Users' evaluation of digital libraries (DLs): Their uses, their criteria, and their assessment

Hong Iris Xie

University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA

Abstract

Millions of dollars have been invested into the development of digital libraries. There are many unanswered questions regarding their evaluation, in particular, from users' perspectives. This study intends to investigate users' use, their criteria and their evaluation of the two selected digital libraries. Nineteen subjects were recruited to participate in the study. They were instructed to keep a diary for their use of the two digital libraries, rate the importance of digital library evaluation criteria, and evaluate the two digital libraries by applying their perceived important criteria. The results show patterns of users' use of digital libraries, their perceived important evaluation criteria, and the positive and negative aspects of digital libraries. Finally, the relationships between perceived importance of digital library evaluation criteria and actual evaluation of digital libraries and the relationships between use of digital libraries and evaluation of digital libraries as well as users' preference, experience and knowledge structure on digital library evaluation are further discussed.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Digital library evaluation; Digital library evaluation criteria; User perspectives; Digital library use

1. Introduction

The availability of the Internet brings dramatic changes to millions of people in terms of how they collect, organize, disseminate, access, and use information. Perceptions of digital libraries vary and evolve over time, and many definitions for digital libraries have been proposed. The concept of a digital library means different things to different people. Even the key players in the development and use of digital libraries have different understanding of digital libraries. To librarians, a digital library is another form of a physical library; to computer scientists, a digital library is a distributed text-based information system or a networked multimedia information system; to end users, digital libraries are similar to the world wide web (www) with improvements in performance, organization, functionality, and usability (Fox, Akscyn, Furuta, & Leggett, 1995). Borgman's (1999) two competing visions of digital libraries stimulate more discussions on the definition of a digital library

---

by researchers and practitioners. The common elements of a digital library definition identified by the Association of Research Libraries (1995) are more acceptable to researchers of digital libraries:

- The digital library is not a single entity.
- The digital library requires technology to link the resources of many.
- The linkages between the many digital libraries and information services are transparent to the end users.
- Universal access to digital libraries and information services is a goal.
- Digital library collections are not limited to document surrogates: they extend to digital artifacts that cannot be represented or distributed in printed formats.

The Digital Library Initiative 1 & II funded by National Science Foundation (NSF) and other federal agencies have advanced the technical as well as the social, behavioral, and economic research needed to design and develop digital libraries. Millions of dollars have been invested into the development of digital libraries. Many unanswered questions related to whether users use them, how they use them, and what facilitate and hinder their access of information in these digital libraries. These questions cannot be answered without the evaluation of the existing digital libraries. We need to assess the usability of digital libraries in order to evaluate the full potential of digital libraries (Blandford & Buchanan, 2003). Moreover, user model of digital libraries and digital library model of users are different (Saracevic, 2004). It is important to understand users' perspectives of digital libraries. In addition, needs assessment and evaluation is also essential for the iterative design for digital libraries (Van House, Butler, Ogle, & Schiff, 1996). In order to answer these questions, we need to first clarify the following questions: (1) what are the digital library (DL) evaluation criteria? (2) who determines these criteria? or how these criteria are determined?

Just as definitions for digital libraries vary, so do the DL evaluation criteria. A major challenge for DL evaluation is to identify what to evaluate and how to evaluate non-intrusively at low cost (Borgamn & Larsen, 2003). DL evaluation is a challenging task due to the complicated technology, rich content and a variety of users involved (Borgman, Leazer, Gilliland-Swetland, & Gazan, 2000; Saracevic & Covi, 2000). The most recognized DL evaluation criteria are derived from evaluation criteria for traditional libraries, IR system performance, and human-computer interaction (Chowdhury & Chowdhury, 2003; Marchionini, 2000; Saracevic, 2000; Saracevic & Covi, 2000). Very few studies actually apply all the DL evaluation criteria to assess a digital library. Many of the studies focus on the evaluation of usability of digital libraries. After reviewing usability tests in selected academic digital libraries, Jeng (2005a, 2005b) found that ease of use, satisfaction, efficiency and effectiveness are the main applied criteria. Some of the evaluation studies extend to assess performance, content and services of digital libraries while service evaluation mainly concentrates on digital reference (Carter & Janes, 2000). Other evaluation studies also look into the impact of digital libraries (Marchionini, 2000).

Little research has investigated users' evaluation of digital libraries, in particular, their criteria and their actual assessment of digital libraries. Xie (2006) pointed out that even though researcher have developed DL evaluation criteria, and conducted actual studies to evaluate existing digital libraries or prototypes of digital libraries, there is a lack of user input regarding evaluation criteria. User evaluation of digital libraries by applying their own criteria is needed. The remaining questions related to DL evaluation are: What is the relationship between users' use and evaluation of digital libraries? And what is the relationship between users' perceived DL evaluation criteria and their actual evaluation?

2. Research problem and research questions

Evaluation of digital libraries is an essential component for the design of effective digital libraries. Digital libraries are designed for users to use. However, most of the research on evaluation of digital libraries has applied criteria from researchers themselves. In particular, these studies focus on the usability studies. Little has been done on the identification and ranking DL evaluation criteria from users' perspectives. Furthermore, less has evaluated digital libraries by applying criteria developed from users themselves.

This study continues the author's previous research (Xie, 2006) on digital library evaluation, and intends to apply DL evaluation criteria developed by users. To be specific, the following research questions are investigated for this study:

(1) How do users use digital libraries?
(2) What are users' perceptions of the important DL evaluation criteria?
(3) What are the positive and negative aspects of digital libraries from users' perspectives?

By answering these research questions, we are able to understand more about relationships between the relationships between users' use and evaluation of digital libraries and relationships between users' perceived DL evaluation criteria and their actual DL evaluation.

3. Literature review

The emergence of digital libraries calls for the need for the evaluation of digital libraries. Evaluation is a research activity, and it has both theoretical and practical impact (Marchionini, 2000). An evaluation is a judgment of worth. The objective of DL evaluation is to assess to what extent a digital library meets its objectives and offer suggestions for improvements (Chowdhury & Chowdhury, 2003). Even though there are no standard evaluation criteria and evaluation techniques for DL evaluation, DL evaluation research has been conducted on different aspects. Previous research on DL evaluation related to this study can be summarized mainly into the following categories:

- General DL evaluation framework and evaluation criteria.
- Specific DL evaluation framework and evaluation criteria, e.g. usability.
- Usability studies.
- Evaluation studies on other aspects.

While more research is on specific issues of DL evaluation, there is less research on general DL evaluation criteria. Moreover, these criteria are derived from evaluation criteria on traditional libraries, human-computer interaction, IR system performance and digital technologies. Digital libraries are extensions and augmentations of physical libraries. Marchionini (2000) suggested applying existing techniques and metrics to evaluation digital libraries, such as circulation, collection size and growth rate, patron visits, reference questions answered, patron satisfaction, financial stability. Reviewing evaluation criteria for libraries by Lancaster (1993) and library and information services by Saracevic and Kantor (1997), Saracevic (2000) and Saracevic and Covi (2000) offered more detailed evaluation variables related to traditional library criteria on collection consisting of purpose, scope, authority, coverage, currency, audience, cost, format, treatment, and preservation; on information including accuracy, appropriateness, links, representation, uniqueness, comparability, presentation; and on use comprising of accessibility, availability, searchability, and usability; and standards. Referring to evaluation criteria for IR systems by Su (1992), they further suggested the applications of traditional IR criteria related to relevance with precision and recall, satisfaction and success into digital library evaluation. Finally, citing human-computer interaction research by Shneiderman (1998), human-computer interaction/interface criteria were discussed regarding usability, functionality, efforts as well as task appropriates and failures.

