ENHANCING IMAGE FINDABILITY THROUGH A DUAL-PERSPECTIVE NAVIGATION FRAMEWORK

by

Yi-ling Lin

B.A. Information Management, National Sun Yat-Sen University, 2002

MBA, Information Management, National Taiwan University, 2004

M.S. Management Information Systems, University of Arizona, 2006

Submitted to the Graduate Faculty of

The School of Information Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

UNIVERSITY OF PITTSBURGH

THE SCHOOL OF INFORMATION SCIENCES

This dissertation was presented

by

Yi-Ling Lin

It was defended on

July 25, 2013

and approved by

Lora Aroyo, Associate Professor, School of Computer Science, Free University of Amsterdam

Rosta Farzan, Assistant Professor, School of Information Sciences

Michael Spring, Associate Professor, School of Information Sciences

Dissertation Advisor: Peter Brusilovsky, Professor, School of Information Sciences

Dissertation Advisor: Daqing He, Associate Professor, School of Information Sciences

Copyright © by Yi-Ling Lin

2013

ENHANCING IMAGE FINDABILITY THROUGH A DUAL-PERSPECTIVE NAVIGATION FRAMEWORK

Yi-Ling Lin, PhD

University of Pittsburgh, 2013

This dissertation focuses on investigating whether users will locate desired images more efficiently and effectively when they are provided with information descriptors from both experts and the general public. This study develops a way to support image finding through a human-computer interface by providing subject headings and social tags about the image collection and preserving the information scent (Pirolli, 2007) during the image search experience.

In order to improve search performance most proposed solutions integrating experts' annotations and social tags focus on how to utilize controlled vocabularies to structure folksonomies which are taxonomies created by multiple users (Peters, 2009). However, these solutions merely map terms from one domain into the other without considering the inherent differences between the two. In addition, many websites reflect the benefits of using both descriptors by applying a multiple interface approach (McGrenere, Baecker, & Booth, 2002), but this type of navigational support only allows users to access one information source at a time. By contrast, this study is to develop an approach to integrate these two features to facilitate finding resources without changing their nature or forcing users to choose one means or the other.

Driven by the concept of information scent, the main contribution of this dissertation is to conduct an experiment to explore whether the images can be found more efficiently and effectively when multiple access routes with two information descriptors are provided to users in the dual-perspective navigation framework. This framework has proven to be more effective and efficient than the subject heading-only and tag-only interfaces for exploratory tasks in this study.

This finding can assist interface designers who struggle with determining what information is best to help users and facilitate the searching tasks. Although this study explicitly focuses on image search, the result may be applicable to wide variety of other domains. The lack of textual content in image systems makes them particularly hard to locate using traditional search methods. While the role of professionals in describing items in a collection of images, the role of the crowd in assigning social tags augments this professional effort in a cost effective manner.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

First and foremost, I would like to thank my advisors, Dr. Peter Brusilovsky and Dr. Daqing He, for invaluable advice and patience throughout my PhD training and all the support they provided during my graduate studies. I also greatly appreciate the other members of a top-notch committee: Dr. Lora Aroyo, Dr. Rosta Farzan, and Dr. Michael Spring, who provided me with helpful comments, suggestions, and supports.

I am also very fortune to have a group of friends, Jon Walker, Erh-hsuan Wang, Claudia López, Yi-huang Kang, all the other friends and labmates of Paws lab and IRIS lab, who took time from their busy schedules to discuss with me for different problems with me and provides me alternative points of view on statistical analysis, study design, writing issues, and presentation. In addition, I want to specially thank the Carnegie Museum of Art for the use of the Teenie Harris image collection.

Last, yet the most important, I am deeply indebted to my family for their endless love, support, and encouragement. It was you - my parents, my sisters, my relatives, and my dearest fiancé, Li-Wei Lin – who gave me the strength and confidence to accomplish my PhD program.

1.0 INTRODUCTION

Currently, the most common way to find an image is using keywords on Web-based image search engines, such as Google image search[1], or using metadata in museum archives. However, there are several limitations in utilizing this existing image retrieval approach. With the rapid rise of digital image sharing online and the miscellaneous content of images, extracting relevant keywords to describe an image becomes increasingly more challenging. First, most images are without any textual content so it is hard to index images with keywords automatically as is done in the traditional text indexing. Second, manual metadata generation requires significant effort that is not practical when applied to large-scale image collections. Third, classic metadata generation relies on users being able to assign professional structured vocabulary terms under certain rules and disciplines. It is difficult for automatically generated metadata to match the quality of that created by professional indexers. Last but not least, manual indexing and automatic extraction techniques may be somewhat effective when searching for specific items with their names or titles but cannot support browsing as a self-contained access to images (Markkula & Sormunen, 2000). With the increasing volume of content and increasing pressures of time, money, and competition, improving techniques for findability (Morville, 2005) of images is getting more attention.

Cultural heritage institutions have made efforts to enhance image findability by eliciting

[1] https://www.google.com/imghp?hl=en&tab=wi

descriptors of items with subject headings, a type of controlled vocabulary, to capture the essential properties of an image. Subject headings are used to identify and generate common information (topic) facts about a given subject. Professionals usually take the "presupposition" approach to assign subject headings to state an item's *aboutness* (Hutchins, 1977) based on their perception and understanding with different field-sharing paradigms (Hjerland, 2001). The cultural heritage sector ensures that subject headings are of high quality (high degree of agreements among professionals) and to explicitly describe the *aboutness* of items using terms from the experts' point of view. However, from the general users' point of view, they usually perceive the *aboutness* of items in a different manner. According to an epistemological view of *aboutness*, general users, without professional training and experience with a field-sharing paradigm, might select different subject headings to describe an item's *aboutness* (Hjerland, 2001). When general users don't have any author's/creator's presupposed knowledge, indexers are recommended in order to adopt a "summary" approach to "*aboutness"* (Hutchins, 1977). Yet providing a comprehensive summary of an item's content without including all possible descriptors is nearly impossible to accomplish.

With the increased availability of online access and the inherent human need to organize personal information and resources, social tagging has become an alternative mean to generate different valid *aboutness* descriptors of an item by a large crowd of people. Sites like Flickr, Delicious, and others provide functionalities that allow users to organize content with free-form keywords called tags. Tags created by users are easily understood by the general public. The more people who annotate an item with diverse tags, the more likely one of these tags will come to mind for a potential searcher and increase the item's findability. It is this diversity of input sources that gives social tagging its tremendous power to assist the general users in finding

images from the users' point of view. However, some limitations might decrease the accuracy of the *aboutness* represented by social tags. For example, users might give a tag to an item before fully understanding the content or users might be affected by others when they tag an item because mass psychology often plays a critical role in how people perceive an item. Furthermore, social tags can be seen as discounting the *aboutness* of an item from the author's perspective. A hybrid approach including both experts' and general users' descriptors to representing the *aboutness* of an item might increase the completeness of the *aboutness* from diverse points of view and further enhance the item's accessibility and findability.

In the cultural heritage domain, many researchers have realized the value of social tags and proposed the integration of folksonomies with controlled vocabularies for enhancing item's accessibility and reaching users of the future (Hayman & Lothian, 2007; Rolla, 2009; Steele, 2009; Weller, 2007). Therefore, this study is particularly interested in the descriptors created by both experts and general users. This study supports end users through interfaces that provide two types of information related to an image's *aboutness*, subject headings and social tags, from experts' and general users' points of view, respectively.

1.1 FOCUS OF THE STUDY

The focus of this study is to investigate whether users will locate desired images more efficiently and effectively when they are provided information descriptors from experts and the general public (subject headings and social tags). This study develops supporting image finding through an efficient and effective human-computer interface by providing more information about the image collection and preserving the information scent (Pirolli, 2007) during the image searching

experience.

In order to improve search performance, most proposed solutions integrating these two types of descriptors, experts' annotations and social tags, focus on how to utilize controlled vocabularies to structure folksonomies which are taxonomies created by multiple users (Peters, 2009). These solutions tend to focus on a variety of approaches including algorithms (Begelman, Keller, & Smadja, 2006; Hotho, J, Schmitz, & Stumme, 2006), feature combinations such as query expansion with WordNet (Laniado, Eynard, Colombetti, & Milano, 2007; Laniado, Eynard, & Colombetti, 2007), and semantic relations between metadata and social tags (Al-Khalifa & Davis, 2007a, 2007b). However, these proposals do not actually lead to an integration of the two sources of information but rather merely add terms from one domain into the other domain without taking into account the inherent differences between the two. Presently, many websites reflect the benefits of using both professional index terms and social tags and provide multiple types of navigational support with these features on their websites for users' various information needs. These websites display different types of information representations on interfaces by placing them in different tabs or pages called a multiple interface approach (McGrenere, Baecker, & Booth, 2002). However, this type of navigational support from two types of information representations uses two mutually exclusive sources of information. By contrast, the main contribution of this study is to develop an approach to integrate these two features to facilitate resource finding without changing their nature or forcing users to choose one means or the other.

Information scent guides users' information seeking behavior through a site according to information foraging theory (Pirolli, 2007). In the Web space, users usually forage for information by navigating through pages along Web links which present proximal cues to distal targets by means of snippets of text or icons (Chi, Pirolli, Chen, & Pitkow, 2001). These

proximal cues can provide users with an orientation of how to access the distal content (resources at the end of the link) and guide them to interesting items. When the cues appear to lead users to what they are looking for, it strengthens the information scent of the site; when the cues appear to point to unrelated targets, it weakens the information scent.

Sites that fail to provide a strong information scent are subject to frustrate the users with unnecessary non-productive actions, such as back tracking (Card et al., 2001; Jimmy, 2013) which could mean that users thought the interface's information scent was directing them to a target item but when they got there, they failed to find the target item and had to back track to search in a different area. However, in the exploratory mode, back tracking could also be the process of exploring the variety of items in the collection. Futile search, when a query is issued and no item is returned, is an additional unnecessary non-productive action caused by misleading information scent either failing to provide any scent-bearing keywords or leading the person to believe an incorrect scent-bearing keywords would be affective. Sites that provide a strong information scent are good at guiding users to the distal content they require. Based on the notion of information scent, I assume that users' navigational choices are not random, and the website with a strong information scent is good at guiding users to content.

Some design principles (Nielson, 2003; Spool, Perfetti, & Brittan, 2004) show that strong information scent can help users quickly and easily identify the best option available from the site, lead users successfully to work their way straight to their desired content, and give users confidence to use (click) the cues (links) to reach their goal. The principles around creating stronger information scent are to provide diverse hints (information descriptors) multiple meaningful and useful accesses points to information with more contexts. However, providing

feature-rich interfaces may not always be efficient and may lead to information overload for many users (Findlater & McGrenere, 2007; Marchionini, 1993).

Driven by the concept of information scent, I designed and carried out an experiment to explore whether the images can be found more efficiently and effectively when multiple access routes with two contexts of information descriptors (subject headings and social tags) are provided to users in a combined manner, the dual-perspective navigation framework. In light of the definitions of findability (Morville, 2005), this study contributes to the enhancement of image findability by supporting people's ability to find their way to their desired items in an efficient and effective manner. Effectiveness, efficiency, and satisfaction of the ISO 9241 standard for usability (1998) are adopted to evaluate the usability of the dual-perspective navigation framework. According to the ISO definition of usability, effectiveness is the accuracy and completeness with which users achieved specified goals, efficiency is the resources required in relation to the accuracy and completeness with which users achieved the goals, and satisfaction is the positive attitudes toward the user of the product (ISO, 1998).

1.2 RESEARCH QUESTIONS

Figure 1 shows the focus of this study "Whether images can be found more efficiently and effectively when users perform search with the dual-perspective navigation framework?" following by three main research questions (RQ) and the corresponding hypotheses (H) of each research question.

- RQ1: Does the dual-perspective navigation approach provide better information to help users achieve their goals in an effective manner than the single-perspective navigation approaches?
- RQ2: Does the dual-perspective navigation approach guide users to their targets with fewer resources required than the single-perspective navigation approaches?
- RQ3: Does the dual perspective navigation approach make users confident of their finding ability and have a positive perception of the approach?

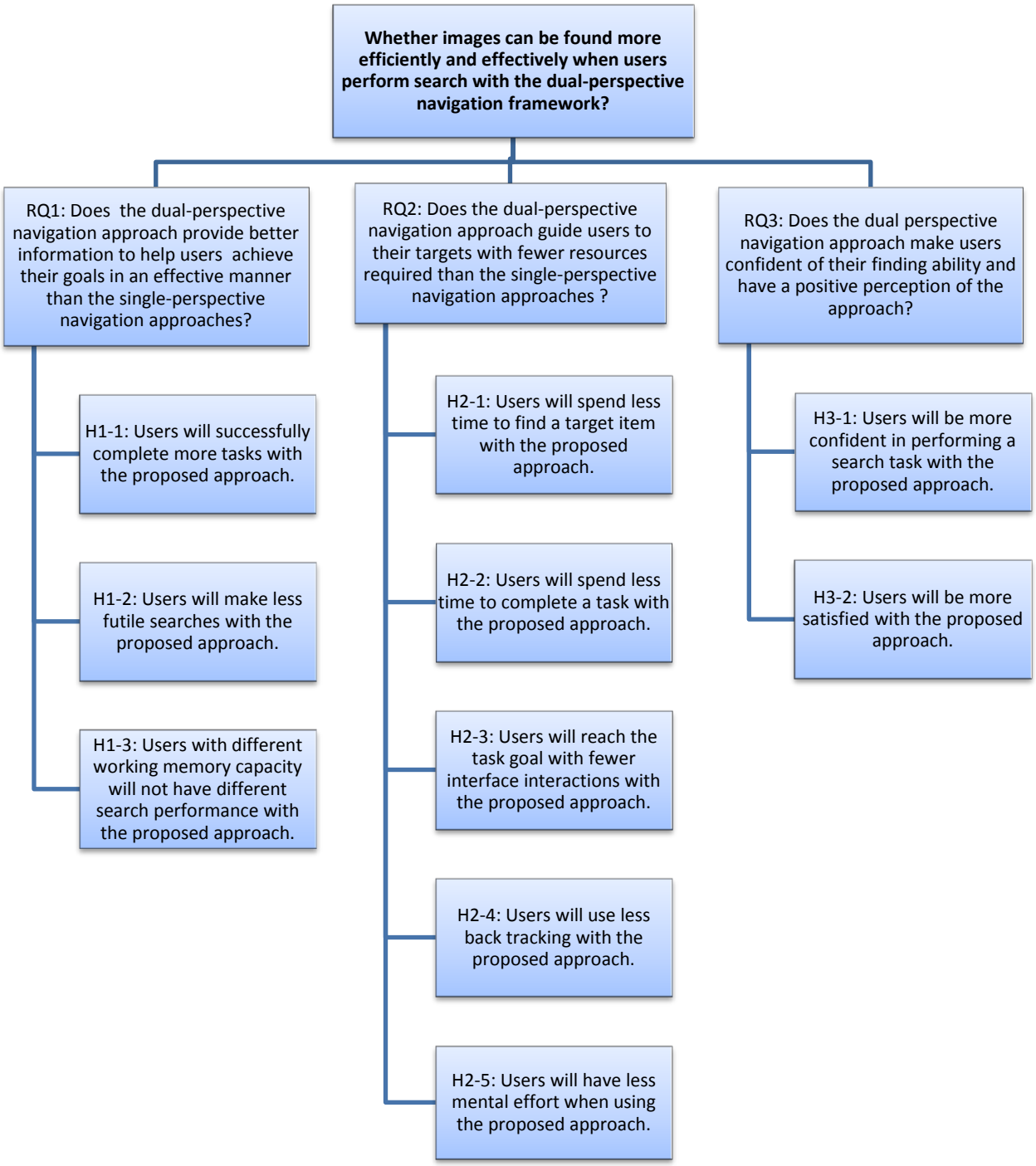


Figure 1. The research structure

1.3 LIMITATIONS AND DELIMITATIONS

Although this study was designed with careful consideration, a few limitations of this study need to be acknowledged. The first limitation relates to the image collections used in this study. One of the collections, the Teenie Harris collection, has limited subject headings and tags. The subject headings of this collection can be only classified into three facets: people, activities and objects. The location facet is missing because all of the images in this collection were taken in the greater Pittsburgh area and were annotated with limited location info. Since this collection is not published on any social platform such as Flickr and the museum exhibits' official website did not provide any tagging functionality to collect general users' tags, the tags used for this collection were created by recruiting Mechanical Turkers. In contrast to the typical social tags, created by the users in a community, the taggers were recruited from the Amazon Mechanical Turk. This might be a concern when I claim any effectiveness of social tags. However, the crowdsourcing tags were able to enhance image findability according to our preliminary study (Trattner, Lin, Parra, & Brusilovsky, 2012). This study has another collection crawled from the Flickr website and the social tags of the Flickr collection are generated with the general users in the Flickr community.

Second, the Teenie Harris collection only contains black-and-white pictures. Compared to the Flickr collection, users might not be able to search pictures with the color attribute. However, it only affects color relative accesses, but won't influence the main goal to explore users' information seeking behavior. Besides, the Flickr dataset contains both black-and-white and color images that might make this study's results more relevant to other image collections.

The third limitation is that the studied interfaces may not be innovative in creating all kinds of functionality representations. The studied interfaces are limited to facet browsing and

tag-cloud representations. Although there are many different interfaces available to test the goals of this study, I chose facet browsing and tag-cloud representations because they are the most popular applications in current practice and related research areas. The interface selection is informed by studies of the search interface literature of different resources.

In addition, to fulfill the purpose of this study, I applied three interfaces focusing on single-perspective and dual-perspective navigation support from experts' and general users' descriptors. To make interfaces comparable, I implemented the three interfaces based on some existing studies on displaying subject headings and social tags instead of applying some platforms in current practice, such as Flickr or Google image search as my baseline. However, I was unable to test all potential positions to display subject headings and tags in the interface. To make the three interfaces display these features in an identical manner and consider the space of the interface, I designed facet browsing with subject headings on the left hand side of the screen and the tag cloud on the top middle of the interface. To compensate the space issue, I provided the 'show more' function for each feature. An additional limitation of the study interface was that all three were created by the author of this study. However, the dual-perspective navigation interface only used elements that were also in the other interfaces. The subject heading-only interface and the tag-only interface had been tested in my preliminary studies and were found to be more effective than the traditional keyword search interface.

The fifth limitation is that this study applied both a within-subject and between-subject design. Although a within-subject design is an optimization to test interface differences among all the conditions with less variance, it was hard to conduct an experiment with twelve conditions within the two-hour time limit. In order to control for three interfaces, two search types, and two collections, this study conducted the experiment that each participant was tested under each

interface with each search type in only one collection (3x2x1) and the between-subject design was at the collection level (i.e. Each subject only worked with images from one of the collections).

Furthermore, this study was only conducted with users from the Pittsburgh area. The participants' age ranged from 18 to 45 (40% 21-25 and 33% 26-30). Most of them are students at local universities. Although the unit of analysis in this study is primarily on individuals, some of the results from this study may be unique to this population.

1.4 DEFINITION OF TERMS

1.4.1 Findability

Findability is defined in (Morville, 2005) as 1) the quality of being locatable or navigable, 2) the degree to which a particular object is easy to discover or locate, 3) the degree to which a system or environment supports navigation and retrieval. The speed of accessing or locating users' interested items can be used to measure findability in IR domain.

1.4.2 Taxonomies

Taxonomy is a hierarchical classification system that establishes parent-child, or broad-narrow relationships between terms. Taxonomy as defined by (Garshol, 2004a) is that a taxonomy is a subject-based classification that arranges the terms in the controlled vocabulary into a hierarchy. It provides more information about concept. For example, knowledge representation is part of

knowledge management, and there are several topic maps under knowledge representations such as XTM, TMQL, TMCL, and HyTM. Here, taxonomy is defined as a hierarchical classification system.

1.4.3 Subject headings

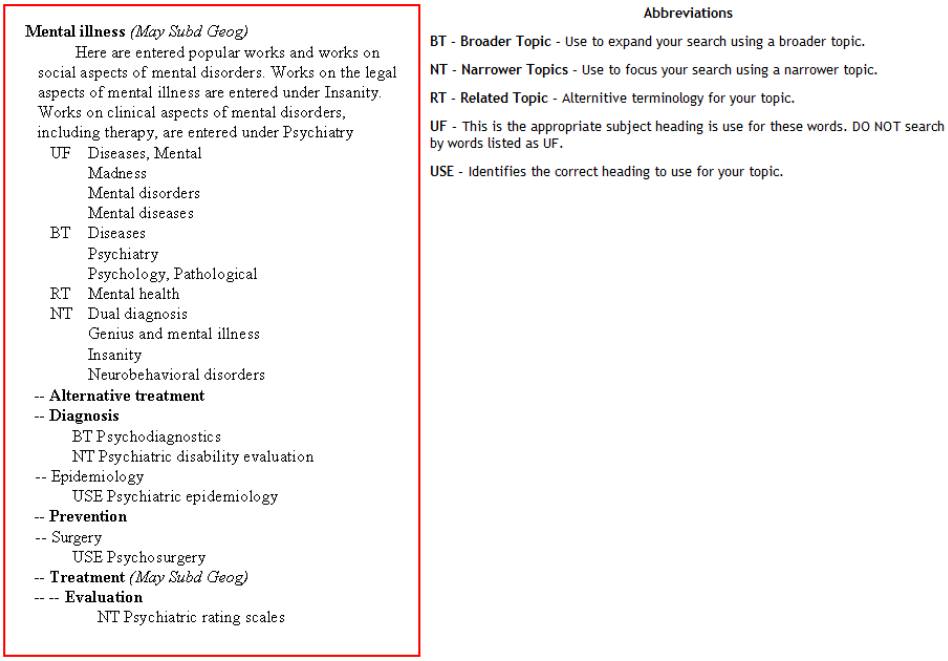
Subject headings are a type of controlled vocabulary used to capture the essence of the topic of a subject in order to take the guesswork out of searching by applying a single term in IR (University of Mississippi Libraries, 2004). By applying controlled vocabulary, all objects about the same topic can be assigned to one heading with consistent subdivisions and relevant headings. There is no duplication of topics in subject headings under the generating rules. It can be used to complement traditional title search for some items with poor title descriptor. For example, a book titled ”Have a seat, please? Barbers? Furniture?” could be about barbers or furniture, but it is actually related to capital punishment. In this case, the subject headings could be assigned to the book and allow users to find this book by its topic whereas its title is ambiguous.

The Library of Congress determines the vocabulary of subject headings (over 270,000 usable headings, and over 490,000 headings that are used for cross referencing) used by the majority of libraries in the United States since 1898 (University of Mississippi Libraries, 2004). In the Library of Congress’s system[2], subject headings are grouped by types: topical, form, chronological, and geographic. The Library of Congress provides different thesauri for different types of resources. The Thesaurus for Graphic Materials[3] (TGM) is designed for indexing visual materials by subject and genre/format. There are more than 7,000 subject terms used to index

[2] http://www.loc.gov/

[3] http://www.loc.gov/rr/print/tgm1/

topics shown or referenced in images, as well as 650 genre/format terms applied to index types of photographs, prints, design drawings, ephemera and other categories. There are new terms being added in the TGM continually. TGM allows users to search through the Prints and Photographs Online Catalog (PPOC). Figure 2 shows the reference structure includes unauthorized terms (UF), broader terms (BT), narrower terms (NT), and related terms (RT) under each authorized subject heading. For each relationship in the thesaurus, terms are in reciprocal relationship. In order to improve the efficiency of using subject headings, the Library of Congress chooses only one to be the official subject heading.



Mental illness *(May Subd Geog)*
Here are entered popular works and works on social aspects of mental disorders. Works on the legal aspects of mental illness are entered under Insanity. Works on clinical aspects of mental disorders, including therapy, are entered under Psychiatry
UF Diseases, Mental
Madness
Mental disorders
Mental diseases
BT Diseases
Psychiatry
Psychology, Pathological
RT Mental health
NT Dual diagnosis
Genius and mental illness
Insanity
Neurobehavioral disorders
-- **Alternative treatment**
-- **Diagnosis**
BT Psychodiagnostics
NT Psychiatric disability evaluation
-- Epidemiology
USE Psychiatric epidemiology
-- **Prevention**
-- Surgery
USE Psychosurgery
-- **Treatment** *(May Subd Geog)*
-- -- **Evaluation**
NT Psychiatric rating scales

Abbreviations

BT - Broader Topic - Use to expand your search using a broader topic.

NT - Narrower Topics - Use to focus your search using a narrower topic.

RT - Related Topic - Alternitive terminology for your topic.

UF - This is the appropriate subject heading is use for these words. DO NOT search by words listed as UF.

USE - Identifies the correct heading to use for your topic.

Figure 2. The sample of a subject heading in the LCSH

1.4.4 Social tags

Social tags, also known as social annotations or collaborative tags, have widely been used for personal information organization and retrieval since the social bookmarking system, del.icio.us, was launched in 2003. In most social applications of Web 2.0, users are allowed to annotate free-form keywords on resources such as photos, websites, blogs, and videos. In spite of personal incentives, (Zollers, 2007) pointed out that taggers are motivated by a number of personal and social incentives simultaneously during the tagging process. The study of (Zollers, 2007) showed that opinion expression, performance and activism are emerging social motivations for tagging as well.

1.4.5 Folksonomy

Folksonomy, a portmanteau of the words folk and taxonomy, is a term coined by Thomas Vander Wal (Vander Wal, 2007). It refers to a taxonomy generated by users applying their personal free-format tags to information and objects for their own future retrieval. This kind of user generated taxonomy stands in contrast to professionally created ontologies and taxonomies for discovering and retrieving resources in traditional knowledge organization systems. The value of the folksonomy is derived from that people are allowed to use their own vocabulary to give a resource an explicit meaning, which could infer to each user's own understanding of the information/object.

1.4.6 Aboutness

Aboutness is a term coined by R. A. Fairthorne (Fairthorne, 1969), which is widely used in library information science, linguistics, philosophy of language and philosophy of mind. “Aboutness” has been considered synonymous with subject. The common view of the “aboutness” of a collection is that information associated with a collection of documents on the site, given clues into the aboutness of that collection. Users will get the overall semantic picture of the collection and use the provided information to find their target need better.

2.0 LITERATURE REVIEW

This chapter presents topics that are relevant for exploring how to enhance image finding experience through an interface support based on the theories: aboutness and information scent. This review of relevant literature is divided into the following four sections:

2.1 Aboutness. This section provides an overview of the context of aboutness in which information retrieval is situated, positions this study within the contemporary area, and discusses the implications for this study.

2.2 Information scent. This section summarizes the concept of information scent from the information foraging theory and reviews the design principles around how to create stronger information scent cues. Implications for this study are also provided.

2.3 Indexing from subject headings to social tags. The section first discusses the advantages and disadvantages of subject headings (one type of controlled vocabulary categories) and social tags as indexes in the information retrieval sector. Then, reviews the work devoted to integrating controlled vocabulary categories and social tags in both academia and current practice. Contrary to the contemporary research and applications, this dissertation develops an approach without changing different descriptors' nature or forcing users to choose one means or the other.

2.4 Search interfaces. This section reviews the work devoted to facilitating search through interfaces in which image search interfaces and exploratory search interfaces are

the essential parts. Particularly, this study reviews two technologies in the section of exploratory search interfaces, faceted browsing and tag-based navigation, for the purpose of representing structured vocabulary and social tags together in the proposed framework.

2.1 ABOUTNESS

Aboutness is a term coined by R. A. Fairthorne (Fairthorne, 1969), in library and information science (LIS), and is widely used in linguistics, philosophy of language and philosophy of mind. It has been considered synonymous with subject. The common view of the "aboutness" of a document is that the index descriptors assigned to the document represent or indicate what the document is about. R.A. Fairthorne (1969) discussed aboutness with "intentional aboutness" and "extentional aboutness". "Intentional aboutness" includes the holistic notions of the total document and its purpose, while "extentional aboutness" constitutes the individual elements of a document such as paragraphs, heading, and general syntactic style.

In the information retrieval sector, subject analysis is a complex segment in which an indexer judges whether a document is (or is not) about some given subject. It is usually hard to say precisely what actually happens when we make a judgment of aboutness. Hutchins proposed an alternative concept of "aboutness" with a combination of linguistic and discourse analyses of a document to build a sounder foundation for indexing (Hutchins, 1977). Hutchins introduced themes and rhemes to define the concept of aboutness stating, "The thematic part of the text expresses what the text is 'about', while the rheme expresses what the author has to say about it" (Hutchins, 1977). Hutchins therefore drew attention to the notion that the process of subject

analysis is highly multi-faceted and the indexing process is often influenced by the subjective perception of the work in question.

Maron and Studies (1977) offered an operational definition of 'about' which interprets the term in relation to search behavior with three types of aboutness: S-about, O-about and R-about. S-about (subjective about) is a relationship between a document and the inner experience of the user. O-about (objective about) is a relationship between a document and the assigned index descriptors. R-about is a generalization of O-about to a specific user community. Maron further constructed a probabilistic model of R-aboutness to support users' information need.

Hjørland (1992) also pointed out that "aboutness" is hard to capture objectively. He asserted that "Neither the author's, the reader's, librarian's information specialists, not any other person's (for example the publisher's) point of view or subjective can have any certain objective knowledge about the subject of a document, nor defined the concept of subject." Along with this line of reasoning, we could argue that no process of subject analysis could ever be considered totally objective or unbiased. Aboutness is best captured when we use multiple diverse perspectives to present it.

This study of the dual-perspective navigation framework for supporting image finding in a more effective and efficient manner is situated within the broad context of the aboutness paradigm. It practices the dual-perspectives of aboutness from experts' and general users' opinions to explore interface support on the image finding.

2.2 INFORMATION SCENT

Information foraging is an important concept emerging from the human computer interaction (HCI) domain since 1993 (Pirolli, 2007). It uses the analogy of a wild animal searching food to analyze how users gather information online. The World Wide Web (Web) environment could be referred to as a jungle that is full of different resources. Users like animals forage for information in the Web environment. They have to constantly decide what to look for, whether to stay or leave, which path or link to follow, whether find different information instead, and when to stop the search.

The most essential concept in the information theory is information scent that refers to the descriptors or use of cues, such as Web links. Information scent plays an important role in guiding users to find their desired items and also providing users a sense of the content. With a list of options provided, only the clearest indication (strongest scent) can quickly lead users closer to the information they require. Sites with weak information scents will cause users to spend longer time to evaluate the options and then users have very high chance to abandon the search.

To give users the clearest indication (or strongest scent) on the search interface, many websites tried to avoid weak information scents by listing content in multiple places; however, critics argue that this approach leads to a cluttered site very quickly. Another simple solution is providing users with extra context to give them more chances to select a good option. Yet, questions remain regarding how much more context a designer has to provide and whether they would overwhelm users.

Nielson and Spool, among others, have put effort into recommending how to enhance information scent on the Web. Nielson (2003) used the metaphor of animal spoor, a track of an

animal, to present the concept of information scent as cues. Users estimate a given hunt's likely task from the spoor by assessing whether the cues on the path relate to their needs. Nielson (2003) proposed guidelines on increasing information scent and fostering faster interaction including:

- Ensure that links and category descriptions explicitly describe users' desired outcome
- Don not use made-up words or your own slogans as navigation options
- Provide feedback about where the user is in the site's hierarchy

Spool, Perfetti, and Britten (2004) investigated the scent-following behavior of Web users and things that block information scent using the framework of information foraging as a base for providing guidelines for Web design. They found that users searched for a scent trail and followed it toward their content. Specifically, "As the scent got stronger, they grew more eager. When they lost the scent, they backtracked until they picked it up again." A sample of the "Tao of Scent" assertions and guidelines is derived from empirical studies from Spool et al. (2004).

- The design communicates information scent through links.
- Trigger words lead users to click and reach the destination page. If the destination page contains the trigger words, scent is the strongest.
- Links work best when they are between 7 and 12 words long which is the perfect length to attract users' attention and be easily found.
- Links should lead to information that is more specific.
- Links should accurately describe what the next page contains.
- Avoid jargon and cute marketing terms that might confuse users.

