Wright State University

CORE Scholar

[Browse all Theses and Dissertations](#)

[Theses and Dissertations](#)

2009

Integrated Faceted Browser and Direct Search to Enhance Information Retrieval in Text-Based Digital Libraries

Shea-Tinn Yeh

Wright State University

Follow this and additional works at: https://corescholar.libraries.wright.edu/etd_all

Part of the [Operations Research, Systems Engineering and Industrial Engineering Commons](#)

Repository Citation

Yeh, Shea-Tinn, "Integrated Faceted Browser and Direct Search to Enhance Information Retrieval in Text-Based Digital Libraries" (2009). *Browse all Theses and Dissertations*. 955.

https://corescholar.libraries.wright.edu/etd_all/955

This Thesis is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.

INTEGRATED FACETED BROWSER AND DIRECT SEARCH TO ENHANCE
INFORMATION RETRIEVAL IN TEXT-BASED DIGITAL LIBRARIES

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Engineering

By

SHEA-TINN YEH

B.A., National Cheng Kung University, 1983
M.L.S., University of Maryland, 1985
B.S., Franklin University, 1995

2009

Wright State University

WRIGHT STATE UNIVERSITY
SCHOOL OF GRADUATE STUDIES

September 3, 2009

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY
SUPERVISION BY Shea-Tinn Yeh ENTITLED Integrated Faceted Browser and Direct
Search to Enhance Information Retrieval in Text-Based Digital Libraries BE
ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF Master of Science in Engineering.

Yan Liu, Ph.D.
Thesis Director

S. Narayanan, Ph.D.
Department Chair

Committee on
Final Examination

Yan Liu, Ph.D.

Misty Blue, Ph.D.

Pratik Parikh, Ph.D.

Joseph F. Thomas, Jr., Ph.D.
Dean, School of Graduate Studies

ABSTRACT

Yeh, Shea-Tinn. M.S.E.. Department of Biomedical, Industrial and Human Factors Engineering, Wright State University, 2009. Integrated Faceted Browser and Direct Search to Enhance Information Retrieval in Text-Based Digital Libraries.

Browsing and searching are two prominent paradigms in information retrieval. As digital libraries grow in size and types of collections, their user bases are expanding and include more non-expert users. For those who are not familiar with what the content of a specific text-based digital collection is or its data structure, the browse functionality offers the entry point to explore and obtain an overview. In current digital library implementations, exploratory browsing is sometimes not available as an option. When offered, it is commonly presented as an alphabetical listing of chosen categories depending on the scope of the digital collections. Studies have revealed that information seekers feel fatigued easily and that they are willing to scan at most 200 items of citations in electronic environments. In addition, when users are ready to search with a specific query during the browsing activity, they have to switch to another information space where the search box is separated from the browsing lists. A consequence of such design is that users have to toggle back and forth between browsing and searching states.

Within this context, faceted navigation has been discussed and a conceptual model of an integrated faceted browser and direct search utilizing faceted navigation for text-based digital libraries is also proposed in this research. A faceted browser provides users with many entry points at first glance which enable them to receive an overview of the content of the digital libraries effectively and efficiently. A faceted browser integrated with

direct search ensures the switching between browsing and searching seamlessly. While searching, users are able to work with the recognized structure of the collection right from the browsing overview. An experiment was conducted to test five hypotheses derived from the conceptual model of the integrated faceted browser and direct search for text-based digital libraries. The experiment results show that four hypotheses are supported. For both browsing and searching approaches in information retrieval in a text-based digital library, compared to an alphabetical browser, the faceted browser can significantly improve the effectiveness by (30.8%, $p = 0.015$) and efficiency by (11.3%, $p = 0.001$) of information retrieval. Also, compared to un-integrated alphabetical browser with direct search interfaces, the integrated faceted browser with direct search interfaces can significantly improve the effectiveness in information retrieval by (35.7%, $p = 0.03$) and bring users greater satisfaction (by 34.8%, $p < 0.03$) with the process.

TABLE OF CONTENTS

TABLE OF CONTENTS.....	v
LIST OF FIGURES	vii
I. INTRODUCTION.....	1
1.1 What are Digital Libraries and Why are They Useful?	1
1.2 Limitations in Current Digital Library Implementations.....	2
1.3 Research Objectives and Scope	6
1.4 Significance of the Research.....	6
1.5 Organization of the Thesis.....	7
II. LITERATURE REVIEW.....	8
2.1 History of Digital Libraries.....	8
2.2 Key Components of Digital Libraries.....	9
2.3 Information Retrieval in Digital Libraries	10
2.4 What is browsing?	11
2.5 What would be a Good Design for Browsing?	13
2.6 What is Faceted Navigation and What are its Strengths?	14
2.7 Integrated Faceted Browsing and Direct Search in Image-based Collections.....	17
3.1 Conceptual Model.....	19
3.2 Derivation of Hypotheses	20
IV. EXPERIMENT	23
4.1 Methodology	23
4.1.1 Participants.....	23

4.1.2 Apparatus	24
4.1.3 Procedure	24
4.1.4 Independent Variables.....	25
4.1.5 Dependent Variables	28
4.1.6 Statistical Model	30
4.2 Statistical Analysis.....	30
4.2.1 Hypotheses I and II	30
4.2.2 Hypotheses III and IV	33
4.2.3 Hypotheses V	36
V. CONCLUSIONS AND FUTURE RESEARCH.....	38
5.1 Findings and Contributions.....	38
5.2 Limitations	40
5.3 Future Research	41
Appendix A Pre-study Questionnaire	42
Appendix B Hypotheses I and II: Tasks on alphabetical browser and faceted browser...	44
Appendix C Hypotheses III and IV: Tasks on un-integrated alphabetical browser and direct search and integrated faceted browser and direct search	48
Appendix D Post-study questionnaire	53
List of References	55

LIST OF FIGURES

Figure	Page
Figure 1 A Screenshot of the alphabetical title list from the Chopin's Collection at the University of Chicago Library	3
Figure 2 A Screenshot of the direct search functionality on the search screen in the Chopin's Collection at the University of Chicago Library	5
Figure 3 Structure of current digital library implementation with separate screens for alphabetical browser and direct search	5
Figure 4 A Screenshot of FLAMENCO's Nobel Prize winners' collection.....	16
Figure 5 The Conceptual model of the proposed integrated faceted browser and direct search	20
Figure 6 A Screenshot of the Title list in the AB	26
Figure 7 A Screenshot of the facets in the FB	27
Figure 8 A Screenshot of the search functionality in the un-integrated AB and direct search interface	28
Figure 9 Scattered plot of the number of correctly answered questions with the FB subtracted by that with the AB.....	32
Figure 10 Scattered plot of task completion time with the FB subtracted by that with the AB.....	32
Figure 11 Scattered plot of the number of correctly answered questions with the integrated FB and direct search interfaces subtracted that with the un-integrated AB and direct search interfaces.....	35

Figure 12 Scattered plot of the task completion time with the integrated FB and direct search interfaces subtracted by that with the un-integrated AB and direct search interfaces

..... 35

LIST OF TABLES

Table	Page
Table 1 Summary of experiment results for hypotheses I and II.....	31
Table 2 Summary of experiment results for hypotheses III and IV.....	34
Table 3 Summary of experiment results for hypothesis V	36

ACKNOWLEDGEMENTS

My adviser, Dr. Yan Liu, has earned deep appreciation for her three special qualifications: solid knowledge of the subject matter, dedication to teaching, and her passion for excellence in challenging me to discover my real potential. With her advice, I have not only concluded this study, but also learned valuable research skills to enable future contributions.

