	2536	211.33	23.38	54.89%
unique w/ ref.	727	60.58	39.16	15.74%
nonunique	988	82.33	43.01	21.39%
Nonunique w/ ref.	369	30.75	24.43	7.99%

When classified using the Jörgensen attributes, queries were clustered around a small portion of the attribute types (see table 3 for definition of attributes). The most frequently submitted type of query was *Art historical information*; 60.65% of all queries submitted were classified as containing art historical information. The next largest categories were *People, Locations,* and *Objects*. These top four categories comprise 92.62% of all queries.

Table 10. Frequency of Jörgensen attributes across samples.

Jögensen Attributes Across Samples				
	Total	Average per sample	Std. Dev.	Percentage
Art historical info	2802	233.50	38.99	60.649%
People	639	53.25	19.78	13.831%
Locations	458	38.17	14.13	9.913%
Objects	380	31.67	8.66	8.225%
Description	247	20.58	6.82	5.346%
Content/story	227	18.92	9.43	4.913%
Abstract concepts	105	8.75	5.75	2.273%
People-related	57	4.75	3.11	1.234%
Visual Elements	14	1.17	1.19	0.303%
Color	5	0.42	0.67	0.108%
External Relation	2	0.17	0.39	0.043%
Viewer Response	0	0.00	0.00	0.000%



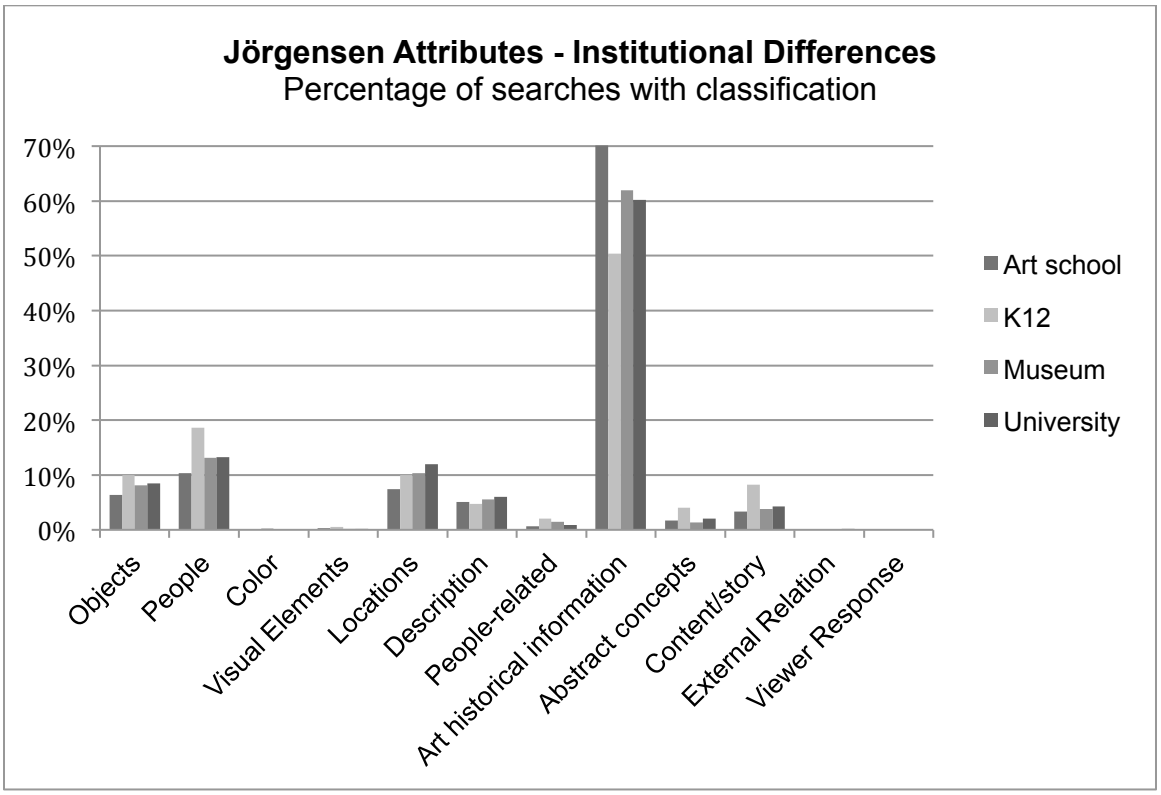
The popularity of such types of queries seems somewhat unsurprising since these categories are largely comprised of objective and/or standardized types of information and are more likely to be recorded as within cataloging

metadata than some of the other types of attributes such as *Visual Elements, People-related, External Relation,* or *Viewer Response.* The categories of terms used to search the ARTstor image library largely correspond with the typical types of information cataloged in images' metadata. For example, an image such as *The Coronation of Napoleon* painted by Jacques Louis David would be very likely to be catalogued with general art historical information like artist name, title, repository, media, and size, but it might also be catalogued with subjects that contain the other top categories: Napoleon I, Emperor of the French (*People)*; Paris, France, Palace (*Locations)*; Crown, Scepter (*Objects).* The search log alone can only indicate what types of terms the users employed when searching, but not necessarily what the users might *wish* to use in order to find images.

6.4 Variations across institutions

Differences between the searches originating from different types of institutions were quite small. The museum and university types had very similar types of queries. The Art School queries on average tended to be more focused toward *Art Historical Information* with 70.13% of all queries falling into this category compared to 50.39% of k12 queries, 61.90% of museum queries and 60.17% of university queries. K12 queries consisted of more *Objects* (9.97%), *People* (18.61%)*,* and *Content/Story* (8.23%). Given the educational mission of K12 institutions, it is not surprising that image queries would be less directly focused on specific art historical terms and more focused on broader topics.

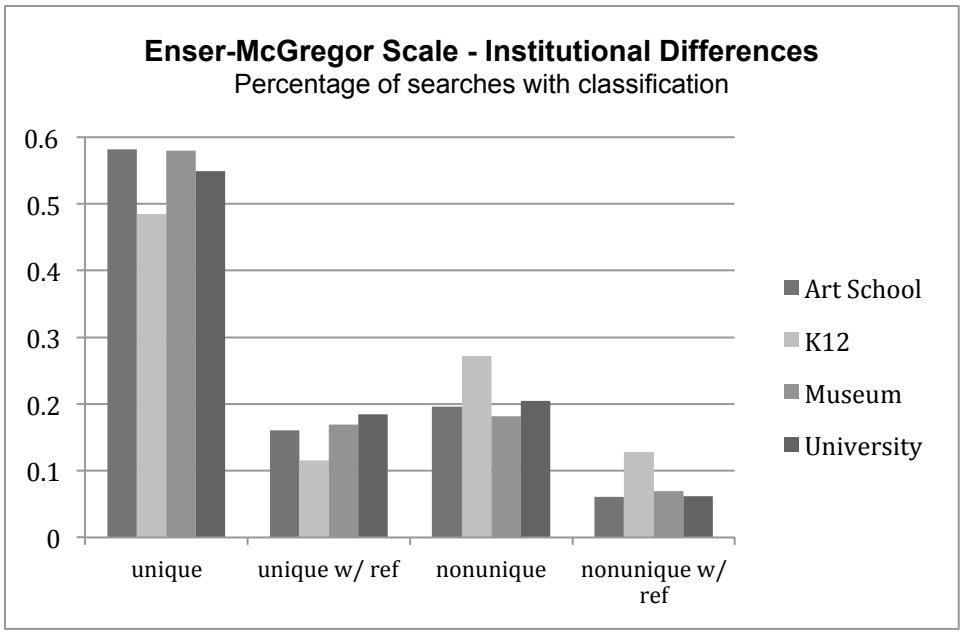
Table 11. Institutional differences in Jörgensen attributes.



A similar trend can be seen when we look at queries using the Enser-McGregor uniqueness scale. Queries originating from K12 institutions were the most divergent from the other institutions. The majority of queries from each institution type were queries for *unique* entities. However, among queries from K12 institutions, the percentage of *unique* and *unique with refiners* queries were lower than the averages for the other institutions while the percentage of nonunique and nonunique with refiners queries comprised a higher percentage of k12 queries than the other institutions. When the categories are combined to reveal whether a searcher is looking for a specific item or a general item, all unique queries form 60.00% of K12 queries (48.48% *unique*, 11.52% *unique with refiners*), while the all other institutions had nearly three-quarters of their queries made up of some kind of unique request. Art schools had 74.29% specific item

queries (58.18% *unique*, 16.10% *unique with refiners*), museums had 74.89% specific item queries (58.01% *unique*, 16.88% *unique with refiners*), and universities had 73.33% specific item queries (54.89% *unique*, 18.44% *unique with refiners*).

Table 12. Enser-McGregor classifications across institutions.



In parallel, the K12 institutions had higher rates of general searches (e.g. *nonunique* and *nonunique with refiners*). The total general search queries for k12 institutions were on average 40.00% of all queries (27.19% *nonunique*, 12.81% *nonunique with refiners)*. Conversely, art schools, museums, and universities' queries were comprised of about one-quarter general searches (25.71%, 25.11%, and 26.67% respectively). Of the general queries, the breakdown between *nonunique* and *nonunique with refiners* was very similar across the remaining three institutions: art schools queries were 19.65% *nonunique* and

6.06% *nonunique with refiners* in total, museum queries were 18.18% *nonunique* and 6.93% *nonunique with refiners*, and university queries were 20.52% *nonunique* and 6.15% *nonunique with refiners.*

Table 13. Distribution of queries with Enser-McGregor classifications across Jörgensen attributes by institution.