These principles provide guidelines to designers on how to increase user visits and to increase the success rate of finding a user's need. Although these guidelines can help designers create a site with strong scents, conducting usability testing alongside the design of the site is still essential to evaluate the real impact of stronger information scents. How to support users with the clearest indication (or strongest scent) on the search interface is still unknown. How can we design an interface that provides both experts' and general users' understanding (interpretations) of large collections of information items? How does one build an interface that successfully supports perceptual cues with subject headings and social tags to assist users in finding the information they require?

This dissertation addresses on the aforementioned issues. Although feature-rich interfaces provide multiple accesses points to information for users and require less working memory to process, they may not always be as efficient to use as simple interfaces and may be overwhelming for many users (Findlater & McGrenere, 2007; Marchionini, 1993). To balance the ease of learning with access to powerful features for users, this study proposes a dual-perspective navigation framework demonstrating that images can be found more efficiently and effectively when two types of information descriptors (subject headings and social tags) are provided to users in a combined manner.

2.3 INDEXING FROM SUBJECT HEADINGS TO TAGS

To retrieve information effectively, most retrieval systems represent the items in the collection with index terms to compare with the words (queries) provided by the user and find the matched items for his/her needs. Index terms retrieve documents in an information system by capturing

the essence of the aboutness of a document or an item. They are created by manual subject indexing or automatic indexing with sophisticated methods of keyword extraction.

2.3.1 Subject headings

Subject headings, one type of index terms, are designed to be a controlled vocabulary for capturing the *aboutness* of a subject in order to take the guesswork out of searching by applying a single term in information retrieval (IR) (University of Mississippi Libraries, 2004). The Library of Congress Subject Headings (LCSH) comprises a controlled vocabulary of subject headings, hosted and maintained by the professionals of the United States Library of Congress since 1989. Libraries and the cultural heritage sector have a long history of using LCSH to facilitate access to their materials and collections from an expert's point of view.

Although subject headings provide many benefits and opportunities for searching or browsing, index terms created by applying controlled vocabularies have shown limited adequacy for online resource discovery (Macgregor & Mcculloch, 2006). First, this is one of the fundamental obstacles which prevents wide deployment of controlled vocabularies and restricts access to professionally indexed resources by general users (Trant, 2009). Second, professional index terms limit users' expression of their needs to a set of professional terms that users might not understand hindering users from generating efficient search queries (Furner, 2007). Third, the carefully crafted subject headings defined by professionals are often misunderstood or incomprehensible to general users who lack the professionals' domain knowledge. Last but not least, anticipating the ways users will try to navigate to a resource is also challenging. Different users might issue different queries to look for the same resource. Unless the resource contains all possible index terms, the resource remains difficult to find by different users applying

idiosyncratic search queries. The terminology gap between experts and novices has long been criticized. Users have displayed difficulty selecting the correct terminology from sets of index terms defined by experts. To bridge the gap, social tags have been suggested as a possible solution. Social tagging, beyond the vocabulary constraints, is considered a useful way to facilitate resource discovery and knowledge organization in the World Wide Web.

2.3.2 Social tags

The Social Web incrementally enhances the end user's ability to create and access information. Sites like Flickr, Delicious, and others have functionality that allows users to organize content with free-form keywords (called tags). This user-generated content provides valuable evidence for learning the taxonomy that general users in a community apply to organize the content. The distinction between social tags and subject headings is that users are the main indexers who assign free-form keywords to information resources based on their own preference and understanding of the resources. The social tags, commonly used to organize information within the personal information space, are also capable of being shared with other users and then providing further search and navigational support. From the point of view of IR, social tags are the common language between the indexer and the searcher since taggers are simultaneously indexers and searchers. Several studies have proved that social tags can be used to improve search performance and support users to access information more efficiently (Begelman et al., 2006; Bischoff, Firan, Nejdl, & Paiu, 2008; Kammerer, Nairn, Pirolli, & Chi, 2009). Since users could quickly adapt tags according to the changes in their needs and vocabulary, social tags are able to reflect emerging concepts and assign terms to the resources. Social tags are the index terms general users can understand easily. Although researchers have devoted time to creating

subject headings to help search, subject heading searches are still used less frequently than general keyword searches (Rolla, 2009).

In addition, creating a set of social tags is relatively inexpensive compared to creating a set of subject headings. Researchers have indicated that the creation of professional index terms requires considerable time and expertise to implement and also is difficult to maintain (Cardinaels, Meire, & Duval, 2005; Garshol, 2004b; Greenberg, 2004; Park & Lu, 2009). Additionally, scaling a framework of subject headings is expensive (Duval, Hodgins, Sutton, & Weibel, 2002). A large effort has to be made for each appropriate term assignment (Heymann & Garcia-Molina, 2006). The economic factor emphasizes the potential importance of applying social tags to facilitating finding items. However, with the nature of freedom that users could annotate according to their preferences and knowledge, social tags also present a number of challenges because of its potential to be noisy, shallow, ambiguous, inconsistent, and sparse (Peters, 2006; Plangprasopchok, Lerman, & Getoor, 2010).

Many researchers have realized the value of social tags and proposed the integration of folksonomies with controlled vocabularies (Weller, 2007). Macgregor and Mcculloch (2006) emphasized that librarians and information professionals have to learn to respect social aspects of information consumption and to engage users with information management. Rolla (2009) also pointed out that tags could be used to enhance subject access but not entirely replace controlled vocabularies like LCSH. Steele (2009) stated that using both traditional professional index terms and social tags is a valid method for reaching users of the future. Another similar statement from Hayman argued that tags and standardized subject headings are not mutually exclusive (Hayman & Lothian, 2007).

2.3.3 Integration

Most proposed solutions integrating diverse vocabularies, social tags and professional index terms, focus on how to utilize controlled vocabularies to structure folksonomies from different points of view, algorithms (Begelman et al., 2006; Hotho et al., 2006), feature combinations such as query expansion with WordNet (Laniado, Eynard, Colombetti, et al., 2007; Laniado, Eynard, & Colombetti, 2007), and semantic relations between metadata and social tags (Al-Khalifa & Davis, 2007a, 2007b) to improve search performance. These proposals do not integrate the two sources of information; they merely add terms from one domain into the other domain without considering the inherent differences between the two domains. For example, social tags, which are inherently non-hierarchical, are often forced into the hierarchy of an existing ontology created by professional indexers. For instance, Etsy, an e-commerce site, sells one-of-a-kind craft items. Etsy predefines a set of general product categories for its site and generates subcategories for its site navigation based on tags provided by users; however, the Etsy team manages the tags to ensure that subcategories are created from valid, high quality tags. Etsy has both predefined categories and social tags to facilitate the item findability, but it still needs human power to classify tags into their predefined hierarchical structure.

Currently, many websites have noticed the benefits of both professional index terms and social tags and provide multiple types of navigational support using these features on their websites for users' various information needs. These websites display different types of information representations on interfaces by placing them in different tabs or pages called multiple interface approach (McGrenere et al., 2002). Each time users can choose one type of navigational support to discover the resources on one interface and switch to the other through a specific link; however, the navigation support from two types of information representations are

exclusive to each other. For example, Technorati, a blog aggregator, uses tags to discover trends across the blogosphere. By using the multiple interface approach, Technorati provides several ways for users to navigate the site by browsing the directory, top 100, tags, people and so on (See Figure 3). However, the navigational support from two types of information representations is mutually exclusive.

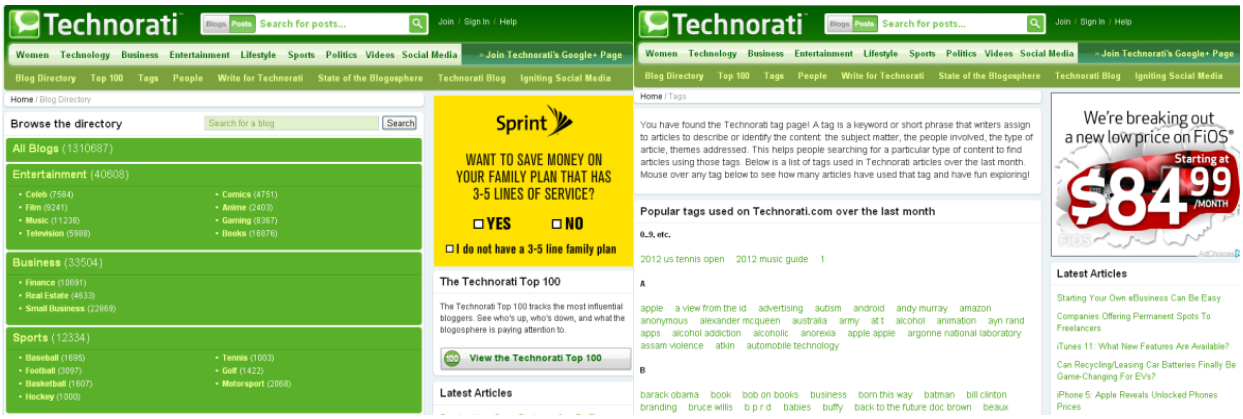


Figure 3. The Technorati interface: (a) browsing by the directory (b) browsing by tags

By contrast, this study develops an approach to integrate these two features to facilitate resource finding without changing their nature or forcing users to choose one means or the other. This study addresses offering a variety of indexing terms that cover different types of users' perspectives: subject headings created by professionals with specific domain knowledge, and social tags created by a crowd without any specific domain knowledge or training. To integrate the two sources from experts and the general public, the approach proposed in this study represents a more complete description of the items in the collection, which will allow users to successfully search for and retrieve the desired information.

2.4 SEARCH INTERFACES

Search is becoming an essential part of everyone's online life. Nowadays, people use search engines for dealing with a wide range of needs and desires on a daily basis. To fulfill an enormous variety of information needs, researchers investigate diverse interfaces to support users with a better search experience. Diverse research topics surrounding user search interface span query specification, clustering/grouping retrieval results, navigation of information collection, query reformulation, search personalization, etc.

2.4.1 Image search interface

Currently, the most familiar image search is that used by Web image search engines which are based on manual annotations (metadata) or automatically extracted information surrounding an image, such as web page content. In that, users can issue keyword terms to retrieve relevant images back. However, there are several limitations of utilizing this kind of text-based image retrieval approach. First, manual metadata generation requires significant effort which is impractical when applied to large-scale image collections. Second, it relies on users being able to assign professional structural vocabulary under certain rules and disciplines, which increases the difficulty for automatic metadata generators. Third, with the rise of digital image sharing online and rapid increase in images with miscellaneous content on the Web, extracting relevant content of images becomes increasingly more difficult. Last but not least, these systems are often effective for searching specific items but not supportive for browsing and exploratory tasks well (Markkula & Sormunen, 2000).

To overcome these limitations, content-based image retrieval (CBIR) was proposed in

early 1990s (RyumT., Huang, Ortega, & Mehrotra, 1998). On the collection with limited textual content or without annotations, this approach has been widely used with various visual properties, such as color, texture, and shape, for searching images (Markkula & Sormunen, 2000; Vasconcelos, 2005). The approach calculates the similarity between the images based on the features of the image context and return ranks of object images according to their relevancy of the images in the query (Jiang, Wan, Zhang, & Zhou, 2008; Marques & Furht, 2002). Recently, Google has a new content-based image search application called 'Google Goggles' for Android mobile phones, which enables users to upload a photo of a target object and get more information about the object (Jamaal, 2010). However, this kind of image retrieval techniques is not as friendly as text-based image search for users. It aims to return highly similar image content using feature comparisons but because the result is not guaranteed to be semantically relevant to the query issued by the end-users, they might not intuitively understand why the image is in the result set.

By engaging users during the search process, relevance feedback techniques are applied to improve the performance of image retrieval. Users are able to provide their preferences from their feedback to refine the query iteratively based on the previous search results and improve search performance (Lew, Sebe, Djeraba, & Jain, 2006). The interaction between users and the system leads researchers to learn more about users' behavior during the search process in order to enhance the search performance and users' search experience. Spink and Greisdorf (2001) proposed a three-dimensional spatial model including levels of relevance, regions of relevance and time dimension of relevance to support user interactive search for text retrieval. Campbell (2000) focused on the time dimension, and proposed the Ostensive Model (OM) including four ostensive relevance profiles (decreasing, increasing, flat and current profiles) to indicate the

degree of relevance relative to when a user marked relevant information from the results set. Ruthven et al. (2003) adapted two dimensions from Spink et al.'s model, regions of relevance and time, for ranking query expansion in the text-based IR and applied OM to the time dimension with iteratively feedback. Liu et al. (2009) introduced an adaptive four-factor user interaction model (FFUIM) for content-based image retrieval including relevance region, relevance level, time and frequency.

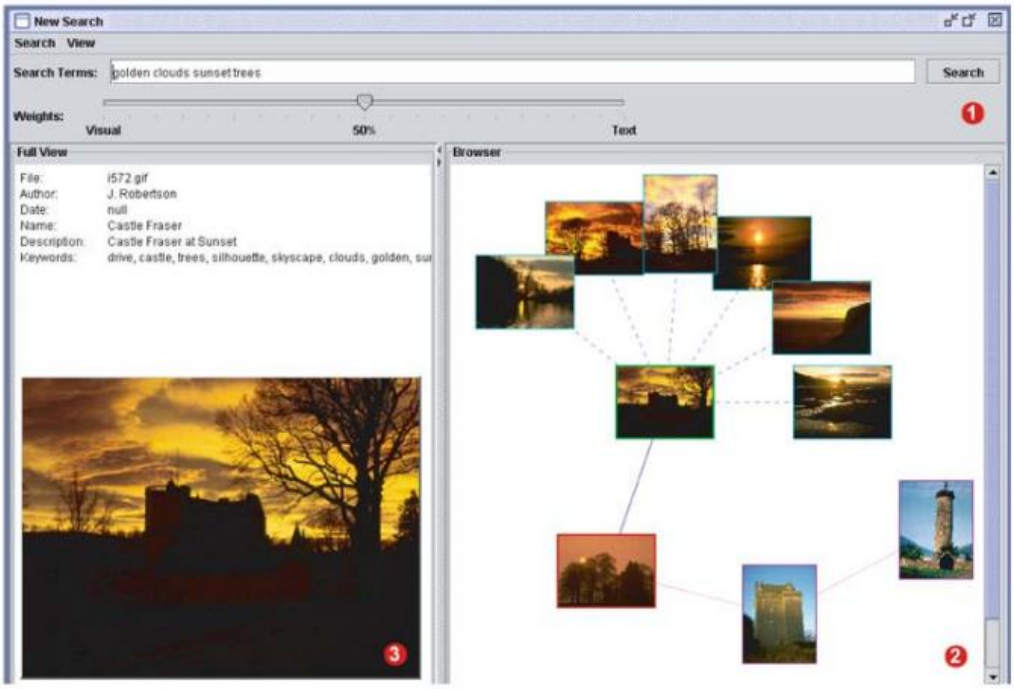


Figure 4. An interface based on the ostensive model

Some interactive user interfaces have been developed to deliver the interactive models visually and to further improve the user interaction. Heesch and Rüger (2003) proposed a query-by-example search interface with adjustable search weights for the multitude of image content. Urban et al. (2006) utilized OM to build a browsing-based image search system by applying a

dynamic tree view to display the query path and results (see Figure 4). Another personalized image search and management tool proposed by Urban and Jose (2006) allowed users to provide feedback by grouping the results (see Figure 5). Liu et al. (2009) proposed an interactive CBIR retrieval system, uInteract (see Figure 6), to deliver the FFUIM and to allow users to manipulate the model effectively.

Figure 5. The Ego interface

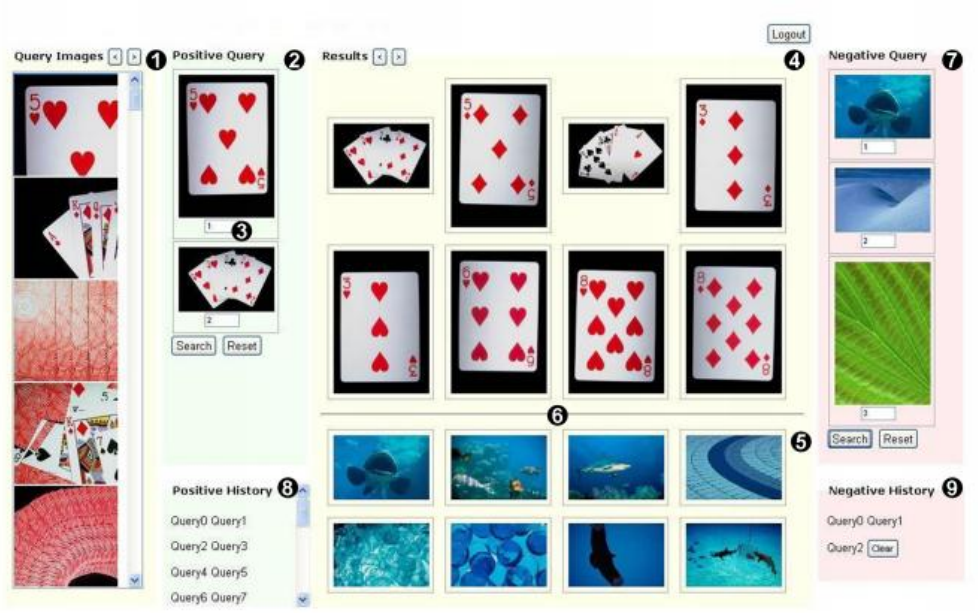


Figure 6. The uInteractive interface

2.4.2 Exploratory search interfaces

In the interactive image search, information seeking tasks involve various levels of exploration in line with different users' contexts which can be very different between users (Liu, Little, & Ruger, 2011). People are different and naturally have different searching start points when they initiate any search task. Exploratory search is emerging to support information seeking by providing more finding guidance during the seeking process (White & Roth, 2009).

Exploratory search is hard to define, but one definition is that it is a search requiring both querying and browsing strategies to meet users' information goals (Marchionini, 2006; White & Roth, 2009). In the aspect of finding an optimal path to an information resource, White and Marchionini (2007) pointed out exploratory search is related to information foraging theory (IFT) (Pirolli, 2007). Over search processes, several user-oriented factors such as information

goals, searching strategies, and information usage, are highly considered in the exploratory search. Mulholland et al. (2008) have shown that IFT can interpret the effectiveness of the exploratory search technologies and their findings are a concrete step toward supporting the exploratory search.

Technologies for exploratory search have been developed to support navigation (or browsing) as part of the search process when a user's goal is more exploratory and less directed. Several methods integrating navigation structure with the keyword search have been proposed including category systems (flat, hierarchical, and faceted), TOC views, and automated clustering techniques. Of particular interest in this study is faceted browsing and tag-based navigation support. The following sections will review the faceted browsing and tag-based navigation for supporting exploratory search in more detail.

2.4.2.1 Faceted browsing

Faceted browsing emerged as an attractive alternative to "text box" search in the exploratory search context (Hearst, 2006; Karlson, Robertson, Robbins, Czerwinski, & Smith, 2006; Marchionini & Brunk, 2003; Yee, Swearingen, Li, & Hearst, 2003). It was an expansion of an older hierarchical browsing paradigm that was considered an alternative to Web search. With hierarchical browsing, users navigate a single extensive hierarchy to narrow their choices. Both Yahoo directory[4] and Open Directory Project[5] are examples of classic hierarchical browsing. Faceted browsing integrates browsing with the classification of objects along several dimensions called facets. With faceted browsing, users progressively narrow down the list of results, making choices in several taxonomies that classify different aspects of the objects of interest. The

[4] http://dir.yahoo.com/
[5] http:// dmoz.org/

presence of these multiple facets allow the users to search more flexibly and to specify their interests more precisely than one dimension of classification. To further guide the users' choices and help them make sense of results, modern faceted browsing interfaces such as Flamenco (M. A. Hearst, 2006; Yee et al., 2003) or Relation Browser (Capra & Marchionini, 2008; Marchionini & Brunk, 2003) display query previews, which show the number of documents available for every facet category.

In research on the Flamenco project, Hearst et al. (M. A. Hearst, 2006; M. Hearst, 2008; M. Hearst et al., 2002; Stoica, Hearst, Berkeley, & Richardson, 2007) proposed faceted classification systems for web site navigation (see Figure 7). The Flamenco project aimed to support flexible navigation, seamless integration of browsing with directed keyword search, fluid alternation between refining and expanding, avoidance of empty results sets, and allowing the user to retain a feeling of control and understanding. Meanwhile, the Flamenco project also aimed to promote the idea of faceted navigation in online systems, as an alternative to the hierarchical focus of Web site structure. The term "faceted" was chosen by this project to reflect the underlying spirit of the idea from library science. Ranganathan (1933) is often credited with introducing the idea with his Colon Classification System, which suggested describing information items by multiple classes, and Bates (1988) advocated for faceted library catalogue representations in the 1980's. However, Ranganathan didn't focus on how to use such systems in user interface in 1933, and Bates's 1988 work was restricted to TTY-based interfaces. Another noted system, the Dewey Decimal system, often used in local libraries, has aspects of facet analysis by combining multiple categories into one description string (Maple, 1995), but it does not allow for the flexible application of ordering and combination of categories that online faceted navigation affords.

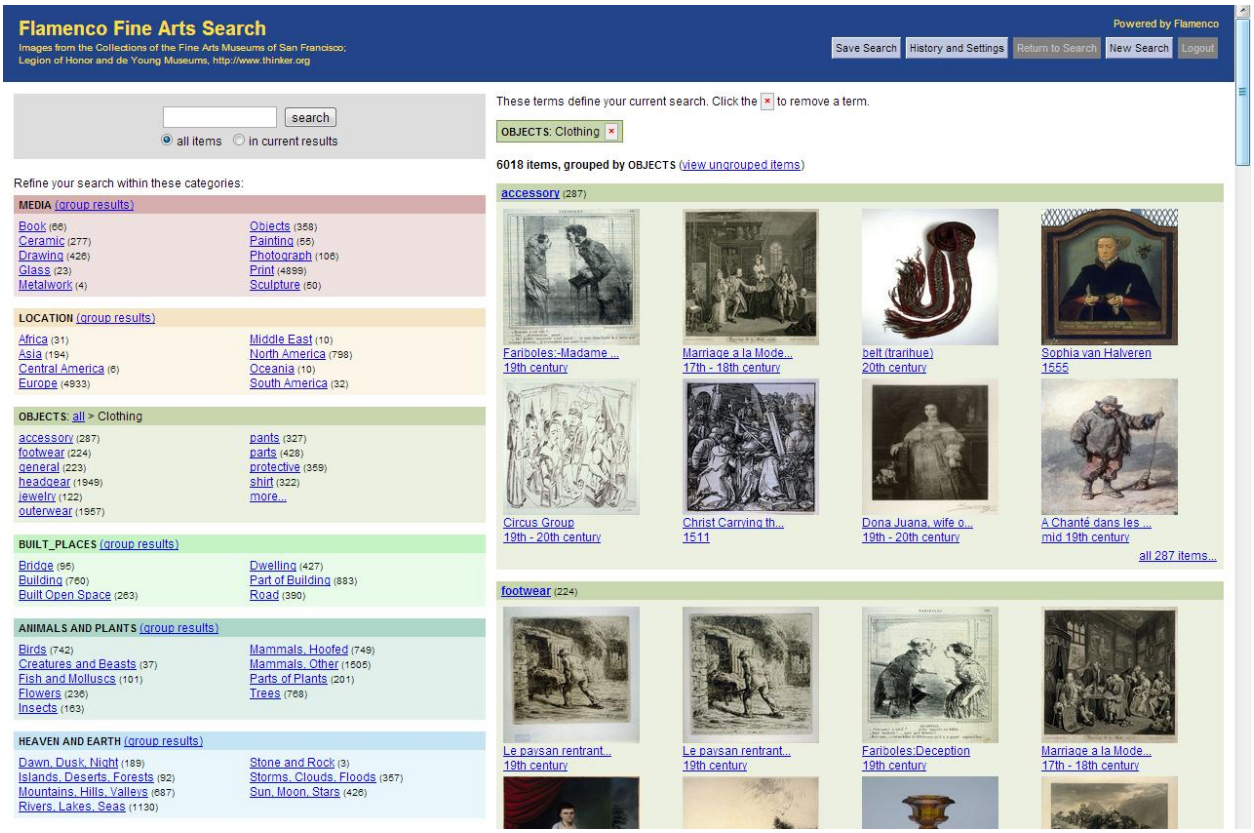


Figure 7. The Flamenco interface

Shraefel et al. (2006) proposed a general interaction model and software framework, mSpace (see Figure 8), by supporting multiple ways of exploring the information itself. Within mSpace the user navigates the content through a series of information slices representing different perspectives of content with levels of the hierarchical relationship moving from the left to right. For example, for a classic music resource, the first "Era" slice has restricted the instances that appear on the second Composer column to those composers of the selected era. The slices scaffold the user in searching and navigating the content even if unfamiliar with the vocabulary of the classical music domain. mSpace is built on exploratory search technologies and uses metadata of the content to construct the information slices.

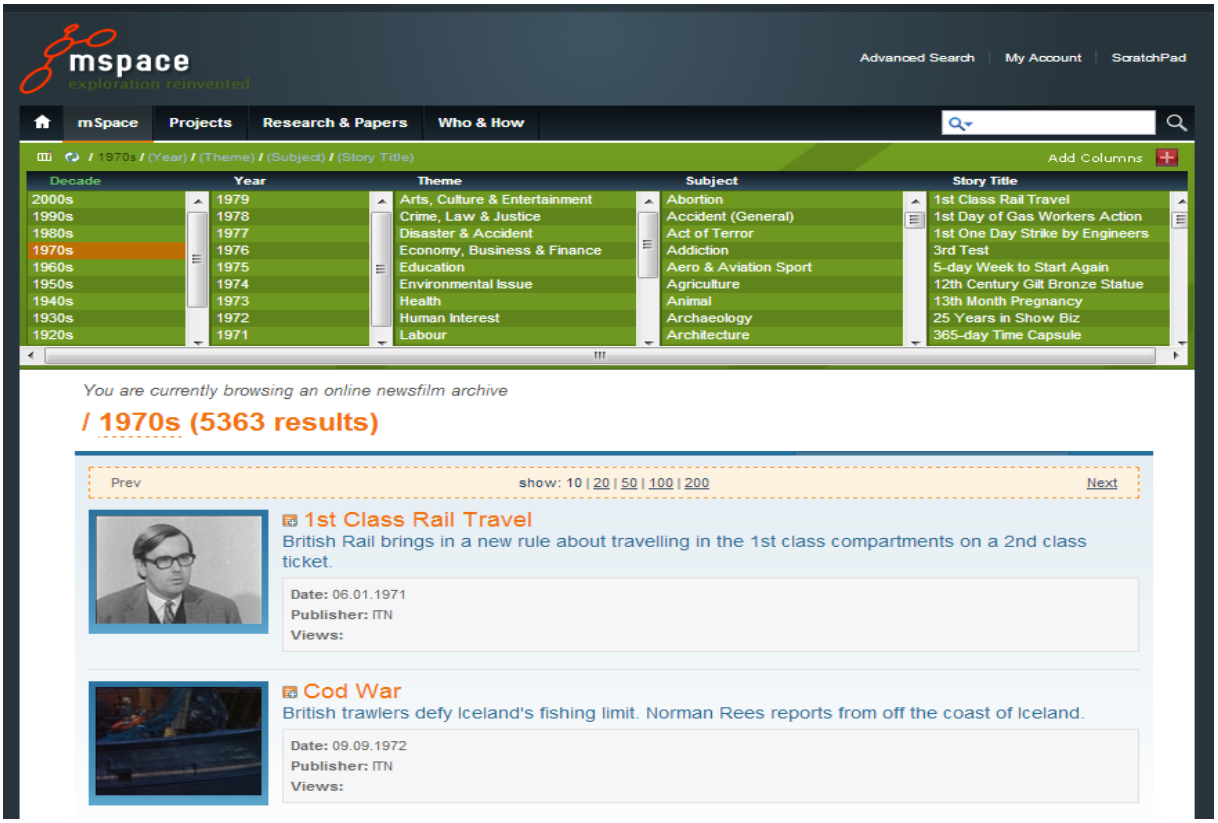


Figure 8. The mSpace interface

Faceted browsing interfaces have been shown to be helpful and preferred by users over the traditional search interface (Yee et al., 2003). However, its application in its standard form was limited to domains where objects of interest are classified along several dimensions of metadata – e.g., price, year, brand, and other object-specific aspects. Thus, classic faceted search cannot replace traditional search in domains where multiple classification facets are not established, or where the objects are not classified along multiple facets. To resolve this problem, we conducted a pre-study to automatically generate a specific faceted search interface. We explored the specific faceted search interface with name entities, called ImageSieve, for exploring a collection of images using associated textual descriptions. The details will be addressed in Chapter 3.

2.4.2.2 Tag-based navigation

The availability of social tags has greatly enhanced access to information. Social tags are unlike other information representations used for navigation. There is no parent-children structure, no hierarchy, and no relationships within tags, as well as no categories or facets to anchor information in many cases (Smith, 2008). Tag clouds emerged as a new "social" way to find and visualize information with a snapshot of the "aboutness" of a tagged collection and a simple one-click access to it. A large volume of research has investigated diverse tag artifacts for information access such as tag clouds (Bateman, Gutwin, & Nacenta, 2008; Rivadeneira, Gruen, Muller, & Millen, 2007; Venetis, Koutrika, & Garcia-Molina, 2011), clustered/classified tag clouds (Hassan-montero & Herrero-solana, 2006; Knautz, Soubusta, & Stock, 2010; Zubiaga, García-Plaza, Fresno, & Martínez, 2009) and tag hierarchies (Candan, Di Caro, & Sapino, 2008; Helic & Strohmaier, 2011; Trattner, Körner, & Helic, 2011).

Many Social Web sites, such as Flickr, Delicious, and others have the distinct advantage of adopting tags to support users' Web browsing and navigation. Figure 9 shows the results of a search on the tag "flower" from the Flickr website. In this view, the most recent pictures are shown and the related tags are shown beside the result set. A cluster function is provided as well. By selecting the function, users can view the results in different groups including nature, macro, color, types as shown in Figure 9.