I am grateful to my committee members, Dr. Misty Blue and Dr. Pratik Parikh for their time and advices to ensure the satisfactory completion of this research work.

A special thank goes to Dr. Frank Ciarallo for his encouragement when I hit the Great Wall of self-doubt during my studies in one of his classes and was then able to move forward with confidence.

I thank Amy McCrory, the Digital Imaging Specialist at the Ohio State University Libraries, for granting me permission to generate Dublin Core XML files as the dataset for the experiment.

I would like to thank all the individual volunteers who participated in the Experiment - your dedication contributed to the success of the data collection.

And most precious, I thank my husband Larry Owens for his many proof edits, discussions, and unconditional support. I also thank him for his trust in my ability to succeed.

For my nine year old son, Perry, for his consideration and independence that allow me to
continuously pursue my dreams.

For my mother in Taiwan and father in Heaven who worked hard throughout their lives
to provide me with the best education I could ever ask for.

I. INTRODUCTION

1.1 What are Digital Libraries and Why are They Useful?

Since the announcement of Digital Library Initiative (DLI) projects jointly sponsored by the National Science Foundation (NSF), the Department of Defense Advanced Research Projects Agency (DARPA), and the National Aeronautics and Space Administration (NASA) in 1993, digital libraries have taken information retrieval into a new era. The terms *digital library*, *electronic library*, *virtual library*, *library without walls*, and the most recently *hybrid library* have been used interchangeably in the literature and in different countries. The original working definition of digital libraries was provided by Digital Library Federation (DLF) in 1998: “Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, and preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily available for use by a defined community or set of communities.” (Digital Library Federation, 1998, Home, para. 1)

Advances in information technology have made materials available in standardized electronic formats at a much lower cost, compared to printed forms. In addition, digital libraries can hold more information with much less physical space and preserve bodies of knowledge for longer periods. With the growth of the Internet and the advancement in the World Wide Web (WWW) technology, a digital environment can facilitate cross-community interactivity and collaboration, regardless of the physical locations of the users (Schwartz, 2000). Digital libraries have changed the way humans

interact with information - as long as there is an Internet connection available, the requested information can be accessed from anywhere at anytime (Kani-Zabihi et al., 2006).

1.2 Limitations in Current Digital Library Implementations

Browsing and searching are the two principal paradigms in information retrieval. In this thesis, the term browsing refers to *exploratory browsing* rather than examination of the results from information retrieval, and the term searching specifically refers to *direct searching* with a single search box, not including advanced search functions that contain multiple search boxes. Users who are not familiar with the content of a digital library collection may desire to acquire an overview of the collection and the relationships and structure of the data content. When the overview is presented in a structured manner, users can have a better understanding of the information to be searched and thus formulate further queries more effectively and efficiently. Once a query is formed usually as a list of keywords, with or without Boolean operators, the user then enters the paradigm of *direct searching*. Either paradigm by itself is generally not sufficient to achieve complex information goals and may lead to an unsatisfying user experience or result sets (Olston & Chi, 2003). By integrating exploratory browsing with direct searching methods, it becomes possible to transcend the boundary between these two paradigms.

In current digital library implementations, exploratory browsing is mostly not offered, or is offered as an alphabetical listing of items organized in a sequential manner for specific categories depending on the scope of the collections. Categories of author,

title, and subject are the most commonly deployed. For the purpose of exploratory browsing to obtain an overview of the collection, users may feel lost in the midst of pages of listings. Imagine a user browsing through a text list of 25 pages in length, what sense could the user make of those pages? This type of alphabetical browsing is illustrated in the screenshot shown in Figure 1. Here users browse through the alphabetical title list of composer Chopin's work at the University of Chicago Library.

Figure 1 A Screenshot of the alphabetical title list from the Chopin's Collection at the University of Chicago Library

Although browsing and searching in the information retrieval process have been extensively studied, the research efforts have been mostly parallel. There has been little

effort in integrating exploratory browsing and direct search. In the previous example, suppose the user discovers a topic of interest during the exploratory browsing phase and desires to start searching with a query. With the typical current digital library setting, the user has to remember the term(s) of interest or copy it/them from the browsing window, switch to a separate screen that contains the direct search functionality, and then input the term(s) of interest into a search box. A screenshot of this type of the direct search implementation is illustrated in Figure 2 where users enter a query or queries about composer Chopin's work in a search box on a separate search screen.

Also, the structure of the dominant information retrieval paradigms: exploratory browsing and direct searching on separate information on two different information space can be illustrated in Figure 3. Although browsing and searching have been viewed and implemented as two different processes to be switched between, the switching between them is not necessarily linear in informational retrieval; rather, users generally seek information in an iterative manner (McKay et al., 2004).

Figure 2 A Screenshot of the direct search functionality on the search screen in the Chopin's Collection at the University of Chicago Library

Figure 3 Structure of current digital library implementation with separate screens for alphabetical browser and direct search

1.3 Research Objectives and Scope

To address the above mentioned limitations of information retrieval in current digital libraries, the goal of this thesis is to design a framework of integrated exploratory browsing and searching interface that will enable users to move through large information spaces effectively and efficiently in a flexible manner. In particular, two objectives are pursued in this research:

- (1) Design a faceted browser that can facilitate users in browsing text-based digital libraries effectively and efficiently.
- (2) Design an integrated faceted browser and direct search interface that can further improve the effectiveness and efficiency in retrieving information from text-based digital libraries and can provide users with greater satisfaction in the process.

Faceted browser is a software tool that uses faceted navigation for information browsing. Faceted navigation can be defined as “a technique for accessing a collection of information represented using a faceted classification, allowing users to explore by filtering available information” (“Faceted_browser,” 2009, para. 1). Previous studies suggest that compared to traditional alphabetical browsing, faceted browsing can lead to more effective and efficient information seeking (Yee et al., 2003; Hearst, 2006).

1.4 Significance of the Research

This research has proposed an integrated framework of exploratory browser and direct search interface which can facilitate more effective and efficient information retrieval. An experiment was conducted to test the merit of the framework. In addition,

this research offers valuable practical implications on how to design a user-centered text-based digital library.

1.5 Organization of the Thesis

The organization of this thesis is as follows. Chapter 2 reviews the history and key components of digital libraries, development of information retrieval in digital libraries, and previous studies on browsing in general and faceted browsing in particular. In Chapter 3, a conceptual model of the integrated faceted browser and direct search interfaces is proposed, and five hypotheses are derived from the conceptual model. Chapter 4 describes an experiment conducted to test the proposed hypotheses and the analysis of experiment results. At last, in Chapter 5, conclusions of this research are offered, limitations of this research are specified, and future work is proposed.

II. LITERATURE REVIEW

2.1 History of Digital Libraries

Between 1994 and 1999, over 68 million of federal research awards were offered to the universities in the USA through the DLI programs in two phases, with \$24 million in the first phase (DLI-1) and \$44 million in the second phase (DLI-2) (Fox, 1999).

DLI-1 focused on research of developing a testbed for varieties of data, such as documents, reports, digital images, videos, maps, and sensor data. It also focused on services including interoperability among existing collections, networking, and accessibility. The efforts of DLI-1 were led by six teams from the University of California at Berkley and Santa Barbara, Carnegie Mellon University, University of Illinois at Urbana-Champaign, and University of Michigan. With the visible achievements of DLI-1, DLI-2 was joined by additional sponsors: the National Library of Medicine, the Library of Congress, and the National Endowment for the Humanities. The number of awards was increased, and the breadth of coverage in DLI-2 was broadened to support research in many disciplines such as anthropology, computer science, fine arts, and additional types of contents consisting of bibliographic records, engineering education, and ePrints. The research also aimed at diverse technical areas including 3-D modeling, agents, information visualization, speech processing, and text analysis (Fox, 1999). The last project of DLI-2 was concluded on September 30, 2005 with the creation of “Variations2” which provides online access to selected recordings and scores from the Indiana University Cook Music Library.