Distribution of Queries with Enser-McGregor Classifications Across Jörgensen attributes by institution								
	Unique				*Unique with refiners*			
	Art Sc.	K12	Muse.	Univ.	Art Sc.	K12	Muse.	Univ.
Objects	0.30%	1.07%	0.60%	0.95%	1.61%	2.26%	2.56%	5.16%
People	6.70%	16.25%	12.69%	12.15%	3.76%	9.02%	5.64%	6.10%
Color	0.15%	--	--	--	--	--	--	--
Visual Ele.	--	--	--	--	--	--	--	--
Locations	8.18%	11.25%	11.49%	14.83%	6.45%	15.04%	10.77%	8.92%
Description	0.74%	1.61%	1.94%	1.26%	0.54%	2.26%	4.10%	1.88%
People-related	0.30%	0.54%	0.60%	0.47%	0.00%	1.50%	1.03%	0.00%
Art hist. info	83.04%	66.61%	72.54%	70.35%	93.01%	72.93%	86.67%	84.98%
Abst. concepts	0.15%	0.54%	0.15%	0.16%	0.54%	1.50%	0.00%	0.00%
Content/story	1.64%	5.00%	1.94%	1.89%	1.08%	13.53%	2.05%	5.16%
External Rel.	--	--	--	--	--	--	--	--
	Nonuniuqe				*Nonunique with refiners*			
	Art Sc.	K12	Muse.	Univ.	Art Sc.	K12	Muse.	Univ.
Objects	19.38%	24.20%	24.29%	22.78%	34.29%	20.27%	42.50%	38.03%
People	25.55%	27.07%	25.71%	24.05%	12.86%	18.24%	2.50%	8.45%
Color	--	1.27%	--	--	--	--	--	--
Visual Ele.	1.76%	1.27%	0.95%	0.84%	--	1.35%	--	--
Locations	4.41%	3.82%	5.24%	7.59%	12.86%	13.51%	12.50%	9.86%
Description	13.66%	9.24%	13.81%	19.41%	31.43%	8.78%	17.50%	16.90%
People-related	1.76%	3.82%	5.24%	2.11%	1.43%	4.05%	0.00%	2.82%
Art hist. info	19.38%	14.65%	12.86%	14.35%	50.00%	44.59%	41.25%	47.89%
Abst. concepts	7.93%	8.92%	4.76%	8.86%	0.00%	9.46%	5.00%	1.41%
Content/story	8.81%	10.51%	8.57%	6.33%	8.57%	10.81%	11.25%	15.49%
External Rel.	--	--	--	0.42%	--	--	--	1.41%

The high percentage of *unique* queries seems to align with the also high level of *Art Historical Information*. On average, queries judged to be *unique* by the researchers were also judged to be *Art historical information* 73.46% of the time. Across institutions there were slight variations in what type of Jörgensen

attributes were applied to *unique* queries. *Unique* queries from art schools were comprised of art historical information 83.04% of the time, for k12 66.61%, museums 72.54%, and universities 70.34% of the time.

Locations and *People* Jörgensen attributes were applied to *unique* queries at the next highest rates to *Art historical information* (11.40% and 11.75% respectively). The rate of *unique* queries comprised of *Locations* and *People* were higher among queries from K12, museum, and university type institutions than from art schools. *Unique* queries from art schools were comprised of *People* and *Location* information at rates of 6.70% as *People* and 8.19% as *Locations* while *unique* queries from K12 institutions had higher rates of *People* (16.25%) and *Location* (11.25%) information. Museums and universities had rates of *unique* queries also classified as *People* at a rate of 12.69% and 12.15% respectively and as *Locations* at rates of 11.94% and 14.85% respectively. The slightly higher rate of *unique* queries also classified as *People* originating from K12 institutions and classified as *Locations* originating from universities may be tied to their educational missions. Lesson plans and student interest in images at K12 institutions may be focused more on specific notable people throughout history, for example image of George Washington or Cleopatra. Similarly due to the wider fields of study at universities, there may be more use of the database from non-art focused disciplines like anthropology and archaeology. Within these disciplines, specific geographic sites (and therefore *unique location* information) may be of greater interest in instruction and research.

Though queries from K12 institutions were more frequently *nonunique* than queries from other institutions, largely the types of *nonunique* queries submitted were comprised of a similar percentage of Jorgensen attributes across institution type. For *nonunique* queries, the most frequently co-assigned Jörgensen attribute was *People.* Queries for *People* made up about a quarter of the *nonunique* queries across institutions: 25.55% at art schools, 27.07% at K12 institutions, 25.71% at museums, and 24.05% at universities. Among the *nonunique* queries, the query attribute that varied the most was the *Description* attribute. Among art schools and museums, the rate was very similar 13.66% and 13.81% respectively. *Nonunique* queries from K12 institutions were less likely to be a *Description* with only 9.24% of the *nonunique* queries being a *Description.* Universities had a higher rate with nearly one-fifth (19.41%) of *nonunique* queries being comprised of *Descriptions.* There does not appear to be a single reason for the higher rate of *Description* attributes among the *nonunique* queries originating from universities. However, it may be the result of more nonsensical searches (e.g. the, Br) that were classified as *nonunique* and as a *Description* because they did not clearly fall into any other Jörgensen attribute categories. These types of nonsensical queries might be increased in the university environment where students and faculty might be likely to be searching for images in the context of some other assignment and therefore switching back and forth between applications. This would only account for a small portion of the increase in *nonunique description* classified queries from universities. There may be other reasons for increased description such as queries from disciplines like

communications (e.g. “celebrity headshots”) or better user training that may be provided by university libraries (e.g. “Hosted collections”).

Some further differences between institutions can be seen in the types of *Art Historical Information* that was submitted in queries. When a query was judged to be *Art Historical Information,* it was very likely to be an artist name. Across all institutions the percentage of *Art Historical Information* was between 50-65% comprised of artist names with 64.06% from art schools, 61.96% from museums, 55.16% from K12 institutions, 51.15% from universities. A more striking difference appears when examining the use of artist names across all queries. Artist information comprises 45.20% of all queries submitted from art schools. This percentage of artist information across other institutions was still high but was much lower than that of art schools. Artists comprised 27.79% of K12 queries, 38.36% of museum queries and 30.74% of university queries. This difference between art schools and other institution types may align with the way studio students are instructed. From the author’s anecdotal experience, comparisons between a student’s work and a well-known artist’s work are often made during class time and critiques. Students may be frequently instructed by their professors to seek out images of artwork by particular artists.

Table 14. Types of art historical information in queries.

Types of Art Historical Information in Queries												
	Art school			K12			Museum			University		
	#	% AH 808	% all 1,155	#	% AH 582	% all 1,155	#	% AH 715	% all 1,155	#	% AH 694	% all 1,155
Media	10	1.24	0.87	5	0.86	0.43	6	0.84	0.52	4	0.58	0.35
Technique	7	0.87	0.61	3	0.52	0.26	10	1.40	0.87	3	0.43	0.26
Artist	522	64.60	45.19	321	55.15	27.79	443	61.96	38.35	355	51.15	30.74
Style/Period	32	3.96	2.77	61	10.48	5.28	24	3.36	2.08	43	6.20	3.72
Title	77	9.53	6.67	62	10.65	5.37	55	7.69	4.76	103	14.84	8.92
Size	--	--	--	--	--	--	--	--	--	--	--	--
Work type	32	3.96	2.77	31	5.33	2.68	27	3.78	2.34	26	3.75	2.25
Combination	128	15.84	11.08	99	17.01	8.57	150	20.98	12.99	160	23.05	13.85

The other two categories of *Art Historical Information* that comprised large portions of the attribute were *Title* and *Combination.* Queries from universities that contained *Art Historical Information* were comprised of 23.05% combinations of art historical information (13.85% of total queries) and 14.84% of titles (8.92% of total university queries). Both of these were higher percentages than seen at any of the other institutions. The increased searches titles may be partially explained by art history instruction and studying. Whereas students at art schools or museum professionals may be more interested in artists, students and faculty members engaged in art historical instruction and learning may have a list of art works for a particular course. This may also partially explain why the combinations of types of *Art Historical Information* were increased in queries from universities.

Overall there were few striking differences among the queries originating from different institution types. Art schools, museums and universities shared the most similarities while K12 institutions' searches were slightly more generalized and spread across non-art specific interests. Art schools had a higher percentage *Art historical information,* particularly artist, queries than the other three institutions, but overall the percentages of query type across institutions was similar. This particularly becomes clear when we examine the queries across the three time periods.

6.5 Variations of queries across time.

The most striking differences among the samples arise when one compares the query samples across the three time periods. Even though there are differences among queries from each type of institution, trends emerge across the four types of institutions. Overall, queries became slightly more specific and complex in 2007 and then became much more generalized and shortened in 2009.

Table 15. Word count by year.

Average Word Count by Year			
	2005	2007	2009
Art schools	2.01	2.43	1.31
K12	2.06	2.30	1.26
Museums	1.89	2.18	1.28
Universities	2.11	2.47	1.28
Totals	2.02	2.35	1.28

The average word count rose in 2007 and then dropped somewhat dramatically in 2009. In 2005, the average word count per query was 2.02 words per query. In 2007, this average rose to 2.35 words per query and then dropped to 1.28 words per query in

2009. This trend occurred across all of the institution type samples. Interestingly, the standard deviation of word count across institutions was smallest for 2009 at .020 words per query compared to .133 for 2007 and .093 for 2005. The query length also went from being predominantly spread across lengths of one, two and three words in 2005 and 2007 to being heavily weighted on one-word queries. One word queries comprised between 29% to 41% across all samples from 2005 and 2007, and then the frequency of one-word queries jumps to between 79% to 83% in 2009. This seems to be a clear shift away from lengthier queries to shortened ones and well above levels of statistical significance (for 95% confidence value and 5% error rate).