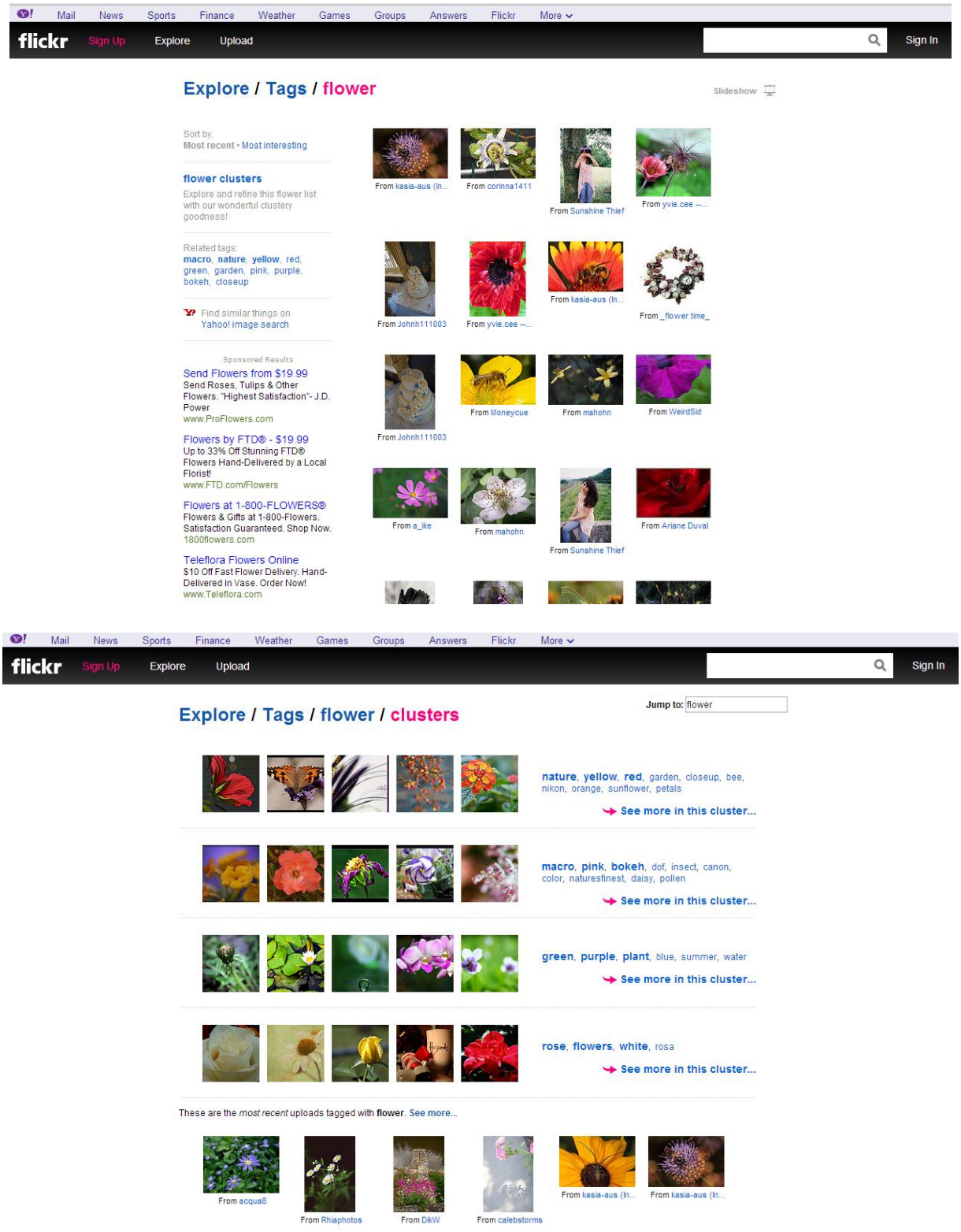


Figure 9. Flickr output on the tag flower (top) and cluster output on the tag flower (bottom)

The majority of research on tag information access focused on an information- or network-theoretical approach to evaluate the quality of different tag constructs in terms of search and navigation but they ignored the user side of the interface. There is a lack of user studies comparing users' performance by applying different tag-based browsing constructs in a set of realistic search tasks, along with the effectiveness of various tag constructs against simple search-based access to tagged collections. Therefore, we attempted to bridge this gap by comparing several types of tag-based information access in a controlled user study. The details will be addressed in Chapter 3.

3.0 PRELIMINARY ANALYSIS

3.1 INTRODUCTION

The goal of this study is to investigate whether dual-perspective information descriptors, subject headings and social tags, can support users to find their desired items more efficiently and effectively. To capture the essence of the usefulness of descriptors provided by experts and users with their corresponding representation, two pre-studies investigated different combinations to enhance image findability. In the first pre-study, we discovered the effectiveness of applying faceted browsing with expert-related descriptors (named entities extracted from image description generated by experts) for exploring images. In the second study, we examined whether social tags (image annotations created by general users) are helpful to support users in finding images and compared the performance of applying different representations of social tags to enhance image findability. The summary of the overall studies is shown in Table 1.

The result of the first pre-study showed us that faceted browsing was useful for enhancing image findability in a large-scale image collection. Facets did give the user more insights about the collection and provide somehow meaningful descriptors (not weak scent) to guide the user to the desired images easily. The reason we claimed that facets provide somehow meaningful descriptors is that some of misclassified named entities in facets might mislead our participants sometimes. Some of the participants ignored the misclassified named entities and

moved on with others but some were frustrated after several misleading entities and gave up. The idea of using structural content descriptors along with faceted browsing was pointed out in the first pre-study in order to provide a better navigational support to users.

Table 1. The summary of the studies

Study	Content aboutness providers		Information representation		Focus of the study
	Expert	User	Facets	Tag cloud	
Exploring images with named entity - based faceted browsing	Named entities (experts' image description)		Faceted browsing		What is an effective representation to enhance image findability with museum provided content?
Tag-based information access in image collections		Social tags	Faceted tag clouds	Traditional tag cloud	Can user provided content be used to help image finding? What is the best representation of it?
Final study	Subject headings	Social tags	Faceted browsing	Traditional tag cloud	Whether images can be found more efficiently and effectively when users perform search with the dual perspective navigation framework?

The second pre-study informed us that we could use social tags created by general users to facilitate image finding. By comparing user performance with different tag-based interfaces, a

simple tag-cloud interface outperformed other interfaces. According to the participants' comments, the faceted tag-cloud interface gave them more information about the content than other interfaces, yet they felt distracted by the multiple similar representations activated in parallel and with misclassified tags. A simple tag-cloud interface was the best representation for social tags because it allows the users to easily catch the importance of tags with different font size in an alphabetic order.

This chapter begins by summarizing key observations made on the faceted browsing system, imageSieve, which can help users to narrow down a variety of search criteria and find related images more efficiently in the exploratory search. Following by applying social tags on the search/browsing interface, users could find images with tag-based interfaces in an effective way and had better performance with a simple tag-cloud interface than with other compared interfaces (search only and faceted tag cloud interfaces). After learning the effectiveness from different interface designs, the dual-perspective navigational support approach will be introduced in the next chapter.

3.2 EXPLORING IMAGES WITH NAMED ENTITY-BASED FACETED BROWSING

Faceted browsing emerged as an attractive alternative to a traditional "text box" search in the exploratory search context (Hearst, 2006; Karlson, Robertson, Robbins, Czerwinski, & Smith, 2006; Marchionini & Brunk, 2003; Yee, Swearingen, Li, & Hearst, 2003) and have been shown to be helpful and preferred by users over the traditional search interface (Yee et al., 2003).

Faceted browsing can be successfully integrated with traditional search forming faceted search. Faceted search replaces the traditional ranked list of results with a faceted browsing interface, providing a superior approach to make sense of and explore search results. Combining the features of search and browsing, faceted search emerged as a strong alternative to the "text box" and became a de-facto standard for interaction on multiple e-commerce sites. Instead of using predefined category structure or metadata as shown in modern e-commerce or objects of art in Flamenco (Yee et al., 2003), this study proposed a novel way to automatically generate a specific faceted search interface. We explored the specific faceted search interface with experts' description of images by extracting name entities (NEs), which are the words or phrases referring to names of people, places, organizations, etc. We presented an implementation of named entity-based faceted browsing in ImageSieve, an experimental interface for exploring a collection of images using associated textual descriptions.

3.2.1 ImageSieve

The faceted browsing system, ImageSieve (Figure 10), offers several benefits: (1) The search results become more transparent to the user – the most critical information in the form of NEs contained in hundreds of retrieved images was brought to the surface. This helps users to make sense of the search results. (2) The system uncovers critical people, locations, and organizations relevant to the users' tasks by showing the main NEs related to the user's original search terms, even with no additional functionality, the extracted NE can help users to formulate new queries. (3) The system allows users to narrow down the list of retrieved results by progressively focusing on specific NEs of interest.

Figure 10 shows the ImageSieve interface loaded with the Teenie Harris Collection of the Carnegie Museum of Art[6]. A brief scenario can demonstrate how the interface works. Imagine a user is searching for some images about political history. The user starts with a query "president" and the system returns a set of images that have matching descriptions. The result set was presented in a traditional information retrieval style: 10 items per each page with image ID, image thumbnails, and textual surrogates. The textual surrogates are descriptions about the images' content. When a user clicks on the thumbnails or IDs, a new window pops up with higher resolution image display, so that the user can visually examine it in more detail. The user can further explore these results using panels in the control area on the right hand side of the screen: (1) Query Term Panel, (2) Named Entity Panel, and (3) Shoebox Panel. The Query Term Panel shows each term in the current query accompanied by the number of images in the result list containing the respective term. Users can turn a query term filtering on or off by clicking on it. When a query term filter is turned on (this is the default state indicated by term highlighting), the images in the result set are updated and all items without the selected term are filtered out. When a term filter is turned off, all relevant items will be shown whether or not the term exists in the description of an image. For example, if a user turns off a filter "banquet" from the original query "president banquet", the new result list increases from 12 to 172 documents. The number of images increases because the Boolean post-filtering was reduced from two terms ("president AND "banquet") to one ("president"). The updated number of documents is redisplayed next to the term in the Query Term Panel.

The Named Entity Panel is the core feature of the system. The system extracts and displays NE from the list of images' descriptions on the left hand side of the interface. The NEs are organized into four "editor's W" tabs according to their types. The size and color of the

[6] http://web.cmoa.org/

displayed NEs are determined by image frequency. More frequently occurring NEs in the retrieved images are rendered in a larger font and brighter color than less frequent ones. This is ImageSieve's domain-adapted analogy for showing the number of matching documents for each category in traditional faceted search. NEs remain unselected waiting for the user to examine and select them based on the user's preference. When the NE filter selection is complete, the user clicks the "Apply Filter" button constituting one faceted browsing step. In response, the system returns an updated image list. The updated list is post-filtered from the original list and includes only the images that contain all of the selected names.

Starting from the situation displayed in Figure 10, the user examines the NE list, selects the important location name "Pittsburgh" and clicks "Apply Filter" to narrow down the current retrieval list. When the filter is applied with "Pittsburgh", the number of documents in the list is reduced to 19, and the list of NE is updated accordingly. The user examines the updated NE list and decides to look for images with the name "Kennedy". After selecting the NE "Kennedy" and applying the filters again, only 7 images remain in the list to be examined by the user in detail. The selected filters can be turned off again anytime, so that the search process using the NE filters is as flexible as possible.

To help users remember which NE filters are turned on within the four tabs, the number of selected NEs is displayed and the tab background changes to yellow. This also follows traditional faceted browsing, although our approach to show the list of selected features is different from the design recommendation. The label of the active tab, Who, in the screenshot is rendered in red (foreground) and dark yellow (background), because the user selected the NE "Kennedy". From the Where tab label, we can see there is another selected name within that tab, a location name "Pittsburgh". In order to distinguish itself from the active tab, the background is

rendered in light yellow. Below the box, all selected NEs are displayed in a smaller font size, followed by the count, giving the user an overview of the exploration process outcomes.

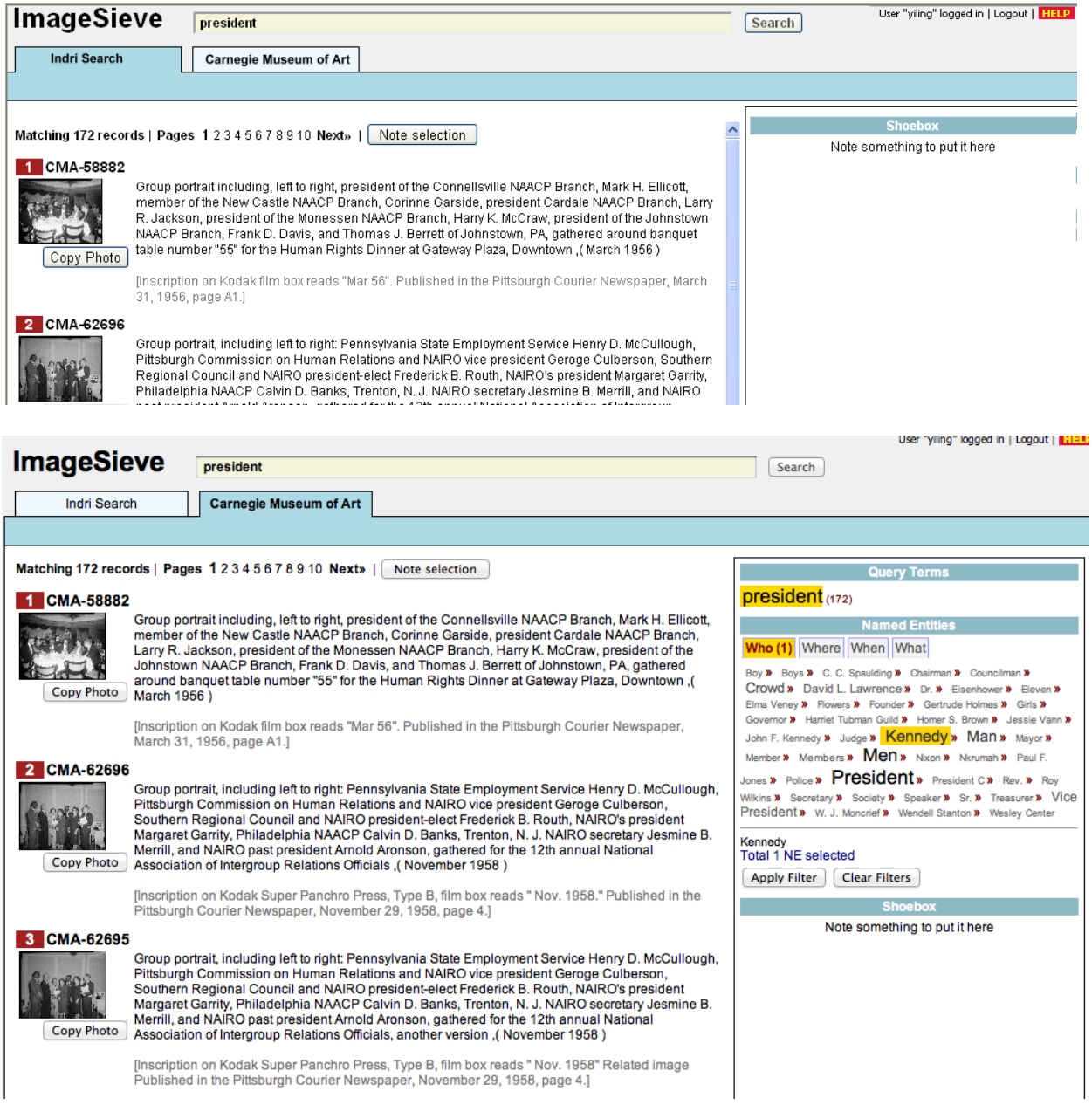


Figure 10. ImageSieve Interface (the baseline is on the top, and the experimental is on the bottom)

To produce a multi-faced NE exploration interface, we used an NE detector developed by IBM (Florian, et al., 2004). It is based on a statistical maximum-entropy model and can recognize 32 types of named, nominal, and pronominal entities (such as PERSON, ORGANIZATION, FACILITY, LOCATION, OCCUPATION, etc.) and 13 types of events. The IBM extractor has a very important feature – it can distinguish different forms of the same entity within and across the documents. For example, it can resolve that the pronoun "he" indicates "President Kennedy" in a specific image description. At the same time, it can give a consistent ID to the entities that have same meanings. It can tag the entities like "JFK", "John F. Kennedy", and "Mr. Kennedy" as "PERSON: JOHN_F_KENNEDY" across multiple documents. This information is very valuable to calculate the exact frequency distribution of a specific entity across the entire corpus and helps to improve ranking the search results.

3.2.2 Experimental design

The goal of our study was to assess the usefulness and the value of ImageSieve faceted search for finding images. The study was designed to answer the research question: "What is an effective representation to enhance image findability with museum provided content?" Our hypotheses was that ImageSieve, a named entity-based faceted browsing, would represent the museum content to users in an effective way so that the users would be interested to use it and get its help to achieve higher performance in image finding.

The study compares two systems. The experimental system was a full-fledged version of the ImageSieve interface. There was also a baseline ImageSieve system in which the filtering functionality and named entity viewer were disabled. To make a fair comparison, we worked

with the museum curators to develop several realistic task scenarios. Working on each scenario, users are expected to collect images that fulfill the scenario's requirements.

3.2.2.1 Participants

Sixteen participants were recruited from the University of Pittsburgh's School of Information Sciences (SIS) to participate in the experiment. To ensure that participants fit the profile of information professional, we only recruited participants who are graduate-level information science students with training in information access (i.e. a course in information retrieval.) Eight of the 16 participants were from Library and Information Sciences Graduate Program, and the other eight were from Graduate Program in Information Science. Three of the 16 participants were female, seven participants were native speakers and all participants fell into the age range from 25 to 55.

3.2.2.2 Procedure and Tasks

The experiment was conducted during one 90-minute session, consisting of a 5-minute introduction of the experiment, a 10-minute training session for each of two systems, a 5-minute break before real tasks, two 20-minute search tasks for two different topics with another 5-minute break between two tasks, 10 minutes for post-task questionnaires, and a 5-minute post-session interview. The training topic scenario was focused on Jazz events in the Pittsburgh area, and the main task topics selected for the study were related to sport and political events. Each main task contains two subtasks. The first subtask of the sport task was about local Pittsburgh baseball teams, and the other was about Pittsburgh professional baseball in 1960s. The first subtask of the political tasks requires finding images of U.S presidents who have visited the

Pittsburgh area, while the second subtask was focused on the images of racial or minority rights activists.

The order of tasks and systems was randomized to avoid possible learning effects. At the end of each search task (including training tasks), participants were required to print out collected images and associate each image with the part of task requirement fulfilled by the image. While participants were working on the tasks, the system logged their activities and performance for each assigned task. After each of two "real" tasks, participants completed a post-task questionnaire to assess their level of satisfaction with the systems used. Finally, after both tasks were completed, a brief interview was conducted with participants to assess their views on the experimental system's features.

3.2.3 Evaluation results

The impact of the system on user performance and satisfaction was accessed by both log analysis and the analysis of user responses to the questionnaires. The log analysis attempted to compare general patterns of user activities as well as their performance on the task (i.e., how well the users were able to collect required images). To assess the performance, all images collected by the users were processed by two human annotators who examined the relevancy of retrieved images to task requirements. Among the 350 images retrieved by 16 participants in our experiments, 183 were judged as relevant to sport task and 167 are for political task. To answer our main research question "What is an effective representation to enhance image findability with museum provided content?" we got the findings (F) as following.

$F_{1\text{-}1}$: Participants used ImageSieve's named entity exploration functionality in their exploratory searches.

- Participants used the filter 342 times in total, and around 20 times on average. Among 18 users, 15 used the filters more than 10 times during the search sessions.
- Participants clicked on the tabs 459 times in total (over 20 times per user). The most frequently clicked tabs were Who (262) and What (84). Where (13) was the least used tab in the sport task, while When was the least used tab in the politics task.
- Participants switched tabs, which may reflect their interest in the NEs and the activities to locate relevant entities, with What (31) and Where (24). Even though "Who" occupied the minimum number (10), this tab was displayed initially by default.

$F_{1\text{-}2}$: A search system equipped with named entity exploration functionality could better support users in finding relevant information.

- The performance of the baseline was very low, less than 0.25 for both rank 5 and 10. ImageSieve showed almost double the precision scores than the baseline (paired Wilcoxon signed rank test henceforth, $p<0.01$).
- The items opened when the NE filter was applied showed that ImageSieve had significant higher improvement than the baseline ($p=0.026$)
- The precision of the saved images showed that ImageSieve had significantly better performance than the baseline (0.66), whether overall (0.74; $p=0.044$) or when the NE filter was used (0.79; $p<0.01$).

$F_{1\text{-}3}$: Users were satisfied with the interface with named entity exploration functionality in their exploratory search.

- Users had positive opinions to ImageSieve. Thirteen of the 18 participants noted that larger font sizes for higher ranked named entities, grouping named entities by

"editor's 4W", and NE filtering were all very helpful in locating important information.

- Users also reported that misclassified named entities occasionally distracted their attention.

3.2.4 Summary and discussion

ImageSieve, a named entity-based faceted browser, is a novel approach to support users' navigation in a large-scale image collection. The goal of this interface was to help users in their search processes including sense making, query formulation, and manipulating search results. We were able to demonstrate that ImageSieve was actively used and positively evaluated by the participants. ImageSieve was able to help users save more relevant images than the baseline system. ImageSieve improved the ability of the search system by bringing relevant documents to the surface and attracting users' attention to them. With relatively simple text processing approaches, ImageSieve can automatically generate a faceted browsing interface that can deliver significant performance improvement over traditional search. Although some of named entities were misclassified, users still strongly agreed that NE did give them more ideas to refine their search and get more precise results for uncertainties especially when they were not familiar with the topics.

3.3 TAG-BASED INFORMATION ACCESS IN IMAGE COLLECTIONS

The majority of research on tag information access focused on an information- or network-theoretical approach to evaluate the quality of different tag constructs in terms of search and navigation but, in most cases, they ignored the user side of the interface. There is a lack of user studies comparing users' performance by applying different tag-based browsing constructs in a set of realistic search tasks, along with the effectiveness of various tag constructs against simple search-based access to tagged collections. This study attempted to bridge this gap by comparing several types of tag-based information access in a controlled user study. The study was performed in the context of image search where the presence of tags is known to be most valuable. To increase the value of the study, we compared the performance of three types of information access interfaces in two commonly recognized types of search tasks – a lookup search and an exploratory search. The tag-based interfaces explored in the study include a search-based interface that plays the role of a baseline and two types of tag-based browsing interfaces: a regular browsing interface using traditional tag clouds and a faceted browsing interface using classified tag clouds. We selected the faceted tag cloud interface from among other advanced tag-based browsing approaches because our previous study (Lin et al., 2010) in the image search domain revealed that faceted search interfaces helped users to better explore large collections of images.

3.3.1 Experimental design

The goal of this tag study was to investigate the effectiveness and value of three interfaces (with and without tag-based browsing support). The study was designed to answer the research

question: “Can user provided content (tags) be used to help image finding? What is the best representation of user provided content to support image finding?” Our hypotheses are that tags are helpful to support users to find images. Tag-based interfaces will be more supportive than the baseline search interface in terms of assisting image finding.

The work compared three tag-based information access interfaces. We designed a within subject study to compare their effectiveness. In this design, each of our participants evaluated the three different search interfaces during one study session. To determine when tag-based support is most effective; each interface was examined in the context of two kinds of search tasks.

3.3.1.1 Interfaces

We implemented three interfaces as illustrated in Figure 11, the baseline search-only interface, tag cloud interface and faceted tag cloud interface, to support users during the search of the Teenie Harris images in two different types of search tasks. As a baseline, we utilized a simple search box-based interface that offers the look and feel of well-known search engines. We provide our users with a search box to issue a query, a thumbnail preview of the resulting images sorted by relevance, and the functionality to click on the image in order to get a detailed view of the large-size image resource. Apache Lucene[7] is the back-end search engine that utilizes the tags of each image to create the search index.

The third interface is a faceted tag cloud interface that can be considered one of the most innovative tag-based search interfaces currently available. Yahoo first introduced the interface in 2009 (Sigurbjornsson & Zwol, 2010) in order to search for images in the social tagging system, Flickr. Although there are very few implementations of this type of interface, there is a great deal of current research in this area (Bischoff et al., 2008; Böhnstedt, Lehmann, Rensing, &

[7] http://lucene.apache.org/java/docs/index.html

Steinmetz, 2011; Overell, S., Sigurbjörnsson, B., and Zwol., 2009; Wartena, 2010). Similar to the tag cloud interface, this type of interface provides users with the functionality to view the tags of the retrieved images in a visually appealing representation. In contrast to the traditional tag cloud interface, where all tags appear in a tag cloud in an unstructured way, this interface classifies tags into several categories.

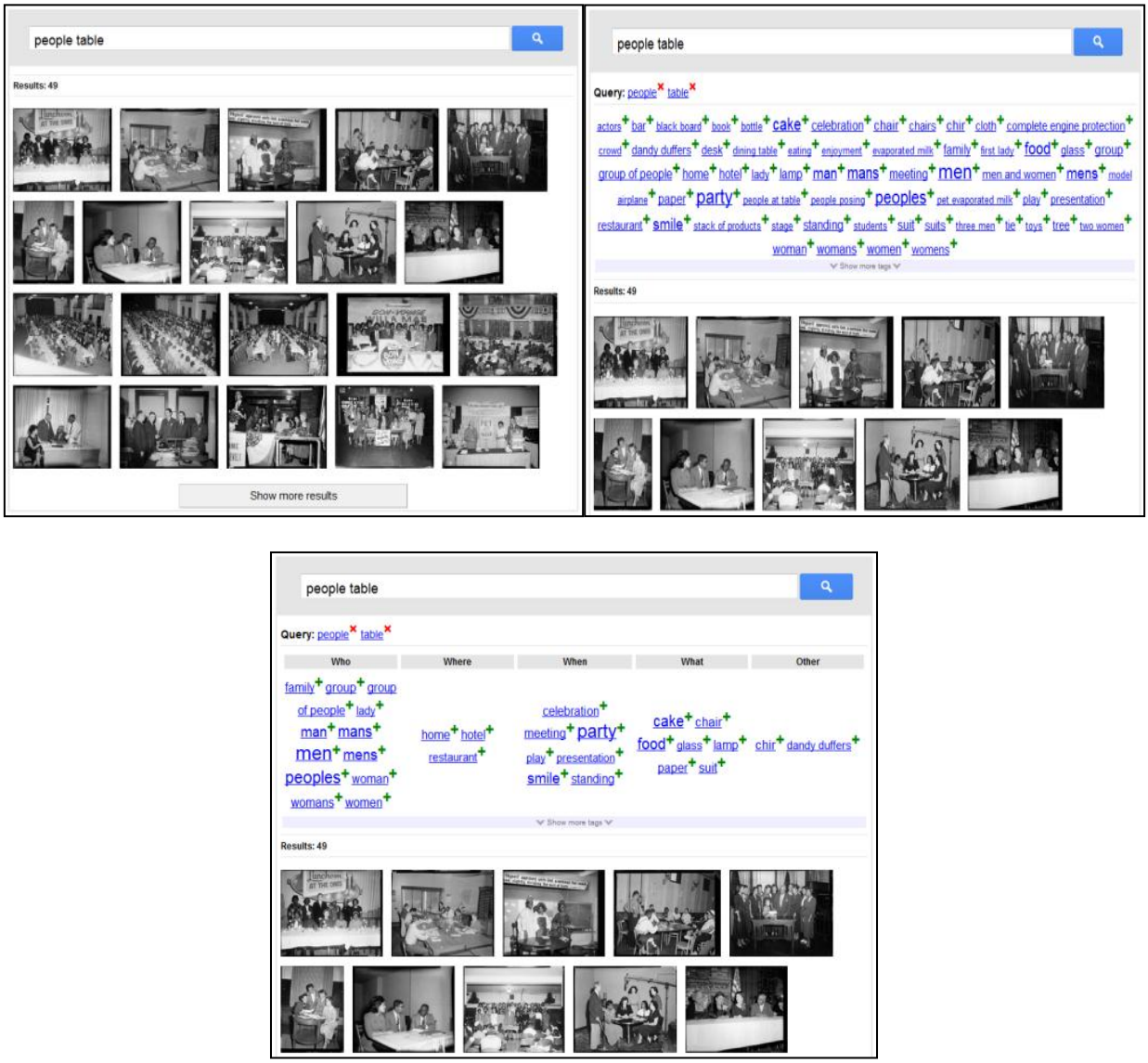


Figure 11. Three types of search interfaces: baseline (top-left), tag cloud (top-right), and faceted tag cloud (bottom)

To decide which classification schema to utilize, we performed an extensive literature survey on currently available tag classification approaches (Böhnstedt et al., 2011; Cantador, Konstas, & Jose, 2011; Overell, S., Sigurbjörnsson, B., and Zwol., 2009; Sigurbjornsson & Zwol, 2010; Wartena, 2010). At the end, we selected "Who" (people, groups or individuals), "Where" (location or places), "When" (time, activities or events), "What" (objects, food, animals or plants) and "Other" (unknown, not classified) classification schema. This schema was found to be effective in classifying tags in the image domain (Sigurbjornsson & Zwol, 2010) as well as in our earlier user studies (Lin et al., 2010).

To classify our tags for this type of interface, we also used Amazon Mechanical Turk. The classification procedure itself was independent of image context as none of the currently available tag classification approaches take into account context information such as resource information, user information or other tags for the same or similar resources. To ensure that the workers (Turkers) on Amazon Mechanical Turk would classify our tags in a meaningful way, we provided them detailed instructions of how to select those tags that fit into the one of the five given categories. The guidance included a sample screenshot of three different types of tags classified into one of the five categories and a detailed explanation of how to apply these categories. Overall, three Turkers were assigned to classify a particular tag. After the first classification round, we noted that 11% of tags were not classified since the Turkers could not agree on which of the five given categories to use. Therefore, we decided to initiate a second round with additional six Turkers for each non-classified tag. In the end, 22% of the tags were classified as "Who", 16% as "Where", 23% as "When", 34% as "What" and only 5% of the tags as "Other". We had 86 different Turkers for the first classification round and 35 Turkers for the second. The mean inter-rater agreement per tag over all Turkers was substantial (75%). Similar

to the tag cloud interface, users have the opportunity to issue a query by clicking on a tag, to expand a query by clicking on the "+" sign or shrink the query by utilizing the "x" sign in the query string beneath the search box in this interface.

3.3.1.2 Participants

We involved twenty-four participants (8 females, 16 males), who were recruited via email and flyers distributed throughout the University of Pittsburgh campus. The participants were from a variety of disciplines ranging from law to computer science. Four of them had earned a bachelor's degree, 16 a master's degree and four a PhD degree. The average age of the participants was 30.6 years old (min=22, max=61, SD=7.59 years). Almost all (except two participants) reported using computers for more than five hours a day. All participants (except two) rated their search engine skills as high and indicated using Google, Yahoo! or Bing frequently. A significant number (19) reported that they were familiar with tagging or used search tagging systems such as BibSonomy, Delicious or Flickr regularly. Four participants reported that they were familiar with the history of Pittsburgh, the rest of our participants stated that they were not. On average, each user study session lasted 90 minutes.

3.3.1.3 Tasks

The work separately evaluated the effectiveness of the three interfaces in the two primary types of search tasks known as lookup search and exploratory search. As indicated by its name, lookup search is a commonly performed task. To study lookup search behavior, we created nine different lookup search tasks. To account for the differences in difficulty, a variety of pictures were selected ranging from "easy" to "hard" to find. To classify images by difficulty, we calculated the mean search time for each image in the image collection based on lookup searches

performed with Amazon Mechanical Turk. Then, we selected nine images ranging from “easy” to “hard” to find in the Teenie Harris image collection.

To study exploratory search behavior, we designed three exploratory search tasks as shops, sports, and music tasks. To ensure the balance between each type of user interface and also to capture the difficulty, we designed the exploratory search tasks carefully with a variety of additional search criteria and attributes. For instance, to capture balance with the faceted search interface, we tried to tune our search tasks to utilize as many facets as possible. We did that by asking the participants to search for several different topics as well as various search criteria such as different locations. To capture the property of familiarity with the search tasks, we asked our participants in the post-questionnaire to rate their expertise level on the given topic or search item.

To be sure that our search tasks were meaningful, we performed several trial searches with Amazon Mechanical Turk and we conducted a pilot study.

3.3.1.4 Procedure

Each subject had to undertake two different search tasks using three different search interfaces within one user study session. During the study, each subject was assigned to perform nine different lookup and three different exploratory search tasks that were the same for the duration of the whole experiment. To counter the impact of fatigue and learning, the order in which the search tasks and system interfaces were used were rotated using a Latin square design. In addition, the lookup and the exploratory search tasks were randomized among all three interfaces to make sure that each of them was evaluated under different search interface conditions.

3.3.2 Evaluation results

The main goal of this work was to investigate the value of using social tags to assist users in image finding as well as the effectiveness of tag-based browsing navigational support. The log analysis of the initial experiment attempted to compare users' performance and the general patterns of user activities on different tasks with different interfaces. Several findings (F) are described in the following two sub-sections.

$F_{2.1}$: Participants had different search performance in two types of search tasks and with three different interfaces.

- The main difference in user performance was observed between the types of search tasks, a lookup search and an exploratory search. As we expected, an exploratory search required much more time and resulted in more actions than a lookup task.
- Users had better performance by using both tag-based interfaces than the baseline interface but only the tag cloud interface significantly outperformed the baseline search-only interface in terms of search time.
- The differences between the two types of tag-based browsing interfaces explored in our study are not as clear in the log analysis.
- Users who are not familiar with the topic performing a task with a medium level of difficulty completed the task faster with the tag-cloud interface than the baseline interface.