2.2 Key Components of Digital Libraries

Three key components constitute the theoretical framework underlying digital libraries, namely: people, information resources, and technology (Shiri, 2003).

Depending on the type and purpose of a digital library, its users can range from those belonging to some special communities where their access is restricted by authentication to the Internet users at large with varying levels of expertise and skills.

A distinction should be made between resources that are born-digital as created in a digital format and those that are results of conversion from a medium by digitizing. For example, a satellite imagery map is born-digital, while a newspaper article scanned and made into an image is a product of digitalization. Collections of both types in subject matters including arts, humanities, literature, medicine, music, and science have been exploited in digital libraries.

Digital libraries are characterized by the large amount of information they store and by the wide distribution of their documents from different locations and system (Kapidakis et al., 2000). Ample storage is an essential requirement for the digital library development. A typical storage needs can be represented by the recent project by the National Digital Information Infrastructure and Preservation Program (NDIIPP) partnered with the San Diego Supercomputer Center (SDSC) that outlines an architecture to provide for reliable, redundant, and geographically dispersed copies of the digital contents. The storage was divided into three parts that are functionally separated from each other: an ingestion machine for hosting incoming data, the back-end storage for acting as an archival storage systems, and access machine for information access

(Ashenfelder et al., 2009). The other characteristic of “wide distribution” presents challenge to the performance of digital libraries that should be addressed by three elements: the computing power required for indexing; the network system that provides interconnectivities between systems as well as between users and systems; and the software technology that makes use of the hardware components.

In short, digital libraries have become one or more electronic collections of digital materials in one or multiple disciplines that are distributed across networks to be accessed online by selected communities or communities at large.

2.3 Information Retrieval in Digital Libraries

It is widely believed that the term information retrieval was first coined in Mooers (1951, p. 25): “Information retrieval embraces the intellectual aspects of the description of information and its specification for search, and also whatever systems, techniques, or machines that are employed to carry out the operation.” This describes the process and method through which a prospective information user converts his/her quest for information into a collection of references that are useful to the user (Gupta & Jain, 1997). Traditionally, this process has been carried out by professional librarians, especially in specialized communities such as science and medicine, interacting with card catalogs and printed materials. In the 1980s, card catalogs were replaced by Online Public Access Catalog (OPAC), trained reference librarians would interact with online services to provide query scientists with the results (Schartz, 1997). As we entered the 21st century, Internet, WWW, browser-based OPAC, and digital libraries coupled with

the prevalence of personal computers, have broadened the information retrieval process to be performed directly by querying individuals in the absence of trained librarians.

2.4 What is browsing?

We have all experience browsing; for examples, we browse for items in a grocery store or on a magazine stand, and we browse the newspaper for headlines. Browsing is common; in fact, Poulter (2003) argued that browsing is such a common daily behavior that we lack words to describe different browsing experience as it is intuitively self-evident. In the information world, however, even though browsing is considered casual and incidental, it is recognized as an important information seeking technique (Bates, 2007).

During the past 20-25 years, there have been many discussions of browsing in the literature in library and information science. Generally speaking, browsing can be categorized in terms of users' perspectives and their browsing strategies. Users' perspectives can be further divided on the degree to which they have clear goals of their browsing activities. Hildredth (1982) identified three levels of browsing from the goal perspectives of a library user: (1) directed browsing – a structured activity when the desired goal and its location is known; (2) semidirected browsing – a loosely structured activity occurring periodically or repeatedly to discover new material of interest; and (3) undirected browsing – an unstructured and almost random activity where neither goal nor location is known. Cove and Walsh (1988) also proposed three stages of browsing: (1) search browsing, a directed and structured browsing activity with a known goal; (2) general purpose browsing, a browsing activity that occurs on a regular basis and the

sources are specified as they likely contain items of interests; and (3) serendipity browsing that is undirected and is random and unstructured. Salomon (1990) found browsing patterns as either casual browsing where the user looks through files casually or goal-directed browsing where the user discovers a goal in the course of browsing.

Other researchers, however, have suggested that browsing is goalless. For instance, Marchionini and Shneiderman (1988, p. 71) defined browsing as “an exploratory, information-seeking strategy that depends on serendipity. It is especially appropriate for ill-defined problems and for exploring new task domains.” This definition suggests that no goal is present and the intention is to define a goal via exploratory browsing.

In terms of browsing strategies, Liebsert and Marchionini (1988, p. 224) stated that many users would desire formulating a simple search query that places them “somewhere in the ball park”, and they would prefer visually scanning search results for the information of interest rather than examining the results in detail. They highlighted that fact that such a strategy utilizes “recognition”, one of the cognitive functions that are significantly faster than the “recall” function prevalently used in formulating complicated query searches. Marchionini (1989, p. 56) also labeled this strategy as “scan and select”, while Rice (2001, p. 8) categorized this as the main sub-dimension to browsing behaviors. Liebsert, Marchionini, and Rice all have maintained that browsing can be simply characterized as the process of scanning, a “point-by-point observation or checking”. Bates (2007, The Proposed Definition, para. 1), nevertheless, questioned whether browsing is indeed intentional scanning of an object; instead, she proposed a

definition of browsing in combinations with cognitive, motivational and behavioral patterns: “Browsing is the activity of engaging in a series of glimpses, each of which may or may not lead to closer examination of a (physical or represented) object, which examination may or may not lead to (physical and/or conceptual) acquisition of the object.” Based on the definition, Bates further argued that current functionality of alphabetical browsing in digital libraries facilitates scanning but not browsing.

2.5 What would be a Good Design for Browsing?

Literature review reveals two concrete suggestions for good design for browsing. Cox (1992) proposed three functional requirements for effective browsing systems: (1) enabling users to position themselves in an area of interest; (2) enabling users to recognize appropriate directions to further the search; and (3) enabling users to move quickly and efficiently throughout the system. In addition, Bates (2007, Implications for information system design, para. 3) pointed out that “Good browseable interfaces would consist of rich scenes, full of potential objects of interest, which the eyes can take in at once (through *massively parallel processing*) and then select items within the scene to give closer attention to.”

As far as we know there was no alternative to alphabetical browsing functionality in digital libraries development until the year of 2000 when FLExible information Access using MEtadata in Novel COmbinations (FLAMENCO) began its development by Professor Marti Hearst at the University of Berkeley with the sponsorship from National Science Foundation (NFS abstract #9984741). The open source FLAMENCO search interface framework utilizes faceted navigation for browsing and searching and attests to

the good design for browsing in Cox (1992) and Bates (2007). Digital image collections have since been developed with FLAMENCO, and faceted navigation has become popular in digital libraries by enterprise servers, such as Solr and Content Management System such as Drupal in particular (Hearst, 2008).

2.6 What is Faceted Navigation and What are its Strengths?

Faceted classification was first used in the Colon Classification in the early 1930s by S. R. Ranganathan, an Indian librarian and classificationist, as mentioned in Wynar and Taylor (1992, p. 320). The Colon Classification, the same as the Dewey Decimal Classification, divides the wealth of knowledge into a number of main classes. However, unlike other systems which simply subdivide the main classes into a series of subordinate classes, a Colon system divides each main class into facets by particular characteristics (Essay, 1984). Wynar and Taylor (1992, p. 320) defined a facet as “clearly defined, mutually exclusive, and collectively exhaustive aspects, properties or characteristics of a class or specific subject.”