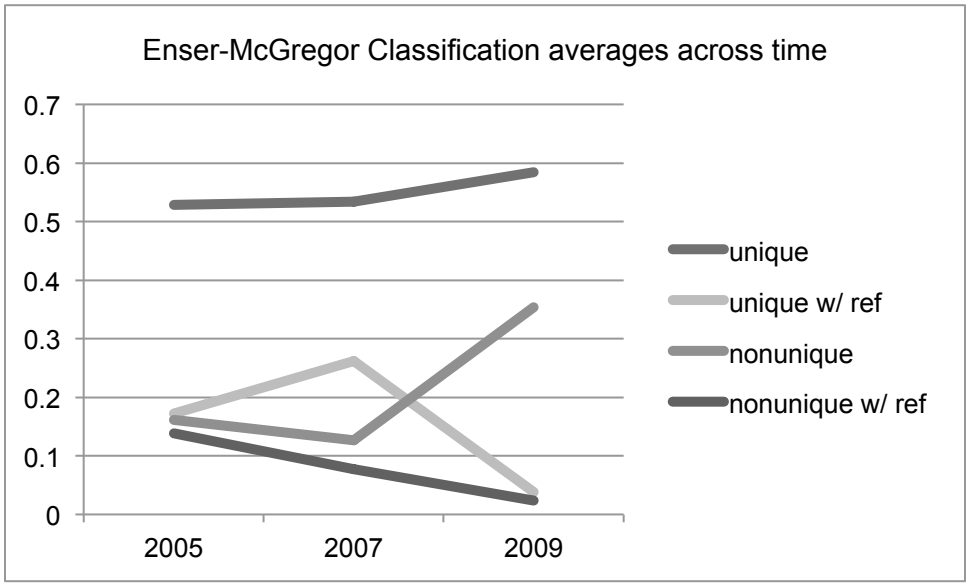
Table 16. Query length across institutions and time.

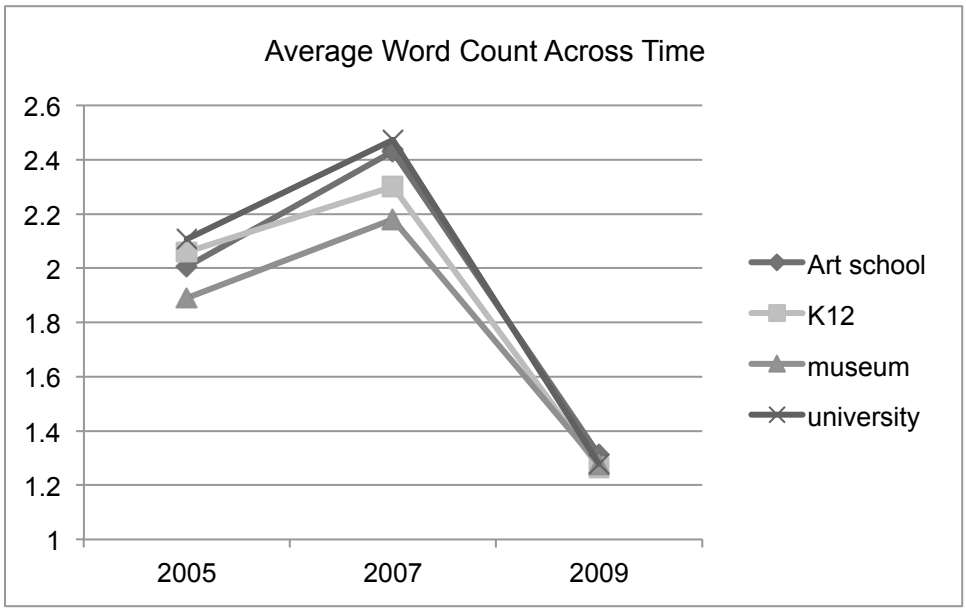
Percentage of Queries with Word Length												
	2005				2007				2009			
	Art Sc.	K12	Muse.	Univ	Art Sc.	K12	Muse.	Univ	Art Sc.	K12	Muse.	Univ
1	34.55	33.77	41.04	34.29	30.13	36.10	32.21	29.09	78.44	81.82	83.12	81.82
2	44.94	38.96	39.22	37.92	40.26	33.77	41.82	35.58	13.77	11.69	10.39	12.47
3	12.73	19.48	13.77	16.88	10.13	15.32	15.32	15.06	5.97	4.94	3.12	3.64
4	3.64	4.94	3.90	6.75	9.61	7.27	4.94	9.87	1.82	1.30	2.86	1.04
5 +	4.16	2.86	2.08	4.16	9.87	7.53	5.71	10.39	0.00	0.26	0.52	1.04

Concurrent to the shift toward shorter queries was a shift toward less complex queries. Though queries judged to be *unique* according to the Enser-McGregor classification consistently comprised the majority of queries, there were shifts in the specificity and complexity of queries in the other categories. In 2007, it seems that queries became slightly more specific and complex while in 2009 there was a significant shift to more general and simplified queries. Queries

judged to be *unique with refiners* increased from an average of 17.21% of 2005 queries to 26.17% of 2007 queries and then sunk to a just 3.83% of queries in 2009. Conversely queries judged to be *nonunique* decreased in 2007 and then increased in 2009. In 2005, *nonunique* queries made up 16.10% of queries in 2005, 12.66% of queries in 2007 and then jump to 35.39% of queries in 2009.

Table 17. Comparison of Enser-McGregor and word count across time.





The changes across time in the Enser-McGregor classifications seem to align with the word count. Most telling may be the converse increase in *unique with refiners* queries and decrease in *nonunique* queries in 2007 as word count increases. Then as word count decreases in 2009, *unique with refiners* queries decreased significantly as *nonunique* queries increased. It seems logical that as word count decreases both types of queries with refiners would decrease as well. It is hard to imagine a *unique with refiners* or a *nonunique with refiners* query that would contain only one word.

When examining the distribution of Jörgensen attributes to queries across the time samples, it seems that for the most part the percentage of the 12 attributes stayed somewhat consistent over time. Samples have average percentages of 8 out of the 12 attributes that stay within 2 percentage points across the three time periods. The remaining four attributes that saw higher variations across samples over time were *Objects, People, Locations,* and *Art historical information.* The largest changes occurred in the *People* and *Art historical information* attributes. A similar converse trend to word count and the *unique with refiners* and *nonunique* Enser-McGregor classifications occurs with *People* and *Art historical information.* There is a decrease in 2007 for queries with the *People* attribute and then an increase in 2009 (averages were 12.92% in 2005, 9.87% in 2007, and 18.70% in 2009). Conversely, there was an increase in *Art historical information* in 2007 and a decrease in 2009 (averages were 58.05% in 2005, 69.16% in 2007 and 54.74% in 2009).

Examining the samples from 2009, it appears that there may be a trend emerging in queries where searchers are moving away from using full names of artists and toward using portions of a personal name to search. Thus, a searcher who may have used “Andy Warhol Disasters” in 2007 may then try searching just on “Andy” in 2009. The shifts in classifications seem to suggest what the researcher has seen anecdotally. When examining the change in Jörgensen terms cross-referenced to Enser-McGregor classifications across the time period samples, there appears to be a spike among *nonunique* queries that are also identified as *People.* This aligns with the researcher’s observations because a query that was comprised of the single word ‘Andy’ could not necessarily be linked to ‘Andy Warhol’, ‘Andy Griffith’ or ‘Andy Kauffman’ and would then be classified as a *nonunique* and *People* query.

Table 18. Art historical information of queries over time.

Art Historical Information of Queries						
	2005		2007		2009	
	% of AH	% of Total	% of AH	% of Total	% of AH	% of Total
Media	1.46%	0.84%	0.56%	0.39%	0.71%	0.39%
Technique	0.90%	0.52%	0.85%	0.58%	0.71%	0.39%
Artist	49.61%	28.77%	54.37%	37.53%	73.55%	40.26%
Style/Period	5.82%	3.38%	4.89%	3.38%	6.64%	3.64%
Title	12.65%	7.34%	12.04%	8.31%	6.64%	3.64%
Size	--	--	--	--	--	--
Work type	4.70%	2.73%	2.35%	1.62%	5.81%	3.18%
Combination	24.86%	14.42%	24.93%	17.21%	5.93%	3.25%

Though a shift toward greater use of general names in queries seems to exist, this doesn’t point toward a vast shift away from the use of specific artist names. In fact, there seems to be an overall shift toward using personal names whether generally as in “Andy” or specifically as an artist name as in “Andy

Warhol." Over time, there is an increase in the use of artist names in queries. Over all queries sampled, the average percent of queries judged to be *artist* information was 28.77% in 2005, 37.53% in 2007 and 40.26% in 2009. The increased use of artist names in queries is even more striking when we examining just the queries judged to be *Art Historical Information.* Of these queries, artist information made up 49.61% of *Art Historical Information* in 2005, 54.37% in 2007, and 73.55% in 2009. The increase in artist information used in 2009 is accompanied by a decrease in combinations of types of *Art historical information* between 2007 and 2009 (24.93% of *Art Historical Information* queries in 2007 to 5.93% in 2009, and 17.02% of all queries in 2007 to 3.25% of all queries in 2005). These shifts suggest that while some users may be shifting away from queries using a specific name like 'Andy Warhol' to using general names like 'Andy', other users may also be shifting from queries that use a combination of *Art historical information* such as 'Andy Warhol silkscreens' or 'Andy Warhol Disasters' to simply the artist's name.

7. Discussion of results

7.1 Comparisons to previous studies

The results show some similarities to other studies, but there is also evidence that there may be different methods of searching within a specialized cultural heritage image database versus other types of web searching. In addition, the results hint that there are significant shifts in user search query construction over time suggesting that we need to continually monitor how users interaction with visual information systems.