Researchers have also developed digital library evaluation framework from different dimensions and levels. For example, DELOS Network of Excellence has conducted a series of research concerning the evaluation of DLs. Fuhr et al. (2007) developed a DL evaluation framework based on a large-scale survey of DL evaluation activities. This framework is derived from conceptual models for EL evaluation. Fuhr, Hansen, Mabe, Micsik, and Sølvberg (2001) proposed a scheme for digital library evaluation which contains four dimensions: data/collection, system/technology, users, and usage. Data/collection assessment mainly focuses on content, description, quality/reliability attributes, and management and accessibility attributes. System/technology assessment is related to user technology, information access, system structure, and document technology. Users and their uses are represented by the types of users, what domain areas users are interested in, how they seek information, and the purpose of seeking information. Tsakonas, Kapidakis, and Papatheodorou (2004) further examined the interactions of DL components. The analysis of the relationships between user-system, user-content, and content-system led to the following evaluation foci: usability, usefulness, and system performance, respectively. Fuhr et al. (2007) integrated Saracevic's (2004) four dimensions of evaluation activities

(construct, context, criteria, and methodology) and essential questions regarding DL evaluation (why evaluate, what to evaluate, and how to evaluate?) together. The why question focuses on making strategic decisions related to the constructs, the relationships, and the evaluation. The what question concerns the major construct of digital libraries and their relationships. The how question offers guidance regarding procedures to perform evaluation. The major contribution of DELOS Network's work is that researchers not only illustrate but also justify why they evaluate, what they evaluate, and how they evaluate.

Lesk (2005) also identified some specific issues related to general DL evaluation. For usability, he discussed the issue of accessibility for users with special needs. For retrieval evaluation, he emphasized the guidance for uses to retrieve too little or too much. For collection, he conferred the problems of item quality. Chowdhury and Chowdhury (2003) reviewed DL evaluation criteria for libraries, IR systems and user interface. They further stressed the need to assess the overall impact of digital libraries on users and society. Nicholson (2005) created a conceptual framework for the artifact-based evaluation in digital libraries to have an in-depth understanding of digital library services and users. However, there is a gap between evaluation theorists and evaluation practitioners identified by Saracevic (2004) since these DL criteria are not always applied in practical studies.

Within the DL evaluation criteria, usability is the one that has been most investigated. Accepted definitions of usability have been focused on multiple usability attributes such as learnability, efficiency, memorability, errors, and satisfaction (Nielsen, 1993). Usability was also extended to performance measures, such as efficiency of interactions, avoidance of user errors, and the ability of users to achieve their goals, affective aspects, and the search context (Blandford & Buchanan, 2002). Blandford and Buchanan (2003) also examined the classical usability attributes in the context of digital libraries, and they suggested adopting many of these attributes to the evaluation of digital libraries. Some of them, such as learnability, need to be modified because users treat the library system as a tool, not as an object of the study. They are more concerned with building a user perspective into the design cycle than with final evaluation. After reviewing the research on usability, Jeng (2005a, 2005b) concluded that usability is a multidimensional construct. She further proposed an evaluation model for assessment of the usability of digital libraries by examining their effectiveness, efficiency, satisfaction, and learnability. User satisfaction covers ease of use, organization of information, labeling, visual appearance, content and error correction. The evaluation model was tested, and the results revealed that effectiveness, efficiency, and satisfaction are interrelated. Dillon (1999) proposed a qualitative framework (termed TIME) for designers and implementers to evaluate usability of digital libraries which focuses on user task (T), information model (I), manipulation facilities (M) and the ergonomic variables (E). Buttenfield (1999) suggested two evaluation strategies for usability studies of digital libraries: the convergent method paradigm that applies the system life cycle into the evaluation process and the double-loop paradigm that enables evaluators identify the value of a particular evaluation method under different situations. Even though usability is widely discussed, it is important to identify the uniqueness of usability attributes for the assessment of digital libraries.

The majority of DL evaluation studies are usability studies. Researchers themselves make decisions regarding the selection of appropriate usability criteria for evaluation. Design elements are one of the popular components for usability studies, for example Van House et al. (1996) focused on query form, fields, instructions, results displays, and formats of images and texts in the iterative design process for the University of California Berkeley Electronic Environmental Library Project. Some usability studies examine specific design for interfaces, such as organization approaches of digital libraries. Meyyappan, Foo, and Chowdhury (2004) measured the effectiveness and usefulness of the alphabetic, subject category, and task-based organization approaches in a digital library, and the results showed that the task-based approach took the least time in locating information resources. Nature and extent of use represent another type of usability studies. Evaluating a digital library testbed, Bishop et al. (2000) investigated the extent of use, use of the digital library compared to other systems, nature of use, viewing behavior, purpose and importance of use, and user satisfaction. In Cherry and Duff's (2002) longitudinal study of a digital library collection, they focused on how the digital library was used and the level of user satisfaction with response time, browse capabilities, comprehensiveness of the collection, print function, search capabilities, and display of document pages. Hill et al. (2000) tested user interfaces of the Alexandria digital library (ADL) through a series of studies. The following requirements were derived from user evaluation: unified and simplified search, being able to manage sessions, more options for results display,

offering user workspace, holdings visualization, offering more Help functions, allowing easy data distribution, and informing users of the process status. User evaluation plays a key role in the evolvement of the ADL system. Bertot, Snead, Jaeger, and McClure (2006) adopted a broad understanding of usability, including satisfaction, in addition to ease of use, efficiency, and memorability for the iterative evaluation of Florida Electronic Library. They also brought in functionality and accessibility as major DL evaluation criteria.

Some researchers concentrate on specific digital libraries and specific users, in particular, educational digital libraries and learners. Focusing on digital libraries for teaching and learning, Borgman et al. (2000) conducted formative evaluation in formulating design requirements and summative evaluation in judging learning outcomes. Yang (2001) examined learners' problem-solving process in using the Perseus digital library by adopting an interpretive and situated approach. The findings of the study helped designers develop and refine better intellectual tools to facilitate learners' performance. Kassim and Kochtanek (2003) performed usability studies of an educational digital library in order to understand user needs, find problems, identify desired features, and assess overall user satisfaction.

In addition to usability studies, DL evaluation studies also cover system performance and content. Applying multifaceted approaches to the evaluation of the Perseus Project, Marchionini, Plaisant, and Komlodi (1998) concentrated on evaluating learning, teaching, system consisting of performance, interface, and electronic publishing, and content comprising of scope and accuracy. Hill et al. (2000) collected feedback about the users' interaction with the interfaces of Alexandria Digital Library, the problems of the interfaces, the requirements of system functionality, and the collection of the digital library. Service, which does not get enough attention, is another aspect for DL evaluation. Based on user perceptions and user preferences, some researchers proposed a framework and combined methodologies for judging the outcomes, impacts or benefits of digital library services (Choudhury, Hobbs, Lorie, & Flores, 2002; Heath et al., 2003). However, actual large-scale studies by applying the combined methodologies are needed. The most evaluated DL service is reference services. For example, Carter and Janes (2000) analyzed logs of over 3000 questions sent to the Internet Public Library regarding how those questions were asked, handled, answered, or rejected. Electronic journals service of the digital library is another area for evaluation. In these studies, the evaluation emphasizes more on characteristics of users and their usage patterns related to preferred databases, preferred electronic journals, and their frequency of use (Atilgan & Bayram, 2006; Monopoli, Nicholas, Georgiou, & Korfiati, 2002).