$F_{2\text{-}2}$: Participants had different usage profiles for each of the interfaces and tasks.

- Users had the action "Search" more frequently with the baseline interface, $p=.006$.
- Users clicked "Show More Results" more frequently with the baseline interface, $p=.015$.

- The action “Add Tag”, which was used to narrow the results by adding tags to the query, was used significantly more frequently with the faceted interface than a simple tag-cloud interface, $p=.006$.
- Users relied more on the search box, the “Add Tag” and “Remove Term” functionalities in lookup tasks than in the exploratory search task.

$F_{2\text{-}3}$: Participants were satisfied with the support from tag-based interfaces.

- Participants judged the support provided by the simple tag-cloud interface significantly better than that provided by the baseline, $p<.001$.
- Participants were significantly more confident in the ability to find relevant information with the simple tag-cloud interface, $p=.05$, and the faceted tag-cloud interface, $p=.037$, comparing to the baseline interface.
- Users found tag-related features significantly more useful with the tag-cloud interface than with the faceted tag-cloud interface.

3.3.3 Summary and discussions

The tag-based interfaces provide a snapshot of the “aboutness” of the collection, guiding the user to a more successful choice of a search term or tag. When users performed exploratory search tasks, they needed more support with the interfaces. Tag-based interfaces could give users sufficient navigational support. However, the analysis of objective data (performance and action profile) and of subjective data (questionnaires) delivered slightly different results. From the users’ perspective, both tag-based interfaces enhanced support for both types of search tasks and gave users higher confidence to find more relevant information. From the performance and log analysis perspective, only simple the tag-cloud interface had significant improvement over the

baseline interface in terms of time and actions. In short, tag-based interfaces are helpful to assist users' navigation but the simple tag-cloud interface might be clearer for users due to the non-structural nature of tags.

4.0 RESEARCH DESIGN

The goal of this study is to explore whether images can be found more efficiently and effectively when two types of information descriptors (subject headings and social tags) are provided to the users in a combined manner, the dual-perspective navigation framework. Enhanced image findability was achieved by providing comprehensive navigational support and combining expert and novice content labeling artifacts to serve a highly diverse user base. To support diverse users' information seeking behaviors, the search interface was created based on well-recognized design guidelines (M. Hearst, 2009) with a seamless integration of multiple browsing features and keyword-based search and effective experts' and novices' information representations.

According to the practices in the cultural heritage domain (Chan & O'Neill, 2010), the community has recognized that there was a strong need for subject access methods to handle the rapid growth of digital resources since late in the 1990s. The Online Public Access Library Catalogs (OCLC) was founded in the 1960s and has been responsible for many innovations in information storage and retrieval and exploration. This organization has explored approaches to subject analysis and representation to enhance resource findability. The development of "Faceted Application of Subject Terminology (FAST)" was proposed by OCLC to extend the use of the Library of Congress Subject Headings (LCSH) and to facilitate subject access to a large volume of material, which indicates that subject headings can be accessed with facets more effectively

and also gives us a clear idea how to effectively represent subject headings to people with less training and experience in the domain.

The preliminary study, "Exploratory images with named entities-based faceted browsing", confirmed that a faceted browser is useful to assist users in exploring in a museum context and to find their desired items more effectively. Although we didn't apply subject headings in that study instead used named entities, it showed us that a faceted browser could give users a quick overview and lead them to their interested items. Regarding to users' post-experiment comments, we learned that uncontrolled data in facets sometimes gives users a mistaken concept about the content and leads them off on an incorrect direction when we applied an automatic method to extract named entities and classify the entities. If we could apply controlled data in facets, such as subject headings, the faceted browser could provide more benefit to users.

With the increased availability of online access and inherent human need to organize personal information and resources, social tagging has become an alternative to validate the different aboutness of an item by a large crowd of people. Another preliminary study," Tag-based information access in image collections" obtained several insights about using tags to enhance image findability. That study first confirmed that artificial tags created by the Mechanical Turk are useful to inform users what the items in the collection are about. Second, the classic tag-cloud interface outperformed other types of interfaces, which tells us that a classic tag cloud is a better layout for representing tags than using a faceted layout. We believe it is because of the nature of tags is that they are non-hierarchical terms assigned to items by different annotators. According to the post interviews, participants told us that the faceted tag-cloud

interface gave them more information but also got them lost in many cases so that they preferred the classic tag-cloud interface with its clear overview of the content.

Based on these preliminary studies, enhancing image findability can be demonstrated by using a faceted browsing with a structured vocabulary or by applying a classic tag cloud with social tags. However, combining both mechanisms to capture the essence of the context's aboutness from both experts and users hasn't been examined. This study addresses whether the combination is helpful to support end-users to find images more effectively and efficiently in image collections. Therefore, this study focuses on the following research questions:

- Does the dual-perspective navigation approach provide better information to help users to achieve their goals in an effective manner than the single-perspective navigation approaches?
- Does the dual-perspective navigation approach guide users to their targets with fewer resources required than the single-perspective navigation approaches?
- Does the dual-perspective navigation approach make users more confident of their image finding ability and leave them with a positive perception of the approach?

4.1 DATASET

As a dataset for this study, I utilized two collections of images. One collection is from a "Teenie" Harris archive belonging to the Carnegie Museum of Art in Pittsburgh, Pennsylvania. This archive collection contains more than 80,000 images taken by Charles "Teenie" Harris, a photographer for an influential African-American newspaper, the Pittsburgh Courier. The collection, which catalogs a 40-year period of Pittsburgh history through the eyes of an African-

American journalist and amateur historian, offers a good opportunity to explore the value of applying different index features, subject headings and social tags, to the interactive search interface. I employed 1,864 of these images, of which 986 have been featured in a recent exhibition (Oct, 2011-April, 2012) at the Carnegie Museum of Art. The remaining images were included in this study as they provide a finer-grained overview of the entire collection. For the 1,864 images, I collected user tags through the portal of the Amazon Mechanical Turk. Amazon Mechanical Turk (MTurk) is a micro-task market platform where researchers can post a variety of tasks, called HITs ("Human Intelligence Task"), and recruit thousands of anonymous non-expert users, turkers, for a small fee (Rashtchian, Young, Hodosh, & Hockenmaier, 2010). MTurk had been applied for tasks ranging from labeling images with keywords (Nowak & Rüger, 2010; Sorokin & Forsyth, 2008) to judging the relevance of search results (Grady & Lease, 2010).

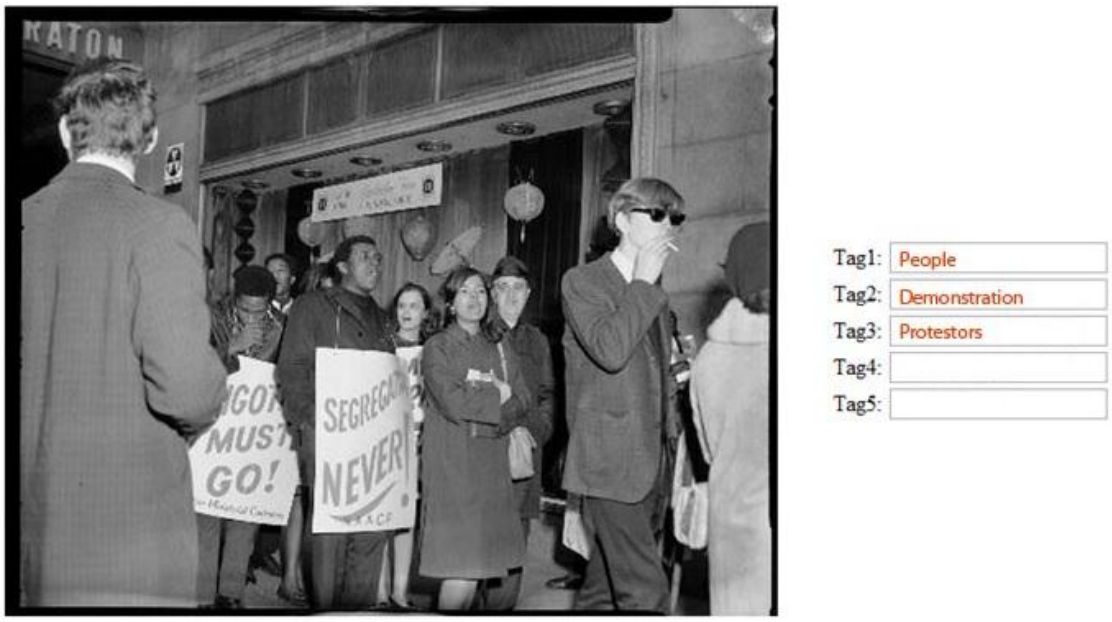


Figure 12. The tagging interface in MTurk

I gathered 5,634 unique tags created by 256 users for the 1,864 images. In order to keep the nature of free-style tags, this study only had few requirements for tag collecting. First, I asked turkers to apply at least 2 tags, which had to be less than four words long. For example "this is a bad tag" is not allowed since it is composed of 5 words. Second, I requested turkers provide tags that contained shareable information for others. For example "my stuff" is a not meaningful for other users who couldn't use that tag to do a search. In order to show them how to generate meaningful tags, some suggestions were shown to the turkers that they imagine what kind of keywords a user on a search engine such as Google or Yahoo! would use for the tag they proposed when trying to find that image. Beyond that, since the whole image collection is "Black and White" so that all possible terms to describe the fact like "without color", "black", "white", and "black and white" are not valid to our tag assignment and were filtered out. (This was a reference to the photos being black and white photos rather than the race of the participants in the photos.) Each image has been tagged by three turkers. A sample of the HIT's interface is shown in Figure 12.

Table 2. The summary of the two collections (The numbers are unique numbers)

	No. of unique SHs	Avg. SHs per image	No. of unique tags	Avg. tags per image	No. of images	No. of terms
Teenie Harris	607	6.35	5,634	17.23	1,864	26,888
Flickr	1,596	5.45	12,896	15.30	5,281	76,684

Another data set I obtained by crawling Flickr, which are the images uploaded by the Library of Congress[8] in January, 2013. It has around 15,194 images in which almost all images (except two) are tagged with the "Library of Congress" that is annotated by the Library of Congress when they uploaded the images. Around 83% of images (12,541) have more than one tag. Overall, there are 1,216,318 tags provided by the Library of Congress and Flickr's users and there are 12,896 unique tags. Flickr has a maximum limitation of 75 different tags for each image. In the crawled dataset, the maximum of number of tags founded was 73. There are around 7,000 (6,923) images that have been assigned 27,232 subject headings by the Library of Congress. Among the subject headings, there are 1,596 unique ones. There are 5,281 images with both subject headings and social tags which are taken to create a dataset for our experiment for testing the interface with both subject headings and social tags. The summary of the two data sets is shown in Table 2.

4.2 PRE-PROCESSING OF DATA

Without a controlled vocabulary, users might choose different tags in various forms to describe the same resource. Researchers (Angeletou, Sabou, Specia, & Motta, 2007) tried to bridge the gap between tags and control vocabularies but they couldn't deny that a broad range of semantic relations exists between tags which are hard to capture from knowledge sparseness. Since the complexity of transforming tags into meaningful presentations, many applications in current practice apply tags with their original form without further processing, such as Flickr tags[9] ,

[8] http://www.flickr.com/photos/library_of_congress/
[9] http://www.flickr.com/photos/tags/

tumblr[10], stackoverflow[11], citeulike[12], etc. According to the analysis from (Syn, 2010) and the purpose of this study, I didn't apply extensive cleaning or semantic clustering on tags except limited pre-processing described below. I maintained the original form of tags and kept a phrase tag as a phrase without decomposing it into terms. I believed that with enough people creating tags, a picture that relates to a specific event, such as World War Two, will be tagged with all kinds of tags that people think of when thinking about World War Two. When a user clicks on WW2, almost all of the pictures returned would also be returned with WWII , World War 2, or World War II, because different users will have used all those tags to describe the same picture. For instance, in our Flickr collection, there are 337, 359, 316, and 327 images assigned with the tags of "WW2"," WWII", "World War 2", and "World War II" respectively. There are 305 images assigned with these four tags together. Some simple filtering and normalizing of tags were taken before using the datasets in the experiments.

1. **Correcting spelling mistakes.** Neither Flickr nor I provide a spell checker when users enter tags but Flickr and I require users to enter tags individually into the text field. It is easier for users to understand how to separate tags unlike some websites such as del.icio.us, with a "space" delimiter to separate multiple tags. Meanwhile, Flickr and I didn't provide any tag suggestions so that users can issue whatever they have in mind at the moment. It is simple and straightforward for users to enter any tag but it is also easier to have typos carelessly entered. To lower the possibility of giving searchers incorrect information hints with typos, I simply apply free software, called Ginger[13], to check the spelling of tags and make simple suggestions to correct the

[10] http://www.tumblr.com/explore
[11] http://stackoverflow.com/questions/tagged/json
[12] http://www.citeulike.org/group/1710/tag/todo
[13] http://www.gingersoftware.com/

typos. Therefore, I could apply those corrected tags to the experimental systems to investigate users' behavior.

2. **Removing tags that are difficult to define their meanings as they appear relatively high frequencies in the collection.** Popularity is one way to judge the quality, value, and importance of the tag I encountered. When the popularity of a tag is high, it has higher chance to be shown on the interface. If the tag is not meaningful term, such as "picture" in our Teenie Harris image collection, showing the tag is not useful for supporting users' navigation. I manually checked the top 100 tags in each collection to ensure that the tags contained useful information scents for users. However, very few tags actually were needed to be removed.
3. **Removing images that don't have any tag or only the specific tag "Library of Congress".** In practice, it is possible that some image objects don't have any tags assigned if tags are not required by the system when the item's owner uploaded it to the system. In order to provide a fair comparison, I removed all images that did not have any tag attached. I also removed tags that are only annotated with the specific tag "Library of Congress", since all our images crawled through Flickr were aimed to be published by the Library of Congress.
4. **Removing non-English tags.** Tags in Flickr might be annotated in different languages, such as Farsi, Chinese, Japanese, and Spanish.

In addition, to provide navigation support from both subject headings and social tags and further examine their differences, I removed images without any subject heading as well.

4.3 INTERFACES

In order to examine if users will find images more efficiently and effectively when they are provided information descriptors from experts and the general public, I designed three interfaces, one with experts' descriptors only, another with the general public's descriptors, and the other with the combination of the descriptors from both sides. The dual-perspective navigation interface, which integrates the descriptors from the experts and the general public, was developed specifically for this study. The other two baseline interfaces were designed with single descriptors, either subject headings or tags, based on (Lin et al., 2010; Trattner et al., 2012). In these preliminary studies, I compared these baseline interfaces with a Googlesque keyword search only interface. This study focuses on comparing interfaces with single and dual perspective navigation support so the most basic, commonly used, search interfaces were implemented. This decision eliminated advanced pre-existing image search systems, such as google image search, and sophisticated a tag preprocessing and semantic analysis. For the subject-headings only interface, I adopted the faceted browsing to represent structural subject heading vocabulary along several dimensions, which have been shown to be helpful and preferred by users over the traditional search interface (Yee et al., 2003) and adopted by different studies(Chan & O'Neill, 2010; English, Hearst, Sinha, Swearingen, & Yee, 2002; Sigurbjornsson & Zwol, 2010; Yee et al., 2003). In terms of tag only interface, although there are other tag interfaces optimized for particular applications, I utilized a tag cloud to represent tags in an alphabetic order which is currently the most popular type of tag-based browsing in general social tagging studies(Hassan-montero & Herrero-solana, 2006; Helic, Trattner, Strohmaier, & Andrews, 2010; Seifert, Kump, Kienreich, Granitzer, & Granitzer, 2008; Sinclair & Cardew-Hall, 2007). For example, Flickr has diverse tag browsing interfaces including the general tag

cloud[14] similar to the one used in this study and also an interface[15] optimized for their particular market.

To assist users' diverse information needs, this study adopts current faceted search approaches which imply a precise information need by narrowing down search results on the search paradigm. The Lucent package was applied to implement the search focus on "and" operation to calculate the intersection with diverse filters (facets) for the base query. The "or" operation was provided when a user issues query terms in the search box which was demonstrated in the training section of the experiment. To make the three interfaces fairly compared, I adopt the same search mechanism in three interfaces. The following sections introduce each interface and its functionalities.

4.3.1 Dual-perspective navigation interface

The dual-perspective navigation interface developed for this study is shown in Figure 13. As indicated by its name, this type of search interface contains two information descriptors, subject headings and tags. This interface contains three important elements. First, it provides a search text box for a basic keyword search that offers the look and feel of well-known search engines. Users can issue a query and get a thumbnail preview of the resulting images sorted by relevance. Apache Lucene is the back-end search engine that utilizes all the image content including subject headings, tags, and descriptions to create the search index. Second, this interface provides facet browsing along with subject headings on the left hand side of the screen. The subject headings are classified into four facets: activities, objects, locations, and people. Along with each subject

[14] http://www.flickr.com/photos/tags/
[15] http://www.flickr.com/photos/tags/flowers/clusters/

heading, the number of images related to each subject heading is shown beside it. The subject headings are presented in each facet in frequency-ordered. Third, a social tag cloud is designed on the top of the screen. The tag cloud is alphabetically ordered and tags are presented with different font sizes according to its popularity. The more popular a tag, the larger the font size is. To generate the tag cloud, I utilized a simple popularity-based tag cloud algorithm.

The following elaborates the design elements applied to this search interface. The design can be discussed through search process. In the early stage of the search process, an interface might intensify an overwhelming problem by attempting to give users a broad overview of the entire collection and to provide users with an idea of how to construct their exploration paths. The opening page of this interface (Figure 13) provides a search text box for a basic keyword search with query suggestions, facets along with top-level subject headings, a social tag cloud of the entire dataset on the top, and image examples. This immediately facilitates users' familiarity with the high-level information aboutness of the entire collection and also gives users the freedom to choose the starting point from any one of the searching and browsing mechanisms. By selecting a subject heading, a tag or issuing a query, the user initiates the search and gathers relevant results back for further refinement. Then users can begin their exploration of the content.

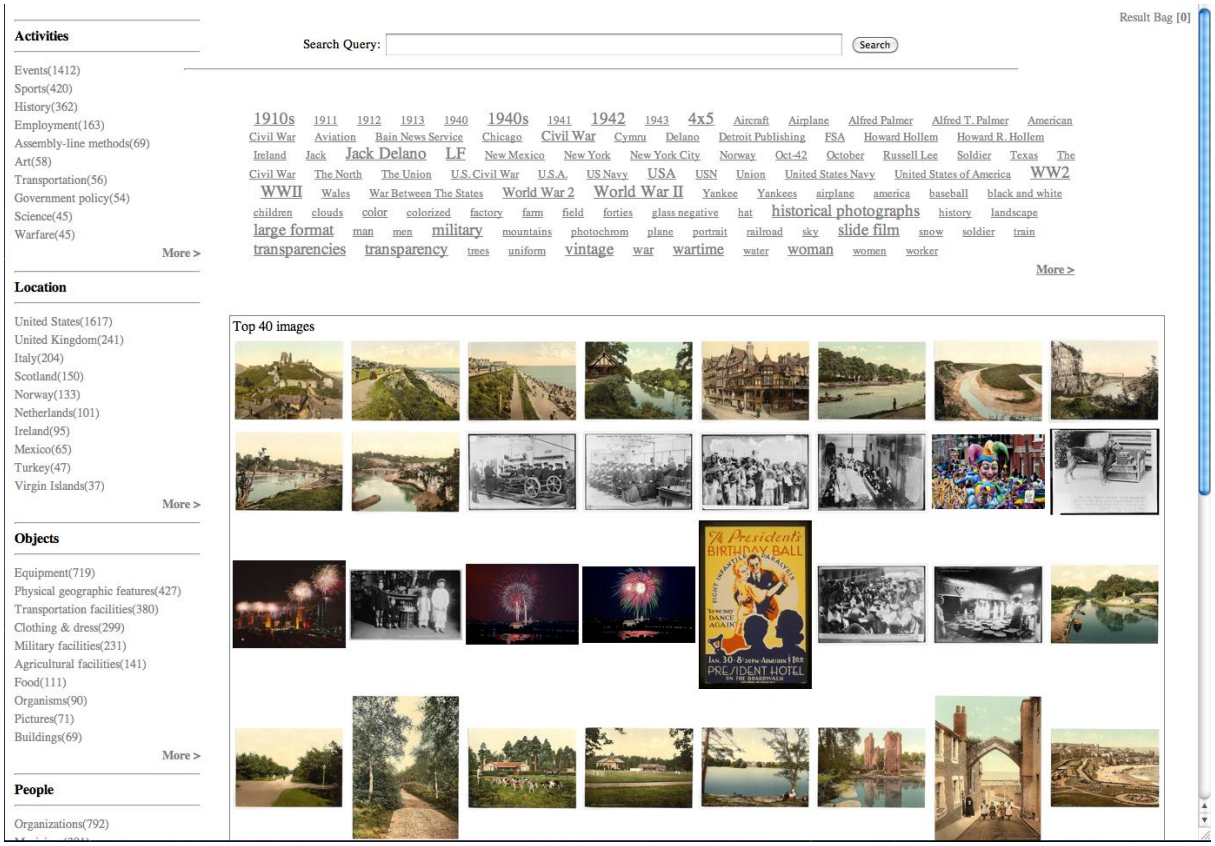


Figure 13. The dual-perspective navigation interface – opening page

In the middle of the search process, users evaluate the returned results and discover possible solutions to reach their goals. The result page (Figure 14) contains a set of images that match the request, the subject headings associate with those images are listed as facets, and the social tags associate with those images are displayed in the result set tag cloud. The current query is shown underneath of the search box to help users keep track of the current selection criteria. The search box remains available for searching within the entire collection. At this stage, a flexible method of refining search is the key concepts in the design. For example, a user select "Men" to start the search, the result page returns in a view with the subcategories of "Men" in the "People" facet, namely father, grooms and so on, along with Men's corresponding social tag

cloud. The user can refine the search by entering a new query to reissue a search by adding a subject heading from the facets, or by adding a social tag to narrow the result set. Users are allowed to remove any selected subject heading, social tag, and keyword in the query at any time to broaden the result set. When the user selects an image, the system takes the user to the item page.

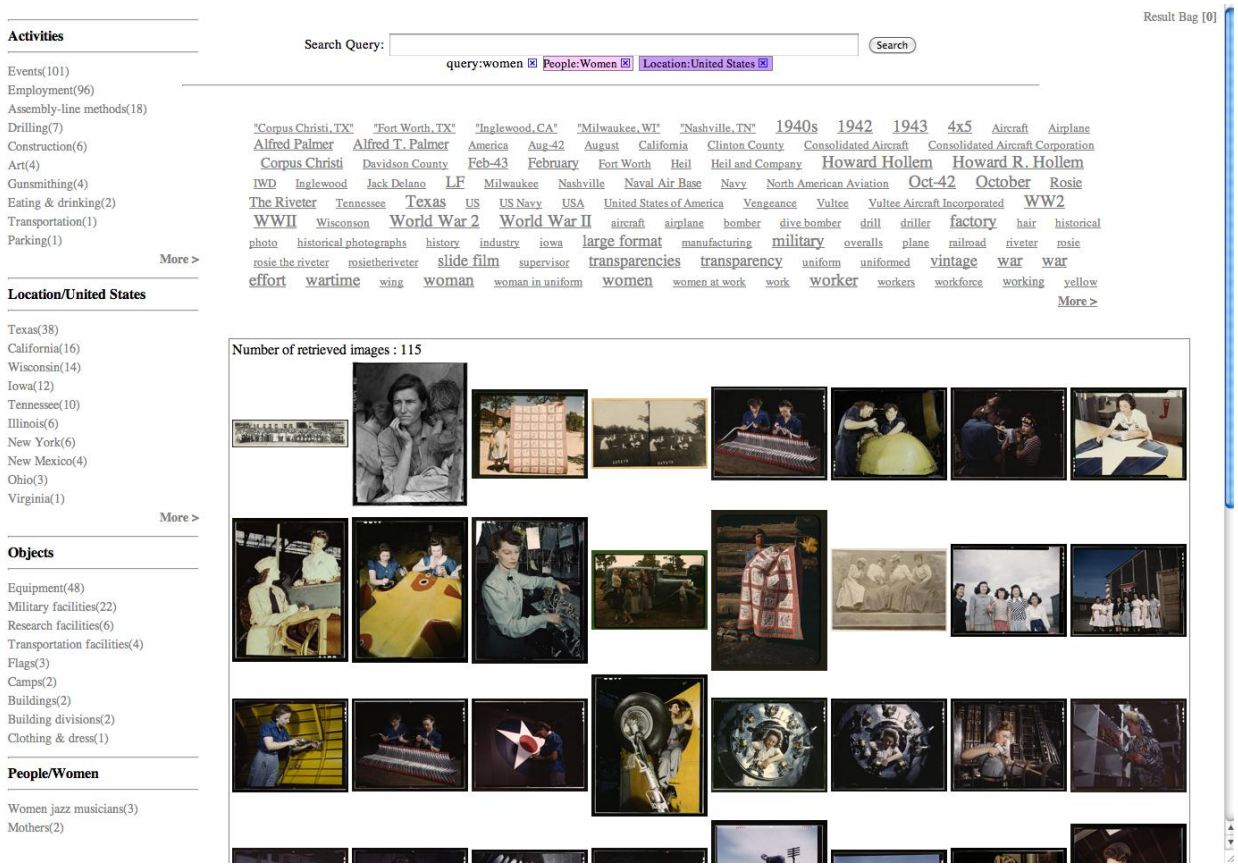


Figure 14. The dual-perspective navigation interface – result page

The item page (Figure 15) displays an individual image with document surrogate (title, all of the subject headings and social tags assigned to the item) and the search path with all the

query selections. All the subject headings and social tags assigned to the item are provided with hyperlinks, which allow users to select any one of them and switch to a new query with the selected subject heading or social tags and gather all items associated with the new query.

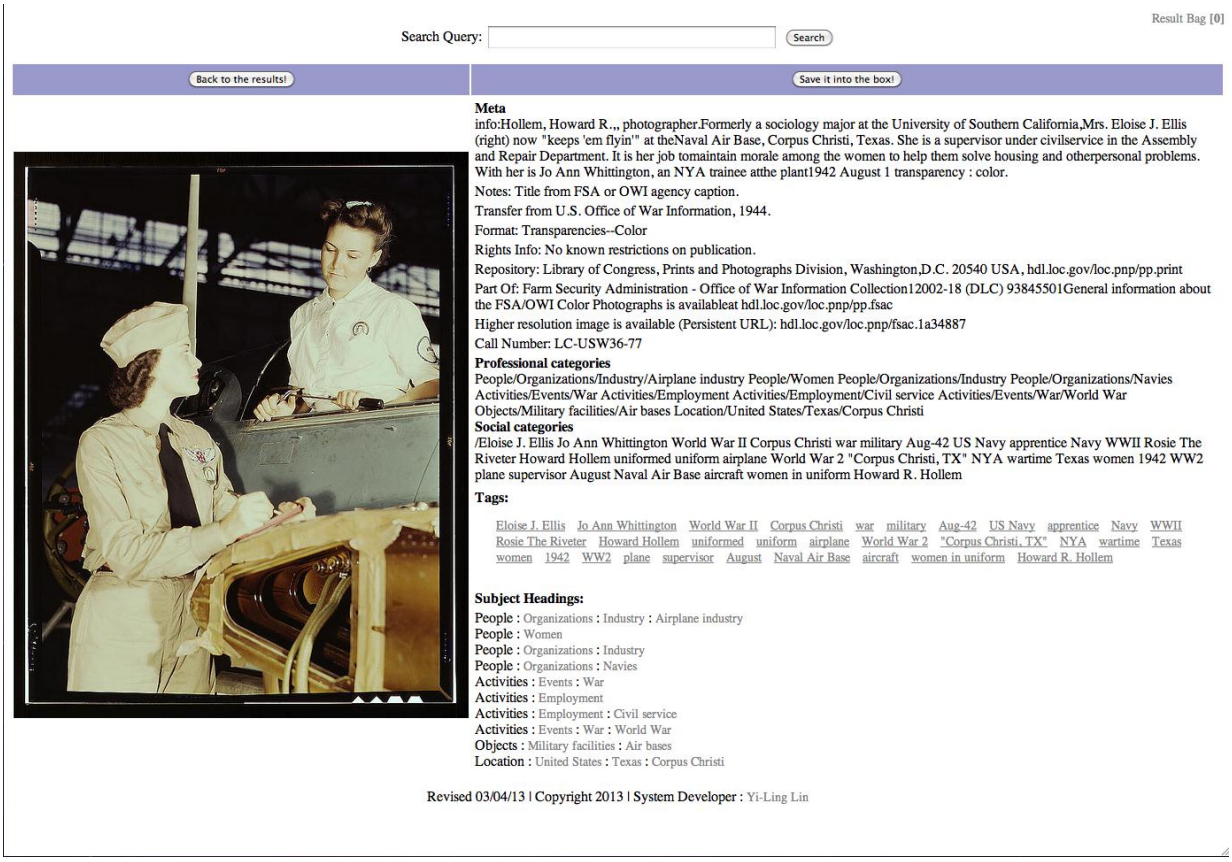


Figure 15. The dual-perspective navigation interface – item page

4.3.2 Tag-only interface

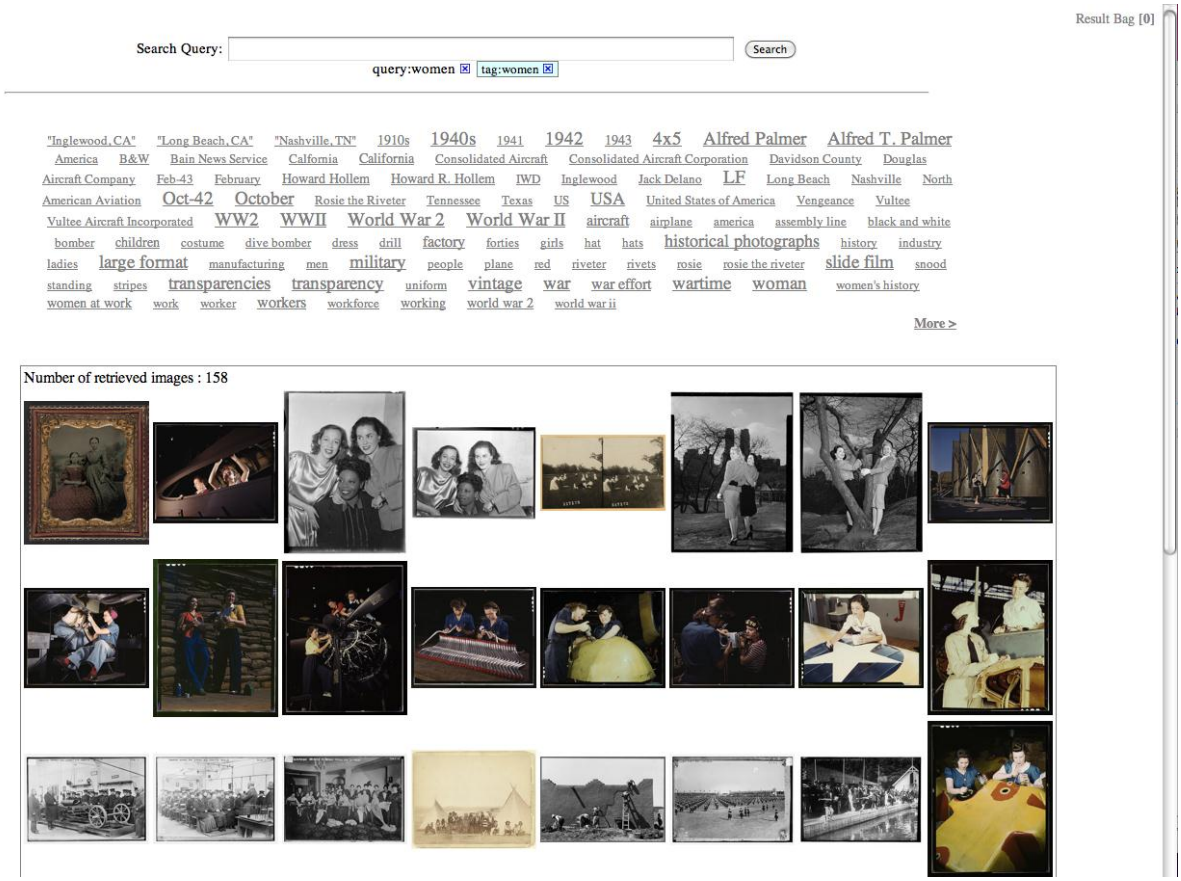


Figure 16. An example of tag-based image finding system

One of the baselines in this study is the interface containing social tags only. This interface has been proved more effective than a Googlesque interface in a preliminary study (Trattner et al., 2012). All the available elements in this interface are designed in the same way as they are in the dual-perspective navigation interface. This interface provides a search text box for a basic keyword search with query suggestions, a social tag cloud of the entire dataset on the top, and image examples underneath (Figure 16). It facilitates users' familiarity with the high-level information aboutness of the entire collection and gives users the freedom to choose the starting

point from any one of the searching and browsing mechanisms. Selecting a tag or issuing a query initiates the search and gathers relevant results back for further refinement and then users can explore the content.