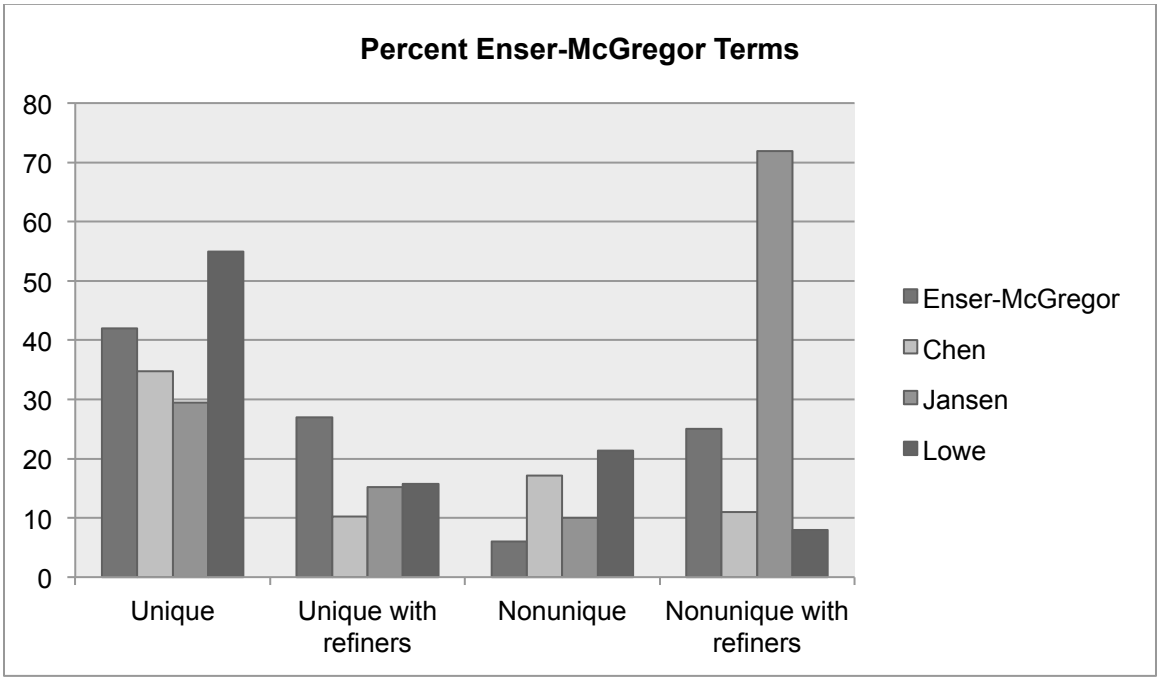
Beginning with the most basic of analysis units, query length, the queries in these data sets appear to be shorter than previous analyses of image searching whether on the web or in image databases. Jansen, Goodrum and Spink found that image searches submitted to the Excite search engine were slightly longer than non-multimedia searches at 3.46 terms per query (2000, 252). The ARTstor queries were much shorter on average over the five-year time period at 1.88 terms per query (with an even smaller average for 2009 at 1.23 terms per query). There may be multiple explanations for this shift. The Jansen, Goodrum, Spink study was conducted on search terms from 1997, so user practices may have shifted over this period of time. Additionally, the means by which image queries were determined to be such included the use of image-related words within the query thereby inherently lengthening the average query. However, even when accounting for the use of an image term within a query, the average length is still much higher than the query length found among the ARTstor queries.

There are some indications that the shortening of queries may be a trend that has evolved over time as search engines improve and users adjust to new information retrieval environments. Studies from much more recent data sets show shorter lengths for queries. Bendersky and Croft found that searches on the MSN search engine site averaged 2.4 terms in length and 90.3% of the queries were less than 4 terms long (2009, 9). Sample sets of queries from ARTstor still seem to be slightly shorter but the percent of queries less than or equal to four terms long is similar (95.95% for all samples, 99.55% for 2009). Other studies have shown query length of web searches to be of similar length to the Bendersky and Croft study, such as 2.25, 2.28, and 2.32 across demographic groups in Weber and Castillo from a 2008 Yahoo! search log (2010, 528).

Interestingly, one study shows that expert users may be more likely to submit shorter queries. Fang et al found that experts' queries averaged 1.96 while web queries averaged 2.35 terms per query (2011, 1189-1190). Perhaps because ARTstor is generally provided through the context of an educational institution, it may be that there are a greater percentage of experts using the service and thus lowering the average terms per query. Additionally, Jörgensen and Jörgensen found that within a professional stock image gallery, query length was also shorter than web searches. The two samples in the Jörgensen and Jörgensen study have mean terms per query of 1.418 and 1.364 (2005, 1352), which is much closer to the overall average found across the ARTstor samples at 1.88 terms per query. Jörgensen and Jörgensen also examined the preview and download rate and this may give some clues to the shortened queries. Preview

rate per query was about 1.37 and 1.27 images per query across the two samples. However, of images previewed only about 5-6% of those images were downloaded. Given the wide use of browsing in image-seeking behaviors, queries may be shortened by users to retrieve a greater number of images in the results. These results may then be viewed to determine whether or not a user wishes to download the image.

Table 19. Enser-McGregor terms comparison across studies.



Results from various studies using Enser-McGregor				
	Enser-McGregor	Chen	Jansen	Lowe
Unique	42%	34.73%	29.5%	54.89%
Unique with refiners	27%	10.26%	15.2%	15.74%
Nonunique	6%	17.16%	10.0%	21.39%
Nonunique with refiners	25%	11.02%	71.9%	7.99%

Across various studies using the Enser-McGregor classifications for specificity and complexity, results have varied greatly. All four of the studies discussed in this paper, Enser and McGregor (1992), Chen (2001), Jansen (2008) and the current study drew their data sets from different types of

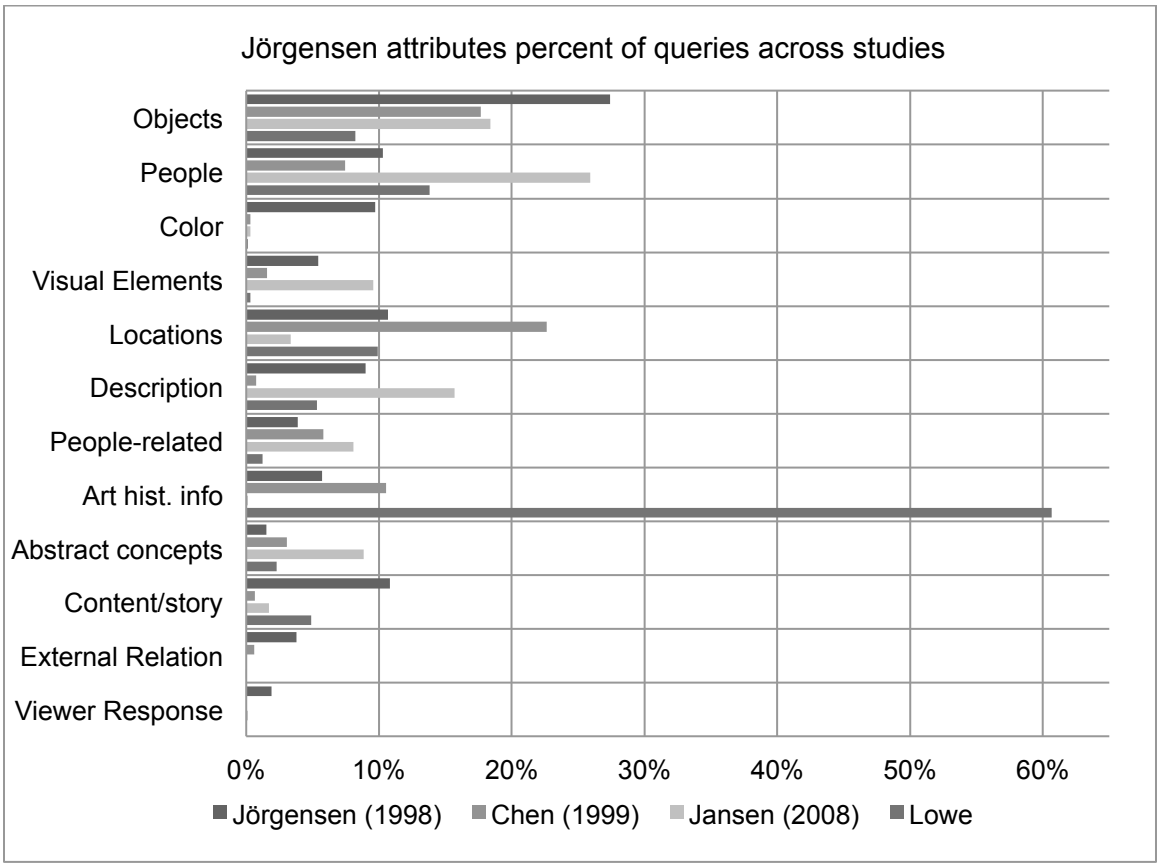
information environments**.** Enser and McGregor (1992) used image requests from a stock imagery collection, Hulton Deustch Collection**.** Chen (2001) drew his sample searches from university art history students image searches for a class assignment. Jansen's search queries were drawn from a 2001 dataset from the transaction log of Excite from April 30, 2001. Each of these studies can give us clues as to the differences and similarities of these image describing and seeking cases. Though there are some trends across the four studies, such as the fairly limited use of both nonunique queries and unique with refiners queries, each case differs from all of the others in some way.

Perhaps the most interesting comparisons can be made between the Jansen (2001) study of general web image searching and the current study of cultural heritage image searching. Most notably the use of refiners is drastically different. In the current study users relied on *nonunique with refiners queries* only about 8% of the time while Jansen found that web users rely on the same type of queries about 72% of the time. The Jansen study did allow for the use of more than one refiner, so his percentages are slightly more inflated than any of the other three studies. However, even given the difference in evaluation, the data suggests that users employ far different strategies when searching the web for images than when searching a specialized image database. Given the character each resource and the reasons a searcher might be drawn to one over the other, it seems that the informational need as well as the information ecosystem may be driving how searchers create queries for images on the open web versus

ARTstor. The more specialized image resources have a far greater use of specific queries than the web.

Table 20. Assignment of Jörgensen attributes across studies.

Jörgensen attributes percent of queries across studies				
	Jörgensen (1998)	Chen (1999)	Jansen (2008)	Lowe
Objects	27.4%	17.65%	18.40%	8.23%
People	10.3%	7.46%	25.89%	13.83%
Color	9.7%	0.33%	0.30%	0.11%
Visual Elements	5.4%	1.57%	9.55%	0.30%
Locations	10.7%	22.62%	3.34%	9.913%
Description	9.0%	0.75%	15.67%	5.35%
People-related	3.9%	5.80%	8.06%	1.23%
Art hist. info	5.7%	10.52%	0.12%	60.65%
Abstract concepts	1.5%	3.06%	8.86%	2.27%
Content/story	10.8%	0.67%	1.73%	4.91%
External Relation	3.8%	0.58%	--	0.04%
Viewer Response	1.9%	--	0.11%	--



Comparing the ARTstor samples to other studies that have used the Jörgensen attributes to describe image queries, the ARTstor searches are significantly different than the searches in the different image environments from other studies: a random selection of American illustration images in Jörgensen (1998), a local image collection in an art history department in Chen (1999), or the web search engine Excite in Jansen (2008). The largest distinction of queries within the ARTstor image database was the reliance of users on *Art historical information.* Over sixty percent of searches to ARTstor contained *Art historical information* far more than Jörgensen (5.7%), Chen (10.52%), and Jansen (0.12%). While it seems unsurprising that there would be few *Art historical information* type queries submitted to the open web as in Jansen (2008) and even Enser and McGregor (1992), it does seem somewhat surprising that the art history students in Chen (1999, 2001) relied on art historical terminology so infrequently compared to the ARTstor queries.