Interaction between users and digital libraries is also an important component for evaluation. Budhu and Coleman (2002) identified the key attributes of interactivities: reciprocity, feedback, immediacy, relevancy, synchronicity, choice, immersion, play, flow, multidimensionality, and control. They evaluated interactivities in a digital library with regard to the following aspects: interactivities in resources, resources selection, description of interactivities in metadata, and interactivities in interface. Their study highlights the importance of the evaluation on interactivity that enhances learning. DL evaluation can also reveal the factors affecting users' acceptance of a digital library. Thong, Hong, and Tam (2002) identified the determinants of user acceptance of digital libraries, and among them, perceived usefulness and ease of use are the major ones which are predicted by interface characteristics (terminology clarity, screen design and navigation clarity), organizational context (relevance and system visibility) and individual differences (computer self-efficacy, computer experience, and domain knowledge). Finally DL evaluation extends to the impact of digital libraries. Derived from usage analysis of log data, some researchers evaluated the impact of a digital collection and characterized the user community of a specific digital library, and they further identified research trends in user communities of digital libraries over time (Bollen & Luce, 2002; Bollen, Luce, Vemulapalli, & Xu, 2003).

Previous evaluation research has assessed digital libraries from different aspects. Even though these studies have involved real users and their usage in their evaluation process, it is the researchers who decide what to evaluate. In these studies, users are only the subjects of the evaluation. They do not have input in terms of what to evaluate and how to evaluate. In some of the studies, researchers do solicit user perception on some of the DL evaluation criteria. In Jeng's (2005a, 2005b) study, the evaluation also detected users' perceptions of ease of use, organization of information, terminology, attractiveness, and mistake recovery. For example, ease of use is considered 'simple," "straightforward," "logical," "easy to look up things," and "placing common tasks upfront." Very few researchers conducted DL evaluation studies from users' perspectives. Xie (2006) investigated DL evaluation criteria based on users' input. Users developed and justified a set of essential criteria for the evaluation of digital libraries. At the same time, they were requested to evaluate their own selected

digital libraries by applying the criteria that they were developing. After comparing evaluation criteria identified by users and researchers, and applied in previous studies, the author found that there is a commonality in the overall categories of the evaluation criteria. However, users emphasized more on the usefulness of the digital libraries from their perspectives and less from the perspectives of developers and administrators. For example, cost, treatment and preservation are the important criteria for DL evaluation from researcher's perspectives, but the participants of that study did not identify these criteria as the important ones. For the same reason, participants placed more emphasis on the ease of use of the interface and high quality of the collection in evaluating digital libraries. Xie (2006)' study offers not only detailed information about evaluation criteria but also why they are important to users.

What can we draw from previous research on DL evaluation? Researchers have developed frameworks to systematically evaluate DL constructs and their associated relationships. The commonly agreed DL evaluation criteria and variables are related to the usability of interface, the value of collection, and the system performance. These criteria and variables are the results of research in evaluating human-computer interaction, information retrieval system performance, and traditional libraries. Previous research offers guidance for researchers and professionals to conduct evaluation studies. At the same time, researchers and professionals have conducted a variety of evaluation studies for the assessment of prototypes or actual digital libraries. These studies have largely focused on the usability aspects of digital libraries. In these studies the criteria or variables applied in the evaluation are determined by researchers or professionals themselves depending on their purpose of evaluation. Compared to the evaluation criteria developed by researchers, the criteria and variables applied in the studies are more detailed and normally on one specific area. These studies contribute significantly to the improvement of the design of digital libraries.

However, there are two gaps in DL evaluation research and studies: first, while researchers have developed DL evaluation criteria and variables from different levels and dimensions, the evaluation studies in general only focus on one aspect of the digital library, such as usability studies. There are very few studies that assess every component of digital libraries. There is also a need to promote communication between researchers and professionals so DL evaluation research can be applied to practical studies. Second, while researchers and professionals have actively engaged in DL evaluations, users are mainly the passive subjects of these studies. Their feedback is mostly limited to what researchers or professionals define for them. There is a lack of users' involvement in determining DL evaluation criteria and associated variables. In order to gain a complete picture of users' assessment of DLs, we need to engage users in every aspect of DL evaluation from defining DL evaluation criteria, their uses, and their assessment.

This study attempts, to some extent, to fill in the gaps discussed above, and further enhances the author's previous (Xie, 2006) study in three ways: (1) the importance of DL evaluation criteria developed by the previous study are further rated and justified, so we can have an in-depth understanding of the most important DL evaluation criteria proposed by users. (2) Subjects of this study searched and used the two selected digital libraries, and their uses were recorded. Therefore the relationships between their use and their evaluation can be explored. (3) Subjects of this study also assessed the two selected digital libraries by applying the most important criteria identified by themselves. The relationships between their perceived DL evaluation criteria and their actual evaluation can be further clarified.

4. Methodology

4.1. Sampling

Subjects were recruited from School of Information Studies, University of Wisconsin–Milwaukee. Two reasons were considered for the recruitment: (1) These subjects have a need to understand digital libraries and have some experience with the use of digital libraries. (2) These subjects are the targeted audience for similar types of digital libraries. (3) This study is the extension of the author's previous study (Xie, 2006). That is why it is important to recruit subjects who share similar characteristics with subjects in the previous study. Nineteen subjects were recruited for this study. They were graduate students who were interested in the learning and using of digital libraries. Female (58%) and male (42%) subjects are pretty close in the composition of

the subject pool. All of them had enough knowledge of digital libraries, and have used and searched digital libraries before this study.

4.2. Data collection

Multiple methods were employed to collect data: diary, questionnaire, and survey. Two digital libraries that represent different contents and designing styles were selected for subjects to evaluate: American memory (AM) http://lcweb2.loc.gov/ammem/ammemhome.html and University of Wisconsin Digital Collections (UWDCs) http://uwdc.library.wisc.edu/. There are two reasons for the selection of these two digital libraries for assessment. First, these two digital libraries represent two types of design and collections of digital libraries. American memory presents comprehensive materials related to American history and creativity. It consists of browsing, searching and help mechanisms at the general level as well as at the individual collection level. University of Wisconsin Digital Collections is created based on the teaching and research needs of the UW community and State of Wisconsin, and it provides access to rare or fragile items of broad research value. It does not have much general level of search, browsing and help mechanisms, but it does have these mechanisms at individual collection level. Second, the subjects recruited for this project are the targeted users of the selected digital libraries. Figs. 1 and 2 illustrate the different designs of the two selected digital libraries. At the study period, UWDCs did not contain overall search function. The screenshot of UWDCs is a current version in which search function is added after hearing feedback from users.