In the middle of the search process, users evaluate the returned results and discover possible solutions to reach their goals. Similar to the previous interface, the current query is shown to help users keep track of the search. The user can refine the search by entering a new query or clicking a tag. When the user selects an image, the system takes the user to the item page, which is displayed in the same way as the one in the previous interface excluding the part of subject headings.

4.3.3 Subject heading-only interface

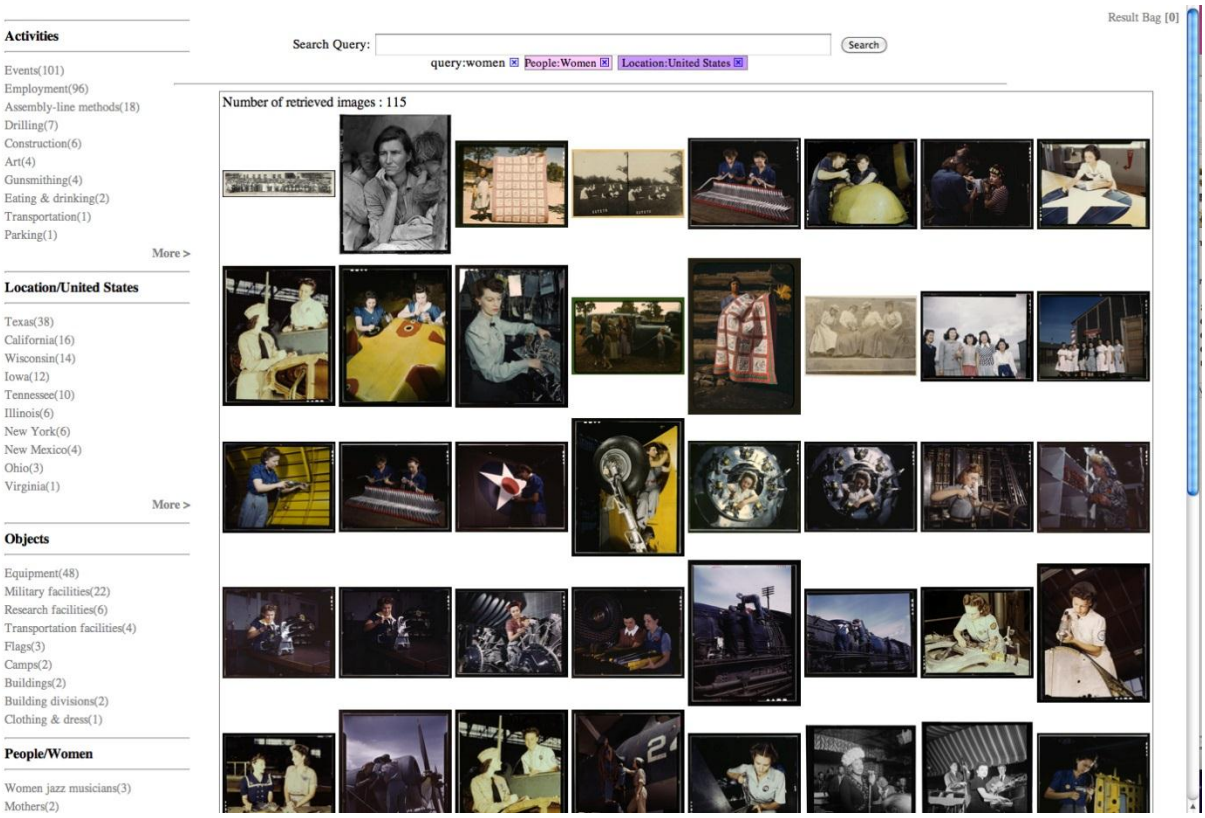


Figure 17. An example of the subject heading-based image finding system

The other baseline in this study is the interface containing subject headings only. This main feature, faceted browsing, has been proved more effective than a Googlesque interface in supporting image finding for the exploratory search in a preliminary study (Lin et al., 2010). All the available elements in this interface are designed in the same way as they are in the dual-perspective navigation interface. The procedure of a search is similar as the tag-only interface. Instead of providing a tag cloud, the interface provides facets with subject headings at the left hand side (Figure 17). If a user issues a search with any subject heading, the corresponding facet shows the subcategories of the subject headings when the system returns the image results. If a user selects an image, the system takes the user to the specific item page as the one in the previous approach replacing the part of tags with subject headings.

4.4 EXPERIMENTAL DESIGN

To compare the three systems (the dual-perspective navigation image finding system, the tag-based image finding system, and the subject heading-based image finding system), I designed a within-subject study. In this design, each of our participants evaluated the three different search interfaces during one experimental session. To determine which navigational support is most effective in what kind of search tasks; each interface was examined in the context of two types of search tasks.

4.4.1 Participants

This study recruited sixty-two participants (eight for the pilot study) from the great Pittsburgh area. Participants are paid at the rate of $12 per hour. The estimated duration of the experiment is two hours.

4.4.2 Apparatus

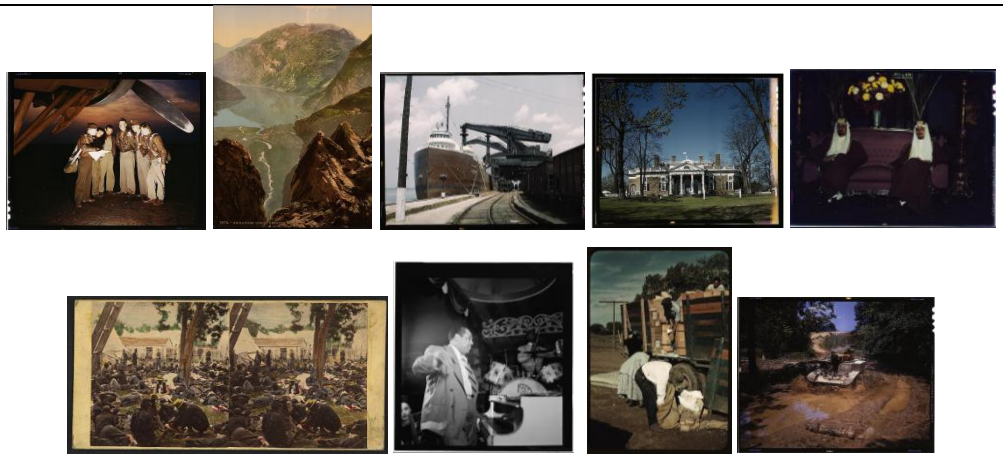
The experiment was conducted at the lab with an independent space located at the University of Pittsburgh's School of Information Sciences. Data was recorded with multiple methods: (a) system logs, (b) a pre-test (working memory capacity test, and user background survey), (c) post questionnaires after each task, each interface, and at the end of the sessions, and (d) a structured interview. I combined data from all of the sources in order to create a completed profile of each participant in each task session.

4.4.3 Search tasks

Lookup search tasks are considered to be relatively simple and most frequently involve using a traditional search interface, which is usually referred to known-item search (Bystrom & Jkrvelin, 1995; Diriyey, Blandfordy, & Tombrosz, 2010; Marchionini, 2006). To study lookup search behavior, I selected nine different images from each collection. I imitated a known-item search by showing the user one of the images and asking him/her to find that exact picture with the selected interface.

By contrast, exploratory search assumes that the user has some broader information need that requires multiple searches interwoven with browsing and analysis of the retrieved information (Lin et al., 2010; Marchionini, 2006). I designed three exploratory search tasks for each collection. Table 3 is the example of the tasks designed in the Flickr collection. I tried to ask participants to search for diverse topics such as religion or sports.

Table 3. Samples of search tasks and descriptions of the Flickr dataset

Search Tasks	**Search Task Descriptions**
Lookup	Find the following picture
	 (only one of the images was presented to the user at one time)
Exploratory	Background: You would like to add a new chapter to a travel book with some historical pictures about Europe. You are looking for images from the Library of Congress Flickr Commons collection The new chapter will include photographs of natural scenery, landmarks or buildings, and events in Europe. You want to present 4 countries. For each country, you will collect one representative picture of its natural scenery, one for its modern facilities, and one for its activities. All three pictures have to be in the same location (e.g., in the same region, state, province, or city of the country). You should gather 12 photos from this search.

4.4.4 Design and procedure

This study conducts a within-subjects investigation with two datasets. Each subject has three different lookup tasks and an exploratory task on each interface. Within the duration of the whole experiment, each subject has to perform the required tasks on the three different interfaces. To counter the impact of fatigue and learning, the order of search tasks across the three interfaces and the order of using the different interfaces are rotated by using the Latin square design as shown in Table 4. The datasets are randomly switched among participants.

Table 4. Latin square design of the experiment rotation

Participants (P) \ Interfaces	Baseline 1	Baseline 2	Experimental
P1	Task1	Task2	Task3
P2	Task2	Task3	Task1
P3	Task3	Task1	Task2

A pilot study was carried out to test the feasibility of the proposed study design. The task descriptions and requirements were adjusted based on the participants' suggestions from the pilot study. Incorporating pilot data, I refined the overall experiment procedure as follows (Figure 18): First, I inform the recruited participants of the objectives of the study and get their consent for the study (~10 minutes). Each participant completes a short background survey (~2 minutes) and takes a working-memory capacity test (~3 minutes). Before running the official tasks with each interface, I train each participant to use the features of the testing interface with a detailed explanation of the different requirements of the search tasks. The participant has sufficient time to become familiar with the interface and the two types of search tasks with each interface (~5

minutes per interface). In each official testing session with an assigned interface, a participant is required to complete three lookup tasks and one exploratory task.

a. Lookup task: According to the pilot data, six participants spent an average of 104.69 seconds (SD =71.91 seconds) to find an image, a limit of 3 minutes (180 seconds) was given to find any image for each task in the main study. The 180-second limit can from rounding off the average plus one standard deviation. In a normal distribution, this would approximately include 68 % of the sample in this study; table 8 will show approximately 75% of the search attempts were completed within 3 minutes. After three lookup tasks, a post-task questionnaire is given to the participant to elicit disposition toward the system interface.
b. Exploratory task: A description of the task is given to the participant before starting the task. The participant is allowed to ask any question about the task except how to search for relevant images. Based on the pilot result, six participants spent an average of 506.29 seconds (SD=94.15 seconds) to achieve the requirements of each exploratory task. The main study gives a limit of 10 minutes (600 seconds) to each participant to complete the assigned exploratory task. Participants were told to complete the task as quickly as possible within the 10-minute limit. A post-task questionnaire is presented upon completion of the assigned task.

After the post-task questionnaire, with questions about the difficulty of finding images with the corresponding interface (2 minutes), a NASA-TLX workload survey (Hart & Staveland, 1988) is applied to assess the participant's workload for interacting with each interface (3 minutes). After the three interface sessions, I have a final post-experiment questionnaire to assess the differences among the three search interfaces in terms of the user's preference, perception, etc. This survey is followed by a structured interview to insure that the participants' responses are not misunderstood by the user. The overall experiment takes about 120 minutes to complete.

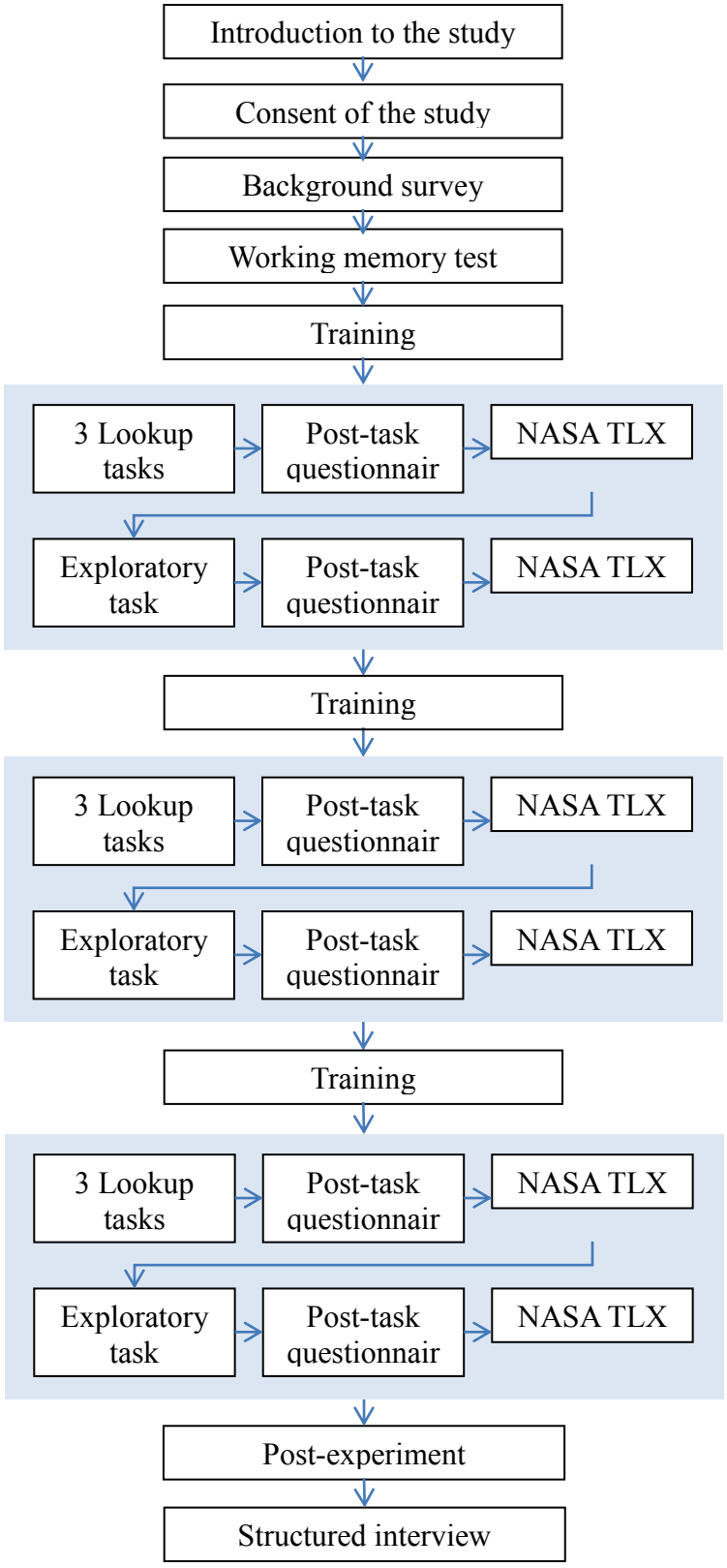


Figure 18. Experiment procedure

4.4.4.1 Pre-test (working memory capacity test) and background survey

An interface with different designs might affect users' performance with different required working memory load during the cognitive processing. An effective use of working memory required to conduct computer-based tasks require alternative interface designs (Gevins, Smith, & Leong, 1998). Therefore, I designed a pre-test (an individual working memory capacity test) to group our participants into high WMC and low WMC. With the groups, I was able to understand whether the difference of working memory load between two groups would affect users' interactions with each interface.

Figure 19. The interface of cognitive fun!

The working memory test I selected for the pre-test is the n-back test from "cognitive fun[16]" (Figure 19). A participant is requested to click on the hit box when the current picture repeats what he/she saw two items ago. If there is no repeat, the participant does nothing and waits for the next item to appear. During the test, each picture appears for 3-5 seconds and then jump to the next one. The whole test lasts for around one minute depending on the speed of a subject's performance. The sign O will appear beside the picture when the participant clicks correctly. If the participant click when there is no repeat, or don't click when there is a repeat, the red X will appear.

The background survey elicits users' background information, computer related experience, personal interests, and experience with tagging and faceted browser (the questionnaire can be found in Appendix A). The examples of tagging systems and faceted browser (Figure 20) are provided while the user fills in the background survey.

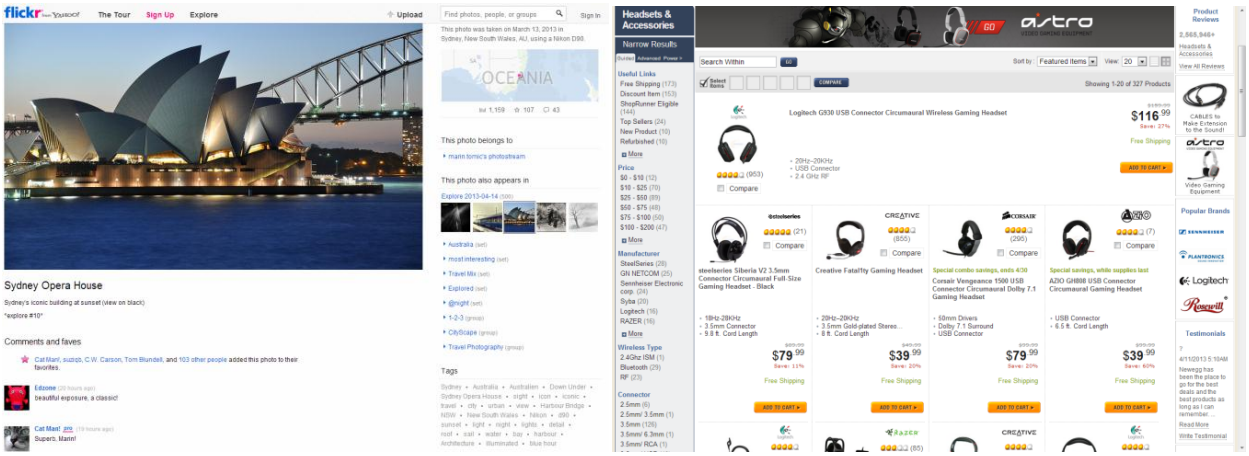


Figure 20. Examples of a tagging system (left) and a faceted browser (right)

[16] http://cognitivefun.net/test/4

4.4.4.2 Post-task questionnaire and workload survey

The post-task questionnaire was given after each type of tasks with each interface. Each post-task questionnaire of lookup tasks was given after the user finished three lookup tasks with each interface. The questions are about the difficulty of finding each of the three images with the corresponding interface. The post-task questionnaire of an exploratory task was give after completing the exploratory task with each interface. One specific question for this task is about the familiarity of the type of the search task. Another is about whether the user was confident in using the system's functionality to find useful information on the specific topic. Some general questions are for both post-questionnaires:

- Did the interface provide enough support for this task?
- Were some of the interface features unnecessary for this task?

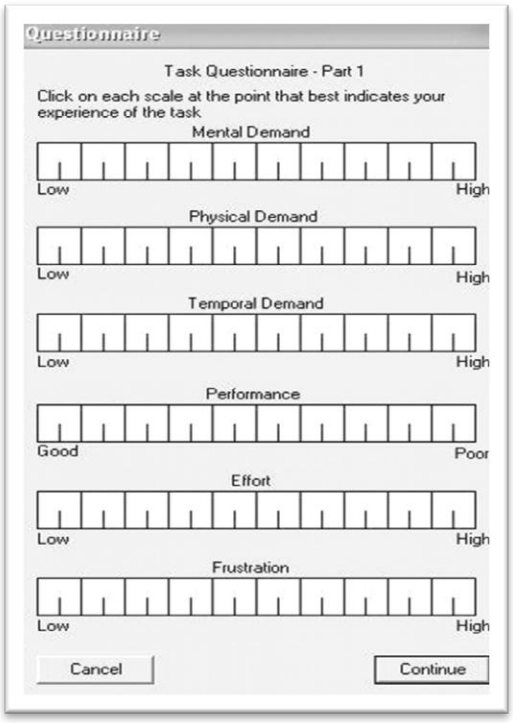


Figure 21. NASA-TLX questionnaire

In order to understand users' subjective experiences to different type of tasks with each interface, I gave the NASA-TLX workload survey (Hart & Staveland, 1988) to assess workload after interacting with each interface. It is a subjective workload assessment tool that contains a multi-dimensional rating procedure (Figure 21). It derives an overall score from the weighted average of ratings on six subscales including mental demands, physical demands, temporal demands, own performance, effort and frustration. It has been applied successfully in HCI research and user interface evaluation (Hornbæk, 2006).

4.4.4.3 Post-experiment questionnaire and structured interview

A final post-experiment questionnaire assesses the differences among the three search interfaces based on three dimensions, preference, satisfaction, and perception of users following by a structured interview to insure that the participants' responses are not misunderstood by the experimenter. While they were filling in the post-experiment questionnaire, three interface screenshots were provided to remind them the differences among three interfaces.

- Preference
    - Which one of the interfaces did you like most?
    - Which one of the interfaces would you prefer to use for lookup search?
    - Which one of the interfaces would you prefer to use for exploratory search?
    - Which one of the interfaces would you suggest museums to provide to their visitors?
- Satisfaction with the interface
    - How would you rate your experience with each interface?
    - How would you rate the functionality of each interface?
    - How would you rate the visual layout of each interface?

- What would you suggest to improve any of the interfaces?
- How did you come out the keywords to describe the kind of photos you memorized after using each interface?

4.4.5 Hypotheses

The tasks and measures are designed to test the following hypotheses about the proposed interface, the dual-perspective navigational image finding system, verses other interfaces (the tag-based and the subject heading-based image finding systems) according to the three research questions (Table 5). The metrics associated with the different hypotheses will be described in section 5.1 and 5.2.

Table 5. The research questions, hypotheses, and measurements

Research Questions	Hypotheses	Metrics
RQ1: Does the dual-perspective navigation approach provide better information to help users achieve their goals in an effective manner than the single-perspective approaches?	• H1-1: Users will successfully complete more tasks with the proposed approach. • H1-2: Users will make less futile searches with the proposal approach • H1-3: Users with different working memory capacity will not have different search performance with the proposed approach.	• Task success • Number of selected pictures • Number of futile searches Interaction by working memory capacity and interface

RQ2: Does the dual-perspective navigation approach guide users to their targets with fewer resources required than the single-perspective navigation approaches?	• H2-1: Users will spend less time to find a target item with the proposed approach. • H2-2: Users will spend less time to complete a task with the proposed approach. • H2-3: Users will reach the task goal with fewer interface interactions with the proposed approach • H2-4: Users will use less back tracking with the proposed approach. • H2-5: Users will have less mental effort when using the proposed approach.	• Time to find a targeted item • Time to complete a task • Number of the actions to reach the task goal • Number of back tracking used • NASA's Task Load Index questionnaire
RQ3: Does the dual-perspective navigation approach make users confident of their image finding ability and have a positive perception of the approach?	• H3-1: Users will be more confident in performing a search task with the proposed approach. • H3-2: Users will be more satisfied with the proposed approach.	• Post-questionnaire • Structured interview

5.0 RESULTS

In this study, fifty-four participants were recruited for exploring whether the dual-perspective navigation framework is an efficient, effective, and user-oriented method to support image finding. The experimental conditions include interfaces, collections, search tasks, interface order and task order (only for lookup tasks). Several covariates were collected from different resources, including gender, major, speaking language from the background survey, and working memory capacity from the working memory test. The analysis of data is grouped into two sections: (1) Effect on users' performance; (2) Effect on users' subjective perception under different conditions. The system log was applied to investigate users' performance and the participants' response to the questionnaires was used to evaluate subjective perception from the participants.

5.1 PERFORMANCE ANALYSIS

5.1.1 Study variables

This study has two search types, lookup and exploratory, designed to investigate how users interact with different interfaces for different information needs. For each lookup task, this study requires each participant to find each required picture in *three minutes*. For the exploratory task,

this study asks each participant to find twelve pictures, which fulfill the requirements of the assigned task in *ten minutes*. The outcomes include task success (task_success: 1 if successful; 0 if not), number of selected pictures (selected_picture), the total time (search_time) spent to find a specific image for the lookup task and to find the required images within ten minutes for the exploratory task, and the number of different interactions (total_action: the number of overall interactions; navigation_action: the number of clicks on subject headings or tags; search_action: the number of searches by issuing a search query; futile search_action: the number of futile searches which is getting an empty list of results for the search; back tracking_action: the number of back tracking actions used) to complete the task. A back tracking action in our context is going back to the result set after examining a specific picture or deleting a search query.

Table 6. The summary of the variables for the performance analysis

<table>
<tr><td colspan="2">Dependent Variables (DV)</td><td colspan="2">Independent Variables (IV)</td></tr>
<tr><td>Performance</td><td rowspan="2">• Success
- task success
- selected_picture (for exploratory tasks)
• search_time
• interactions
- total_action
- navigation_ action
- search_action
- futile search_action
- back tracking_action</td><td>• interface :
○ Subject Heading (SH)
○ Tag (Tag)
○ Dual-perspective (Dual)
• search_type :
○ Lookup
○ Exploratory (Exp)
• collection :
○ Teenie Harris (TH)
○ Flickr
• interface_order
• task_order (only for lookup tasks)</td><td>Experimental conditions</td></tr>
<tr><td></td><td>• working_memory
• native
• gender
• major</td><td>Subject demographics</td></tr>
</table>

To discover the effectiveness and efficiency of participants' performance, this study includes several independent variables and constructs in the analyses. The independent variables are all categorical as shown in Table 6. Interface and search_type, are repeated measurements, which indicate that each participant has to experience all interfaces and all types of search tasks in the experiment. There are three levels of interface, including the subject heading-only interface, the tag-only interface, and the dual-perspective navigation interface. Two types of search tasks, lookup and exploratory, are defined as the levels of search_type. The collection variable has two values indicating which dataset (Teenie Harris collection or Flickr collection) was assigned to the participant. Although I employed randomized design to decrease the effect of order, the effect of order, the interface order (interface_order) and task order (task_order), was included in the model. The working_memory variable is defined with two values indicating the level of a participant's working memory (high=1; low=0). The participants who scored higher than the mean (73.43%) of the working memory scores from all participants are defined as the group with the high level of working memory. The native variable has two values representing whether the participant is a native English speaker or not (native=1; non-native=0), gender with two values (female=0; male=1), and major with two values indicating whether the participant is from a computer related major or not (computer related=1; non-computer related=0). The relationships of variables are shown in Table 6. In order to investigate whether different working memory capacities affect users' search performance with different interfaces, the interaction between the working_memory and interface is checked in all the following performance analyses.

Among the participants, Table 7 reports the characteristic of subject demographics by collections. In this study, the participants are distributed evenly according to gender. Although

the percentage of the native English speakers is slightly lower than the non-native English speakers, this study ensures that at least one third of the participants are native English speakers in both collections. Although high and low working memory are almost equally distributed (13 vs. 14 in both collections), the high low distinction was made without considering the collection a user was assigned to.

Table 7. Demographics of the participants by interfaces in two collections

Characteristic	Teenie Harris (N=27)	Flickr (N=27)
Working_memory		
- Low	14(52%)	14(52%)
- High	13(48%)	13(48%)
Native		
- Non-native English speaker	16(59%)	17(63%)
- Native English speaker	11(41%)	10(37%)
Gender		
- Female	13(48%)	13(48%)
- Male	14(52%)	14(52%)
Major		
- Non-computer related	13(48%)	17(63%)
- Computer related	14(52%)	10(37%)

*Values represent the number of cases (the percent of cases) in each category

5.1.2 Models and results

The generalized estimating equations (GEE) is applied to model and analyze the data in this study (Liang, Zeger, & Apr, 2007). GEE can not only manage the analyses of longitudinal data, in which participants are measured at different points in time, but also clustering data, in which measurements are taken on participants with a common characteristic. In this study, each participant is measured at multiple points in time with different interfaces and search tasks. GEE can specify the repeated measures on two variables, interface and search_type, in an appropriate manner. In addition, it allows us to define the distribution and link function to model different types of outcome variables, such as linear, Gamma with log link, ordinal logistic, Poisson loglinear, binary logistic, etc. It also provides many correlation structures and produces model-based and empirical estimates. The proportional odds model (Pedhazur, 1982) is available in GEE which can make it easier to interpret the associations found in the data.

5.1.2.1 Success

To access the first hypothesis, "H1-1: Users will successfully complete more tasks with the proposed approach" I measured participants' task success with the requirements:

1) For each lookup task, a participant has to find the required picture within three minutes. The variable, task_success, is coded as zero for those who didn't find the assigned picture within three minutes and as one for those who found the assigned picture within the time limit. For testing each interface, this study requires each participant to perform three lookup tasks per interface.
2) For the exploratory task, a participant has to find twelve pictures for fulfilling the requirements in the assigned task within ten minutes. The variable is coded as zero for

those who didn't find twelve pictures for all the requirements within ten minutes, and one for those who completed the requirements within ten minutes.

Table 8 shows the number of successful and failure cases in two search types with three interfaces. From the observations, users have different performance between two search types. Table 8 shows that there is ~25 % failure rate in the lookup task whereas the failure rate in the exploratory task ranges from 0 to ~10%.

Table 8. The numbers of non-success and success observations in interfaces and search types

Search_type	SH		Tag		Dual	
	Non-success	success	Non-success	success	Non-success	success
Lookup	39	123	38	124	40	122
Exploratory	5	49	9	45	0	54

A binominal distribution with log link function is applied in GEE to evaluate the association between success and two within-subject variables, interface and search_type. Since fifty four participants were all successful while using the dual-perspective navigation interface for the exploratory tasks (Table 8), it is invalid to apply the binary logistic to model success with zero observation in this condition. The GEE analysis of task_success among interface was limited to the cases with the lookup search tasks.

In addition, these two types of search tasks are designed quite differently with different time limitations (3 minutes for each of the three lookup tasks and 10 minutes for the exploratory task) and requirements (1 picture for each of the three lookup tasks and 12 pictures for the

exploratory task) for each interface. Therefore, the cases are split based on the search type for the following performance analyses using GEE.

To predict success in the lookup tasks, I controlled for experimental conditions, subject demographics, and the interaction effect between interface and working_memory. There is no significant effect of interfaces. The model has significant effects of collection $\chi(1)^2=25.757$, $p<.001$, and working_memory $\chi(1)2=5.516$, p=.019. The estimated result of the analysis is shown in Table 9. When comparing the Flickr and Teenie Harris collections for successful lookup task completion, lookup tasks performed on the Teenie Harris collection were 3 times more successful than on the Flickr collection (OR=3.00, $p<.001$). The participants were expected to locate a specific picture easier in the Teenie Harris collection due to the total number of images in the Teenie Harris collection (1,864 images) being relatively smaller than those in the Flickr collection (5,281 images). However, the results differ from what I expected.