Because there is such a large gap in the time frame of these previous studies and the current samples, as well as a lack of user input, it is difficult to determine the core reasons that may cause this difference between the Chen (1999, 2001) study and the ARTstor samples. The information environment may play a major role in the differences. In the Chen study, students recorded the terms they planned to use and terms they actually used in searching for images across the entire resources of a university art library. Conversely, users within the information environment of ARTstor are restricted to the information available in ARTstor.

There have been relatively few restrictions on the amount of metadata contributed to ARTstor from institutions contributing to its image collection. Many entries may only have creator, date, and title information with few added fields (Wees, 2013). Therefore, users may be more likely to rely on the types of information they believe will be within the system. For example, a student may search the index of an encyclopedia for a broader range of associated terms (Baroque painters, Absolutism, Chiaroscuro, as well as Rembrandt or Caravaggio), but the same type of university students may limit themselves to types of terms usually seen alongside images when searching in an image-only database such as ARTstor (e.g. Rembrandt, Caravaggio, and other title or creator information). This theory may be somewhat bolstered by the high percentage of artist names used within queries to ARTstor (35.51% of across all samples).

When comparing the samples from ARTstor with the Jansen (2008) study of web image searching, it is clear that there are many differences among the two search environments. Jansen believed that the web and academic or art image databases were such different environments that extra categories needed to be added to the Jörgensen classifications to accommodate these differences. Using suggestions from Chen (2001) Jansen added categories to the Jörgensen attributes that may be more appropriate for open web image searching: cost, URL, and collections (93). These added categories appeared in web searches at a significant rate with queries being comprised of 31.2% collection, 5.8% cost and 0.8% URL attributes. Jansen considered queries like “snow blizzard

photography" and "stock photography" to be collection queries (2008, 98). While these attributes clearly seem applicable to the web environment, it is not clear that they would have extensive use within ARTstor or other similar cultural heritage image collections. Both URL and cost seem like highly unlikely searches within ARTstor because the images are a self-contained collection and are free to download for educational use.

Within a cultural heritage image context, Jansen's suggested collection attribute may be more useful than URL or cost but becomes problematic when compared to the *art historical information* and *location* attributes. Were a user to submit the query 'Caravaggio Uffizzi' would this be a query solely of the *art historical type* or also the *collection* type? It seems that both the creator or title information as well as repository information might be considered *art historical information* or as *location* as in a *unique with refiner* query: 'Caravaggio' (unique art historical information) 'Uffizzi' (a refiner with *location* information). Using a collection attribute may also confuse data about queries in cultural heritage image databases because the images may be separated into individual collections themselves that are aggregated together through a type of union catalog as happens in many university digital image collections.

In addition to the additional types of queries observed in web image queries, the focus and information need may also be different. Considering the two information environments, the marked differences between the queries potentially have logical explanations. Web queries encompass a much wider range of human information than queries to a targeted subscription-based

educational might. One such example is the high occurrence of queries for pornographic or otherwise sexually explicit within image search engines on the web. Aside from sexual arousal, there are countless other reasons that might prompt a user to submit a query for an image on a web search engine from scholarly inquiries to illustrations for a flyer to entertainment purposes. However, the reasons one might visit a resource such as ARTstor are much more limited. Given that ARTstor must also be accessed through a subscription, the range of information needs a user might have are likely to be even further restricted. It seems that the reason such a large percentage of queries to ARTstor are *Art historical information* is that the database is designed to fit the needs of users who are seeking that particular type of information.

While there are trends across the different search environments, some trends do become evident. There is a very low use of certain types of information- even in web searches. *Viewer response* and *External relation* attributes were rarely if ever used across the four studies. *Color* queries were observed very infrequently across the four studies with the exception of the Jörgensen study where these types of searches were observed 9.7% of the time compared to 0.33% in Chen (2001), 0.30% in Jansen (2008), and 0.11% in ARTstor. One possible explanation for the increased use of color terms in the Jörgensen study was the information environment. Students were searching the description provided by other students, and therefore may be more likely to try to rely on color descriptions than they might be were they searching a digital database. The extremely low rate of color queries in Chen, Jansen, and the

ARTstor samples suggests that users may rarely have a desire to filter such results by color.

7.2 Practical Implications

Though there is still a need for further inquiry, the results of the query classification do have some practical implications ARTstor's design as well as other cultural heritage online image databases. The heavy reliance on art historical terms within queries should be some cause for relief among image cataloguers and art historians. While the search log cannot confirm or deny whether users actually wish to search predominantly using art historical information, clearly among the institutions surveyed in this study users due often employ in their searches the same type of information that is often catalogued about an image (e.g. title, creator, media). Therefore, at least in the context of a tool like ARTstor much of the information considered to be 'minimal cataloging' is actually useful to users. There may be some question about whether this will be true to the same extent when the search logs of online museum collections are studied.

One concrete method of improving user success might be to improve the browsing structures for artist names. At the time of this study and when the search log was recorded, there is no method of browsing by artist name within ARTstor. Making the creator field a hierarchical structure similar to the existing geography or work type browsing structures might be extremely useful for users.

However, a rigid alphabetical list of artist names may not be the most useful way of facilitating user searches for artists' works. Given the high prevalence of searches containing only a first name (and often the first name of a famous artist), ARTstor and other cultural heritage image databases may want to consider employing the use of co-occurrence suggestive algorithms. Such tools would allow a user that types in 'Roy' to select the correct spelling of 'Roy Lichtenstein' or other artists names with Roy in them such as 'Roy Decarava'. ARTstor already uses algorithms to suggest terms when a query is likely to be misspelled. Because such an algorithm does not alter the structure of the backend database or extensive cataloging, providing the additional functionality of suggestive searching may not be prohibitively costly or time-consuming.

A more complex long-term improvement of the user interface would be to create linked records across the images. Therefore if a user were searching for 'Elizabeth', they might browse the results for the correct 'Elizabeth' among the results that return multiple images by artists like 'Elizabeth Murray' and 'Elizabeth Peyton' as well as many paintings of Queen Elizabeth. The user might then be able to browse to find an image she recognizes to be like the ones she is seeking. For example, she might be seeking works by an 'Elizabeth' that paints portraits rather than abstractly-shaped canvases. The user could then click on the artist name in the image record to be taken to all of the images with 'Elizabeth Peyton' in the creator field. Though this improvement would greatly expand the usability of ARTstor, it may be much more difficult to achieve. Given that ARTstor's collection is an aggregation of images from a wide array of

institutions from museums to libraries to universities and even personal image collections, the metadata is not standardized. Some such standardization might be achieved through automatic means such as using controlled vocabularies to group records with name variations. However, many of the records might need to be individually corrected by hand. This would be an almost overwhelming task to be carried out by a small staff across a collection of over a million records.

ARTstor might be able to take incremental steps toward such linked records. It might be possible to set new standards for metadata with format and content standards for all new acquisitions. For example, they might require that creator names be submitted with references to outside controlled vocabularies such as the Union List of Artist Names. While this wouldn't completely solve the problem of a large backlog, it might make it possible to work toward a manageable solution between methods of automatic aggregation and new acquisitions.

While fully implementing browsable creator structures into ARTstor may be a daunting task, smaller more homogenous collections might have an easier task creating linked metadata. For institutions like museums and universities that might publish their image collections to the web, they may be more likely to set such flexibility for browsing and searching on creator names more easily. Cataloging standards are much more likely to be compatible within a single institution than across the many hundreds of institutions' collections that ARTstor aggregates. There may be some difficulty in making records that use different controlled vocabularies linked such as records that might use both the Library of

Congress Names Authority File and the Union List of Artist Names. However, solving this problem within an institution that will continue to create and publish images may be worthwhile since users clearly depend on names whether specific full artist names or partial general names.

Beyond making creator names a more flexible and usable metadata field within a record, ARTstor and other cultural heritage institutions might want to consider making use of other metadata. Three other categories of the *Jörgensen* attributes comprised another near third of the queries: *People* (13.83%)*, Locations* (9.91%)*,* and *Objects* (8.23%)*.* These categories are also attributes that are often captured in traditional cataloging through subject, location, media, work type or other fields. For example, an image of the fresco *The School of Athens* if cataloged within the context of a university image collection would include title, creator, and date information but would also likely include some subject cataloging as well. An image cataloguer would be likely to include Aristotle and Plato (*People)*, the repository: the Apostolic Palace, Vatican City (*Location*) and might also include some nameable objects like books, globes, scrolls *(Objects)*. While standardizing all of these across even a smaller collection might be difficult, it may be possible to take cues from tools like Flickr and simply make these fields a linkable tag. Therefore, any image with Plato in the record might be accessed by clicking on Plato in the subject metadata of a record.

For other collections or those who might be creating their own cataloging recommendations, it might be wise to use available controlled vocabularies for

descriptive metadata that might be able to capture some of these frequently used types of metadata. Geographic place names, whether as a repository, archaeological or historical site, might be linked to controlled vocabularies like *Thesaurus of Geographic Names* or similar hierarchies. This might enable better browsing structures allowing users not simply to browse to a current country (as in the ARTstor interface at the time of this study) but also to regions within the country or historical kingdoms.