The data collection procedures are as follows: First, subjects were instructed to find information for six questions for each of the two digital libraries selected for this study including their own questions. The questions for these two digital libraries are in the similar structure for easy comparison later. For example, subjects were instructed to find a film titled "A Day with Thomas A. Edison" and its copyright owner in American Memory. At the same time, they were also asked to find a picture of Cutler–Hammer Building, and the photographer for the picture in UWDCs. In another question, subjects need to identify two approaches to find information about Jackie Robinson and a picture of him in AM and identify two approaches to find a picture of a ship named "Lotus" and the type of ship and its original name. At the same time, the subjects were asked to record their search process in their *diaries* including: (1) time used (by minutes), (2) the whole search process (any queries used, any categories browsed, any help used, and any problems encountered), and (3) The answers to each of the questions. The subjects could work on the digital libraries at any locations that they felt comfortable.
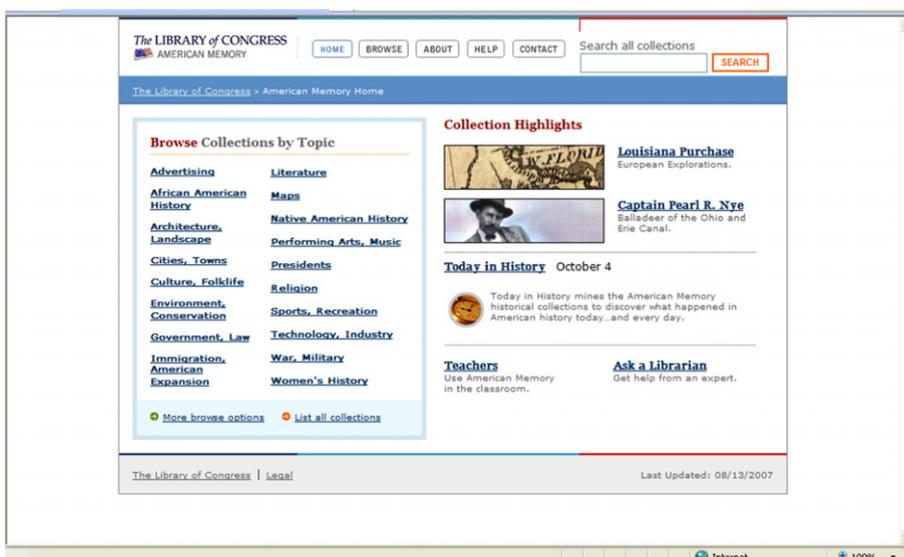

Fig. 1. Screenshot of American Memory.

Fig. 2. Screenshot of University of Wisconsin Digital Collections.

Second, they were instructed to rate the importance of the digital library evaluation criteria in a close-ended *questionnaire* and add new ones if needed. They were asked to use the following scale for the rating: 1 = not at all important, 2 = a little, 3 = some, 4 = very, and 5 = extremely important. This questionnaire was designed based on a set of essential criteria for the evaluation of digital libraries developed and justified by 48 subjects in the previous study (Xie, 2006). The subjects of this study were also asked to justify their choices of criteria that they consider very or extremely important for the evaluation of digital libraries.

Third, they were instructed to apply the evaluation criteria that they identified as important to evaluate the two digital libraries that they have searched information in the first step. Evaluate and rate the quality of American Memory Digital Library and University of Wisconsin Digital Collections. In the open-ended *survey*, they were asked to discuss the strength and problems of American Memory Digital Library and University of Wisconsin Digital Collections based on their rating of the criteria and sub-criteria of the digital library with justification of their ratings, examples, and how to overcome the problems.

4.3. Data analysis

Both quantitative and qualitative methods were used to analyze the data. Table 1 summaries the data collection and data analysis. Quantitative methods were employed to conduct descriptive data analysis, such as frequency and mean. Content analysis (Krippendorff, 2004) is applied to develop categories and sub-categories

Table 1
Summary of data collection and data analysis

Research questions	Data collection	Data analysis
(1) How do users use digital libraries?	Diary	Identify time used, no. of initial queries used, no. of paths in browse, no. of levels in browse, no. of queries in browse, no. of times help used and types of problems encountered
(2) What are the perceptions of the importance of the digital library evaluation criteria?	Questionnaire	Calculate the mean and standard deviation of perception of the importance of each DL evaluation criteria; Identify types of reasons for their choices of important criteria
(3) What are the positive and negative aspects of digital libraries from users' perspectives?	Survey	Calculate mean and standard deviation of evaluation of AM and UWDCs; Identify types of positive and negative aspects of digital libraries

of evaluation criteria, and categories and sub-categories of positive and negative aspects of digital libraries. According to Krippendorff (2004), categorical distinctions define units by their membership in a category by having something in common. In this study, each category and sub-category of the positive and negative aspects of digital libraries is defined and discussed by citing responses directly from the participants. Instead of creating an existing structure, the positive and negative aspects of digital libraries were derived directly from users.

The data analysis procedures contain the following steps: For use of digital libraries, (1) for each subject and each question, identify time used, initial queries used, paths in browse, levels in browse, queries in browse, help used and types of problems encountered (see Table 2); (2) then, calculate the average time used, number of initial queries used, number of paths in browse, number of levels in browse, number of queries in browse, and number of times help used. For perception of DL evaluation criteria, (1) calculate the mean and standard deviation of the perception of the importance of each DL evaluation criteria based on the questionnaire; (2) identify types of reasons for their choices of important or extremely important criteria. For actual evaluation of digital libraries, (1) calculate the mean and standard deviation of their evaluation for AM and UWDCs; (6) identify types of positive aspects of the selected digital libraries; (7) identify types of negative aspects of the selected digital libraries.

In order to identify types of positive and negative aspects of the selected digital libraries, the researcher took the following steps: (1) Each subject's responses in relation to positive and negative aspects of the two selected digital libraries were extracted from the survey. (2) All the responses that shared the same properties were classified into one category. For example, all the responses about the use of interface of digital libraries were classified into one category, and all the responses about the content of digital libraries were classified into another category. (3) After all the responses were classified into general types of evaluation criteria, all the responses of one type of criterion were further classified into different positive and negative aspects of digital libraries based on their properties. For example, in usability, all the responses regarding different perspectives of simplicity of interface design were classified into one aspect. (4) A name for each aspect was assigned based on the content of the responses, such as simplicity versus attractive interface, etc. The same procedures were followed for the analysis of the justification of important or extremely important DL evaluation criteria. To save space, examples of positive and negative aspects of digital libraries and justification of important DL evaluation criteria are discussed in the Results section.

5. Results

The results present answers to the three research questions in relation to users' use, their evaluation criteria and their actual evaluation of the two selected digital libraries. Fig. 3 illustrates the structure of the results.

5.1. The use of the two selected digital libraries

The best way to evaluate digital libraries is to actually use them. Nineteen subjects tried to find information related to six questions including their own questions in American Memory and University of Wisconsin Digital Collections. Their experience of using these two digital libraries greatly affects their perception of the importance of DL evaluation criteria and their evaluation of the two digital libraries. Table 3 presents the summary data of American Memory use. Table 4 shows the summary data of University of Wisconsin Digital Collections use.

First of all, the design of the digital libraries affects how users search them. In particular, the availability of unavailability certain features affect how uses interact with these digital libraries. University of Wisconsin Digital Collections did not have search functions; therefore there are no initial queries for all the questions.

Table 2
Coding sheet for digital library use

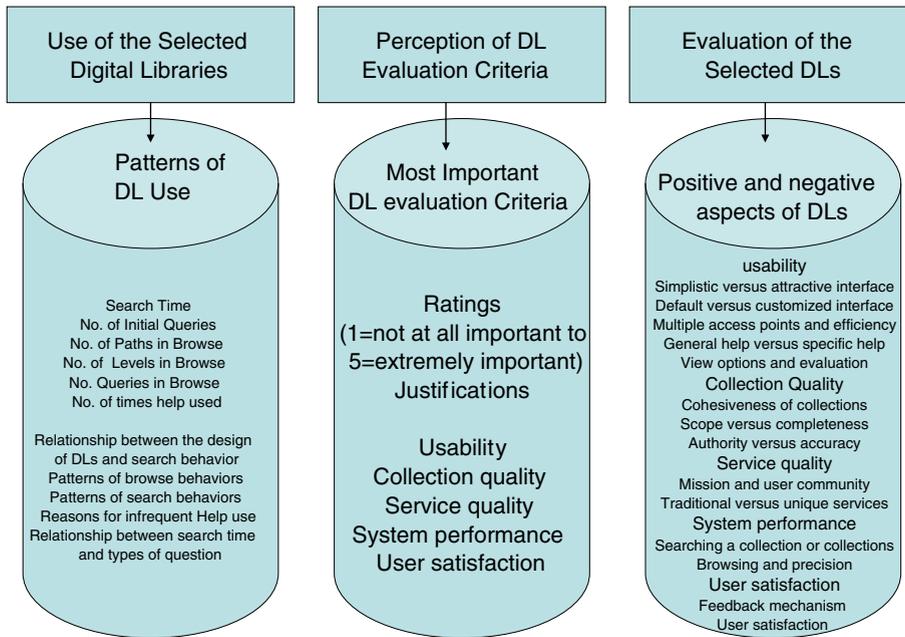| Subject no.	Question no.	Search time	Initial queries	Paths in browse	Levels in browse	Queries in browse	Help used

Fig. 3. Structure and summarization of the results.