A faceted browser provides browsing functionality by utilizing navigation that is driven by the faceted properties rather than classes. A facet can be flat and hierarchical. In a faceted browser, each facet is associated with a set of sub-categorical labels, and each item is assigned with multiple labels from the set. This method of multiple label assignment is different from a strictly hierarchical system in which each item is assigned to a single category (Hearst, 2008). In this aspect, facets are like different axes into which each document can be classified with one or more axes (Garshol, 2004). Furthermore, each axis can be viewed as an entry point, and each item may possess many

entry points. An example of the image collection using faceted navigation can be found on the FLAMENCO's demo site (<http://orange.sims.berkeley.edu/cgi-bin/flamenco.cgi/nobel/> *Flamenco*). It is a collection of Nobel Prize winners classified using a gender facet, a country facet, an affiliation facet, a prize facet, and a year facet. Among them, affiliation is a hierarchical facet arranging each Nobel Prize winner's affiliated organizations under the countries and cities to which they belong. Subcategories at different levels are all distinct, and items can be assigned to them at any level. For example, Mother Teresa was assigned to subcategory "female" in gender facet, subcategory "peace" in prize facet, and subcategory "India->Calcutta->Missionaries of Charity" in affiliation facet. To find Mother Teresa in this collection, any of the above mentioned subcategories can be an entry point during the browsing activity. Figure 4 illustrates a screenshot of the FLAMENCO's Nobel Prize winners' collection.

Figure 4 A Screenshot of FLAMENCO's Nobel Prize winners' collection

Marchionini (1995, p. 117) commented that browsing is challenging to information seekers because an entry point must first be identified. As demonstrated in FLAMENCO's Nobel Prize winners' collection, a browser based on faceted navigation offers as many entry points as possible for information seekers; therefore, it is naturally advantageous over alphabetical browser.

With many entry points presented, the faceted browser also reflects Bates' suggestion for good browsing by providing as many potential objects of interest as possible for the eye to take in all at once (Bates, 2007). In addition, being able to view all entries enables users to position themselves in an area of interest as Cox (1992) suggested. Hearst (2006) also claimed that compared to alphabetical browsing, the

information seeking approach with faceted categories can reduce users' mental work because "recognition" (as opposed to "recall") was involved in performing the tasks. She also found that with hierarchical faceted browsing, users encounter "logical but unexpected alternatives at every turn."

2.7 Integrated Faceted Browsing and Direct Search in Image-based Collections

The frequency of the toggling actions of browsing and searching varies among users and is dependent upon many variables, such as the goal of the user's search, the user's search strategies, and the scope of the digital library collection. However, studies indicate that users frequently deploy a combination of browsing and searching in information seeking and prefer to combine browsing and searching. Empirical studies have demonstrated that the combination of browsing and searching is effective in information seeking (Bonder et al., 2001; McKay et al., 2004).

In the implementation of FLAMENCO, faceted browsing and searching are integrated seamlessly in the same information space. Such a design supports flexible movement and fluidity of switching between browsing and searching as noted in Hearst (2008). At each point in the browsing or searching process, it also provides users with a preview of the next step and therefore a feeling of control and understanding without confusion (Hearst, 2008). A collection of recipes of 13,000 items, an architecture image collection of 40,000 items, and a fine arts image collection of 35,000 items have been implemented in FLAMENCO. All of these collections are "born digital" images. To the best of our knowledge, there does not exist an integrated faceted browser and direct search tool for text-based digital libraries. In addition, usability study of the faceted

navigation in digital libraries has been rather limited. Literature research reveals only one study that evaluated users' satisfaction with faceted metadata approach to imaging browsing and searching without any objective measures of their task performance (Yee et al., 2003).

III. CONCEPTUAL MODEL AND HYPOTHESES

3.1 Conceptual Model

Figure 5 illustrates the conceptual model of the proposed integrated faceted browser and direct search interfaces for text-based digital libraries. As discussed in section 2.6, faceted browser gives users the ability to find items based on more than one dimension and may involve several levels of information exploratory. As shown in Figure 5, the top level, as facet I, facet II, etc. serves to provide users with appropriate facets for a quick overview of the library's content. Each facet can have multiple subcategories of labels, such as subcategory I, subcategory II, etc. as level two. The number of levels in the facet browsers can be further expanded as needed, shown in Figure 5. Based on the information acquired from the top level overview, users can drill down on any subcategory for further exploration. In addition, the faceted browsing and direct searching functionalities are offered at the same screen. As a result, at any point during the browsing activity, users can enter search queries without having to switch to a different screen.

Figure 5 The Conceptual model of the proposed integrated faceted browser and direct search

3.2 Derivation of Hypotheses

Five hypotheses are derived from the conceptual model of integrated faceted browser and direct search interfaces described above.

Hypothesis I: Compared to an alphabetical browser, a faceted browser can improve the effectiveness of information retrieval in text-based digital libraries.

Hypothesis II: Compared to an alphabetical browser, a faceted browser can improve the efficiency of information retrieval in text-based digital libraries.

Even though browsing is a natural activity in information retrieval, it has its limitations. Browsing demands information seekers' attention and time, and when the tasks are repetitive, fatigue sets in quickly. "Studies suggest an upper limit of 200 items as the number of citations that users are willing to scan in electronic environments" (cited in Marchionini, 1995, p. 117). When users have to browse through page after page of listings with an alphabetical browser, they may become distracted, confused, or frustrated. A design "that is flexible so as to allow the user to modify their information need and information-seeking strategy at will" can support browsing effectively (McKay, 2004, p. 284). The proposed faceted browser presents entire facets of a collection all at once such that users can obtain an overview timely. In addition, when pertinent information is selected from the top level facet, users can move around the subcategories at will to explore the collection further.

Hypothesis III: Compared to un-integrated alphabetical browser and direct search interfaces, integrated faceted browser and direct search interfaces can improve the effectiveness of information retrieval in text-based digital libraries.

Hypothesis IV: Compared to un-integrated alphabetical browser and direct search interfaces, integrated faceted browser and direct search interfaces can improve the efficiency of information retrieval in text-based digital libraries.

Hypothesis V: Compared to un-integrated alphabetical browser and direct search interfaces, integrated faceted browser and direct search interfaces can bring users greater satisfaction in performing the information retrieval tasks.

Browsing and searching are two common approaches in information access and they each offer advantages and disadvantages when utilized in information retrieval. Browsing is effective to be used when information problems are ill defined or the goal of information seeking is to gather an overview (Machionini, 1995). However, browsing alphabetical listings especially introduces challenges and may cause disorientation, confusion, or “cognitive overhead” to the users.

Searching is effective to be used when information need is known a priori so that users can formulate queries in the search box immediately. On the contrary, queries are difficult to define and the users may experience undesirable outcomes from searching alone: too many results, too few results, zero result, or irrelevant results. Mackinlay & Zellweger (1995) encouraged researchers to combine these two approaches synergistically. The proposed integrated faceted browser and direct search interfaces allows users to either browse or search depending on the information needs and to switch between these two activities at anytime seamlessly during the information retrieval process.

IV. EXPERIMENT

A one-factor within subject experiment was designed to test the five hypotheses described in section III: compared to an alphabetical browser, participants can retrieve information from text-based digital libraries more effectively and efficiently by using a faceted browser; compared to un-integrated alphabetical browser and direct search interfaces, participants can retrieve information more effectively and efficiently by using the integrated faceted browser and direct search interfaces, and experience greater satisfaction with the information retrieval process. The methodology employed and results obtained from the experiment are presented in sections 4.1 and 4.2.

4.1 Methodology

The methodology applied in this experiment is presented in the following order: participants, apparatus, procedure, and experiment design.