Similarly, descriptive terms whether as subject entries or as work types might be standardized to improve access to object information. For example, *The Art & Architecture Thesaurus* might be used to aid in the selection of object descriptions. This might allow the user to move from a general *object* term like 'amphora' to either a narrower term like 'bail amphora' or to a broader *object* term like 'storage vessels'. Such integration of a widely used vocabulary might not be possible in all institutions. Such extensive hierarchical linking of metadata might not be necessary, but institutions might wish to create their own in-house vocabularies or choice lists for certain fields. This will make linking records in an online database much easier.

For the most part, the results from the ARTstor queries of the types of searches submitted are good news for the image cataloging community. Largely *Art historical information, People, Location,* and *Object* information are assigned by cataloguers. Therefore, we should take these results as an encouragement to make better use of the metadata that is traditionally ascribed to images by

implementing interfaces that make query suggestions, offer wider browsing structures, and link records via metadata.

7.3 Research Implications

While the results of this study of the ARTstor query logs is encouraging because it points to concrete improvements we might be able to make in image retrieval systems, many of which tie into the way many institutions already catalog their digital images. However, there are many questions that the current study leaves unanswered or raises. This study merely gives a window into the types of queries submitted, how they have trended over time and how they trend across specific institution type.

One problem with transaction log studies, even robust studies with ample data points, is that they are only a trace of the actual human computer interaction. Often many interactions such as hitting the back button or moving to a separate window to investigate another resource cannot be recorded in a transaction log. Transaction logs cannot capture the thoughts, feelings, or actual needs of the user. They are only the trace of a user acting out a search to fulfill those needs. While we can make a guess at the informational goals of users, we must remember that this is still conjecture. Simply because we observe a move away from searches for ‘Andy Warhol’ and toward ‘Andy’, we may be able to guess that users are trying to expand their results, but there may be multiple reasons that prompt the searcher to take these actions that would be completely out of the realm of the imagination of the researcher.

Several types of studies can give us a greater understanding of the user experience and desires from users' perspective. Client-side transaction log studies can record multiple interaction not only with the information retrieval system in question but also if a user leaves the system to inquire into other resources. Such studies generally take the form of a specialized browser that a client may download and allow for the researcher to collect their interactions with the browser. Using such a client side method of collecting human computer interactions while the user searches for art and cultural heritage images might give researchers a greater understanding of how several tools may work together. Is the user accessing other tools like Oxford Art Online, using ARTstor while also using classroom management systems like blackboard or watching YouTube videos in between searches? Understanding how ARTstor or other cultural heritage image databases fit into the larger information environment might create new opportunities for partnerships or cross-linking between resources. Client-side transaction studies still retain some drawbacks. It may be more difficult to obtain a properly randomized and representative sample of ARTstor users, so the information gleaned may lean toward one particular user group. Additionally, although a browser might be downloaded onto an individual's computer, it is still not completely the native search environment for the user, so this may slightly skew how the user interacts with ARTstor or another target image database.

Talk-aloud studies may be one way to gain insight into the motivations and reasoning involved in searches to ARTstor. A talk-aloud study would focus on a

user engaging in some interaction with ARTstor or a selected database while speaking aloud her thoughts to a researcher. The researcher could inquire about the motivations for the search, for altering a query, and for deeming a session successful or abandoning it. Through such a study, researchers might be able to ascertain whether users actually desire to use *Art historical information* at such a high level or whether they use it because they believe it will be more likely to return successful results. Whether query reformulation becomes shorter, longer, more complex or simplified, a user can give their motivations for altering their search strategy in this manner. Additionally, having separate user groups that have the ARTstor interface explained to them versus users that are trying it without training may give an indication as to whether additional tools are useful. Options such as date or geographic filtering offered within ARTstor might be less apparent to new users than to users who have had ample time to learn the interface. The experience of having a user explain their thoughts and points of frustration could reveal roadblocks to successful searches that search log analysis does not uncover.

The web environment has changed drastically over the last ten years. As discovered within this study, users' strategies for an image-seeking task change quickly. However, many studies about user motivation for images were conducted nearly that long ago. New studies outlining the motivating information need should be conducted. These studies could give researchers better metrics by which to judge a search session in ARTstor or any other cultural heritage image database. If motivations are largely to download images for class, to

obtain images for a lecture, or to create a personal educational collection, a session where an image is viewed but not downloaded might not be considered successful. Depending on these motivations there may also be discoveries that could lead to new tools, similar to the flash card mobile tool currently offered by ARTstor. These studies should carefully select participants based on types of user groups. Even within one type of institution, multiple types of users should be considered and selected proportionally. For example, within a museum, a researcher might want to gather information from curators, educators, librarians, photo managers, and other potential users of image databases.

Finally, institutions currently tracking user interactions with their image databases should make better use of these logs. There are numerous types of transaction log studies; each that gives a different insight into the user's information-seeking behaviors and practices. While it is important to understand the basic units of image-seeking behaviors such as queries, much more information can be gleaned from a robust transaction log. In particular, looking at individual querying sessions could shed light into how users reformulate queries when at first unsuccessful, how willing users are to struggle with a search, and when searchers might consider an interaction to be successful or when they might abandon a search.

There is still a great deal of work that needs to be conducting to make the results from transaction log analysis within image databases, particularly those dealing with specialized material such as ARTstor, comparable across studies. For this reason, it is important that the cultural heritage community work together

to form standards for transaction log analysis. Some standards might include types of interactions to record, methodologies for examining specific questions, algorithms or boundaries for image search sessions, and what constitutes a successful search. These metrics might be slightly different across institutions, but uniformity in the types of information recorded and clarity in the purposes for specific metrics would vastly improve how we discuss and consider transaction log studies across various image-retrieval systems.

8. Recommendations for Transaction Log Tracking

Large educational image databases such as ARTstor and museum image collections have been to date largely an untapped resource for discovering more about user efforts to find images. We understand so very little of the actual strategies users employ. For the most part, the studies of image databases have been divided between large general image databases and fairly small or specific museum image collections. With more large museums opening their digital image resources to public web searches and the popularity of educational image databases like ARTstor, it seems that there is now much wider use of cultural heritage databases than ever before, yet our understanding of our users is still so limited. Unlike commercial sites that host images like Flickr or Google, there is little disincentive to share information about how users interact with cultural heritage image databases like museums. The more we understand how users interact with similar types of tools, the better we can make these tools across the entire cultural heritage spectrum. Therefore it makes sense to establish some guidelines as to the types of information we collect in the transaction logs of cultural heritage image search sites. With such standards we may be able to begin to see similarities or differences between the way users interact with a site like ARTstor versus a museum like the Los Angeles County Museum of Art. We may also be able to better detect points of frustration for our users.

In the non-profit world of cultural heritage and educational institutions, we have spent so much time and effort over the last decade to digitize collections that little thought has been given to how these collections might actually be used.

Now that digitization and to some degree storage and preservation issues have been streamlined and codified, the cultural heritage fields should now invest in making their tools more effective for users. One such way is through interface improvement, and one of the most common means of learning ways to improve user experience is through studying transaction logs. Were the cultural heritage community to set standards for data collection in transaction logs, this would make the analysis and comparison of behaviors in collections and trends across time easier to discover.

8.1 Considerations before logging

Before one can analyze a search log there has to be a search log itself and a system within which it operates. Any organization, including ARTstor should carefully consider what they wish to discover through recording the transactions, who will be studying the log, and how often it will be studied. Making these types of decisions brings a focus and purpose for institutions to acquire data about how users interact with their systems.

Decisions about when and who will study transaction logs may seem like a decision that could be put off until enough data has been collected to put together a study. However, depending on the institution who the researchers are who will be able to analyze the search log may effect the types of information you record. For example, if a museum also sells images or prints online, a purchase or inquiry may be considered important information for the institution to track. However, the institution may wish to keep sales information out of the public eye.

Recording such information might mean that the data will need to be cleaned before it can be handed over to outside researchers. Additionally, if an institution decides it does not have the staff to regularly examine the transaction log, relationships with outside reviewer may need to be forged or the data may need to be made publicly available through request.

The question of how often a transaction log should be examined may also depend upon the internal or external researchers and resources available. With the many other concerns of museum and library staff, it may be difficult to allocate frequent time to dedicate to studying a transaction log. However, with some knowledge of SQL, many types of analyses can be done automatically using commands. Lists of sample SQL commands can be found in Jansen (2009, 118-121). Jansen's sample SQL commands will clean very long sessions, identify unique visitors, identify session lengths, number of results, query length, presence of Boolean operators and other simple investigations. Using these simple SQL commands, even small institutions may be able to perform mini-studies of their search logs on a regular basis. Each institution will have its own timelines but should set goals for regular small-scale analysis of their transaction logs with periodic in-depth analysis.

The extent of information an institution collects in a transaction log should be dictated by what they wish to discover from their users. Small institutions may have little ability to greatly alter their web interfaces or backend databases, so they may have little need to investigate the full complexities of a user's session. Conversely, large entities with massive collections such as ARTstor may find

discovering whether user sessions are successful, points of frustration and other more detailed data points about user experience within the system. These findings might enable interface improvements that would increase the use of the service at institutions and insure continued subscriptions.

A final consideration before implementing a transaction log is its integration with the image retrieval system itself and the overall system. Some examples of these considerations would be decided which interactions are important to record and designing accordingly. Conversely, things that might be considered mistakes or corrupted data should not be allowed within the system. For example, a blank search submitted might be considered as an inappropriate or unclear data point. Rather than accept this type of information into the transaction log, the system should be designed to require text to submit a search. If users are employing blank searches to view all results, then a button might be added to the interface to allow the user to view all images. Conversely, programmers may need to insure that some actions such as image views and image downloads can be recorded separately. For a more extensive overview on how to consider user interface design and transaction log analysis, see Muresan (2009).

Further considerations that might need to be built into a system would be different user groups. Museums may wish to look at users from the perspective of local, national and international visitors and might wish to stratify their user groups through the geographic region, which might require a login including the use of a zip code such as the Weber and Castillo (2010) study. Subscription-

based systems such as ARTstor may be able to stratify their users based upon the subscribing institution through which a user accesses the system. Other image collecting institutions may simply wish to describe the behaviors between users that create an account with the institution and users that do not. Many institutions already require the creation of a user account for more advanced features, so the if there are specific types of users an institution would like to target they may wish to include demographic questions within the account registration.