Table 3
Summary data of American Memory use

Question No.	Search time		Initial queries		Paths in browse		Levels in browse		Queries in browse		Help used	
		Mean		Mean		Mean		Mean		Mean		Mean
A1	$n = 17$	3.27	$n = 16$	1.13	$n = 6$	1.83	$n = 9$	2.22	$n = 2$	1.5	$n = 0$	0
A2	$n = 17$	3.32	$n = 5$	1	$n = 15$	1.13	$n = 16$	2.31	$n = 11$	1.09	$n = 0$	0
A3	$n = 17$	13.6	$n = 12$	2.25	$n = 10$	1.9	$n = 11$	2.36	$n = 7$	1.29	$n = 1$	1
A4	$n = 16$	2.38	$n = 12$	1	$n = 18$	1.33	$n = 14$	2.07	$n = 4$	1	$n = 0$	0
A5	$n = 17$	5.8	$n = 2$	2	$n = 13$	1.53	$n = 16$	2.81	$n = 3$	1.67	$n = 0$	0
A6	$n = 13$	3.77	$n = 8$	1.38	$n = 10$	1.2	$n = 12$	2.42	$n = 4$	1	$n = 1$	1

Table 4
Summary data of University of Wisconsin Digital Collections use

Question No.	Search time		Initial queries		Paths in browse		Levels in browse		Queries in browse		Help used	
		Mean		Mean		Mean		Mean		Mean		Mean
B1	$n = 17$	7.18	$n = 0$	0	$n = 17$	2.65	$n = 17$	2.71	$n = 16$	1.25	$n = 0$	0
B2	$n = 15$	15.07	$n = 0$	0	$n = 13$	2.23	$n = 37$	2.24	$n = 13$	1.39	$n = 0$	0
B3	$n = 14$	8.70	$n = 0$	0	$n = 18$	2.22	$n = 31$	2.45	$n = 19$	1.32	$n = 0$	0
B4	$n = 15$	3.38	$n = 0$	0	$n = 19$	1	$n = 19$	3.37	$n = 4$	1.25	$n = 0$	0
B5	$n = 12$	3.42	$n = 0$	0	$n = 16$	1.13	$n = 18$	3.22	$n = 12$	1.08	$n = 0$	0
B6	$n = 12$	4.88	$n = 0$	0	$n = 13$	1.23	$n = 14$	2.79	$n = 6$	1	$n = 0$	0

American Memory allows users to search across collections. That is why the mean of initial queries for each question is from 1 to 2.25. It seems that users did not reformulate queries that much. Most subjects did not even reformulate their queries. If their queries did not generate successful results, they would change from searching to browsing. For some of the subjects who actually reformulated their queries, their reformulations were limited. Their reformulations could be characterized in the following ways: (1) adding more terms, such as from "Edison" to "day with Thomas Edison" to "day with Thomas a. Edison" (S7). (2) using different

forms of a term, for example, from "Gettysburg address translation" to "Gettysburg Address translated" (S15). (3) changing from subject search to actual content search, for example from "Franklin Roosevelt and Pearl Harbor" to "a day which shall live in infamy" (S4). The first two approaches were more frequently applied in this study. In addition, subjects normally extracted terms from the search questions instead of coming up with their query terms themselves.

Second, one uniqueness of digital libraries is their browsing functions. Browsing, to some extent, plays a more important role in facilitating users fine relevant information then searching. Browsing behaviors can be examined at two levels: number of paths and number of levels in browse. The path refers to subject categories that participants accessed in AM and collections that they selected in UWDCs because AM allows users to browse by subjects while UWDCs does not offer that function. Comparatively speaking, the average of paths in browse for answers to questions in UWDCs (three of the questions are higher than 2) is higher than in AM (from 1.33 to 1.83). One potential reason is that users of UWDCs had to browse different collections since they could not search and browse by subjects across collections. In particular, when there are multiple collections available for the subject area of a search task. Subjects could not search across collections, so they could only try different related collections to find the answer. For example, a subject checked the following collections in order to find a picture of household items from China made in Song dynasty: The Arts Collection, Digital Library for the Decorative Arts and Material Culture, South and Southeast Asia Video Archive, Southeast Asian Images and Texts, and Global Perspectives on Design and Culture.

In terms of levels in browse, subjects had to go through several levels before they could find the answers of their search questions. Subjects went more levels in UWDCs (average levels in browse from 2.24 to 3.37) than in AM (average levels in Browse from 2.07 to 2.81). The browsing option enables users to identify the answer without actually searching for it. For example, here is the path for subject 11 and 12 looking for question 1 in AM, Technology, Industry > Edison Companies ~ film and sound Recordings > Browse by alphabetical Title List. In UWDCs, the path starts from an actual collection because there is no subject browse option across collections. Interestingly, most of the subjects followed the same path to search for question 4, Collections > Africa Focus: Sights and Sounds of a Continent > Search the Collection > Greetings. Subjects can only access the greetings items by selecting Search the Collection first.

Third, another benefit for the digital libraries is the availability of a search option within a category or a collection, which is an effective approach for users to find specific information within a specific category or a collection. Even though UWDCs did not offer general search, it did provide search options within specific collections. The mean of number of queries used are pretty much similar in searching AM (ranging from 1 to 1.66) and UWDCs (ranging from 1 to 1.38). Subjects did not formulate their queries much either. Query reformulations here followed the same patterns occurred in the initial query reformulations. It also generated two new patterns: (1) changing from specific to general, for example, from "hi" to "greetings (S7)." (2) deleting terms, e.g. from "Song dynasty China" to "Song dynasty (S17)," from "Song Dynasty" to "Song (S14)." While some subjects browsed to find information, some of them further integrated browse and search strategies together.

Fourth, surprisingly, very few users used Help and other features when they encountered problems in the search process. American Memory does offer several general implicit and explicit help as well as specific help within specific collections to users. Its explicit Help feature presents information related to How to View, Search Help, FAQs and Contact Us. Only two users with each one tried only once in searching two of the questions in American Memory. One subject complained the unusefulness of the Help, "I finally checked the help options and looked around for some search tips. Help was not all too much help for my purpose (S4)." Another subject found the discrepancy between what the Help says and what the actual digital library offers. According to subject 1, "After several unsuccessful queries with the general search and more specific ones in several collections, I went to Help and found out there is a search option called Bibliographic Record Search which allows to search more specifically. However, on the site itself I could not find this search option." Their experience in use Help features of other types of IR systems in the past and these digital libraries affect their perceptions of the importance of Help features in digital libraries.

Finally, it seems that search times vary depending on the question. It takes from 2.38 to 13.6 min and from 3.38 to 15.07 min, respectively for users to find information for their questions in AM and UWDCs. Overall, subjects took less time in finding information in AM than in UWDCs. It is largely affected the design of these

digital libraries. Their use of digital libraries sets up a foundation for their evaluation of these two digital libraries. Their experience in interacting with these digital libraries also affects their perception of the importance of DL evaluation criteria.