4.1.1 Participants

Fourteen graduate students, twelve males and two females with ages between 18 and 40 were recruited for this experiment. They all met the following criteria: (1) had the ability of conducting tasks on personal computer; (2) had no knowledge of faceted navigation; and (3) had experience using digital libraries online. Those criteria were specified during the recruiting process.

4.1.2 Apparatus

The alphabetical browsing interface was created with the eXtensible Text Framework (XTF) on the Linux operating system (OS). XTF is an open source software platform to implement digital contents. The faceted browser interface was created with FLAMENCO on the Linux OS. FLAMENCO is also an open source software providing web-based interface for browsing large collections of items. As far as we know, FLAMENCO has been implemented and tested on digital image collections only. The Linux OS variant was Ubuntu v7.10 running kernel 2.6.22.18 virtualized under the Microsoft Windows XP Professional OS as packaged by andLinux on a 1.8 GHz Intel processor equipped with 2 GB of RAM. All participants performed the tasks on this personal computer equipped with a 17-inch LCD monitor, a standard 101-key US keyboard, and a 3-button mouse.

4.1.3 Procedure

Four sessions were administered in sequence during the experiment: pretest, training, formal tasks, and post-study. Each participant was first asked to fill out a pretest questionnaire to collect information of his/her age, gender, educational background, and familiarity with using online digital libraries and other search engines. During the training session, each participant was demonstrated an alphabetical list browsing interface implemented in XTF and a faceted browsing interface in FLAMENCO. The purpose of the demonstration was to introduce participants with the features offered in

both types of browsers. The dataset used in the training was not related to the formal tasks.

After the training session, each participant was allocated 40 minutes working on information retrieval tasks using both the alphabetical browser (AB) and the faceted browser (FB). The dataset used in those tasks was adapted from the OhioLINK's Archival Finding Aids Repository. Finding aids describe sources in detail and are essential for understanding the content of the archival collection and are text-based. The adapted repository titles were encoded in eXtensible Mark-up Language (XML). To implement alphabetical lists for the AB, corresponding schema from Dublin Core's elements was fetched from each title in the repository and utilized by the XTF software. For the FB, various PERL scripts were written to generate facets for the FLAMENCO software.

Finally, after each participant finished the formal tasks, he/she filled out a post-study questionnaire asking for his/her satisfaction with both the AB and the FB.

4.1.4 Independent Variables

The independent variable of hypotheses I and II was the type of browser, AB versus FB. The AB contained browse entry points of "Repository", "Title", and "Subject" listed in an alphabetical order, which is the conventional design in current digital libraries. Figure 6 illustrates a screen shot of the Title list in the AB.

Figure 6 A Screenshot of the Title list in the AB

The interface for the FB consisted of facets of “Topics”, “Repositories”, “Media Types”, “Geographic Locations”, “Individual Names”, and “Organizations”. The choice of the facets was adapted from the Polar Bear Expedition Digital Collections project by Professor Elizabeth Yakel at the University of Michigan School of Information (http://polarbears.si.umich.edu/index.pl?node_id=1163&lastnode_id=272). According to Professor Yakel, this project was “the first example of rethinking traditional archival finding aids to provide better access to primary sources in the web.” Figure 7 illustrates a screenshot of the top level facets in the FB.

Figure 7 A Screenshot of the facets in the FB

The independent variable of hypotheses III, IV, and V was the type of information retrieval interface, un-integrated AB and direct search interfaces versus integrated FB and direct search interfaces. Under the un-integrated AB and direct search environment, when the participants desired to search with queries, they had to first abandon the current AB information space and then enter another screen containing a search box. Under the integrated FB and direct search environment, on the other hand, participants were able to explore the digital library content with refinement through hierarchical structure using the FB and could enter a search query at any level of the FB without interrupting the interaction flow. The other major advantage of this integrated interaction mechanism is that it could provide participants with a preview of the next step which amounts to a

feeling of control and understanding without confusion (Hearst, 2008). Figure 8 illustrates a screenshot of the search functionality in the AB which is in a different information space from browsing.

Figure 8 A Screenshot of the search functionality in the un-integrated AB and direct search interface

4.1.5 Dependent Variables

The dependent variables of hypotheses I and II were the effectiveness and efficiency of information retrieval via browsing, respectively. A list of 20 yes/no binary choice questions with equivalent difficulty was asked for each participant, 10 for the AB and 10 for the FB. Participants were observed to make their chosen answers based on

performing the browsing activities rather than guessing the answers. To minimize the order effect associated with a within-subject experiment, the order in which the AB and the FB were presented to the participants were counterbalanced. The number of questions that were correctly answered by each participant was recorded as the measure of his/her task effectiveness, while the time spent in completing the questions (within the 10-minute time limit) was recorded as the measure of his/her task efficiency.

There are three dependent variables of hypotheses III, IV, and V in this experiment: effectiveness in information retrieval, efficiency in information retrieval, and users' satisfaction with the information retrieval tool. For hypotheses III and IV, a list of 16 yes/no binary choice and multiple choice questions with equivalent difficulty was asked for each participant, 8 for the un-integrated AB and direct search and 8 for the integrated FB and direct search. Participants were observed to make their chosen answers based on performing the browsing and searching activities rather than guessing the answers. Again the order in which the questions were presented to the participants was counterbalanced to minimize the order effect associated with this within-subject experiment. The number of questions that were correctly answered by each participant was recorded as the measure of his/her task effectiveness, while the time spent in completing the questions (within the 10-minute time limit) was recorded as the measure of his/her task efficiency. To test hypothesis V, a seven-point likert scale post-test satisfaction questionnaire with 5 questions was given to the participants and their responses were recorded as their satisfaction with the information retrieval tool.

4.1.6 Statistical Model

The statistical model of the one-way within subject experiment design is shown as Equation 1.

$$Y_{ij} = \mu + \alpha_i + \tau_s + \alpha\tau_{is} (i=1, 2; s=1,2,\dots,14), \quad (\text{Equation 1})$$

where μ represents the population mean, α represents the effect of the type of browser in the first hypothesis and the effect of type of information retrieval mechanism in the second hypothesis, τ represents the effect of participants, and $\alpha\tau$ represents the interaction between the interaction effect between α and τ .

4.2 Statistical Analysis

Paired T-tests were used to analyze the experiment results to test hypotheses I through IV. Tables 1, 2, and 3 summarize the statistical analysis results of the hypotheses I and II, hypotheses III and IV, and hypothesis IV, respectively.

4.2.1 Hypotheses I and II

H₁ (of hypothesis I): Compared to AB, FB can improve the effectiveness of information retrieval in text-based digital libraries

H₁ (of hypothesis II): Compared to AB, FB can improve the efficiency of information retrieval in text-based digital libraries

Table 1 summarizes the experiment results of the number of correctly answered questions and task completion time using the AB and the FB. It suggests that compared to the AB, the FB can improve the effectiveness and efficiency of information retrieval in

text-based digital libraries. As shown in Table 1, compared to the AB, the number of correctly answered questions with the FB was increased by 2 points (out of the maximum of 10 points) and 30.8% on average (s.e. = 0.82), with $t(13) = 2.44$ (one-tailed), $p = 0.015$, and Cohen's $d = 0.68$. In addition, compared to the AB, the response time with the FB was reduced by 1.09 minutes (out of the maximum of 10 minutes) and 11.3% on average (s.e. = 0.37), with $t(13) = 2.98$ (one-tailed), $p = 0.005$, and Cohen's $d = 0.83$.