In summary, there are several questions an institution may want to answer when considering how to best implement a transaction log and its subsequent analysis:

1.) What types of information does the institution hope to discover?

For example:

- depth of analysis: query, session, user types
- successful interactions vs. unsuccessful interactions
- rates of abandonment

2.) How often and by whom will the log be studied?

For example:

- Quarterly automatic analysis by IT staff
- Release cleaned datasets to outside researchers
- Hire consultants to analyze data

3.) What types of data must be recorded to answer the questions the institution seeks to answer?

For example:

- To determine if searches are successful record image views or download actions
- To determine user frustration points record abandonment or help page access

4.) Can this data be collected within and aided by the image-retrieval system?

For example:

- Use scripts that will allow for the type of records
- Design the system to produce cleaner results in the search log

The answers to these questions may be vastly different depending on the institution. However, being clear about the goals an institution has in tracking data—even if solely to provide information for other researchers—can help steer the design of the image retrieval system, its study, and possibly even future improvements.

8.2 Technical specifications

Once basic planning and goals are set for the transaction logging, pinpointing the exact data points to be recorded can begin. Below is a table listing commonly used or potentially useful actions to record within an image retrieval system:

Table 21. Suggested transaction log fields.

	Field	*What it is*
Minimal	*User ID*	Generally begins as a unique user IP address. During data preparation, these are usually anonymized by assigning an identification number.
	Timestamp	Date when the action occurred. Generally logged with all other actions
	Query (also called Search URL)	The text string submitted in a search
Medium	Results page	Identifiers of the items listed on the results pages
	Image viewed	Records when an image is viewed
	Image downloaded	Records when an image is downloaded
	Advanced search	Records when an advanced search function is used
	Filter	Records which filters are used on results
	Browse	Records browsing actions. May be separated into kinds of browsing (e.g. collection, category, etc).
	Login	Records when a user logs into an account
Extensive	Help accessed	Records if the user clicks on a help page
	Metadata view	If metadata is hidden, records when a user clicks a link to access it.
	View linked image	If related images are suggested on an image page, records when the user clicks to view these.
	Zoom image	Records whether the user zooms the image
	Email record	Records if the user emails an image or record
	View linked page	If a related webpage is listed on the image record, records when a user clicks this link.
	View linked group	If other images are related through linked data, records when the user clicks to access these

The fields listed above are by no means an exhaustive list of the types of data that might be recorded. Depending on the system, there may be many more levels of granularity within the possible fields to be recorded.

While it is unlikely that institutions across the cultural heritage image collection spectrum could agree to a single set of parameters to collect within a transaction log, it is important that when reporting results within the community that we also share the types of data points that are collected and how this is being done. In order to best understand the information ecosystem of our users, we should be able to compare results across institutions even if data is collected slightly differently.

In addition to the types of data collected there are a few other parameters that the cultural heritage image collecting community may need to determine. One of these parameters is the boundaries of sessions. As discussed in the literature review, there are many ways that researchers have decided to automatically determine the boundaries of an individual user's session. Some of these include time limits, use of terms that are syntactically or semantically related, some mix of both of these or other more heuristic methods (for a deeper discussion of various methods of determining session boundaries see Gayo-Avello, 2009). While there is still no accepted standard method of determining session boundaries, further studies on image retrieval databases may be necessary to determine if currently used methods are as effective when users are searching for images as they are on web studies.

Finally, each institution will need to determine the definition of a successful search. Within a museum collection context perhaps a successful session would be one in which the user simply views an image or travels to a page within the museum website. However, this metric may not apply equally well for a system

such as ARTstor where user needs may be much more likely to be linked to downloading or saving an image. Such determinations may also be highly dependent on what the transaction log might be able to record. For example, in an online image collection a user might right-click on an image and download the image through browser or operating system. Though a click on the image might be recorded, the transaction log would be unlikely to capture the download event. Therefore, when considering what determines a successful event, ideal conditions may not be met in all cases.

9. Conclusion

From comparing the results of the ARTstor search log samples to other studies of web image searching and other types of image searching, it is clear that there are differences in the way users search for cultural heritage images within a system like ARTstor. Because there are few similar studies of museum image collections and other cultural heritage digital image collections, it is difficult to know whether these results represent trends across the entire landscape of cultural heritage image searcher or if they are distinct to ARTstor itself.

The results confirm that there is wide use of art historical information as well as a high percentage of reliance on the types of subject and descriptive metadata that is standardly cataloged for images. By far the most frequently used type of search was an artist name search. This points to the need to expand tools that make finding work by particular artists easier for the user. Incorporation of thesauri and linked records as well as browsing structures on artist names may aid users in more effectively finding the images they seek. Whether the use of common types of terms used to catalog images actually reflects the types of images users desire or reflects a strategy users have learned to be effective is yet to be seen. Further studies need to be conducted that investigate information need and desired information-seeking behaviors from the users' perspective.

The distinct changes across the samples over time suggests that the study of information seeking behaviors can be time sensitive and that users' practices may shift widely in a relatively short amount of time. This suggests that researchers should continue to monitor even simplistic measures from

transaction logs such as popular queries, query length, and queries per session. Major shifts in these basic metrics may suggest areas of further research or changes in how the public interacts with image retrieval systems or even the web at large.

This study has only begun to scratch the surface of our understanding in how users search for cultural heritage images. Much more research is required to fill in the gaps that this study leaves. Both user-focused studies as well as more robust types of transaction log studies should be conducted. Toward this end, cultural heritage image collecting institutions have many reasons to work together to share data and standardize practices. Perhaps the most important of these reasons is the improvement of current user interface designs for image databases and cataloging practices. Even if information professionals have some intuitions into how users might search for images, having hard data on these practices can help to confirm or deny these hunches.

The digitization and preservation of image collections has become a somewhat standardized process. Cultural heritage and educational institutions have gone through cycles of digitization and distribution of images to web users. Now that the use of such collections is widespread and interwoven into many institutions missions, it is important to carefully consider how we deliver these images to our users and whether we are most effectively and efficiently meeting our users' needs. This can only be accomplished through continued and expanded research into informational needs, cognitive processes, and information seeking behaviors within image retrieval systems.

Works Cited

Armitage, Linda H., and Peter G.B. Enser. 1997. Analysis of user need in image archives. *Journal of Information Science* 23, no. 4: 287 -299. http://jis.sagepub.com/content/23/4/287.abstract.

Baca, Murtha. 2010. Controlled Vocabularies for Art, Architecture, and Material Culture. In *Encyclopedia of Library and Information Sciences*, ed. Marcia Bates and Mary Niles Maack, 1277-1281. 3rd ed. Taylor & Francis. http://www.informaworld.com/10.1081/E-ELIS3-120044074.

Baeza-Yates, Ricardo, Carlos Hurtado, Marcelo Mendoza, and Georges Dupret. 2005. Modeling User Search Behavior. In *Proceedings of the Third Latin American Web Congress*, 242. IEEE Computer Society.

Baeza-Yates, Ricardo, Liliana Calderón-Benavides, and Cristina González-Caro. 2006. The Intention Behind Web Queries. In *String Processing and Information Retrieval*, 4209:98-109. Lecture Notes in Computer Science. Springer Berlin / Heidelberg. http://dx.doi.org/10.1007/11880561_9.

Bates, Marcia J. 2001. *Information needs and seeking of scholars and artists in relation to multimedia material*. Los Angeles, CA: Getty Research Institute. http://www.gseis.ucla.edu/faculty/bates/scholars.html.

Bendersky, Michael, and W. Bruce Croft. 2009. Analysis of long queries in a large scale search log. In *Proceedings of the 2009 workshop on Web Search Click Data*, 8-14. Barcelona, Spain: ACM.

Broder, Andrei. 2002. A taxonomy of web search. *SIGIR Forum* 36, no. 2: 3-10.

Chen, Hsin-liang. 2001. An analysis of image queries in the field of art history. *Journal of the American Society for Information Science and Technology* 52, no. 3: 260-273. http://dx.doi.org/10.1002/1532-2890(2000)9999:9999<::AID-ASI1606>3.0.CO;2-M.

Chen, Hsin-liang. 2007. A socio-technical perspective of museum practitioners' image-using behaviors. *The Electronic Library* 25, no. 1: 18-35. doi:10.1108/02640470710729092.

Chen, Lin-Chih. 2011. “Term Suggestion with Similarity Measure Based on Semantic Analysis Techniques in Query Logs.” *Online Information Review* 35, no. 1 (February 22): 9–33.

Cobbledick, Susie. 1996. The information-seeking behavior of artists: exploratory interviews. *The Library Quarterly* 66, no. 4: 343-372.

Cochran, William G. 1963. *Sampling Techniques.* New York: Wiley.

Conniss, Lynne R., Ashford, A. Julie and Margaret E. Graham. 2000. Information Seeking Behavior in Image Retrieval: VISOR I Final Report. In *Library and Information Commission Research Report 95* Newcastle upon Tyne: Institute for Image Data Research, University of Northumbria at Newcastle.

Eakins, John, and Margaret Graham. 1999. *Content based image retrieval*. Technical. Joint Information Systems Committee Technology Applications Programme. Newcastle: University of Northumbria, October.

Enser, Peter G. B., and Charles McGregor. 1992. *Analysis of visual information retrieval queries*. Report on Project G16412 to the British Library Research and Development Department. London: British Library.

Enser, Peter, Christine J. Sandom, Jonathan S. Hare, and Paul H. Lewis. 2007. Facing the reality of semantic image retrieval. *Journal of Documentation* 63, no. 4: 465-481. doi:http://dx.doi.org/10.1108/00220410710758977.

Fang, Yi, Naveen Somasundaram, Luo Si, Jeongwoo Ko, and Aditya Mather. 2011. "Analysis of an Expert Search Query Log." In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval (24 July 2011)*, 1189–1190.