5.2. The perceptions of the importance of the DL evaluation criteria

Table 5 presents mean and standard deviation of the perceived importance of each of the DL evaluation criterion and associated variables generated from author's previous research (Xie, 2006). These evaluation criteria identified from users range from interface usability, collection quality, service quality, system performance to user satisfaction. Digital libraries can be evaluated at two levels. At the first level, interface usability in general, collection quality in general and system performance in general are used to measure which aspect users care the most. At the second level, different variables are used to represent interface usability, collection quality and system performance. For example, collection quality can be assessed by its scope, authority, accuracy, completeness and currency. To users, interface usability in general ($x = 4.7$) and system performance in general ($x = 4.6$) were the most important ones, followed by user satisfaction ($x = 4.3$) and collection quality in general ($x = 4.2$). Comparatively speaking, service quality was rated less important with most of the categories deemed as some important except usefulness of services. This is different comparing with the previous research (Xie, 2006) in which interface usability and collection quality were considered as the important criteria. One reason might be that in previous study, users evaluated a variety of digital libraries developed by different organizations that they did not always trust the authority and accuracy of the content. In this study, users only evaluated two selected digital libraries that were perceived with high authorities. In that sense, the content is less important comparing with the system performance.

In the previous study, users emphasized both usability and collection quality. In this study, usability in general was rated the highest in its importance. One of the reasons might be that these users actually looked for information in these two digital libraries before they rated the DL evaluation criteria. To users usability is essential for their effective use of digital libraries. One subject well explained it, "Obviously, even with

Table 5
Overall perception of the importance of the digital library evaluation criteria ($N = 19$) (1 = not at all important 2 = a little 3 = some 4 = very 5 = extremely important)

Types of criteria	Variables	Mean	Standard deviation
Interface usability	Interface usability in general	4.7	0.45
	Search and browse functions	4.5	0.70
	Navigation	4.4	0.64
	Help features	3.6	1.12
	View and output options	3.8	0.62
	Accessibility	4.2	0.69
Collection quality	Collection quality in general	4.2	0.81
	Scope	3.8	0.92
	Authority	4.3	0.91
	Accuracy	4.6	0.68
	Completeness	3.8	0.83
	Currency	3.8	0.83
Service quality	Mission	3.4	0.90
	Targeted user community	3.5	0.86
	Traditional library service	3.4	0.77
	Unique services	3.6	0.84
	Usefulness	4.5	0.62
System performance	System performance in general	4.6	0.62
	Efficiency and effectiveness	4.5	0.61
	Relevance	4.3	0.75
	Precision and recall	4.2	0.71
User satisfaction	User satisfaction	4.3	0.67
	User feedback	4.1	0.71
	Contact information	4.4	0.76

collections comprised of excellent content, it would be of little use to the user if the level of usability is poor. As Ranganathan noted in his *Five Laws of Library Science*, 'Books are for use.' This is just as valid today, even when the concept is updated to include information in electronic format, comprised of a variety of media (S7)." More important, usability needs to gear to users with different levels of knowledge and skills. A subject pointed out, "[Digital library design] should be easy to use and easy for newcomers to learn the interface (S15)." Another one further emphasized, "It should be based on the skills and needs of the intended audience. If the digital library is not suited to the search skills and content of its intended users, it may be too difficult for them to use (or even too childish or simplistic for their taste) (S17)."

In order to design digital libraries for high levels of usability, interface design interface usability is crucial. The question is what users want for the interface design. According to one subject, "With an effective and easy to use interface a digital library is more usable. It should be simple, effective, accurate, and understandable (S4)." Other requirements include saving time (S7) and good functionality (S11). "A simple no-frills interface that is easy to navigate, that provides search and browse features that work as expected, and that delivers results in a readable format will always be preferred to a 'pretty' page that is difficult to read and navigate," Subject 12 illustrates in more detail.

Search and browse functions ($x = 4.5$) play a key role for the usability of a digital library. To some extent, search and browse are inseparable with each other. They are the "two major ways to locate information (S19)." More important, they can satisfy different users' needs as stated by subject 17, "Both must exist, to satisfy both users with search skills and specific needs as well as those with lesser skills or a less defined need. Having both is a small way to ease the digital divide between those with search skills and those without. Additionally, the browse function allows the user to see what the collection holds…" Navigation is related to the overall interface design. It is regarded as very important ($x = 4.4$) to users of digital libraries. To users, easy navigation determines whether they would return back to the digital library, according to subject 17, "Navigation is closely related to interface design; however, navigation holds a greater importance because an interface can be either sophisticated or simple and yet still allow for easy navigation. …it will likely encourage the user to return to the site." For users, first, they need to navigate within the digital library to go anywhere they want suggested by subject 12, "…every page should contain links to the site's homepage and, when applicable, links to intermediate upper-level pages should also be provided. A user should not have to backtrack through multiple pages to reach major foundational pages." Second, they also need to know where they are, "users need to know where they are within a collection/digital library and how to get back to the start or homepage," pointed out by subject 15. Third, most importantly, they do not want to get lost. Subject 7 suggested, "[Digital library interfaces] should be designed to let the user go to where he wants to go, in a convenient, quick, logical, and consistent manner. It is important to try to prevent the user from getting 'lost' (confused), or from have to engage in cumbersome and excessive clicking and Web navigation."

Accessibility ($x = 4.2$) was considered very important to users. Many of the users associated issues of accessibility with people with special needs, such as people with disability, information poor people, etc. "A digital library must consider the difficulty of presenting visual materials to users with visual impairments, or audio material to those with hearing impairments…extreme importance because digital libraries need to strive to make information freely available to all who seek it," subject 2 stressed. Many of the users echoed the idea. Others also discussed the issue of digital divide with people who cannot use digital libraries because of computer literacy or unable to afford computers and network access. View and output ($x = 3.8$) was rated as some important. Subject 3 well summarized users' opinion, "The ability to view information in a variety of outputs is required to complete the information transaction."

One surprising finding is that help features ($x = 3.6$) were rated the lowest within the usability category. One possible reason is that users were affected by their past experiences in using help features. In their searching for information in these two libraries, very few actually used help features. Interestingly, this is also the category with highest standard deviation which means users do have different opinions. While 31.6% considered help features extremely important, 42.1% regarded them as some important, and about 15.8% thought them a little important. Users have high expectation for help features. For example, subject 7 argued, "Help Features should allow the user to solve various problems, or answer general questions about the utilization of the library and access of the content." "These features are access points. In a virtual setting one does not have the luxury of asking the librarian for instant help in finding a particular item," subject 11 offered her expla-

nation. Moreover, the help features should be obvious to user at any location of a digital library. "Also, a user should most definitely not have to search for help. It should always be obviously and prominently place on each and every page/location of a digital library so a user does not have to get lost backtracking to the last place he or she remembers seeing an option for getting help with something. The help feature should be as in-depth as possible (S2)."

For digital library, usability and collection quality are two sides of the coin. Users normally judge the value of digital libraries by evaluating its content. Subject 4 argued, "If the retrieval and usability through interface design is fantastic but the content is poor, all of the time spent creating the digital library has been wasted. Effective collection management is an integral factor in judging the worth of any library, traditional or digital." Specifically, the quality of content refers to completeness, authority, accuracy, reliable and relevant information, etc., as suggested by subject 18 and 19, "If the collection does not have a high level of completeness in a particular area or if there is any question about the authority or accuracy of any item in the collection, the quality of the entire collection is diminished (S18)." Echoed by subject 19, "Whether the user can find relevant and reliable information is significantly conditioned by the content, quality, scope, authority, accuracy, and completeness of the collection." The ultimate objective of digital libraries is to satisfy user need. Only the targeted users of a digital library can determine the usefulness of a collection. "If the content of the collection is of no use to the user, the user will have little or no reason to even browse the collection," well said subject 13.