Table 1 Summary of experiment results for hypotheses I and II

Browser	# of Correctly Answered Questions	Time-to-Completion
AB	6.50	9.61
FB	8.50	8.52
Percentage improvement	30.8%	11.3%
t-test	$p = 0.015$	$p = 0.001$
Effect size	Cohen's $d = 0.68$	Cohen's $d = 0.83$

Figure 9 shows the scattered plot of the number of correctly answered questions with the FB subtracted by that with the AB, i.e. the increase of task accuracy with the FB compared to that with the AB. Figure 10 shows the scattered plot of task completion time with the FB subtracted that with the AB, i.e. the decrease of task completion time with the FB compared to that with the AB. From these figures, we can see that although hypothesis I and II are supported by significance testing, two participants had higher task accuracy with the AB than with the FB, and one participant finished tasks within a shorter time with the AB than with the FB.

Figure 9 Scattered plot of the number of correctly answered questions with the FB subtracted by that with the AB

Figure 10 Scattered plot of task completion time with the FB subtracted by that with the AB

4.2.2 Hypotheses III and IV

H₁ (of hypothesis III): Compared to un-integrated AB and direct search interfaces, integrated FB and direct search interfaces can improve the effectiveness of information retrieval in text-based digital libraries

H₁ (of hypothesis IV): Compared to un-integrated AB and direct search interfaces, integrated FB and direct search interfaces can improve the efficiency of information retrieval in text-based digital libraries

Table 2 summarizes the experiment results of the number of correctly answered questions and task completion time using the un-integrated AB and direct search and the integrated FB and direct search. It suggests that compared to the un-integrated AB and direct search interfaces, the integrated FB and direct search interfaces can improve the effectiveness of information retrieval in text-based digital libraries. The number of correctly answered questions with the integrated FB and direct search interfaces was increased by 1.54 points (out of the maximum of 8 points) and 35.7% on average (s.e. = 0.75), with $t(13) = 2.05$ (one-tailed), $p = 0.03$, and Cohen's $d = 0.57$, compared to the un-integrated AB and direct search interfaces. However, using the integrated FB and direct search interfaces did not significantly reduce the time-to-completion compared to the un-integrated AB and direct search interfaces, with $t(13) = 0.51$, $p = 0.31$ (one-tailed), and Cohen's $d = 0.14$.

Table 2 Summary of experiment results for hypotheses III and IV

Browser	# of Correctly Answered Questions	Time-to-Completion
AB	4.32	8.29
FB	5.86	8.12
Percentage improvement	35.7%	2.1%
t-test	p = 0.03	p = 0.31
Effect size	Cohen's d = 0.57	Cohen's d = 0.14

Figure 11 shows the scattered plot of the number of correctly answered questions with the integrated FB and direct search interfaces subtracted by that with the un-integrated AB and direct search interfaces. Figure 12 shows the scattered plot of the task completion time with the integrated FB and direct search interfaces subtracted by that with the un-integrated AB and direct search interfaces. From Figure 11, we can see that although hypothesis III is supported by significance testing, four participants had higher task accuracy when using the un-integrated AB and direct search interfaces than with the integrated FB and direct search interfaces.

Figure 11 Scattered plot of the number of correctly answered questions with the integrated FB and direct search interfaces subtracted that with the un-integrated AB and direct search interfaces

Figure 12 Scattered plot of the task completion time with the integrated FB and direct search interfaces subtracted by that with the un-integrated AB and direct search interfaces

4.2.3 Hypotheses V

H₁ (of hypothesis V) : Compared to un-integrated alphabetical browser and direct search interfaces, integrated faceted browser and direct search interfaces can bring users greater satisfaction in performing the information retrieval tasks.

To test this hypothesis, the scores of the participants' answers in the satisfaction questionnaire were first subtracted by 4 to signify that the participants have no preference between the FB and the AB. If the subtracted values are greater than 0, then the participants are more satisfied with the FB than the AB. Otherwise, the participants are more satisfied with the AB. Table 3 summarizes the scores of participants' responses to the satisfaction questionnaire. The table suggests that compared to the un-integrated AB and direct search interfaces, the integrated FB and direct search interfaces can improve users' familiarity with the contents in the digital library (by 30.3%, $p = 0.026$, Cohen's $d = 0.59$), their confidence in the results found (by 35.8%, $p = 0.001$, Cohen's $d = 1.02$), the usefulness and flexibility of the tool (by 37.5%, $p = 0.002$, Cohen's $d = 1.00$), and their overall satisfaction (by 35.5%, $p = 0.005$, Cohen's $d = 0.84$). The Cronbach's alpha of the satisfaction questionnaire was about 0.82.

Table 3 Summary of experiment results for hypothesis V

Statistical Measure	Satisfaction Measures			
	Familiarity with content	Confidence in results found	Usefulness & flexibility	Overall Satisfaction
Average	1.21	11.43	1.50	1.42
Standard error	0.6	0.4	0.4	0.405
Percentage improvement	30.3%	35.8%	37.5%	35.5%
t-statistics	2.14	3.68	3.61	3.02
p-value	0.026	0.001	0.002	0.005
Cohen's d	0.59	1.02	1.00	0.84

V. CONCLUSIONS AND FUTURE RESEARCH

This chapter summarizes the findings and contributions in this study as well as its limitations. In addition, future research directions are discussed.

5.1 Findings and Contributions

The two principle paradigms in information retrieval are browsing and searching. While they compliment each other, when used alone, either paradigm is adequate to achieve information goals (Olston & Chi, 2003). This research identifies the observed shortcomings in exploratory browsing and direct search in current implementations in text-based digital libraries. The mechanism of exploratory browsing is often not offered, or is offered as an alphabetical list of few categories depending on the contents of the digital collection. Exploratory browsing is also not integrated with direct search functionality. In this thesis, a conceptual model of a faceted browser integrated with direct search as a tool for text-based digital libraries is proposed. Five hypotheses regarding the advantages of the conceptual model have been derived and an experiment was conducted to test those advantages.

The findings of this study indicated that compared to an alphabetical browser, the faceted browser for a text-based digital library can improve the effectiveness of information retrieval by 30.8% and the efficiency by 11.3%. In addition, compared to un-integrated alphabetical browser with direct search interfaces, the integrated faceted browser and direct search interfaces can improve the effectiveness of information

retrieval by 35.7%. However, the experiment result does not support the hypothesis that integrated faceted browser with direct search interfaces can improve the efficiency of information retrieval when compared to un-integrated alphabetical browser with direct search interfaces. This may be due to the fact that participants, through their previous experience in conventional digital libraries, had been exposed to and become familiar with the alphabetical browsing interface. In consequences, they had no significant improvement in the task completion time with the integrated faceted browser and direct search interfaces. It may also be due to the speed-accuracy tradeoff. Participants were asked to complete tasks as accurately and as fast as possible. Participants working under the un-integrated interfaces might have tried to finish their tasks fast in the sacrifice of their task performance. Although there was no significant difference in time-to-completion between the integrated and un-integrated mechanisms, the task accuracy was higher in the integrated ones. Furthermore, this study indicates that using integrated faceted browser and direct search interfaces can bring users greater satisfaction in information retrieval process, enable them to become more familiar with the digital library content, enhance their confidence in the results retrieved, and they perceive this tool as useful and flexible. The finding of overall satisfaction in using integrated faceted browser and direct search interfaces in a text-based digital library in this study is consistent with another usability study of a hierarchical faceted browsing and direct search tool in an image-based digital library where participants experienced greater task satisfaction and success (Yee et al., 2003).

As noted in section 2.5, Cox (1992) and Bates (2007) suggested what a good browsing design should be. Results of this research demonstrated that faceted browsers

in according to Cox's and Bates' suggestions provide an overview of the content of a text-based digital library in a rich scene for the eyes to take in all at once. It also provides many entry points for users to recognize the area of interest and for further search. Integrated faceted browser and direct search tool enables users to move quickly and fluidly throughout the system. In conclusion, this thesis has proposed a framework of an integrated browsing and searching mechanism and tested its effectiveness and efficiency with an empirical experiment.