Table 9. Significant effect influencing task **s**uccess for the lookup task

Parameter		S.E.	Wald χ^2	sig	Exp(B)
collection	Teenie Harris	.2165	25.757	<.000	3.000
working_memory	low	.3777	9.971	.002	3.295

When comparing low working memory and high working memory for number of successes, low working memory participants had 3.295 of the successes that the high working memory participants did (OR=3.295, $p=.002$). The participants with low working memory capacity are more successful in completing the lookup task. The result is different with what we usually expect that users who have high working memory capacity might have higher chance to

success in the tasks but it also indicates that working memory capacity won't affect the search performance in our cases. The rest of the variables have no effect on success.

To test the difference of success among interfaces in the exploratory task, I transformed the data and adopted a nonparametric test (Cochran's Q Test) for the binary response analysis. In order to compare the pattern across the interfaces, I obtained all of the pairwise comparisons among the three interfaces. While computing multiple comparisons, this study applies a Bonferroni correction for $p<.05$ rule to protect against Alpha inflation. To do this, a Type I error across the pairwise comparisons is adjusted to be less than a 5% chance, which is accomplished by dividing .05 by the number of comparisons.

The percentage of success that occurred in each interface was 90.74 % with the subject heading-only interface, 83.33% with the tag-only interface and 100% with the dual interface, $Q(2)=10.167$, $p=.006$. The pairwise comparisons using a Bonferroni correct $p=.0167$ (.05/3) indicates that the participants with the dual-perspective navigation interface had more success than with the tag-only interface, $Q(1)=9.00$, $p=.004$, but there was no significant differences found between the subject heading-only and the dual, and between the subject heading-only and the tag-only.

Table 10. Significant effect influencing success based on the number of selected pictures for the exploratory task

Parameter		S.E.	Wald χ^2	sig	Exp(B)
interface	Subject Headings	.0065	6.091	.014	.984
	Tags	.0158	6.164	.013	.962

In addition to "task_success", to test the difference of completing the exploratory tasks among interfaces, I also adopted the number of selected pictures during the exploratory search as another outcome variable of an exploratory task's success. GEE Poisson was applied to predict the number of selected pictures since the Poisson distribution can interpret the distribution of this counting outcome variable well. Table 10 shows that there is a significant interface effect $\chi(2)^2=14.643$, $p=.001$. Compared to the dual-perspective navigation interface, the participants selected 2% fewer pictures (OR=.984, $p=.014$) with the subject heading-only interface, and 3.8% fewer pictures (OR=.962, $p=.013$) with tag-only interface. There is no other effect of the number of selected pictures found in the exploratory task.

In summary, the result shows that participants had higher success rate when they performed the exploratory search task with the dual-perspective navigation interface compared to with the tag-only interface. For the lookup search, different collections and different levels of working memory affected participants' performance.

5.1.2.2 Search time

To assess the hypotheses, "H2-1: Users will spend less time to find a targeted item with the proposed approach" and "H2-2: Users will spend less time to complete a task with the proposed approach," the search time is examined in this section. Table 11 shows the descriptive statistics of search time in two search types with three interfaces.

The outcome of search time is a continuous dependent variable that arises naturally in process. The Gama distribution can well interpret the distribution of search time. Therefore, GEE Gamma was applied to predict search time in the system given the experimental conditions, subject demographics, and the interaction between interface and working_memory.

Table 11. Descriptives (mean±SE) of search time by search type and interface

Search Type	Measure	SH	Tag	Dual
Lookup	Cases	123	124	122
	Search time	63.80±4.20	56.79±3.78	64.00±4.11
Exploratory	Cases	49	45	54
	Search time	354±15.75	402.98±15.96	354.94±13.47

With the lookup task, the result is shown in Table 12. There is a main effect of collection, $\chi(1)^2$=13.182, p<.001. When comparing the Flickr and Teenie Harris collections on time spent to reach the target item, the participants working with the Teenie Harris collection spent 74.2% of the time that the participants did on the Flickr collection (OR=.742, p<.001). Because the lookup task requests to find the required picture in the collection, the total number of pictures in the collection might influence participants' performance in the lookup task. There is neither other main effect nor an interaction effect for the lookup task.

Table 12. Significant effect influencing search time for the lookup and exploratory tasks

	Parameter		S.E.	Wald χ^2	sig	Exp(B)
Lookup	collection	Teenie Harris	.0821	13.182	<.000	.742
Exploratory	interface	Subject Heading	.0827	.322	.570	1.048
		Tag	.0849	.419	.517	1.057

With the exploratory task, the result indicates that there is a main effect of interface, $\chi(2)^2$=6.364, p<.042, when controlling experimental conditions, subject demographics, and the

interaction between interface and working_memory. The details are shown in Table 12. The results of the pairwise comparisons with the sequential Bonferroni adjustment indicates that the participants spent significantly less time, *p*=.020, with the subject heading-only interface (Mean=352.87, S.E.=16.21), and significantly less time, *p*=.050, with the dual-perspective navigation interface (Mean=352.10, S.E.=13.03) than with the tag-only interface (Mean=395.70, S.E.=15.24) to complete a task. There is no other effect of search time found in the exploratory task.

Due to the time limitation of the exploratory task, participants might have same search time up to 10 minutes (the time limitation in the experimental setting) but obtain different numbers of images. For example, a participant might successfully obtain 12 pictures in the required 10 minutes and another participant might only obtain 6 pictures within the time limit. Their performances should be considered differently. As a result, I calculated average time to obtain a required picture as another dependent variable to more accurately predict how fast a participant is in the exploratory search.

There is a main effect of interface, $\chi(2)^2=19.694$, $p<.001$ (Table 13). When comparing the dual-perspective navigation interface against the other two interfaces on average time spent, the participants using the subject heading-only interface spent 20.7% longer (OR=1.207, *p*=.017) and the participants using the tag-only interface spent 31.3% longer (OR=1.313, *p*=.001).

Table 13. Significant effect influencing average time for finding a required picture in the exploratory task

Search type	Parameter		S.E.	Wald χ^2	sig	Exp(B)
Exploratory	interface	Subject Heading	.0788	5.684	.017	1.207
		Tag	.0847	10.364	.001	1.313

In summary, the dual-perspective navigation interface facilitates users' search performance when comparing to the tag-only interface from the search time of the successful cases. The average search time to find a target picture in the exploratory task shows that participants spent less time to locate their desired items with the dual-perspective interface in the exploratory search than with the other two interfaces. For the lookup search, interfaces did not affect users' performance but the effect of collection made differences on users' performance.

5.1.2.3 Total actions

To test the hypothesis, "H2-3: Users will reach the task goal with fewer interface interactions with the proposed approach," the count of the overall interactions, total_action, is used as the outcome variable. Table 14 shows the descriptive statistics of total actions in two search types with three interfaces. Poisson can describe the distribution of this count variable. Poisson distribution with log link function is applied in GEE to predict participants' interactions with the interfaces given the independent variables from the experimental conditions, subject demographics and the interaction between interface and working_memory.

Table 14. Descriptives (mean±SE) of total actions by search type and interface

Search Type	Measure	SH	Tag	Dual
Lookup	Cases	123	124	122
	Total actions	12.93±.67	11.40±.60	12.84±.73
Exploratory	Cases	49	45	54
	Total actions	80.45±2.94	83.64±3.54	78.74±2.53

The analysis result is in Table 15. For the lookup task, the result shows that there are several main effects found collection $\chi(1)^2$=11.393, *p*=.001. When comparing the collections for number of actions, the participants who worked in the Teenie Harris collection took 80.6% of the actions that the participants who worked in the Flickr collection did (OR=.806, *p*=.001).There are no other effects found in the analysis. For the exploratory task, there is no any effect found. If I included failure cases into the analysis (with all cases), the effect of interface $\chi(2)^2$=7.441, *p*=.024 was found. It indicated that users who succeed in the exploratory task didn't have difference among effects but users who failed to achieve the task goal were influenced by the effect of interface. They had 19% more interactions with the subject-heading only interface (OR=1.190, *p*=.005) and 25.2% more interactions with the tag-only interface (OR=1.252, *p*=.001) than with the dual-perspective navigation interface.

Table 15. Significant effect influencing the number of total actions for the lookup task

Search type	Parameter		S.E.	Wald χ^2	sig	Exp(B)
Lookup	collection	Teenie Harris	.0640	11.393	.001	.806

To understand differences in user behavior across tasks and interfaces, I performed a deeper analysis of users' interactions from their behaviors, mainly focusing on navigation and search actions. Table 16 presents usage profile of each action for different interfaces and search types. The following sections will investigate the difference among conditions with each action one by one. The outcome variables in the following sections are the counts of diverse actions, such as search, navigation, back tracking etc. These variables are over-dispersed due to a large number of participants who might not have any related action in a particular section. The

negative binomial with log link can describe the distribution of these variables best so I applied negative binominal regression in the GEE model to predict the counts of these actions.

Table 16. Descriptives (mean±SE) of actions by search type and interface

Search Type	Measure	SH		Tag		Dual	
		All	Successful	All	successful	all	successful
Lookup	cases	162	123	162	124	162	122
	navigation	2.48±.24	1.56±.18	1.25±.14	.83±.13	2.71±.23	1.80±.20
	SH navigation	2.48±.24	1.56±.18	-	-	1.65±.17	1.10±.14
	Tag navigation	-	-	1.25±.14	.83±.13	1.06±.16	.70±.12
	search	3.72±.25	2.77±.20	4.22±.27	2.94±.20	3.63±.24	2.75±.22
	Futile search	.51±.08	.31±.07	.63±.09	.37±.09	.54±.07	.39±.07
	Back tracking	.14±.04	.04±.02	.12±.04	.08±.04	.08±.03	.02±.01
Exp	cases	54	49	54	45	54	54
	navigation	6.22±.79	6.45 ±.87	4.46±.67	3.67±.63	6.67±.73	6.67±.73
	SH navigation	6.22±.79	6.45 ±.87	-	-	4.70±.64	4.70±.64
	Tag navigation	-	-	4.46±.67	3.67±.63	1.96±.40	1.96±.40
	search	5.76±.61	4.96±.53	9.46±.83	8.36±.76	5.20±.62	5.20±.62
	Futile search	1.37±.22	1.09±.17	1.70±.24	1.53±.22	1.04±.19	1.04±.19
	Back tracking	20.78±.98	19.84±.94	21.93±1.34	20.29±1.43	18.61±.87	18.61±.87

(Exp is the abbreviation of exploratory; dashes "-" indicate no test was performed)

5.1.2.4 Navigation actions

The first action I examined, called navigation action, occurs when the participant clicks on any navigation features, a subject heading or tag.

When completing the lookup task, there is a significant effect of interface $\chi(2)^2=13.865$, p=.001. The result is shown in Table 17. When comparing the dual-perspective navigation

interface against the tag-only interface, the participants using tag-only interface applied 39.5% of navigation actions that the participants using the dual-perspective navigation interface did (OR=.392, p<.001). This might be because that the dual-perspective navigation interface provides more features to help users have a better understanding of the context so they tended to use more navigation support.

Table 17. Significant effect influencing the number of navigation actions for the lookup and exploratory tasks

Search type	Parameter		S.E.	Wald χ^2	sig	Exp(B)
Lookup	interface	Subject Heading	.1501	.097	.755	.954
		Tag	.2933	10.010	.002	.395
Exploratory	interface	Subject Heading	.2401	.142	.706	1.095
		Tag	.2894	.7.997	.005	.441
	collection	Teenie Harris	.1679	13.101	<.001	.545

When completing the exploratory task, the effect of interface $\chi(2)^2$=18.479, p<.001, and of collection $\chi(1)^2$=13.101, p<.001 were found. The participants had 56 % less navigation actions with the tag-only interface (OR=.441, p=.005) than with the dual-perspective navigation interface. In comparison with the Flickr collection, the participants performed the exploratory task on the Teenie Harris collection with 58% of the navigation actions that the participants did on the Flickr collection (OR=.580, p=.001).

To understand whether there is any difference of using subject headings or tags between single perspective (subject heading-only or tag-only) and dual perspectives (dual-perspective navigation framework), I also specified the navigation actions into subject-heading navigation

and tag navigation and did further analyses between subject-heading only and dual-perspective navigation framework and between tag-only and the dual approach.

For completing the lookup task, there was no effect found when I only considered the successful cases. When I included all cases, we found the interface effect, $\chi(1)^2=8.728$, $p=.003$ between subject heading-only interface and the dual. The participants had 41.9 % more subject-heading navigations with the subject heading-only interface (OR=1.419, $p=.045$) than with the dual-perspective navigation interface. This indicates that users who failed the task might be affected by interfaces when they utilized the subject heading features. They used more subject headings with the subject heading-only interface but those actions didn't lead them to the successful finding. I didn't find any difference between the tag-only and the dual in successful cases or all cases.

For the exploratory tasks, the interface effect $\chi(1)^2=10.078$, $p=.002$ was only found between tag-only and the dual. The results of the pairwise comparisons with the sequential Bonferroni adjustment indicates that the participants applied significantly more tag navigations, $p=.003$, with the tag-only interface (Mean=2.87, S.E.=.487) than with the dual-perspective navigation interface (Mean=1.63, S.E.=.615) to complete a task. Users applied tags more frequently when they are completing the exploratory task with the tag-only interface compared to when they are performing with the dual-perspective navigation framework. However, I didn't find any significant difference between the subject heading-only and the dual.

5.1.2.5 Search actions

Second, I am also interested in the actions surrounding issuing queries with different interfaces in different tasks. For the exploratory task, a significant effect of interface $\chi(2)^2=28.474$, $p<.001$, was found. Compared to the dual-perspective navigation interface on search actions, the

participants using the tag-only interface increased the number of actions taken on the exploratory tasks by 78.7% (OR=1.787, p<.001). With the result of total action for the exploratory task, it might be reasonable to claim that participants with the tag-only interface required more actions to complete the assigned task and search action is one of the contributing factors. The effect of collection $\chi(1)^2$=15.672, p<.001, was found significant as well. The participants conducted 63.2% (OR=1.632, p<.001) more searches in the Teenie Harris collection when comparing to the Flickr collection. Along with the result of navigation action, I found that the participants applied less navigation actions and more search actions when they worked with the Teenie Harris collection. The navigation support in the Teenie Harris collection might not be sufficient to help the participants to find what they want so they had to apply more search to complete the assigned task. The result of the search action is shown in Table 18.

Table 18. Significant effect influencing the number of search actions for the exploratory task

	Parameter		S.E.	Wald X2	sig	Exp(B)
Exploratory	interface	Subject heading	.1906	.104	.747	1.063
		Tag	.1650	12.383	<.000	1.787
	collection	Teenie Harris	.1238	15.672	<.000	1.632

The analysis above is based on the successful cases to assess the search actions required to assist users to achieve the goal of the task. There is no difference found for the lookup task. When I included all cases in the analysis, there is a main effect of collection $\chi(1)^2$=15.824, p<.001 for the lookup task. The participants performed lookup tasks on the Teenie Harris collection with 74.5% of the search actions that the participants did on the Flickr collection

(OR=.745, p<.001). This result indicates that the difference of the collections didn't affect successful users but affect the users who failed to achieve the goal of the task.

5.1.2.6 Futile search

To discover whether users will make less empty list of search results back with the proposal approach (H1-2: Users will make fewer futile searches with the proposal approach), I defined the number of futile searches which is getting an empty list of results for the search as a dependent variable. To examine the futile searches, the negative binominal log link is used in GEE.

With the exploratory search task, the result is shown in Table 19. There is a significant effect of interface $\chi(2)^2$=7.615, p=.022. Pairwise comparisons with sequential Bonferroni adjustment shows that the participants failed to get returned results significantly more frequently, p=.038, with the tag-only interface (Mean=1.65, S.E.=.256) than with the dual-perspective navigation interface (Mean=.93, S.E.=.142). There is another effect of collection $\chi(1)^2$=5.467, p=.019. Compared to the participants in the Flickr collection, the participants had 57.2% more futile searches than the participants did in the Teenie Harris collection (OR=1.572, p=.019).

Table 19. Significant effect influencing the number of futile search actions for the exploratory task

	Parameter		S.E.	Wald X2	sig	Exp(B)
Exploratory	interface	Subject heading	. 3552	.116	. 733	.886
		Tag	. 3179	.531	. 466	1.261
	collection	Teenie Harris	. 1935	5.467	. 019	1.572

With the lookup tasks, there was no effect found according to successful cases. However, when I included all cases in the analysis, the result shows that there are significant effects of collection $\chi(1)^2$=3.951, *p*=.047 and gender $\chi(1)^2$=4.541, *p*=.033. The results imply that smaller size of collection (Teenie Harris Collection) might reduce around 24% of the chances to make slower participants encounter the futile searches compared to the Flickr collection (OR=.759, *p*=.047). Female users who didn't success the task experienced around 40% more futile searches than male users (OR=1.395, *p*=.033).

5.1.2.7 Back tracking actions

To investigate whether users will need back tracking less frequently with the proposed approach (H2-4: Users will use less back tracking with the proposed approach), back tracking actions (the number of back tracking actions) is examined with the negative binominal log link in GEE.

Table 20. Significant effect influencing the number of back tracking actions for the exploratory task

	Parameter		S.E.	Wald X2	sig	Exp(B)
Exploratory	collection	Teenie Harris	.0538	5.654	.017	1.136

For the lookup search task, there was no significant effect found based on successful cases. When failure cases were included into the analysis, I found the effect of collection $\chi(1)^2$=7.219, *p*=.007. The participants working in the Teenie Harris collection used 30.2% of back tracking actions that the participants working in the Flickr collection did (OR=.302, *p*=.007). The reason might be that the number of total images contained in the Teenie Harris is around two thirds smaller than the number of images in the Flickr. In the lookup task, each

participant is required to find an exact picture so this type of task seems to be easier to complete in the small collection.

For the exploratory search task, an effect of collection $\chi(1)^2=5.654$, $p=.017$ was found. The participants needed 13.6% (OR=1.136, $p=.017$) more back tracking actions to complete the exploratory task in the Teenie Harris collection than in the Flickr. The participants had more search actions, conquered more futile searches, and used more back tracking actions but applied less navigation support (navigation action) in the Teenie Harris. The rest of the effects are not significant and the result is shown in Table 20.

When I included failure cases in the analysis, except the effect of collection, another effect of interface $\chi(2)^2=6.310$, $p=.043$ was found. Compared to the dual-perspective navigation interface, the participants used 31% (OR=1.310, $p=.004$) more back tracking actions with the subject heading-only interface and 30.4% (OR=1.304, $p=.003$) more with the tag-only interface. Users who failed to achieve the task goal use less back tracking with the dual-perspective navigation interface than with the other interfaces.

In summary, participants took fewer total actions, search actions, and back tracking actions to accomplish the exploratory task with the dual-perspective navigation interface. With the dual-perspective navigation interface, participants clicked more subject headings and tags than with other interfaces. Participants had very different behaviors in different collections for different search tasks. For the exploratory search, participants who worked in the Teenie Harris collection took more search actions, futile search actions, and back tracking actions than the ones working in the Flickr collection. For the lookup search, they had totally opposite behaviors search with the behaviors for the exploratory.

5.1.3 Summary of performance analysis

The result of significant effects is shown in Table 21. The effect of interface was mainly for the exploratory search. The dual-perspective navigation interface outperformed than other interfaces for the exploratory search. For the lookup search, the effect of collection is stronger than other effects.

Table 21. The summary of the main effects on performance measures

		interface	collection	working memory(W) or gender(G)
success				
-task_success	lookup		$\chi(1)^2$=25.757, p<.001	W: $\chi(1)^2$=5.516, p=.019
	exp	Q(2)=10.167, p=.006	-	-
-selected_pictures	exp	$\chi(2)^2$=14.643, p=.001		
search time	lookup		$\chi(1)^2$=13.182, p<.001	
	exp	χ(2)2=6.364, p=.042		
average_time	exp	$\chi(2)^2$=19.694, p<.001		
total action	lookup		$\chi(1)^2$=11.393, p=.001	
	exp(all cases)	$\chi(2)^2$=7.441, p=.024		
navigation	lookup	$\chi(2)^2$=13.865, p=.001		
	exp	$\chi(2)^2$=18.479, p<.001	$\chi(1)^2$=13.101, p<.001	
sh navigation	lookup(all cases)	$\chi(1)^2$=8.728, p=.003		
tag navigation	exp	$\chi(1)^2$=10.078, p=.002		
search	lookup(all cases)		$\chi(1)^2$=15.824, p<.001	
	exp	$\chi(2)^2$=28.474, p<.001	$\chi(1)^2$=15.672, p<.001	
futile search	lookup(all cases)		$\chi(1)^2$=3.951, p=.047	G: $\chi(1)^2$=4.541, p=.033
	exp	$\chi(2)^2$=7.615, p=.022	$\chi(1)^2$=5.467, p=.0194	
back tracking	lookup(all cases)		$\chi(1)^2$=7.219, p=.007	
	exp		$\chi(1)^2$=5.654, p=.01702	
	exp(all cases)	$\chi(2)^2$=6.310, p=.043		

(exp is the abbreviation of exploratory; dashes "-" indicate no test was performed; blank indicates no significance found)

5.2 SUBJECTIVE PERCEPTION ANALYSIS

To better understand the participants' perception of each interface, this section focuses on analyzing user feedback from the post questionnaires, which consists of post-task questionnaires and a post-experiment questionnaire. Following each interface session, each participant was asked to take a work load survey, NASA TLX, and to evaluate the interface's supportiveness in each post-task questionnaire. Furthermore, the participant was also requested to evaluate her/his confidence in the just previously used interface. After using three interfaces, the participant was required to self-evaluate his/her satisfaction with the three interfaces on the post-experiment questionnaire. To understand which interface the participant preferred or felt confident to use for different purposes, the post-experiment questionnaire also asks several questions in order to compare the proposed interface with the others at the end of the experiment.

5.2.1 Study variables

To analyze subjective perception, this section classifies the subjective opinions into three parts. From the post-task questionnaire, the first part focuses on each participant's assessment of the workload and the supportiveness of each interface for the different tasks (lookup and exploratory). The outcome of workload assessment is a continuous value calculated by NASA TLX and the supportiveness is rated by the participants on a five-level Likert scale (1= not supportive to 5= very supportive). Apart from the experimental conditions and demographic variables, search time is included as one covariate since users' perception might be affected by their performance. The variables for the first part are shown in Table 22.

Table 22. The variables of the workload and supportiveness analyses

Dependent Variables (DV)		Independent Variables (IV)	
Subjective perception post task	• Workload • Supportiveness	• Search_time	Performance variables

		• interface :
		  ○ Subject Heading only (SH)
		  ○ Tag only (Tag)
		  ○ Dual Perspective Navigation (Dual)
		• search_type :
		  ○ Lookup
		  ○ Exploratory (Exp)
		• collection :
		  ○ Teenie Harris (TH)
		  ○ Flickr
		• interface_order
			Experimental conditions

		• working_memory • native • gender • major	Subject demographics

Second, this study emphasizes how the participant evaluates each interface. The outcomes, satisfaction and confidence, are ordinal variables rated by the participant on a five-level Likert scale (1= not at all to 5=very satisfied/confident). In addition, I collected information on which interface users felt more confident using and why they felt so. Since this part aims at investigating participant's overall experience (3 lookup tasks and 1 exploratory task) with the three interfaces, there are not any performance variables included (Table 23).

The third part of this section targets the subjective perception about the proposed interface. Each participant selects one interface from the set of three interfaces for 1) preferred

interface, 2) preferred interface for different types of tasks (lookup and exploratory), and 3) preferred interface to recommend to cultural heritage institutions.

Table 23. The variables of the satisfaction, confidence, and recall analyses

<table>
<tr><td colspan="2">Dependent Variables (DV)</td><td colspan="2">Independent Variables (IV)</td></tr>
<tr><td rowspan="2">Subjective Perception about the three interfaces (post experiment)</td><td rowspan="2">• satisfaction (ordinal)
• confidence (ordinal)</td><td>• interface :
○ Subject Heading only (SH)
○ Tag only (Tag)
○ Dual Perspective Navigation (Dual)
• collection :
○ Teenie Harris (TH)
○ Flickr
• interface_order</td><td>Experimental conditions</td></tr>
<tr><td>• working_memory
• native
• gender
• major</td><td>Subject demographics</td></tr>
</table>

5.2.2 Models and results

5.2.2.1 Workload and supportiveness.

The first model of the subjective perception analysis takes into account the effect of search time because participants' perception might be affected by their performance. However, two search types are quite different with different time limitations (3 minutes for each of the three lookup tasks and 10 minutes for the exploratory task) and requirements (1 picture for each of the three lookup tasks and 12 pictures for the exploratory task). Therefore, I still keep the same method of

splitting cases based on the search_type. Two outcome variables, workload and supportiveness, were examined in this section. Table 24 shows the descriptive statistics of two measures in two search types with three interfaces.

Workload. To examine the hypothesis **"**H2-5: Users will have less mental effort when using the proposed approach", this study uses NASA TLX to collect users' workload evaluations after they experienced each type of tasks with each interface. The outcome of workload is a continuous dependent variable that calculated by NASA TLX. The Gama distribution can well interpret the distribution of search time. Therefore, GEE Gamma was applied to predict workload in the system given the experimental conditions, subject demographics, performance and the interaction between interface and working_memory.

Table 24. Descriptives (mean±SE) of workload and supportiveness by search type and interface

Search Type	Measure	SH	Tag	Dual
Lookup	cases	54	54	54
	workload	57.81±2.62	55.01±2.93	55.37±2.70
	supportiveness	3.78±.11	3.63±.14	4.06±.11
Exploratory	cases	54	54	54
	workload	49.33±2.27	54.01±2.63	46.30±2.24
	supportiveness	4.26±.10	3.81±.12	4.24±.10

There are significant effects of native $\chi(1)^2$=6.364, p=.012, and search_time $\chi(1)^2$=61.703, p<.001 for the lookup task. The same effects are also found in the exploratory task, native $\chi(1)^2$=6.159, p=.013, and search_time $\chi(1)^2$=25.323, p<.001(Table 25). For both types of search tasks, the participants who are not native English speakers had 23% (OR=1.23,

p=.012; OR=1.233, *p*=.013) more workload to complete either the lookup or exploratory task than the participants who are native English speakers. The language effect influences how users perceive the workload from the assigned tasks.

Table 25. Significant effect influencing workload for the lookup and exploratory tasks

Search type	Parameter		S.E.	Wald χ^2	sig	Exp(B)
Lookup	native	Non-native	.0821	6.364	.012	1.230
Exploratory	native	Non-native	.0844	6.159	.013	1.233

Supportiveness. To examine the supportiveness of each interface, I investigated participants' subjective opinion about system supportiveness by collecting their rating on a five-level Likert scale (1= not supportive to 5= very supportive). Since the outcome of supportiveness is ordinal response, the ordinal logistic can describe the distribution of these variables best so I applied ordinal logistic in the GEE model to predict users' ratings on supportiveness of systems given the experimental conditions, subject demographics, performance and the interaction between interface and working_memory.

The effect of interface_order is found for both lookup $\chi(2)^2$=15.337, p<.001 and exploratory $\chi(2)^2$=8.370, p=.015 tasks. The result in Table 26 reveals that the participants felt 14.8% (OR=1.148, *p*<.001) and 9.2% (OR=1.092, *p*=.004) more supportive in the first attempt than in the last attempt when they tried to complete the lookup and exploratory tasks. The association between interface order and the supportiveness is shown in Figure 22. Since a participant evaluates every interface right after his/her experience with that interface, s/he might

increase her/his expectation as s/he experiences more. The contrast effect might cause them to evaluate the last interface more harshly than the prior ones (Plous, 1993).

Table 26. Significant effect influencing supportiveness for the lookup and exploratory tasks

Search type	Parameter		S.E.	Wald χ^2	sig	Exp(B)
Lookup	interface_order	1	.0386	12.788	<.000	3.606
		2	.0388	1.006	.395	1.355
	interface	SH	.0359	4.390	.027	.455
		Tag	.0381	15.082	.001	.291
	gender	female	.0310	7.947	.035	1.939
Exploratory	interface_order	1	.0300	8.505	.004	2.667
		2	.0310	.781	.091	1.742
	interface	SH	.0284	.375	.544	1.236
		Tag	.0262	5.820	.027	.517
	major	Non-computer related	.0331	4.822	.019	.410

When completing the lookup task, there is a significant effect of interface $\chi(2)^2$=11.138, *p*=.004. The participants using the subject heading-only interface felt 45.5% (OR=.455 *p*=.027), and the participants using the tag-only interface felt 29.1% (OR=.291, *p*=.001) of the supportiveness that the participants using dual-perspective navigation interface did. There is also an effect of gender $\chi(1)^2$=4.466, *p*=.035. In comparison with the male participants, the female participants felt 93.9% (OR=1.939, *p*=.035) more supportiveness from the interfaces.

When completing the exploratory task, an effect of interface $\chi(2)^2$=8.020, *p*=.018 was also found. The participants using the tag-only interface felt 51.7% (OR=.517, *p*=.027) of the supportiveness that the participants using the dual-perspective navigation interface felt. In

addition, the effect of major $\chi(1)^2$=5.486, p=.019 is significant. The participants with a non-computer related major felt 41.0 % (OR=.410, p=.019) of the supportiveness that the ones with a computer related major felt.

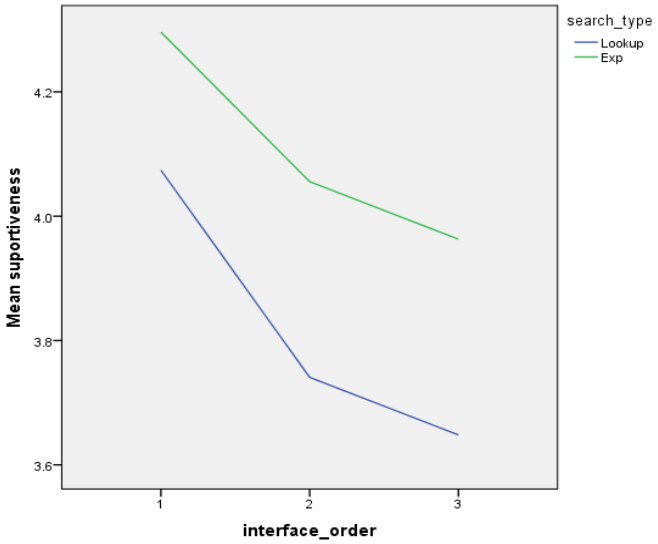


Figure 22. The relationship between supportiveness and interface order

5.2.2.2 Satisfaction and confidence

To understand whether users will be more satisfied with the proposed approach (H3-2: Users will be more satisfied with the proposed approach) and be more confident in performing a search task with the proposed approach (H3-1: Users will be more confident in performing a search task with the proposed approach), two sets of questions were designed for collecting users' satisfaction and confidence after each participant experienced each interface. Table 27 shows the descriptive statistics of satisfaction and confidence with three interfaces.

Table 27. Descriptives (mean±SE) of satisfaction and confidence by interface

Measure	SH	Tag	Dual
cases	54	54	54
satisfaction	4.00±.09	3.20±.14	4.67±.07
confidence	4.06±.10	4.06±.09	4.19±.09

The outcomes, satisfaction and confidence, are ordinal variables rated by the participant on a five-level Likert scale (1= not at all to 5=very satisfied/confident). Therefore, ordinal logistic GEE model was applied to predict users' satisfaction and confidence given the experimental conditions and subject demographics.

Satisfaction. The questions about satisfaction were asked in the post-experiment questionnaire after experiencing all the three interfaces. The participants rated their satisfaction from the scale 1 (not at all) to 5 (very much). The result in Table 28 shows that there is an effect of interface $\chi(2)^2=70.910$, $p<.001$ on satisfaction. The participants using the subject heading-only interface had 12.5% (OR=.125, $p<.001$) and the participants using the tag-only interface had 21% (OR=.021, $p<.001$) of the satisfaction that the participants using the dual-perspective navigation interface had. The participants were very satisfied with the dual-perspective navigation interface. However, there were no other significant effects found.

Confidence. A set of questions about confidence were asked after each interface's experience in the post-task questionnaire and the participants were asked to rate their confidence on a scale from 1 (not at all) to 5 (very much). In addition, this study designed two more confidence questions at the post-experiment questionnaire to get to know which interface the user would feel more confident to use and why they feel so after three different interfaces experiences.