With the quality of collection, accuracy ($x = 4.6$) and authority ($x = 4.3$) were rated the highest. Digital libraries exist on the Internet, which share the same accuracy and authority problems that information on Internet might have. These two criteria are important because "they are indicators to a user of the validity of the information (S11)." To be specifically, according to subject 12, "We need to know that the information comes from a reputable and trustworthy source and that it is either true (concerning verifiable facts) or that it represents an accurate representation of someone's opinion." Interestingly, users had different opinions regarding the relationships between accuracy and authoritative and which one is more important. Subject 17 claimed that, "Ideally a collection needs to have correct information so that the users can trust it, and accuracy is usually a byproduct of authority. This will attract users, which ultimately could attract funding." However, subject 10 disagreed, "Accuracy, I would say, is the most important aspect of content, even above authority. One can have all the authority in the world, but if the information on the site is not accurate, then what difference does it make?"

Scope, completeness and currency share the same ratings ($x = 3.8$) in this study. Scope needs to be clearly defined because potential users are defined by the scope of a digital library. Subject 11 justified the importance of offering clear scope for a digital library, "[Scope] must be clearly defined to enable the user to decide if it is a collection that meets their needs. The unusual feature of a digital library is the potential patron is defined by the scope of the collection and therefore driven by the collection creator." Simultaneously, it is also essential to balance the scope of a collection. "If the scope of the collection is too wide, it may have less depth. Also, if the scope is narrow, the parameters are more easily defined and the searcher may be able to tell at a glance if the collection can fulfill the information need," explained subject 17. Completeness of a collection is essential because "A partial collection is of no use to a searcher who has indiscriminate information needs (S3)." Updating materials in a digital library keeps it alive. Just as subject 7 stated, "it is necessary to periodically update the digital library contents, in order to retain the currency of the information that is provided." Currency also includes weeding out old materials and materials are no longer available. "Updates are important because dated information is not very useful of if certain pages no longer exist, it makes the digital library incomplete and unprofessional looking," described subject 15.

Surprisingly, subjects of this study consider digital libraries more as IR systems less as libraries. That might be the reason that service related criteria were rated much lower than system performance related criteria. Usefulness of services ($x = 4.5$) is the only one rated high because "usefulness is quite important simply because if the service does not actually accomplish any use, it does not have a reason to exist other than preservation purposes (S17)." Subject 11 further emphasized that has to be "defined by the user community." In addition to usefulness of services, unique services ($x = 3.6$) were also considered some important. First, it gives identity for a digital library as subject 5 explained, "A digital library needs some uniqueness to separate it form others and to give it its own identity." Second, digital libraries are different from traditional libraries. "Although the digital library should mimic the services of a traditional library, I don't think it can fully accept this role. It is

more likely to have unique services, for a community of people with specific information needs," subject 11 stated. Finally, mission statement of a digital library plays the role of informing users the relevance of the library. Subject 8 well put it, "[Mission] will inform the user if this particular digital library will meet his need."

Compared with services, system performance was regarded much higher. Of course, system performance in general was rated quite higher because the objectives of using digital libraries are to find useful and relevant information. Subject 4 stressed the importance of the retrieval aspect of digital libraries, "At its core, however, a digital library's main point of evaluation must be its system performance. A digital library is worthless if the user cannot retrieve the information it contains. The finest and most complete collection in the world is pointless if there is no suitable retrieval system in place." Moreover, "a digital library is crucial for users to find desired information," subject 19 added. Efficiency and effectiveness ($x = 4.5$) were important to users. It was defined as "this can measure how quickly a user will retrieve the desired information (S17)." Saving time and effort for uses are essential in digital library access. According to subject 5, "The library that performs efficiently and effectively saves the users' time and headaches." Relevance ($x = 4.3$) is highly related to efficiency and effectiveness. "Efficient and effective is almost synonymous with relevance and specifically precision. . . .most users will want to find their answer or answers quickly (and, if possible, only the answer for which they are looking)," subject 10 clarified the relationships between the two. She further pointed out that "not only is relevance an important issue for search results, but the user should have some fair chance of determining a collection's relevance overall before spending time searching sub-categories or particular items." Subject 3 further stressed the benefits of relevant results, "Highly relevant search results spare the searcher from unnecessary sorting." Precision and recall are the measurements for relevance. The majority of subjects preferred precision as justified by subject 10, "Speaking from personal experience, I would rather only get six results with three being relevant or possibly relevant, than the initially hopeful number of, say, thirty, only to find that one or two are semi-relevant. My guess is that this preference is very common." However, some subjects preferred recall. For example, "Sometimes I can learn more about the related and unrelated subjects from relevant and irrelevant results. For this reason I believe that recall is more important then precision. Just because the search engine knows about my query does not mean that I know about it. There have been many times that I found the right things with the wrong terms because of a large recall and not much precision," subject 5 offered his opinion. Subject 12 well summarized it, "[precision and recall] vary in importance depending upon the situation."

With in the category of user satisfaction, user satisfaction ($x = 4.3$) was no doubt highly rated. Subject 2 pointed out, "It is vital for a user to be able to express his or her satisfaction with the digital library features and to be able to make suggestions or complaint." The key issue of user satisfaction is whether the targeted audiences are satisfied. "Creators of a digital library would identify the primary user groups for their collection and establish usability and service features that targeted those groups. The next logical step would then be to assess the success of that goal," subject 3 well illustrated it. Simultaneously, digital libraries need to satisfy users with different levels of knowledge and skills. As subject 10 emphasized, "This is really what it all comes down to. They are not meant to serve only to those who are already well-acquainted with its navigation and utility." User feedback is crucial for the development of digital libraries to satisfy user needs. Subject 4 well put it, "User feedback and how well it is used is important for the evaluation of digital libraries." A communication channel is needed between developers of digital libraries and their targeted users. Subject 15 explained, "These criteria help in terms of knowing how well you are doing as a provider." In order to send feedback to digital library developers, contact information should be easily access to. "In the expression of satisfaction, dissatisfaction, or simple inquiry, a user should always have access to contact information," stressed subject 3. Another subject (S7) added "[It is] vital to have contact information, which will allow the user to directly query appropriate library staff, with respect to a particular issue or concern." Moreover, the availability of contact information "encourages users to give feedback to the creators to assist in redesign and solving potential problems with the site (S18)."

Subjects' rating of the DL evaluation criteria validate these criteria that developed by another group of participants in the previous research (Xie, 2006). In general, the majority of the criteria in usability, collection, system performance, and user feedback were deemed as important or extremely important except in service category. Two possible reasons might be: (1) their use of these two digital libraries did not involve in service part of the digital libraries; and (2) their perceptions of digital libraries are more related to IR systems than

libraries. Of course, not only their use of digital libraries affect their ratings, but also their past experience in using of other types of IR systems influence their perceptions of the DL evaluation criteria, such as the low ratings of help features. These are further discussed in the Discussion section. In addition, subjects of this study also added more DL evaluation criteria, such as consistency for usability, uniqueness and cohesive for collection quality, real time reference for service quality, usefulness for system performance, etc.

5.3. Positive and negative aspect of digital libraries

After conducted six searches for each of the selected digital libraries, subjects were instructed to rate the importance of the DL evaluation criteria developed by the previous user group. Then they were asked to apply these evaluation criteria to evaluate the two selected digital libraries. Tables 6 and 7 show the mean and standard deviation of subjects' evaluation of American Memory and University of Wisconsin Digital Collections.