5.2 Limitations

Due to limited resources, the experiment was conducted with graduate students as participants who may not well represent the entire population of digital library users. Therefore, the results may not be generalized to users of digital libraries with other characteristics, such as professional librarians. Although the gender of participants in this experiment is not equally divided between male and female, the effect of gender is not within the scope of this research.

In addition, the datasets for both training and formal tasks were chosen due to their non-restricted and non-authenticated nature; therefore, they may not reflect the restricted and authenticated text-based digital library content. Although the participants in this experiment were given some time to practice with the interfaces before performing the formal tasks, their behaviors might not reflect what would happen in real applications where users have ample practice time.

5.3 Future Research

This research can be extended to test the merits of the integrated faceted browser and direct research interfaces on general user population of text-based digital libraries in broader disciplines.

The faceted browser studied in this thesis by FLAMENCO project only allows users to “drill down” in a single concept path (Prasad & Madalli, 2007). However, there are times users want to browse further by combining multiple facets at the same time; it is also possible that users want to combine more than one facet and specific queries simultaneously. Such flexible combination of facets and queries may enable users to find information even more effectively and efficiently.

The trend in digital library development is to integrate heterogeneous digital information resources. In this respect, the “large-subject-space” presents a two-fold challenge to the faceted navigation paradigm: (1) faceted subcategories become hard to agree upon, and (2) the number of facets becomes unsuitable in size for a screen. For the first challenge, taxonomy can be used and acted as authority to be agreed or disagreed with. As to the second challenge, it has been shown that information visualization can be very effective for large information space; therefore, it should be a technique to research further. In addition, the combination of adaptive user interfaces (i.e. interfaces that tailor its content and interaction behaviors with consideration of the individual needs of its users) and agents with faceted navigation can be considered to provide users with the personalized sets of facet for more flexible, effective, and efficient information retrieval.

Appendix A Pre-study Questionnaire

We need to know a little about you in order to best interpret and analyze your reactions to the user interfaces we will be asking you to use. Please answer the following questions.

1. How old are you?

- ☐ 18-25
- ☐ 26-40
- ☐ 41-55
- ☐ Over 55

2. Are you:

- ☐ Male
- ☐ Female

3. Describe your educational background:

- ☐ Attending high school
- ☐ Attending trade or vocational school
- ☐ Attending college
- ☐ Attending graduate school
- ☐ Other (please specify degree(s)) _____

4. How frequently do you use search engine such as google and yahoo?

- ☐ Not at all
- ☐ Infrequently (only a few times monthly)
- ☐ Frequently (several times every week)

- Very frequently (daily)
- Other (please explain) _____

5. How frequently do you use digital libraries in general?

- Not at all
- Infrequently (only a few times monthly)
- Frequently (several times every week)
- Very frequently (daily)
- Other (please explain) _____

Appendix B Hypotheses I and II: Tasks on alphabetical browser and faceted browser

1. Browse and find a collection on Covered Bridge. Did you find any?

Yes No

2. Can you find any collection offered by Cleveland State University?

Yes No

3. Browse and find whether there is any memorabilia about Franklin Roosevelt?

Yes No

4. Browse and find whether there is any newspaper clipping about Paul Dunbar from Dayton, Ohio?

Yes No

5. Browse and find whether there is any collection on Rockwell International Corporation?

Yes No

6. Browse and find whether there is any collection on Schweinfurth House?

Yes No

7. Browse and find whether Dr. Samuel Saslaw was an oncologist?

Yes No

8. Browse and find whether Jack Klumpe's negatives documents history of Columbus Ohio?

Yes No

9. Browse the collections by John Carroll University; what subject is mentioned most often?

Choir Opera Alumni of the University

10. Browse the collections by Ohio Historical Society; what subject is mentioned most often?

Automobile Racing Baseball Statehouses

11. Browse and find a collection on jazz dance. Did you find any?

Yes No

12. Can you find any collection offered by Oberlin College?

Yes No

13. Browse and find whether there is any photograph about Wright Brothers?

Yes No

14. Browse and find whether there is any review about Bain Murray, the composer?

Yes

No

15. Browse and find whether there is any collection on The Dayton Woman Suffrage Association?

Yes

No

16. Browse and find whether there is any collection on Dayton Bach Society?

Yes

No

17. Browse and find whether Preston Fettrow's collection depicts Burma during World War II?

Yes

No

18. Browse and find whether Eunice Bennett's papers depict missionary work in China?

Yes

No

19. Browse the collections by Cleveland Public Library; what subject is mentioned most often?

Architects

Chess Players

Painters

20. Browse the collections by Dayton Metro Library; what subject is mentioned most often?

Lawyers

Legislators

Poets

Appendix C Hypotheses III and IV: Tasks on un-integrated alphabetical browser and direct search and integrated faceted browser and direct search

1a. You have heard that the 1st combined medical school and hospital was built in Ohio and wanted to know the name of the school. What did you find?

1. Medical College of Ohio
2. Starling Medical College
3. The Ohio State University Medical Center
4. University of Toledo Medical Center
5. None of the above

1b. Were you able to find the name of the founder?

Yes No

1c. Were you able to find more than one collection on that school?

Yes No

2a. You need to find out who founded the Poetry Center at Cleveland State, who is he?

1. Alberta Turner
2. Lewis Turco
3. Paul Dunbar
4. Wesli Court
5. None of the above

2b. Can I find any manuscript of his?

Yes

No

2c. Of all the poets in the collection, how many of them had served as the director for the Poetry Center at Cleveland State?

1

2

3

4

5

6

7

8

9

10

3a. You are researching the City of Paris in the 1940s. What are you able to find? Circle all that applies.

1. Prints
2. Photographs
3. Scrapbooks
4. Statistics
5. None of the above

4a. You are interested in discovering women professionals in from this collection. Who did you find? Circle all that applies.

1. Barritt, Evelyn
2. Gresser, Gisela
3. Morphy, Lea
4. Newton, Margaret
5. None of the above

1a. You are interested in finding a copy of the song “Heart Cry”. Where can you get it?

1. Cleveland State University
2. Dayton Metro Library
3. University of Akron
4. University of Dayton
5. None of the above

1b. Who wrote the words for the song “Heart Cry”?

1. Bain Murray
2. Karl Flaster
3. Lewis Turco
4. Vittorio Giannini
5. None of the above

1c. Who composed the music for that song “Heart Cry”?

1. Bain Murray
2. Karl Flaster
3. Lewis Turco
4. Vittorio Giannini
5. None of the above

2a. You major in psychology and would like to find collections about psychology. How many did you find?

1 2 3 4 5 6 7 8 9 10

2b. What branches of psychology are contained in what you found? Circle all that applies.

- 1.Cognitive psychology
- 2.Clinical psychology
- 3.Industrial psychology
- 4.School psychology
- 5.None of the above

2c. Of psychologists you found, who was from Australia?

- 1.Arnold, James
- 2.Berg, Irwin
- 3.Leonard, Calista
- 4.Taft, Ronald
- 5.None of the above

3a. Your last name is Hardy. You were told that one of your distance relatives was a famous college basketball player in 1920s. You would like to find out what college he was playing in?

1. John Carroll University
2. Kent State College
3. University of Akron
4. University of Dayton
5. None of the above

4a. You were asked to find the military records of Alden Family of Ohio. You find the collection and read the description. You understand the content of the letters among family members are of a particular war. You would like to find all other collections that are related to this war. How many in all can you find?