Table 28. Significant effect influencing satisfaction and confidence

	Parameter		S.E.	Wald χ^2	sig	Exp(B)
satisfaction	interface	SH	.3797	29.928	<.000	.125
	interface	Tag	.4675	69.016	<.000	.021
confidence	interface_order	1	.2873	4.949	.026	1.895
	interface_order	2	.3183	.009	.926	1.030

For the confidence evaluation from the post-task questionnaire, there is only a marginal effect of interface order $\chi(2)^2$=5.880, p=.053, and the rest of the effects are not significant. The participants' confidence decreases while the interface order increases as shown in Figure 23. Since the questions of confidence were asked right after each interface experience just like the supportiveness questions, the contrast effect might also occur to make users evaluate confidence with different standards along the time.

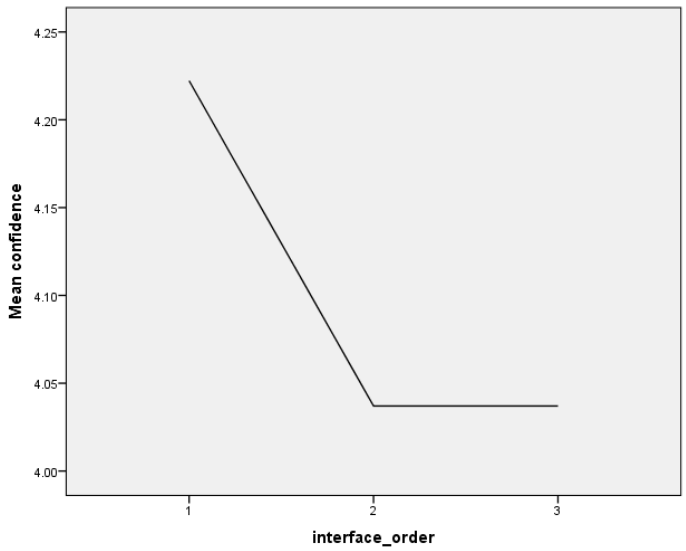


Figure 23. The relationship between confidence and interface order

For the perception of the most confident interface from the post-experiment questionnaire, 85% (47 out of 54) of the participants chose the dual-perspective navigation interface, 11% (6 out of 54) the subject heading-only interface, and only 3.7% (1 out of 54) chose the tag-only interface as the most confident interface for future use.

The comments of the confident interface are grouped by the type of the interface as following:

Confident with the subject heading–only interface. Six participants felt confident to use the subject heading-only interface in the future. Based on their feedback, they attributed their confident interface for the subject heading-only interface to it being more organized to use than the tag-only interface. They also felt that it was easier to follow the structure to find what they want without guessing tags' meanings. Sometimes tags could mislead if the tagger didn't understand the subject matter.

Confident with the tag-only interface. Among the 54 participants, only 1 felt confident to use the tag-only interface. The reason the user gave for choosing this interface was that the tag-only interface gives an overview of the image results and the user can easily capture summary words for searching.

Confident with the dual-perspective navigation interface. Most of the participants (85%) felt confident to use the dual-perspective navigation interface. The reason for this can be categorized into three aspects. First, this interface provided them with diverse information/options to use with different needs. Second, the participants opined that this interface is more powerful to switch their search methods easily between professional categories and social tags. The third aspect is that the subject headings and tags can complement and support each other seamlessly to assist users' search.

5.2.2.3 Preference

In the post-experiment questionnaire, the participants were asked for their opinions about the three different interfaces. At that moment, the participants had gained practical experience with the three interfaces for two types of search tasks. As shown in Table 29, 79.6% of participants preferred the dual-perspective navigation interface, 18.5% the subject heading-only interface, and only 1.9% (one participant) preferred the tag-only interface.

Table 29. The summary of frequencies (percentages) of the diverse preferences

Interface	SH	Tag	Dual
1) Preferred interface (which one of the interfaces did you like/prefer most?)	10(18.5%)	1(1.9%)	43(79.6%)
2a) Preferred interface for lookup tasks (which one of the interfaces would you prefer for lookup search?)	9(16.7%)	10(18.5%)	35(64.8%)
2b) Preferred interface for exploratory tasks (which one of the interfaces would you prefer for exploratory search?)	9(16.7%)	2(3.7%)	43(79.6%)
3) Preferred interface to recommend to cultural heritage institutions (which one of the interfaces would you suggest that cultural heritage institutions provide their visitors for searching their image collections?)	8(14.8%)	4(7.4%)	42(77.8%)

Preferred the subject heading-only interface. Overall, there were ten participants who preferred the subject heading-only interface. Based on their feedback from the open-ended

question on why they preferred this particular interface, they stated that this interface is more organized and easier to use than the tag-only interface.

"It can let me locate one big category, then narrow down to many sub-categories." – P4

"There is no structure of tags. The subject headings has a structure to lead me to sub-categories." -P35

Preferred the tag-only interface. Only 1 participant preferred the tag-only interface. The participant was in favor of this interface because tags can provide direct social path as a reference and inspire him/her to think of additional keywords to search.

"tags are shown in the middle side of the webpage, it is easy for people to find; tags are in different sizes which can indicate the frequency that people are preferred to search."-P29

Preferred the dual-perspective navigation interface. Among the 54 participants, about 78% of the participants preferred the dual-perspective navigation interface. The reason for the preference can be grouped into aspects. First, having subject headings and tags together on the dual-perspective navigation interface can complement these features' weakness.

"Subject headings give structure but tags add missed information on meta-data" – P1

"Provides the clarity, precision, and organization of the subject headings and well as the user-based intuition of the tags." –P52

Second, the participants opined that this interface provided them with more information and inspired them to search.

"Gives multiple options to base search off of. Gave more hints to what to search for next."-P2

"There are more categories and hints that can help me to find objects."-P9

"Because it provides a combination of different supports. More information can help find the most suitable keywords for search and narrow down the scope."-P14

Third, the dual-perspective navigation interface is more powerful by allowing users to switch their search strategies.

"tags give initial thought about the topics, and subject heading gives consistent idea about the relevant images"- P7

"they complement each other. For example, for organized search, subject-headings are quite useful for finding a large number of related images. On the other hand, if I have a specific term in mind already, the tags are useful for rephrasing the term." – P13

"It provides both professional and social information, the former one is highly structured, if I know a lot of details of the things I am searching for, it is efficient to use it; the latter has different fonts to show relevance, which provide better hints when I know less details." – P31

In addition, the preferences on "forward looking" questions were also designed to assess user preferences in future situations for different types of search tasks or for recommending to the cultural heritage sector. For the future preferences, the dual-perspective navigation interface is the most preferred interface for both tasks and for usage in cultural heritage institutions. It is interesting that the participants had divergent preferences when they were asked about preferences in general and for each specific task. The one participant who favored the tag-only interface in the prior question switched to the dual-perspective navigation interface for both tasks to get more support in the future. He suggested the subject heading-only interface to cultural heritage institutions since the visitors might expect more organized information as they usually get from the museums. Eight participants who favored the dual-perspective navigation interface in the prior preference switched to the tag-only interface for lookup search since tags can provide direct feedback to give them ideas to issue suitable queries. Most of these eight participants switched back to the dual-perspective navigation interface for exploratory search. Only one of these eight switched to the subject heading-only interface for exploratory search. The reason he claimed is that he preferred using the tag-only for the lookup search and the subject heading-only for the exploratory search so he chose the dual-perspective navigation interface in general which is more powerful for different purposes.

To recommend to cultural heritage institutions, the dual-perspective is the most preferred interface since it is information rich with both experts' and novices' ideas about the image context. In addition, the dual-perspective navigation interface can satisfy different participants with different background and different needs.

> *"People can get a mix of the factual information associated with the items and socially descriptive tags of the items to see which suits them the best" –P1*

> *"Category information given by professionals is precise however harder to be understood than tags. Use both of them can complement a lot." – P10*

> *"It provides both professional and social information, the former one helps to find information that has obvious different kinds of entities in a hierarchical way, or is highly structured; the latter helps to find information that have not obvious features, and proves more variability to use other search words." – P31*

The second preferred interface to recommend to cultural heritage institutions is the subject heading-only interface. Because cultural heritage is a professional domain and usually provides precise guidance, visitors would expect to use a more organized way to search for information in the museums.

> *"For museum the main reference to help visitor view the page in different category should be the official one. Tag should be more like the place people provide their opinion but to put it on the main is un-necessary." –P35*

> *"More professional to use subject headings. Public tags could be quite inaccurate because a member of the public might title an event or a person as some other place/time/individual"- P54*

Four participants preferred to suggest tag-only interface to cultural heritage institutions because tags are more likely close to visitors' language and easier to be applied to search for the images.

"Tags are an intuitive way for people to classify and think about images, and may be more helpful that specific categories or subjects, and having the latter may confuse some users "-P32

"The tag-only interface is more approachable. The dual may seem overwhelming, in terms of learning curve."-P51

5.2.3 Summary of subjective perception analysis

The result of subjective perception analysis (Table 30) shows that the effect of interface is important on supportiveness, satisfaction and recall. The dual-perspective navigation interface provided significant support for both search tasks and participants were significantly more satisfied with this interface. Although the main effect of interface was not found on confidence, this interface was selected as the interface in which the participants were most confident for future use and also as the top choice to be recommended to cultural heritage institutions.

Table 30. The summary of the main effects on subjective perception measures

		interface	Interface order	native	gender	major	search time
workload	lookup			$\chi(1)^2=6.364$, $p=.012$			$\chi(1)^2=61.703$, $p<.001$
	exp			$\chi(1)^2=6.159$, $p=.013$			$\chi(1)^2=25.323$, $p<.001$
supportiveness	lookup	$\chi(2)^2=11.138$, $p=.004$	$\chi(2)^2=15.337$, $p<.001$		$\chi(1)^2=4.466$, $p=.035$		$\chi(1)^2=30.64$, $p<.001$
	exp	$\chi(2)^2=8.020$, $p=.018$	$\chi(2)^2=8.370$, $p=.015$			$\chi(1)^2=5.486$, $p=.019$	$\chi(1)^2=18.894$, $p<.001$
satisfaction		$\chi(2)^2=70.910$, $p<.001$					
confidence			$\chi(2)^2=5.880$, $p=.053$				

6.0 DISCUSSIONS

In this chapter, I revisit the research questions with the main hypotheses of this study and discuss all the statistical evidence to support each of them.

Table 31. The summary of performance analysis

	Task type	**interface**	**collection**	**working memory**	**gender**
success					
-task_success	lookup		**T>F*****	L>H*	
	exp	D>Tag**	-	-	-
-selected_pictures	exp	D> SH*, D>Tag**			
search time	lookup		**T<F*****		
	exp	D<SH*, D<Tag**			
average_time	exp	D<SH*, D<Tag**			
total action	lookup		T<F**		
	exp(all cases)	D<SH**, **D<Tag*****			
navigation	lookup	D>Tag**			
	exp	**D>Tag*****	**T<F*****		
sh navigation	lookup(all cases)	D<SH**			
tag navigation	exp	D<Tag**			
search	lookup(all cases)		**T<F*****		
	exp	**D<Tag*****	**T>F*****		
futile search	lookup(all cases)		T<F*		F>M*
	exp	D<Tag*	T>F*		
back tracking	lookup(all cases)		T<F**		
	exp		T>F*		
	exp(all cases)	D<SH**, D<Tag**			

(D= Dual-perspective navigation interface, SH= Subject Heading-only interface, Tag= Tag-only interface, T= Tennie Harris, F= Flickr, L= Low working memory, H= High working memory, F=Female, M=Male, *=significant at p<.05, **=significant at p<.01, ***=significant at p<.001. "all cases" include successful and non-successful cases, otherwise only successful cases are included)

In this chapter, I revisit the research questions with the main hypotheses of this study and discuss all the statistical evidence to support each of them. In this chapter, I revisit the research questions with the main hypotheses of this study and discuss all the statistical evidence to support each of them. Table 31and Table 32 reveal the results of the performance and subjective perception analyses. They present the significant effects of different predictors (title row) on all measures (title column). The association with α level of .001 is highlighted in bold. Blank indicates there was no significant effect found and dashes indicate no test was performed.

Table 32. The summary of subjective perception analysis

	Task type	**interface**	**Interface order**	**native**	**gender**	**major**
workload	lookup			NN>N*		
	exp			NN>N*		
supportiveness	lookup	SH<D	**1>3*** **		F>M*	
		Tag<D**				
	exp	Tag<D*	1>3*			NC<C*
satisfaction		**SH<D*** **				
		Tag<D* **				
confidence			1>3			

(D= Dual-perspective navigation interface, SH= Subject Heading-only interface, Tag= Tag-only interface, NN= Non-native English speaker, N= Native English speaker, F=Female, M=Male, NC= Non-computer related major, C= Computer related major, *=significant at p<.05, **=significant at p<.01, ***=significant at p<.001)

This chapter is organized into three parts based on the general construct of usability with objective and subjective factors. The dimensions of effectiveness and efficiency contain objective factors such as search times, completion rates and errors. The dimension of users' perception analyzes subjective opinions of satisfaction and confidence to further understand how users perceive their experience with the interfaces and tasks.

6.1 EFFECTIVENESS

This section focuses on the first research question, "Does the dual-perspective navigation approach provide better information to help users to achieve their goals in an effective manner than the single-perspective approaches?" To answer this question, three hypotheses were tested.

Hypothesis 1-1: Users will successfully complete more tasks with the proposed approach. The measure of task success indicates that the dual-perspective navigation interface is significantly more effective in completing the assigned exploratory task. There is no significant effect of interface on the lookup task. Participants had different search strategies to apply to complete the assigned tasks due to different time constraints (three minutes for the lookup search and 10 minutes for the exploratory search) and different requirements for the two types of search tasks. For the lookup search, they usually tried the way there were most familiar with for their daily search. Therefore, they often started with issuing several queries and then used the navigation support if they did not get the right images after a few searches with their own queries. However, it appears that some participants experienced time constraints as more restrictive for the lookup task so they were hesitant to try the navigation features in order to complete the task in time. Although participants were provided sufficient time, according to the pilot study result, to perform each lookup task, in the post-experiment questionnaire and interview, some participants mentioned that they personally felt time pressure while they performed the lookup task.

In terms of completing the exploratory search task, both measurements, task success and the number of selected pictures, indicated that different interfaces had significant effect on users' performance. The participants not only issued more queries but also explored the functionalities on each interface more frequently than when they performed lookup tasks (See Table 33). After

trying their own search method, they usually tended to apply subject headings or tags to get to their desired items. Since the exploratory search requires 12 relevant images, having a better idea about the context usually helps the participants accomplish the task more easily. Therefore, I can speculate that providing more information to express the aboutness of images is more useful for the exploratory tasks.

Table 33. Mean of navigation and search actions among interfaces on two search types

Search type	Measure	SH	Tag	Dual
lookup	Navigation	2.48	1.25	2.71
	search	3.72	4.22	3.63
exploratory	Navigation	6.22	4.46	6.67
	search	5.76	9.46	5.20

Hypothesis 1-2: Users will make less futile searches with the proposed approach. The measure of futile searches was used to examine this hypothesis. For the exploratory search tasks, participants failed to get the retrieved results back more frequently with the tag-only interface than with the dual-perspective navigation interface. As the result in Table 31, this might be correlated to participants' higher level of search actions. Participants issued queries significantly more frequently with the tag-only interface than with the dual-perspective navigation interface. This shows that participants searched a lot when they performed with the tag-only interface. With the tag-only interface, participants completed the exploratory task with more interactions in which searching and receiving empty results are relatively higher than with other interfaces. The information scent, tags, of the tag-only interface was not as supportive as I

expected. Participants tended to issue queries and manipulated with queries when they performed with the tag-only interface.

Hypothesis 1-3: Users with different working memory capacity will not have different search performance with the proposed approach. To examine this hypothesis, I tested an interaction effect by interface and working memory in the model of performance analysis. There is no significant effect among all the measures. However, the effect of working memory is found on task success for the lookup task. Contrary to general expect, participants with low working memory capacity actually performed better for the lookup tasks: They could find the assigned picture in a shorter time and with fewer actions, when they found the picture.

6.2 EFFICIENCY

This section discusses the efficiency dimension to answer the second research question, "Does the dual-perspective navigation approach guide users to their targets with less resources required than the single-perspective approaches?" In information foraging theory, searchers navigate through information patches to find what they need. With diverse information descriptors provided as information scent, only the clearest indication (strongest scent) can quickly lead users closer to the information they require. To see whether the dual-perspective navigation interface provides strong scent, search time and the interactions with scent are important to investigate. In addition, crafting an interface with optimal levels of information scent can reduce the mental effort which users have to expend to find their desired resource. Therefore, the workload of using interfaces is also an important factor to explore. The following hypotheses are tested to address these issues.

Hypothesis 2-1: Users will spend less time to find a target item with the proposed approach. The measure of search time in the lookup search can be used to answer this research question. There is no significant difference found in the effect of interface. However, the effect of collection was found to be significant. It might indicate that the total number of pictures in the collection might influence participants' performance on finding an exactly the same picture in the lookup task. In addition, the measure of average time of selecting a required picture in the exploratory search (section 5.1.2.2) was supporting that users demand less time to find a required picture with the dual-perspective navigation interface than with the other interfaces.

Hypothesis 2-2: Users will spend less time to complete a task with the proposed approach. This hypothesis is supported by the measure of search time for the exploratory search. Participants spent significantly less time with the dual-perspective navigation interface than with the compared interfaces, subject heading-only and tag-only. The dual-perspective navigation interface can assist users to accomplish their goal more efficiently. It might show that users got more support to accomplish their task goal when the system provides two descriptors from both experts' and general public's sides.

Hypothesis 2-3: Users will reach the task goal with fewer interface interactions with the proposed approach. Analysis of log data shows that there was no different interaction found when participants achieved their task goals of either lookup or exploratory tasks among three interfaces. However, when I analyzed users' actions of the exploratory tasks including all cases, the interface effect was found significantly different. It indicated that users who succeed in the exploratory task didn't have difference among effects but users who failed to achieve the task goal were influenced by the effect of interface. To understand differences in user behavior across

tasks and interfaces, I performed a deeper analysis of users' interactions (Figure 24) from the counts of diverse actions, such as search, navigation, back tracking etc.

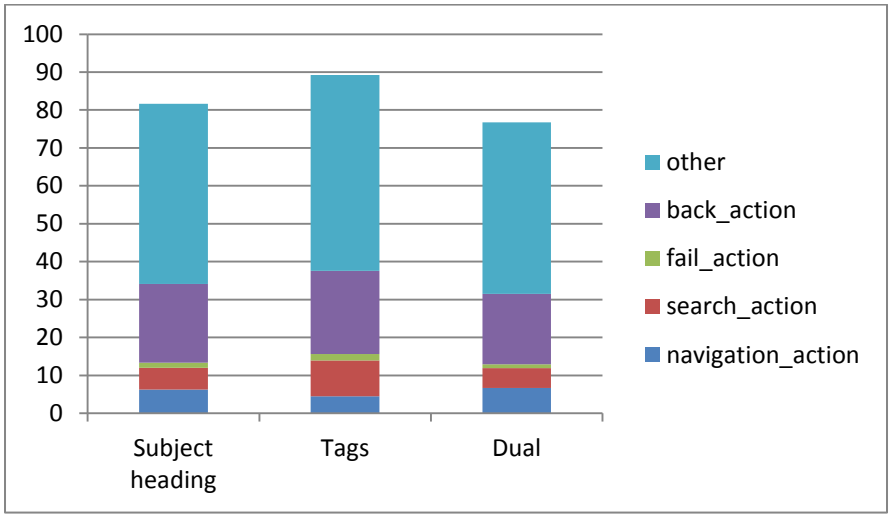


Figure 24. Different types of interactions for the exploratory search task

For the exploratory task, users issued significantly more search queries and significantly less navigation features when they performed the tasks with the tag-only interface than with the dual-perspective navigation interface. It shows that the tag-only interface providing tags only might not be sufficient information scent to guide the users to the target items so users rarely applied the tag navigations.

In terms of the lookup task, I didn't find any significant differences when I only consider successful cases. When I analyzed navigation behavior with all cases, I found there was a significant interface effect between subject heading-only and the dual interfaces on subject heading navigations. This indicates that users who failed the task might be affected by interfaces

when they utilized the subject heading features. They used more subject headings with the subject heading-only interface but those actions didn't lead them to the successful finding.

Hypothesis 2-4: Users will use less back tracking during search with the proposed approach. Information foragers search for a scent trail and follow it toward their desired resource. When they lose the scent, they often return by the same route until they encounter a better scent or give up. If they have to backtrack most of the time, the circumstance might indicate that information scent is not clear enough to provide users with a good direction.

The measure of back tracking action is used to examine whether the proposed approach has clear scent to direct users to what they need. The result reveals that participants applied less back tracking during search with the dual-perspective navigation interface than with the other interfaces. The proposed approach has relatively strong information scent to lead users to the target resources.

Hypothesis 2-5: Users will expend less mental effort when using the proposed approach. An interface with a well-designed information scent can reduce the mental effort users have to expend to find desired images. Mental workload is an essential measure to examine the continuous demand on a user's attention to accomplish a task.

The result of the current study does not provide any evidence that the proposed interface can allow users to have significantly less mental effort for either the lookup tasks or the exploratory tasks. However, the effect of language was found on the measure of workload. Non-native English speakers expended more mental effort than the native English speakers while they used the different interfaces. This might be correlated to the language I applied to present the descriptors. The descriptors are in English and describe the aboutness of the image context so

participants with better English ability might expend less mental effort to consume the information provided from the systems.

6.3 SUBJECTIVE PERCEPTION

Beyond the investigation of users' performance, this section focuses on subjective perception by answering the third research question, "RQ3: Does the dual-perspective navigation approach make users more confident of their image finding ability and leave them with a positive perception of the approach?" Two related hypotheses were tested.

Hypothesis 3-1: Users will be more confident in performing a search task with the proposed approach. Usability studies have found that if users are confident with the information scent provided by a system, they tend to believe they are still on the pathway to their goal so they will keep using the system.

The assessment of users' confidence across the three interfaces is provided by analyzing 1) their explicit rating for each interface on the post-task questionnaire and 2) their confident interface selection and comments on the post-experiment questionnaire. Unfortunately, there is neither an effect of interface nor other effects on confidence found in the rating data. Only a marginal effect of interface order was found. However, a positive result for selecting an interface which users felt confident was found for the proposed interface from the post-experiment questionnaire.

The post-experiment questionnaire offered me an opportunity to get users' perception after they had gained practical experience with both tasks and all the three types of interfaces. When asked the question "Which one of the interfaces would you feel more confident to use for

other search tasks?", 85% of the participants chose the dual-perspective navigation interface, 11% the subject heading-only interface, and only 3.7% chose the tag-only interface as the most confident interface for future use. The reason can be categorized into three aspects from these participants' comments. First, this interface provided them with diverse information/options to use with different needs. Second, the participants felt that this interface is more powerful for transitioning between search methods easily, i.e., between professional categories and social tags. Lastly, the subject headings and tags can complement and support each other seamlessly to assist users' search.

Hypothesis 3-2: Users will be more satisfied with the proposed approach. The measure of satisfaction shows that participants were more satisfied with the dual-perspective navigation interface than with the other two interfaces. Further support for this assessment of users' satisfaction across the three interfaces is found in the preference analysis (section 5.2.2.4). Most of the participants (77.8%) preferred the dual-perspective navigation interface, 18.5% the subject heading-only interface, and only 3.7% (one participant) preferred the tag-only interface in general. Although they had different preferences for different tasks, the vast majority of the participants still preferred to use the dual-perspective navigation interface.

Overall, this study demonstrated that the dual-perspective navigation framework supports users to find images more effectively and efficiently when they perform exploratory search. This framework gives them more useful guidance to find what they want in an unfamiliar context. Meanwhile, users became more confident of their image finding ability when they used the dual-perspective navigation framework and also had a positive experience using it. However, if this study could have focused on only the exploratory task and eliminated the lookup task, this study could have been designed with fewer constraints required by balancing potentially confounding

variables. For the exploratory search, the effect of collection wasn't an essential factor to influence users' performance so eliminate the small collection of the Teenie Harris might be also a way to reduce one of the limitations of this study.

7.0 CONCLUSIONS

7.1 SUMMARY OF RESULTS

This study was conducted to explore whether images can be found more efficiently and effectively when two types of information descriptors (subject headings and social tags) are provided to the users in a combined manner, the dual-perspective navigation framework. This study contributes to the enhancement of image findability by providing comprehensive navigational support and combining experts' and general users' content annotations. Three research questions (RQ) were addressed in this study:

RQ1: Does the dual-perspective navigation approach provide better information to help users achieve their goals in an effective manner than the single-perspective approaches?

RQ2: Does the dual-perspective navigation approach guide users to their targets with fewer resources required than the single-perspective approaches?

RQ3: Does the dual-perspective navigation approach make users more confident of their image finding ability and leave them with a positive perception of this approach?

A controlled experiment was designed to address the above research questions. This study compared user performance and feedback for three types of image finding interfaces in the context of two types of search tasks – lookup and exploratory search. The empirical evidence shows that the dual-perspective navigation interface outperformed the other interfaces (the

subject heading-only interface and the tag-only interface). Both objective performance analysis and subjective perception analysis produced significant findings.

From the performance analysis, significant differences were found for the dual-perspective navigation interface when used in the exploratory search context. The dual-perspective navigation interface was found to be significantly more effective than the tag-only interface in terms of task success and futile search. It was also found to be significantly more efficient than both subject heading-only and tag-only interfaces on search time, total actions, and back tracking actions. The result of this study also shows that participants spent longer time to interact with the tag-only interface including significant larger amounts of search actions, futile searches and back tracking actions compared to the use of the dual-perspective navigation interface.

From the subjective perception analysis, the participants indicated that the dual-perspective navigation interface provided significant support for both types of search tasks and reported significantly higher levels of satisfaction. They also selected the dual-perspective navigation interface as the interface in which they were most confident for future use and the most preferred interface for different tasks. This interface was also the top choice to be recommended to museum professionals. However, the result of this study didn't show that the dual-perspective navigation interface can help users capture more semantic information than the other interfaces.

In summary, the dual-perspective navigation system has proven to be more effective and efficient than the subject heading-only and tag-only interfaces for exploratory tasks. This finding can assist interface designers struggling with what information is best to help users and facilitate the searching tasks. Although this study explicitly focuses on image search, the result may be

applicable to wide variety of other domains. The lack of textual content in image system makes them particularly hard to locate using traditional search method. While the role of professionals in describing items in a collection, the role of the crowd in assigning social tags augments this professional effort in a cost effective manner.

7.2 FUTURE WORK

This study considered several effects to predict a variety of measures. In performance analysis, the significant effect of collection was found mainly in lookup search on task success, search time and total action. An interesting finding is that this effect had different directions between lookup and exploratory search on search action, futile search and back tracking action. To investigate the phenomenon of collection effect, a further analysis on what users saw and used while they performed different search tasks might be required. An eye tracking study might help to gain more insight into the effect of collection on users' performance.

Moreover, although the dual-perspective navigation framework provide two perspectives of descriptors to guide users to find image items effectively and efficiently, this study still can't claim to have uncovered how different information descriptors guide each user to find his/her desired images. Each participant has different background and preferences for using different information descriptors. They might have different preferred search strategies to perform a particular search task with different information needed to execute their preferred strategy in the manner to which they are accustomed. How the participants consumed the different information descriptors during their search process would be interesting to explore in further research.

Further study with eye tracker augmentation might be helpful to learn more about the interaction between users' search behavior and different types of information descriptors.

BIBLIOGRAPHY

Al-Khalifa, H. S., & Davis, H. C. (2007a). Exploring the value of folksonomies for creating semantic metadata. *International Journal on Semantic Web and Information Systems IJSWIS*, *3*(1), 12–38. Retrieved from http://eprints.ecs.soton.ac.uk/13555/

Al-Khalifa, H. S., & Davis, H. C. (2007b). FAsTA: A folksonomy-based automatic metadata generator. (E. Duval, R. Klamma, & M. Wolpers, Eds.)*International Journal*, *2*(27thMarch 2007), 414–419. Retrieved from http://eprints.ecs.soton.ac.uk/14186/

Angeletou, S., Sabou, M., Specia, L., & Motta, E. (2007). Bridging the Gap Between Folksonomies and the Semantic Web : An Experience Report. In *Semantic Web and Web 2.0, SemNet* (pp. 30–43).

Bateman, S., Gutwin, C., & Nacenta, M. (2008). Seeing things in the clouds : The effect of visual features on tag cloud selections. In *the 9th ACM conference on Hypertext and hypermedia* (pp. 193–202).

Bates, M. J. (1988). How to Use Controlled Vocabularies More Effectively in Online Searching. *Online*, *12*(6), 45–56.

Begelman, G., Keller, P., & Smadja, F. (2006). Automated tag custering: Improving search and exploration in the tag space. *Collaborative Web Tagging Workshop at WWW2006 Edinburgh Scotland*, *22*(2), 15–33. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.120.5736&rep=rep1&type=pdf

Bischoff, K., Firan, C. S., Nejdl, W., & Paiu, R. (2008). Can All Tags be Used for Search? In *CIKM* (pp. 203–212). Napa Valley, California, USA: ACM.

Böhnstedt, D., Lehmann, L., Rensing, C., & Steinmetz, R. (2011). Automatic Identification of Tag Types in a Resource- Based Learning Scenario. In *the 6th European Conference on Technology Enhanced Laearing, EC-TEL 2011* (pp. 57–70). Springer.

Bystrom, K., & Jkrvelin, K. (1995). Task complexity affects information. *Information Processing & Management*, *31*(2), 191–213.

Campbell, I. (2000). Interactive evaluation of the ostensive model using a new Test collection of images with multiple relevance assessments. *Journal of Information Retrieval*, *2*(1), 87–114.

Candan, K. S., Di Caro, L., & Sapino, M. L. (2008). Creating tag hierarchies for effective navigation in social media. In *Proceeding of the 2008 ACM workshop on Search in social media - SSM '08* (p. 75). New York, New York, USA: ACM Press. doi:10.1145/1458583.1458597

Cantador, I., Konstas, I., & Jose, J. M. (2011). Categorising social tags to improve folksonomy-based recommendations. *Web Semantics: Science, Services and Agents on the World Wide Web*, *9*(1), 1–15. doi:10.1016/j.websem.2010.10.001

Capra, R., & Marchionini, G. (2008). The relation browser tool for faceted exploratory search. In *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*. Pittsburgh, Pennsylvania, USA.

Card, S. K., Pirolli, P., Wege, M. Van Der, Morrison, J. B., Reeder, R. W., Schraedley, P. K., & Boshart, J. (2001). Information Scent as a Driver of Web Behavior Graphs: Results of a Protocol Analysis Method for Web Usability. In *CHI '01 Proceedings of the SIGCHI conference on Human factors in computing systems*.

Cardinaels, K., Meire, M., & Duval, E. (2005). Automating metadata generation: The simple indexing interface. In *WWW2005*. Chiba, Japan.: ACM, New York. Retrieved from http://hmdb.cs.kuleuven.be/amg/publicationsFiles/paperAMG_www2005.pdf

Chan, L. M., & O'Neill, E. T. (2010). *FAST: Faceted application of subject terminology*. Libraries Unlimited.

Chi, E., Pirolli, P., Chen, K., & Pitkow, J. (2001). Using information scent to model user information needs and actions on the Web. *SIGCHI*, *3*(1).

Diriyey, A., Blandfordy, A., & Tombrosz, A. (2010). When is system support effective? In *Proceedings of the third symposium on Information interaction in context, IIiX '10* (pp. 55–64).