Frank, Polly. 1999. Student artists in the library: an investigation of how they use general academic libraries for their creative needs. *The Journal of Academic Librarianship* 25, no. 6 (November): 445-455. doi:doi: DOI: 10.1016/S0099-1333(99)00077-4. http://www.sciencedirect.com/science/article/B6W50-3YTBW12-4/2/48c3fa3371b9149831388ec598b78f37.

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

Gayo-Avello, Daniel. 2009. "A Survey on Session Detection Methods in Query Logs and a Proposal for Future Evaluation." *Information Sciences: An International Journal* 179, no. 12 (May 30): 1822–1843. doi:10.1016/j.ins.2009.01.026.

Goodrum, Abby, Matthew Bejune, and Antonio Siochi. 2003. A State Transition Analysis of Image Search Patterns on the Web. In *Image and Video Retrieval*, 2728:193-197. Lecture Notes in Computer Science. Springer Berlin / Heidelberg. http://dx.doi.org/10.1007/3-540-45113-7_28.

Goodrum, Abby, and Amanda Spink. 2001. Image searching on the Excite Web search engine. *Information Processing & Management* 37, no. 2 (March): 295-311. doi:doi: DOI: 10.1016/S0306-4573(00)00033-9. http://www.sciencedirect.com/science/article/B6VC8-4292GM8-7/2/5b621304ac151525e28fb14ac26ad7f2.

Hastings, Samantha K. 1999. Evaluation of image retrieval systems: Role of user feedback *Library trends* 48, no. 2: 438-452. http://www.ideals.illinois.edu/bitstream/handle/2142/8276/librarytrendsv48i2k_opt.pdf?sequence=1.

Heidorn, P. Bryan. 1999. Image Retrieval as Linguistic and Nonlinguistic Visual Model Matching. *Library Trends* 48, no. 2 (Fall): 303. doi:Article. http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=2713473&site=ehost-live.

Hollink, L., A. Th. Schreiber, B. J. Wielinga, and M. Worring. 2004. "Classification of User Image Descriptions." *International Journal of Human-Computer Studies* 61 (5) (November): 601–626. doi:doi: DOI: 10.1016/j.ijhcs.2004.03.002.

Hollink, Vera, Theodora Tsikrika, and Arjen P. de Vries. 2011. "Semantic Search Log Analysis: A Method and a Study on Professional Image Search." *Journal of the American Society for Information Science & Technology* 62 (4) (April): 691–713.

Hölscher, Christoph, and Gerhard Strube. 2000. Web search behavior of internet experts and newbies. *International Journal of Computer and Telecommunications Networking* 33, no. 1: 337-346.

Jansen, Bernard J. 2006. Search log analysis: What it is, what's been done, how to do it. *Library & Information Science Research* 28, no. 3 (Autumn): 407-432. doi:doi: DOI: 10.1016/j.lisr.2006.06.005. http://www.sciencedirect.com/science/article/B6W5R-4KSD84F-1/2/d131e3e1b135d3533787b301a941f893.

Jansen, Bernard J., Danielle L. Booth, and Amanda Spink. 2008. Determining the informational, navigational, and transactional intent of Web queries. *Information Processing & Management* 44, no. 3 (May): 1251-1266. doi:doi: DOI: 10.1016/j.ipm.2007.07.015. http://www.sciencedirect.com/science/article/B6VC8-4PMT5X4-2/2/cf258d9e8a14d18e0261d8c91581dde1.

Jansen, Bernard J., and Amanda Spink. 2006. How are we searching the world wide web?: a comparison of nine search engine transaction logs. *Inf. Process. Manage.* 42, no. 1: 248-263.

Jansen, Bernard J., Amanda Spink, and Tefko Saracevic. 2000. Real life, real users, and real needs: A study and analysis of user queries on the Web. *Information Processing & Management* 36, no. 2: 207-227.

Jones, Steve, Sally Jo Cunningham, and Rodger McNab. 1998. Usage analysis of a digital library. In *Proceedings of the third ACM conference on Digital libraries*, 293-294. Pittsburgh, Pennsylvania, United States: ACM.

Jörgensen, Corinne. 1998. Attributes of images in describing tasks. *Information Processing & Management* 34, no. 2: 161-174. http://www.sciencedirect.com/science/article/B6VC8-3TC6SNB-D/2/aee631c184c5f323e314d498e6ecaab2.

Jörgensen, Corine. 2003. Image Retrieval: Theory and Research. Lanham, MA: The Scarecrow Press.

Jörgensen, Corinne, and Peter Jörgensen. 2005. Image querying by image professional. *Journal of the American Society for Information Science and Technology* 5, no. 12: 1346-1359. doi:10.1002/asi.20229. http://dx.doi.org/10.1002/asi.2022.

Krausse, M. G. 1998. Intellectual problems of indexing picture collections. *Audiovisual Librarian* 14, no. 2: 73-81.

Lee, Uichin, Zhenyu Liu, and Junghoo Cho. 2005. Automatic identification of user goals in Web search. In *Proceedings of the 14th international conference on World Wide Web*, 391-400. Chiba, Japan: ACM.

Markey, Karen. 2007. Twenty-five years of end-user searching, Part 2: Future research directions. *J. Am. Soc. Inf. Sci. Technol.* 58, no. 8: 1123-1130.

Matusiak, Krystyna K. 2006. Information Seeking Behavior in Digital Image Collections: A Cognitive Approach. *The Journal of Academic Librarianship* 32, no. 5 (September): 479-488. doi:doi: DOI: 10.1016/j.acalib.2006.05.009. http://www.sciencedirect.com/science/article/B6W50-4KGPPGF-2/2/06a7fa1de1a119497e6d636fe2d10445.

Meister, D., and D. Sullivan. 1967. *Evaluation of user reactions to a prototype on-line information retrieval system: Report to NASA by Bunker-Ramo Corporation*. Oak Brook, IL: Bunker-Ramo Corporation.

Muresan, Gheorghe. 2009. “An Integrated Approach to Interaction Design and Log Analysis.” In *Handbook of Research on Web Log Analysis*, edited by Bernard J. Jansen, Amanda Spink, and Isak Taksa, 227–255. Hershey, PA: Information Science Reference.

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

Penniman, W. D., and W. D. Dominick. 1980. Monitoring and evaluation of on-line information system usage. *Information Processing & Management* 16, no. 1: 17-35.

Pisciotta, Henry A., Michael J. Dooris, James Frost, and Michael Halm. 2005. Penn State's Visual Image User Study. *Libraries and the Academy* 5, no. 1 (January): 33-58. doi:DOI: 10.1353/pla.2005.0011.

Rahman, M. M., S. K. Antani, and G. R. Thoma. “A Query Expansion Framework in Image Retrieval Domain Based on Local and Global Analysis.” *Information Processing & Management* 47, no. 5 (September 2011): 676–691.

Rose, Daniel E., and Danny Levinson. 2004. Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web*, 13-19. New York, NY, USA: ACM. doi:http://doi.acm.org/10.1145/988672.988675.

Searching Museum Collections. 2009. http://museumsearch.pbworks.com/

Shatford Layne, Sara. 1986. Analyzing the subject of a picture: A theoretical approach. *Cataloging & Classification Quarterly* 6, no. 3: 39-62.

Shatford Layne, Sara. 2002. Subject Access to Art Images. In *Introduction to Art Image Access: Issues, Tools, Standards, Strategies*, ed. Murtha Baca. Los Angeles: Getty Research Institute. http://www.getty.edu/research/publications/electronic_publications/intro_aia/layne.html.

Siegfried, Susan, Marcia J. Bates, and Deborah N. Wilde. 1993. A profile of end-user searching behavior by humanities scholars: The Getty Online Searching Project Report No. 2. *Journal of the American Society for Information Science* 44, no. 5: 273-291. http://dx.doi.org/10.1002/(SICI)1097-4571(199306)44:5<273::AID-ASI3>3.0.CO;2-X.

Stephenson, Christie, and Patricia McClung, eds. 1998. *The museum educational site licensing project*. Los Angeles: Getty Research Institute.

Strohmaier, Markus, and Mark Kröll. 2012. “Acquiring Knowledge About Human Goals from Search Query Logs.” *Information Processing & Management* 48, no. 1 (January): 63–82.

Svenonius, Ellen. 1994. Access to nonbook materials: The limits of subject indexing for visual and aural languages. *Journal of the American Society for Information Science* 45, no. 8: 600-606.

Wang, Peiling et al. 2006. Final report on ALISE/OCLC 2005 research grant: Mining web search behaviors: strategies and techniques for data modeling and analysis. Retrieved June 22, 2010 from http://www.oclc.org/research/grants/reports/2005/wang-p.pdf

Wang, Xuanhui, and ChengXiang Zhai. 2007. Learn from web search logs to organize search results. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, 87-94. Amsterdam, The Netherlands: ACM.

Weber, Ingmar, and Carlos Castillo. 2010. “The Demographics of Web Search.” In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 523–530. Geneva, Switzerland: ACM.

Wees, Dustin. 2013. Interview by Heather Lowe.

Wen, Ji-Rong, Jian-Yun Nie, and Hong-Jiang Zhang. 2002. Query clustering using user logs. *ACM Trans. Inf. Syst.* 20, no. 1: 59-81. http://portal.acm.org/citation.cfm?id=503104.503108#.

Wolfram, Dietmar. 2000. A query-level examination of end user searching behaviour on the excite search engine. In *Proceedings of the 28th Annual Conference of the Canadian Association for Information Science*, ed. Hope A. Olson. http://www.caisaci.ca/proceedings/2000/wolfram_2000.pdf.

Wolfram, Dietmar, Peiling Wang, and Jin Zhang. 2009. Identifying Web search session patterns using cluster analysis: A comparison of three search environments. *Journal of the American Society for Information Science and Technology* 60, no. 5: 896-910. http://dx.doi.org/10.1002/asi.21034.

Yun, Gi Woong. 2009. The unit of analysis and the validity of web log data. In *Handbook of Research on Web Log Analysis*, ed. Bernard J. Jansen, Amanda Spink, and Isak Taska, 165-180. Hershey, PA: Information Science Reference.