1 2 3 4 5 6 7 8 9 10

Appendix D Post-study questionnaire

Please answer the following questions by indicating the number that best expresses your feelings and opinion. **Please note: “AB” is the Alphabetical Browser; “FB” is the Faceted Browser**

1. In terms of my overall satisfaction with AB and FB, I

- ☐ Strongly prefer AB
- ☐ Prefer AB
- ☐ Slightly prefer AB
- ☐ No preference
- ☐ Slightly prefer FB
- ☐ Prefer FB
- ☐ Strongly prefer FB

2. In terms of becoming familiar with the contents of the digital library using AB and FB, I

- ☐ Strongly prefer AB
- ☐ Prefer AB
- ☐ Slightly prefer AB
- ☐ No preference
- ☐ Slightly prefer FB
- ☐ Prefer FB
- ☐ Strongly prefer FB

3. In terms of the level of confident in results found using AB and FB, I

- ☐ Strongly prefer AB
- ☐ Prefer AB
- ☐ Slightly prefer AB
- ☐ No preference
- ☐ Slightly prefer FB
- ☐ Prefer FB
- ☐ Strongly prefer FB

4. In terms of usefulness and flexibility of AB and FB, I

- ☐ Strongly prefer AB
- ☐ Prefer AB
- ☐ Slightly prefer AB
- ☐ No preference
- ☐ Slightly prefer FB
- ☐ Prefer FB
- ☐ Strongly prefer FB

5. Generally speaking, if both AB and FB are present as a choice, I

- ☐ Strongly prefer AB
- ☐ Prefer AB
- ☐ Slightly prefer AB
- ☐ No preference
- ☐ Slightly prefer FB
- ☐ Prefer FB
- ☐ Strongly prefer FB

List of References

- Ashenfelter, M., Boyko, A., Mandelbaum, J., Minor, D., McDonald, R., & Moore, R. (2009). NDIIPP Models for mass data transmission and storage. *Library Trends*, 57(3), 541-555.
- Bates, Marcia J. (2007). "What is browsing—really? A model drawing from behavioural science research" *Information Research*, 12(4) paper 330. Retrieved February 2, 2009 from, <http://InformationR.net/ir/12-4/paper330.html>
- Bodner, R. C., Chignell, M. H., Charoenkitkarn, N., Golovchinsky, G., & Kopak, R. W. (2001). The impact of text browsing on text retrieval performance. *Information Processing & Management*. 37, 507-520.
- Cove, J. F., & Walsh, B. C. (1988). Online text retrieval via browsing. *Information Processing and Management*, 24(1), 31-37.
- Cox, K. (1992). Information retrieval by browsing. *Proceedings of the 5th International Conference on New Information Technology*, Hong Kong.
- Digital Library Federation (1998). Retrieved July 31, 2008, from <http://www.diglib.org/about/dldefinition.htm>
- Essays of an Information Scientist (1984, February 13). Retrieved February 1, 2009, from <http://www.garfield.library.upenn.edu/essays/v7p045y1984.pdf>

- Faceted_browser (2009, August 16). In Wikipedia, the free encyclopedia. Retrieved August 19, 2009, from http://en.wikipedia.org/wiki/Faceted_browser
- Fox, E. A. (1999). Digital Libraries Initiative (DLI) Projects 1994-1999. Bulletin of the American society for Information Science, 26(1). Retrieved July 31, 2007. from <http://www.asis.org/Bulletin/Oct-99/index.html>
- Garshol, L. M. (2004). Metadata? thesauri? taxonomies? topic maps! Making sense of it all. Retrieved August 1, 2008, from <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>
- Gupta, A., & Jain, R. (1997). Visual information retrieval. *Communications of the ACM*, 40(5), 70-79.
- Hearst, M. A., Smalley, P., & Chandler, C. (2006). Faceted metadata for information architecture and search. Retrieved March 2, 2009, from http://flamenco.berkeley.edu/talks/chi_course06.pdf
- Hearst, M. A. (2008). UIs for faceted navigation: Recent advances and remaining open problems. *HCIR08 Second Workshop on Human-Computer Interaction and Information Retrieval*. Retrieved March 2, 2009, from <http://flamenco.berkeley.edu/papers/hcir08.pdf>
- Hildreth, C. (1982). In the Concept and Mechanics of Browsing in an Online Library Catalog. *Proceedings of the 3rd National Online Meeting*, New York, 181-196.

- Kani-Zabihi, E., Ghinea, G., & Chen, S. Y. (2006). Digital libraries: what do users want? *Online Information Review*, 30 (4), 395-412.
- Kapidakis, S., Terzis, S., & Sairamesh, J. (2000). A framework for performance monitoring, load balancing, adaptive timeouts and quality of service in digital libraries. *Internal Journal of Digital Libraries*, 3, 19-35.
- Liebscher, P., & Marchionini, G. (1988). Browse and analytical search strategies in a full-text: CD-ROM encyclopedia. *School Library Media Quarterly*, Summer 1988, 223-233.
- McKay, D., Shukla, P., Hunt, R., & Cunningham, S. J. (2004). Enhanced browsing in digital libraries: three new approaches to browsing in Greenstone. *Internal Journal of Digital Libraries*, 4, 283-297.
- MacKinlay, J. D., & Zellweger, P. T. (1995). Browsing vs. search: can we find a synergy? *CHI'95 Mosaic of Creativity*, May 7-11, 179-180.
- Marchionini, G., & Shneiderman, B. (1988). Finding facts vs. browsing knowledge in hypertext systems. *IEEE Computer*, 21(1), 70-80.
- Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. *Journal of the American Society for Information Science*, 40(1), 54-66.
- Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge: University Press.

- Mooers, C. N. (1951). Zatocoding applied to mechanical organization of knowledge. *American Documentation*, 2(1), 20-32.
- National Science Foundation Award Abstract #9984741. Retrieved September 3, 2009, from <http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=9984741>
- Olston, C., & Chi, E. H. (2003). ScentTrails: Integrating browsing and searching on the web. *ACM Transactions on Computer-Human Interaction*, 10(3), 177-197.
- Poulter, A. (2003). Browsing the virtual library. In (M. A. Drake, Ed.) *Encyclopedia of Library and Information Science*, 2nd ed. (pp. 389-396). New York: Marcel Dekker, Inc.
- Prasad, A. R. D., & Madalli, D.P. (2007). Faceted infrastructure for semantic digital libraries. *Library Review*, 57(3), 225-234.
- Rice, R. E., McCreadie, M., & Chang, S.L. (2001). *Accessing and browsing information and communication*. Cambridge, MA: MIT Press.
- Salomon, G. B. (1990). Designing casual-user hypertext: the CHI'89 InfoBooth. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People* (Seattle, Washington, United States, April 01 - 05, 1990).
- Schatz, B. R. (1997). Information retrieval in digital libraries: bringing search to the net. *Science*, 275(17), 327-334.

- Schwartz, C. (2000). Digital libraries: an overview. *Journal of Academic Librarianship*, 26 (6), 385-393.
- Shiri, A. (2003). Digital library research: current developments and trends. *Library Review*, 52(5), 198-202.
- Wynar, B. S., & Taylor, A. G. (1992). *Introduction to cataloging and classification*. Englewood, Co.: Libraries Unlimited.
- Yee, K-P., Swearingen, K., Li, K., & Hearst, M. (2003). Faceted metadata for image search and browsing. *Proceedings of the ACM CHI 2003 Human Factors in Computing Systems*, 401-408.
- Zabihi-Kani, E., Ghinea, G., & Chen, S.Y. (2006). Digital libraries: what do users want? *Online Information Review*, 30(4), 395-412.