Duval, E., Hodgins, W., Sutton, S., & Weibel, S. L. (2002). Metadata principles and practicalities. *D-Lib Magazine*, *8*(4). Retrieved from http://www.dlib.org/dlib/april02/weibel/04weibel.html

English, J., Hearst, M., Sinha, R., Swearingen, K., & Yee, K. (2002). *Flexible Search and Navigation using Faceted Metadata*.

Fairthorne, R. A. (1969). Content analysis, specification and control. *Annual Review of Information Science and Technology*, *4*, 73–109.

Findlater, L., & McGrenere, J. (2007). Evaluating reduced-functionality interface according to feature findability and awareness. In C. Baranauskas, P. Palanque, J. Abascal, & S. Barbosa (Eds.), *Human-Computer Interaction-Interact 2007* (Vol. 9, pp. 592–605). Springer Berlin / Heidelberg.

Furner, J. (2007). User tagging of library resources: Towards a framework for system evaluation. In *World Library and Information Congress: 73rd IFLA General Conference and Council* (pp. 1–10). Durban, South Africa.

Garshol, L. M. (2004a). Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it all. *Journal of Information Science*, *30*(4), 378–391. doi:10.1177/0165551504045856

Garshol, L. M. (2004b). Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it all. *Journal of Information Science*, *30*(4), 378–391. doi:10.1177/0165551504045856

Gevins, A., Smith, M. E., & Leong, H. (1998). Monitoring working memory load during computer-based tasks with EEG patter recognition methods. *Human Factors*, *40*(1), 79–91.

Grady, C., & Lease, M. (2010). Crowdsourcing document relevance assessment with Mechanical Turk, 172–179. Retrieved from http://dl.acm.org/citation.cfm?id=1866696.1866723

Greenberg, J. (2004). Metadata extraction and harvesting: A comparison of two automatic metadata generation applications. *Journal of Internet Cataloging*, *6*(4), 59–82. Retrieved from http://ils.unc.edu/mrc/pdf/automatic.pdf

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (Vol. 1, pp. 139–183). North-Holland.

Hassan-montero, Y., & Herrero-solana, V. (2006). Improving tag-clouds as visual information retrieval interfaces. In *International Conference on Multidisciplinary Information Sciences and Technologies*.

Hayman, S., & Lothian, N. (2007). Taxonomy directed fikjsibinues integrating user tagging and controlled vocabularies for Australian education networks. *World Library and Information Congress: 73rd IFLA General Conference and Council*, (August), 1–27.

Hearst, M. (2008). UIs for Faceted Navigation Recent Advances and Remaining Open Problems. In *the Second Workshop on Human-Computer Interaction and Information Retrieval* (pp. 13–17).

Hearst, M. (2009). *Search User Interfaces*. Cambridge University Press. Retrieved from http://searchuserinterfaces.com/book/

Hearst, M. A. (2006). Clustering versus faceted categories for information exploration. *Communications of the ACM-supporting Exploratory Search*, *49*(4), 59–61. doi:10.1145/1121949.1121983

Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K., & Yee, K.-P. (2002). Finding the flow in Web site search. *Communications of the ACM*, *45*(9).

Heesch, D., & Rüger, S. (2003). Relevance Feedback for Content-based Image Retrieval: what can three mouse clicks achieve. In *Advances in Information Retrieval*. Springer Berlin / Heidelberg. doi:10.1007/3-540-36618-0_26

Helic, D., & Strohmaier, M. (2011). Building directories for social tagging systems. In *Proceedings of the 20th ACM international conference on Information and knowledge management - CIKM '11* (p. 525). New York, New York, USA: ACM Press. doi:10.1145/2063576.2063655

Helic, D., Trattner, C., Strohmaier, M., & Andrews, K. (2010). On the Navigability of Social Tagging Systems. In *Proc. of 2010 IEEE International Conference on Social Computing* (pp. 161–168). Los Alamitos, CA, USA: IEEE Computer Society. doi:http://doi.ieeecomputersociety.org/10.1109/SocialCom.2010.31

Heymann, P., & Garcia-Molina, H. (2006). Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems. *Stanford InfoLab Technical Report*, (2006-10), 1–5. Retrieved from http://ilpubs.stanford.edu:8090/775/

Hjerland, B. (2001). Towards a theory of aboutness , subject , topicality , theme , domain , field , content ... and relevance. *Journal of the American Society for Information Science and Technology*, *52*(May), 774–778.

Hjørland, B. (1992). The concept of "subject" in information science. *Journal of documentation*, *48*(2), 172–200.

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, *64*(2), 79–102. doi:10.1016/j.ijhcs.2005.06.002

Hotho, A., J, R., Schmitz, C., & Stumme, G. (2006). Information Retrieval in Folksonomies : Search and Ranking. (C. Hochberger & R. Liskowsky, Eds.)*Data Engineering*, *4011*, 411–426. doi:10.1007/11762256

Hutchins, W. J. (1977). The concept of "aboutness" in subject indexing. *Aslib Proceedings*, *30*(5), 172–181. doi:10.1108/eb050629

ISO. (1998). *Ergonomic requirements for office work with visual display terminals (VDTs)-Part 11: guidance on usaibility (ISO 9241-11:1998)*.

Jamaal, Q. (2010). Google Goggles – Use Pictures to Search The Web | Telecom News Bulletin | Telecom Magazine. Retrieved September 10, 2012, from http://telecomnewspk.com/google-goggles-use-pictures-to-search-the-web/

Jiang, W., Wan, B., Zhang, Q., & Zhou, Y. (2008). Image Search by Latent Semantic Indexing Based on Multiple Feature Fusion. *2008 Congress on Image and Signal Processing*, (1), 515–519. doi:10.1109/CISP.2008.553

Jimmy. (2013). Using Information Scent to Improve Web Usability. Retrieved from http://jimmysinsights.com/web-usability/information-scent/

Kammerer, Y., Nairn, R., Pirolli, P., & Chi, E. H. (2009). Signpost from the Masses : Learning Effects in an Exploratory Social Tag Search Browser. In *CHI*. Boston, MA.

Karlson, A. K., Robertson, G., Robbins, D. C., Czerwinski, M., & Smith, G. (2006). FaThumb : A Facet-based Interface for Mobile Search. In *CHI'06 Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 711–720). Montréal, Québec, Canada: ACM, New York. doi:10.1145/1124772.1124878

Knautz, K., Soubusta, S., & Stock, W. G. (2010). Tag Clusters as Information Retrieval Interfaces. In *2010 43rd Hawaii International Conference on System Sciences* (pp. 1–10). Ieee. doi:10.1109/HICSS.2010.360

Laniado, D., Eynard, D., & Colombetti, M. (2007). A semantic tool to support navigation in a folksonomy. *Proceedings of the 18th conference on Hypertext and hypermedia HT 07*, 153–154. doi:10.1145/1286240.1286282

Laniado, D., Eynard, D., Colombetti, M., & Milano, P. (2007). Using WordNet to turn a folksonomy into a hierarchy of concepts. *Semantic Web Applications and Perspectives 2007*, (Section 2), 192–201. Retrieved from http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-314/

Lew, M. S., Sebe, N., Djeraba, C., & Jain, R. (2006). Content-based multimedia information retrieval : State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *2*(1), 1–19.

Liang, K., Zeger, S. L., & Apr, N. (2007). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, *73*(1), 13–22.

Lin, Y., Ahn, J.-W., Brusilovsky, P., He, D., & Real, W. (2010). Imagesieve: Exploratory search of museum archives with named entity-based faceted browsing. *Proceedings of the American Society for Information Science and Technology*, *47*(1), 1–10. doi:10.1002/meet.14504701217

Liu, H., Little, S., & Ruger, S. (2011). Multimedia: behaviour, interfaces and interaction. In I. Ruthven & D. Kelly (Eds.), *Interactive Information Seeking, Behaviour and Retrieval* (pp. 221–234). London : Facet Publishing.

Liu, H., Uren, V., Song, D., & Stefan, R. (2009). A Four-factor User Interaction Model for Content-Based Image Retrieval. In *the 2nd international conference on the theory of information retrieval (ICTIR)* (pp. 1–8).

Liu, H., Zagorac, S., Uren, V., Song, D., & Stefan, R. (2009). Enabling effective user interactions in content-based image retrieval. In *the 5th Asia Information Retrieval Symposium*.

Macgregor, G., & Mcculloch, E. (2006). Collaborative tagging as a knowledge organization and resource discovery tool. *Library Review*, *55*(5), 291–300.

Maple, A. (1995). Faceted access: A review of the literature. Retrieved from http://library.music.indiana.edu/tech_s/mla/facacc.rev

Marchionini, G. (1993). Finding facts vs. browsing knowledg in hypertext systems. *Computer*, *2*(11), 70–80.

Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM*, *49*(4), 41–46. doi:10.1145/1121949.1121979

Marchionini, G., & Brunk, B. (2003). Toward a general relation browser: A GUI for information architects. *Journal of Digital Information*, *4*(1). Retrieved from https://journals.tdl.org/jodi/article/viewArticle/97

Markkula, M., & Sormunen, E. (2000). End-User Searching Challenges Indexing Practices in the Digital Newspaper Photo Archive. *Information Retrieval*, *1*, 259–285.

Maron, M. E., & Studies, I. (1977). On Indexing , Retrieval and the Meaning of About. *Journal of the American Society for Information Science*, *28*(1), 38–43.

Marques, O., & Furht, B. (2002). *Content-based Image and Video Retrieval*. Kluwer Academic Publishers.

McGrenere, J., Baecker, R. M., & Booth, K. S. (2002). An evaluation of a multiple interface design solution for bloated software. In *Proceedings of the SIGCHI conference on Human factors in computing systems Changing our world, changing ourselves - CHI '02* (pp. 163–170). New York, New York, USA: ACM Press. doi:10.1145/503403.503406

Morville, P. (2005). *Ambient Findability: What We Find Changes Who We Become*. Retrieved from http://dl.acm.org/citation.cfm?id=1121644

Mulholland, P., Zdrahal, Z., & Collins, T. (2008). Investigating the effects of exploratory semantic search on the use of a museum archive. In *IEEE 2008 International Conference on Distributed Human-Machine Systems*. Athens, Greece.

Nielson, J. (2003). Information foraging: Why Google makes people leave your site faster. Retrieved from http://www.useit.com/alertbox/20030630.html

Nowak, S., & Rüger, S. (2010). How reliable are annotations via crowdsourcing. In *Proceedings of the international conference on Multimedia information retrieval - MIR '10* (p. 557). New York, New York, USA: ACM Press. doi:10.1145/1743384.1743478

Overell, S., Sigurbjörnsson, B., and Zwol., R. van. (2009). Classifying Tags using Open Content Resources. In *the Second ACM International Conference on Web Search and Data Mining* (pp. 64–73). New York, New York, USA.

Park, J., & Lu, C. (2009). Application of semi-automatic metadata generation in libraries: Types, tools, and techniques. *Library & Information Science Research*, *31*(4), 225–231. doi:10.1016/j.lisr.2009.05.002

Pedhazur, E. J. (1982). Multiple regression and behavioral science. *Explanation and Prediction*, *2*.

Peters, I. (2006). Against folksonomies - Indexing blogs and podcasts for corporate knowledge management. In *Online Information* (pp. 93–97). London, Great Britain.

Peters, I. (2009). *Folksonomies: Indexing and Retrieval in Web 2.0.* (p. 154).

Pirolli, P. (2007). *Information Foraging Theory Adaptive Interaction with Information* (p. 221). Cary, NC, USA: Oxford University Press.

Plangprasopchok, A., Lerman, K., & Getoor, L. (2010). Growing a tree in the forest: Constructing folksonomies by integrating structured metadata. In *Proceeding of the 16th International conference on Knowledge discovery and data mining* (pp. 949–958). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1835924

Plous, S. (1993). *The Psychology of Judgment and Decision Making*. New York: McGraw-Hill.

Ranganathan, S. R. (1933). *Colon clasification*. Madras Library Association, Madras.

Rashtchian, C., Young, P., Hodosh, M., & Hockenmaier, J. (2010). Collecting Image Annotations Using Amazon ' s Mechanical Turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk* (pp. 139–147). Retrieved from http://aclweb.org/anthology/W/W10/W10-0721.pdf

Rivadeneira, A. W., Gruen, D. M., Muller, M. J., & Millen, D. R. (2007). Getting our head in the clouds: toward evaluation studies of tagclouds. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (pp. 995–998). New York, NY, USA. Retrieved from http://portal.acm.org/citation.cfm?id=1240775

Rolla, P. J. (2009). User tags versus subject headings: can user-supplied data improve subject access to library collections? *Library Resources & Techinical Services*, *53*(3), 174.

Ruthven, I., Lalmas, M., & van Rijsbergen, K. (2003). Incorporating user search behavior into relevance feedback. *Journal of the American Society for Information Science and Technology*, *54*(6), 529–549. doi:10.1002/asi.10240

RyumT., Huang, T. S., Ortega, M., & Mehrotra, S. (1998). Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, *8*(5), 644–655. doi:10.1109/76.718510

Seifert, C., Kump, B., Kienreich, W., Granitzer, G., & Granitzer, M. (2008). On the Beauty and Usability of Tag Clouds. In *Information Visualisation, 2008. IV '08. 12th International Conference* (pp. 17–25). doi:10.1109/IV.2008.89

Shraefel, M. C., Wilson, M., Russsell, A., & Smith, D. A. (2006). mSpace: Improving information access to multimedia domains with multimodal exploratory search. *Ciommunications of the ACM - Supporting exploratory search*, *49*(4), 47–49.

Sigurbjornsson, B., & Zwol, R. Van. (2010). *TAGEXPLORER : Faceted browsing of flickr photos*.

Sinclair, J., & Cardew-Hall, M. (2007). The Folksonomy Tag Cloud: When Is It Useful? *Journal of Information Science*, *34*(1), 15–29. doi:10.1177/0165551506078083

Smith, G. (2008). Navigation and Visualization. In *Tagging: People-Powered Metadata for the Social Web* (pp. 95–116). New Riders.

Sorokin, A., & Forsyth, D. (2008). Utility Data Annotation with Amazon Mechanical Turk. In *Porceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1–8). IEEE. doi:10.1109/CVPRW.2008.4562953

Spink, A., & Greisdorf, H. (2001). Regions and levels: Measuring and mapping users' relevance judgments. *Journal of the American Society for Information Science and Technology*, *52*(2), 161–173. doi:10.1002/1097-4571(2000)9999:9999<::AID-ASI1564>3.0.CO;2-L

Spool, J. M., Perfetti, C., & Brittan, D. (2004). *Designing for the scent of information*.

Steele, T. (2009). The new cooperative cataloging. *Library Hi Tech*, *27*(1), 68–77. doi:10.1108/07378830910942928

Stoica, E., Hearst, M. A., Berkeley, U. C., & Richardson, M. (2007). Automating creation of hierarchical faceted metadata structures. In *Procseedings of the Human Language Technology Conference* (pp. 244–251).

Syn, S. Y. (2010). *Generation of Classificatory Metadata for Web Resources Using Social Tags*. University of Pittsburgh.

Trant, J. (2009). Tagging , Folksonomy and Art Museums : Early Experiments and Ongoing Research. *Journal of Digital Information*, *10*, 1–41.

Trattner, C., Körner, C., & Helic, D. (2011). Enhancing the Navigability of Social Tagging Systems with Tag Taxonomies. In *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies - i-KNOW '11* (p. 1). New York, New York, USA: ACM Press. doi:10.1145/2024288.2024310

Trattner, C., Lin, Y., Parra, D., & Brusilovsky, P. (2012). Evaluating Tag-Based Information Access in Image Collections. In *Proceedings of the 23rd ACM conference on Hypertext and social media* (pp. 113–122). Milwaukee, Wisconsin, USA.: ACM.

University of Mississippi Libraries. (2004). *Library of Congress Subject Headings Guide*. Retrieved from http://home.olemiss.edu/~tharry/SH/lcshguide.pdf

Urban, J., & Jose, J. M. (2006). EGO : A Personalized Multimedia Management and Retrieval Tool. *International Journal of Intelligent Systems*, *21*, 725–745. doi:10.1002/int

Urban, J., Jose, J. M., & Rijsbergen, C. J. (2006). An adaptive technique for content-based image retrieval. *Multimedia Tools and Applications*, *31*(1), 1–28. doi:10.1007/s11042-006-0035-1

Vasconcelos, N. (2005). Content-Based Image and Video Retrieval. *Signal Processing*, *85*(2), 231–232. doi:10.1016/j.sigpro.2004.10.007

Venetis, P., Koutrika, G., & Garcia-Molina, H. (2011). On the Selection of Tags for Tag Clouds. In *Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11* (pp. 835–844). New York, New York, USA: ACM Press. doi:10.1145/1935826.1935855

Wal, T. Vander. (2007). Folksonomy coinage and definition. Retrieved August 22, 2012, from http://vanderwal.net/folksonomy.html

Wartena, C. (2010). Automatic classification of social tags. In *Proceedings of the 14th European conference on Research and advanced technology for digital libraries* (pp. 176–183). Berlin, Heidelberg: Springer-Verlag.

Weller, K. (2007). Folksonomies and ontologies: two new players in indexing and knowledge representation Folksonomies. In *Online Information* (pp. 108–115).

White, R. W., & Marchionini, G. (2007). Examining the effectiveness of real-time query expansion. *Information Processing & Management*, *43*(3), 685–704. doi:10.1016/j.ipm.2006.06.005

White, R. W., & Roth, R. A. (2009). *Exploratory Search: Beyond the Query-Response Paradigm*. Morgan & Claypool Publishers.

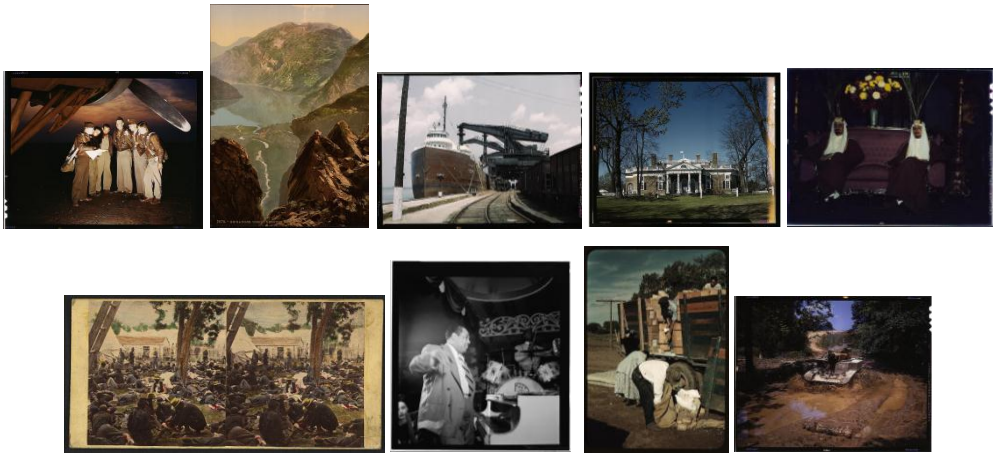
Yee, K.-P., Swearingen, K., Li, K., & Hearst, M. (2003). Faceted metadata for image search and browsing. In *CHI '03 Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 401–408). Ft. Lauderdale, Florida, USA.: ACM, New York. doi:10.1145/642611.642681

Zollers, A. (2007). Emerging motivations for tagging : Expression , performance , and activism. In *WWW2007*. Banff, Alberta, Canada. Retrieved from http://www2007.org/workshops/paper_55.pdf

Zubiaga, A., García-Plaza, A. P., Fresno, V., & Martínez, R. (2009). Content-based clustering for tag cloud visualization. In *2009 International Conference on Advances in Social Network Analysis and Mining* (pp. 316–319). IEEE. doi:10.1109/ASONAM.2009.19

APENDIX A. EXPERIMENTAL TASKS

Flickr Collection

Search Tasks	Search Task Descriptions
Lookup	Find the following picture  (only one of the images will be presented to the user at one time)
Exploratory	1. Background: You would like to add a new chapter to a travel book with some historical pictures about Europe. You are looking for images from the Library of Congress Flickr Commons collection. The new chapter will include photographs of natural scenery, landmarks or buildings, and events in Europe. You want to present 4 countries. For each country, you will collect one representative picture of its natural scenery, one for its modern facilities, and one for its activities. All three pictures have to be in the same location (e.g., in the same region, state, province, or city of the country). You should gather 12 photos from this search.

	2. Background: You are preparing an exhibition about the history of baseball in the late 19th century and the early part of the 20th century. You are looking for images in the Library of Congress Flickr Commons collection. The exhibition will include photographs of teams, trainers, coaches, and fans. In total, you plan to exhibit 12 representative pictures, featuring six different teams, two images of a team trainer or a coach, two depicting primarily the spectators, and two related to the World Series in different years. Also be sure that both the late 19th century and the early 20th century (around 1880-1920) are represented, although it is not necessary to have the same number of images for each period. You should gather 12 photos from this search.

Teenie Harris Collection

Search Tasks	**Search Task Descriptions**
Lookup	Find the following picture  (only one of the images will be presented to the user at one time)
Exploratory	1. Background: You are writing a book chapter about religious history in Pittsburgh. You are looking for images from the Teenie Harris collection. You want to be able to represent a broad range of religious events. You plan to exhibit six **representative pictures** of different events. Each event has 2 pictures

taken in different locations. In total you will collect 12 photos for the book chapter.

2. Background: You are preparing an exhibition on the history of sports in Pittsburgh and are looking for visual documents in the Teenie Harris collection.

The exhibition will be devoted to local baseball teams. You know about the Pirates; however, you also want to showcase other local teams. Use the system to find other baseball teams. In total, you plan to exhibit 6 **representative pictures** of different local teams (**1 picture for each team).** You also want to present pictures from different sports to show the variety of sports in Pittsburgh. For each sport you need to find three pictures showing trainers or coaches, and three pictures showing awards. You want to present 6 other different sports in the exhibition besides baseball.

3. Background: You are making a poster about jazz music history and are looking for images in the Teenie Harris collection.

The poster will use photographs of jazz performers, venues, and performers' outfits. In total, you plan to find 12 representative pictures, of which six are featuring female musicians performing different jazz instruments, three photos of different performing venues, and three different photos showing different types of performers' outfits.

APENDIX B. BACKGROUND SURVEY QUESTIONNAIRE

Participant# :_____

Date :_____

Q1) Major or occupation:

_____

Q2) Gender:

☐ Female

☐ Male

Q3) Age:

☐ 18-20

☐ 21-25

☐ 26-30

☐ 31-35

☐ 36-40

☐ 40-45

☐ 45-50

☐ Above 50

Q4) Native English speaker:

☐ Yes

☐ No

Q5) Highest academic degree earned:

☐ High school

☐ Bachelor's

☐ Master's

☐ PhD

☐ Post-Doctorate

Q6) On average, the amount of time spent per day using a computer:

☐ Less than 1 hour.

☐ 1 hour to less than 3 hours.

☐ 3 hours to less than 5 hours.

☐ 5 hours to less than 7 hours.

☐ More than 7 hours.

Q7) Which of the following search engines have you used before:

☐ AltaVista

☐ Ask.com

☐ Baidu

☐ Bing

☐ Google

☐ Yahoo! Search

☐ Other _____

Q8) How confident are you in your ability to locate specific information using a search engine:

☐ Not at all confident

☐ Slightly confident

☐ Moderately confident

☐ Somewhat confident

☐ Very confident

Q9) What are some of your personal interests:

☐ Arts & Crafts

☐ Games & Puzzles

☐ Geography

☐ History

☐ Music

☐ Politics

☐ Sports

☐ Other _____

Q10) Are you familiar with “tagging” or “tagging systems”, such as Flickr?

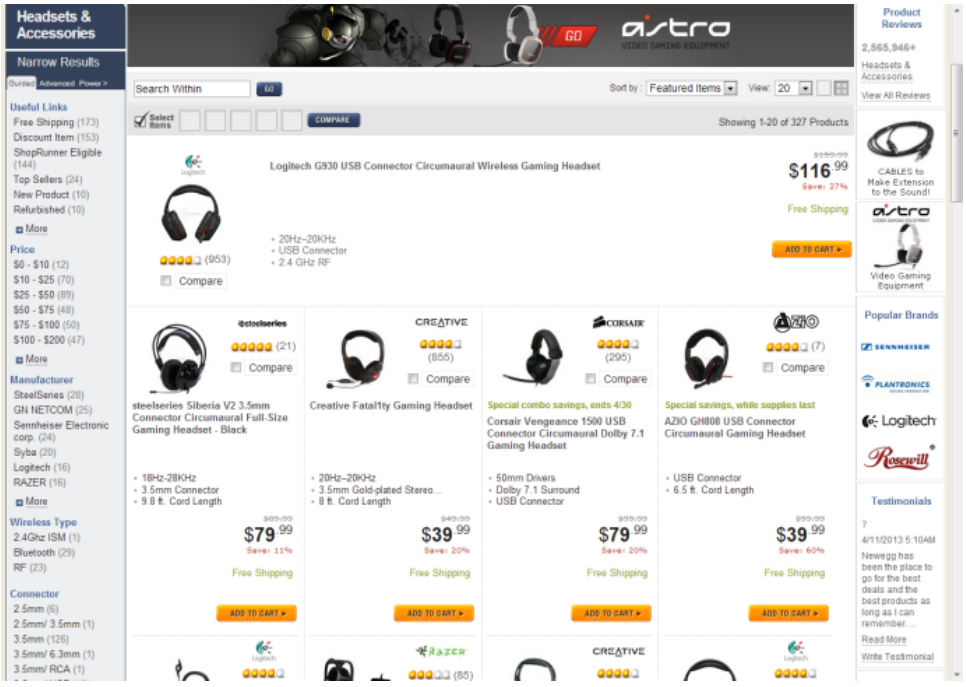
☐ Yes

☐ No

Q11) Are you familiar with “faceted browser”?

☐ Yes

☐ No



APENDIX C. POST-TASK QUESTIONNAIRE – SUBJECT HEADING-ONLY INTERFACE

Participant# :_____

Collection# :_____

Order# :_____

Lookup					
	Not at all	**Slightly**	**Moderately**	**Somewhat**	**Strongly**
1) How well did the interface provide support to this task?	☐	☐	☐	☐	☐
2) How well did you find the subject headings useful in finding the target items?	☐	☐	☐	☐	☐
3) The first image #_____ How clearly did the first image have salient / unique features that make it easy to locate?	☐	☐	☐	☐	☐
The second image #_____ How clearly did the first image have salient / unique features that make it easy to locate?	☐	☐	☐	☐	☐
The third image#_____ How clearly did the first image have salient / unique features that make it easy to locate?	☐	☐	☐	☐	☐
4) How helpful was it to display the subject headings with the numbers according to how many images were assigned with that subject heading?	☐	☐	☐	☐	☐

Task topic# :_____

Exploratory					
	Not at all	**Slightly**	**Moderately**	**Somewhat**	**Strongly**
1) How well did the interface provide support to this task?	☐	☐	☐	☐	☐
2) How well did the subject headings in the result page provide you with a good overview of the returned images?	☐	☐	☐	☐	☐
3) How well did the subject headings in the result page give you good hints about how to proceed to the next step?	☐	☐	☐	☐	☐
4) How well did the subject headings in the detail page guide you to related images?	☐	☐	☐	☐	☐
5) When you clicked to see a full image, was it mostly because you thought the image might be relevant?	☐	☐	☐	☐	☐
6) How familiar are you with this task topic?	☐	☐	☐	☐	☐
7) How confident are you in the system's ability to help you find useful information on other topics?	☐	☐	☐	☐	☐

Please write down any other comments.

_____

APENDIX D. POST-TAST QUESTIONNAIRE – TAG-ONLY INTERFACE

Participant# :_____

Collection# :_____

Order# :_____

Lookup					
	Not at all	**Slightly**	**Moderately**	**Somewhat**	**Strongly**
1) How well did the interface provide support to this task?	☐	☐	☐	☐	☐
2) How well did you find the tags useful in finding the target items?	☐	☐	☐	☐	☐
3) The first image #_____ How clearly did the first image have salient / unique features that make it easy to locate?	☐	☐	☐	☐	☐
The second image #_____ How clearly did the first image have salient / unique features that make it easy to locate?	☐	☐	☐	☐	☐
The third image#_____ How clearly did the first image have salient / unique features that make it easy to locate?	☐	☐	☐	☐	☐
4) How helpful was it to display the tags in differernt font size according to how many images were assigned with that tag?	☐	☐	☐	☐	☐

Task topic# :_____

Exploratory					
	Not at all	**Slightly**	**Moderately**	**Somewhat**	**Strongly**
1) How well did the interface provide support to this task?	☐	☐	☐	☐	☐
2) How well did the tags in the result page provide you with a good overview of the returned images?	☐	☐	☐	☐	☐
3) How well did the tags in the result page give you good hints about how to proceed to the next step?	☐	☐	☐	☐	☐
4) How well did the tags in the detail page guide you to related images?	☐	☐	☐	☐	☐
5) When you clicked to see a full image, was it mostly because you thought the image might be relevant?	☐	☐	☐	☐	☐
6) How familiar are you with this task topic?	☐	☐	☐	☐	☐
7) How confident are you in the system's ability to help you find useful information on other topics?	☐	☐	☐	☐	☐

Please write down any other comments.

_____

APENDIX E. POST-TASK QUESTIONNAIRE – DUAL-PERSPECTIVE NAVIGATION INTERFACE

Participant# :_____

Collection# :_____

Order# :_____

Lookup					
	Not at all	**Slightly**	**Moderately**	**Somewhat**	**Strongly**
1) How well did the interface provide support to this task?	☐	☐	☐	☐	☐
2) How well did you find the subject headings and tags useful in finding the target items?	☐	☐	☐	☐	☐
3) The first image #_____ How clearly did the first image have salient / unique features that make it easy to locate?	☐	☐	☐	☐	☐
The second image #_____ How clearly did the first image have salient / unique features that make it easy to locate?	☐	☐	☐	☐	☐
The third image#_____ How clearly did the first image have salient / unique features that make it easy to locate?	☐	☐	☐	☐	☐

Task topic# :_____

Exploratory					
	Not at all	**Slightly**	**Moderately**	**Somewhat**	**Strongly**
1) How well did the interface provide support to this task?	☐	☐	☐	☐	☐
2) How well did the subject headings and tags in the result page provide you with a good overview of the returned images?	☐	☐	☐	☐	☐
3) How well did the subject headings and tags in the result page give you good hints about how to proceed to the next step?	☐	☐	☐	☐	☐
4) How well did the subject headings and tags in the detail page guide you to related images?	☐	☐	☐	☐	☐
5) When you clicked to see a full image, was it mostly because you thought the image might be relevant?	☐	☐	☐	☐	☐
6) How familiar are you with this task topic?	☐	☐	☐	☐	☐
7) How confident are you in the system's ability to help you find useful information on other topics?	☐	☐	☐	☐	☐

Please write down any other comments.

_____

APENDIX F. POST-EXPERIMENT QUESTIONNAIRE

Participant# :_____

	Subject heaing-only	Tag-only	Dual-perspective navigation
1) Which one of the interface do you like most?	☐	☐	☐
Why do you like the interface most? _____			
2) Which one of the interface would you feel more confident to use for other search tasks ?	☐	☐	☐
Why do you think that you feel more confident to use the interface for other search tasks? _____			
3) Which one of the interfaces would you prefer for lookup search (finding a specific picture)?	☐	☐	☐
4) Which one of the interfaces would you prefer for exploratory search?	☐	☐	☐
5) Which one of the interfaces would you suggest that the cultural heritage institutions , such as museums, provide their visitors for searching their image collections?	☐	☐	☐
Why is the selected interface in the previous question the best for the museum? _____			

	Not at all	Slightly	Moderately	Somewhat	Strongly
6) Overall, how would you rate your experience with the subject heading-only interface?	☐	☐	☐	☐	☐
7) Overall, how would you rate your experience with the tag-only interface?	☐	☐	☐	☐	☐
8) Overall, how would you rate your experience with the dual-perspective navigation interface?	☐	☐	☐	☐	☐

Do you have any suggestions to improve any of those interfaces?

_____