5.3.1. Usability
Comparatively speaking, subjects rated overall interface usability of AM ($x = 3.9$) higher than UWDCs ($x = 3.2$). To subjects, usability of a digital library was evaluated based on its overall interface design, in particular its search and browse function, navigation and view and output formats. Subject 11 well praised the usability of AM, "The American Memory Digital Library, like the Library of Congress is the model of how it should work. The usability of the interface was good. It utilized both search and browse functions. The navigation was painless and the view and output were good." "It has high usability with neatly designed interface, powerful search and browse functions, ubiquitous navigation points and various view and output options," echoed by subject 19. On the contrary, UWDCs was rated lower because the unavailability of the above functions. Just as subject 8 commented, "I gave the lowest scores to the usability features on the UWDCs site. I could easily have given all those categories a "1". Nothing suggested accessibility, including help, searching and browsing, and navigation." Several issues of usability of these two digital libraries emerged from the data.

Table 6
Evaluation of American Memory ($N = 19$) (1 = not at all good 2 = a little 3 = some 4 = very 5 = extremely good)

Types of criteria	Variables	Mean	Standard deviation
Interface usability	Interface usability in general	3.9	0.58
	Search and browse functions	4	0.94
	Navigation	4	0.70
	Help features	4	0.91
	View and output options	4	0.59
	Accessibility	3.8	0.69
Collection quality	Collection quality in general	4.6	0.50
	Scope	4.4	0.60
	Authority	4.5	0.70
	Accuracy	4.4	0.69
	Completeness	3.9	0.78
	Currency	4	0.69
	Copyright	4.5	0.71
Service quality	Mission	4.6	0.69
	User community	4.3	0.87
	Traditional library service	3.7	0.99
	Unique services	4.1	0.74
System performance	System performance in general	4	0.71
	Efficiency and effectiveness	3.9	0.78
	Relevance	3.8	0.71
	Precision and recall	3.7	0.65
	Usefulness	4.1	0.64
User satisfaction	User satisfaction	3.9	0.71
	User feedback	4	0.82
	Contact information	3.9	0.85

Table 7
Evaluation of University of Wisconsin Digital Collections ($N = 19$) (1 = not at all good 2 = a little 3 = some 4 = very 5 = extremely good)

Types of criteria	Types of sub-criteria	Mean	Standard deviation
Interface usability	Interface usability in general	3.2	1.01
	Search and browse functions	2.7	1.13
	Navigation	2.9	0.81
	Help features	2.1	1.12
	View and output options	3.4	0.77
	Accessibility	2.9	1.11
Collection quality	Collection quality in general	3.8	0.71
	Scope	3.2	1.08
	Authority	3.9	1.05
	Accuracy	3.9	0.78
	Completeness	3.4	0.96
	Currency	3.6	0.77
Service quality	Mission	3.2	0.79
	Targeted user community	3.4	0.96
	Traditional library service	2.9	1.08
	Unique services	3.1	0.87
System performance	System performance in general	2.9	0.85
	Efficiency and effectiveness	2.9	0.78
	Relevance	3.3	0.94
	Precision and recall	3.1	0.83
	Usefulness	3.3	0.84
User satisfaction	User satisfaction	2.9	1.16
	User feedback	3	0.98
	Contact information	3.2	1.01

5.3.1.1. Simplistic versus attractive interface. Usability of a digital library is represented by the interface design of that digital library. To some extent, the interface design determines the usability of the digital library. Simplicity of the digital library is also a consideration of usability. According to subject 18, "For the general usability of the site, I gave AM a five because I think the interface design is simple." Subject 12 agreed, "The interface [AM] is clean, uncluttered and easy to navigate." However, not everyone likes the simplicity of interface design, subject 16 argued, "When you go to the introduction page there is not a whole lot to look at. When you go to other digital library sites you are often lured in by many different illustrations or other kinds of pictures that grab the user's attention. However, in this website you start on a very boring and non entertaining webpage. I believe that in order to draw people into a collection and get them excited about it, they have to immediately see something that grabs there interest and makes them want to travel further..."

5.3.1.2. Default versus customized interface. Another issue related to simplicity and attractive interface is default versus customized interface for each collection. Each digital library contains multiple collections. It is important for the designers to make a decision about whether each digital collection is designed in a standard format. AM's ($x = 3.9$) standard interface design was rated higher than the UWDCs' ($x = 3.2$) unique interface design of each collection. Some subjects further required the homepage of the digital library to be consistent with each collection. "A weakness in the usability category is that the design of the collection pages looks a bit different from the overall site design. This confuses navigation because it is somewhat difficult to tell whether you are still in the American Memory site and find the link to get back to the American Memory home," emphasized subject 17. Similarly, one subject offered her reason regarding learning different interfaces at UWDCs site, "The design is not uniform though. Some collections have again a different design such as the Great Lakes Maritime History Project or the Mills Music Library Special Collections. The problem with this is that a user who browses different collections has to get accustomed to different interfaces and basically has to learn over again how a collection is accessed, searched, and browsed (S1)." At the same time, some subjects raised the issue of attractive issue. "The only place that I would down grade the site [AM] on design and usability is for the specific collections. Many of the collections have similar standard designs that are not as attractive or well planned as the American Memory homepage (S18)."

5.3.1.3. Multiple access points versus their efficiency. Search and browse are the two most frequently applied information seeking strategies. Search and browse functions are highly correlated to the overall usability. Subjects liked the following ways they can access information via searching and browsing: (1) find information without accessing specific collections. Subject 1 stated, "What I liked the most about the search option on the homepage was that collections could be searched by keyword without having to access them." (2) organize collections with alphabetic list with brief description. UWDCs was criticized for lacking overall search option, but it was praised for its alphabetic list of collections. Subject 16 offered his opinion, "The first thing that I like about the University of Wisconsin Digital Collections is that it seems to have a very large and detailed browsing option. Unlike the American Memory that only had about eighteen different browsing options, the digital library [UWDCs] has an alphabetical list that consists of 43 different digital libraries. This is the kind of list that I was expecting to find in the Library of Congress..." (3) offer multiple browse options. It is not enough to just browse the collection by title. Users desire to access information from a variety of access points. AM provides a good model for that purpose. "For example, the variety of formats in which a user might arrange the collections, including topical, geographic, and chronological divisions, as well as an opportunity to see all collections in alphabetical order, only enhances the user's ability to browse effectively," as subject 6 explained. (4) integrate search and browse features together. Subject 7 well illustrated how search and browse features together assist users to effectively find relevant information, "It should be noted that the American Memory had search and browse capabilities that worked in conjunction with each other. Thus, a search that was performed might bring the user into a general area of retrieved results which might be relevant. From there, the browse capabilities would allow the user to more directly inspect the specific items to see if they are of ultimate interest or utility. This created a synergy of possibility and allowed for the utilization of "Pearl Growing" techniques." When subject 7 emphasized the effectiveness of search-browse, subject 17 discussed usefulness of search-within-browse, "A handy feature is the search-within-browse because if the user is unfamiliar with the smaller collections within a category, this feature lists relevant collections and allows the user to gain familiarity with them."

While AM's search and browse function was rated 4, UWDCs' search and browse function was rated only 2.7 mainly because it did not have a search feature across collections. "In the usability category, the inability to search and browse across collections is the biggest flaw," stressed subject 17. Subject 18 further pointed out, "The browse feature is fine; the collections are listed alphabetically with brief description. However, if you are not sure what category might include your topic of interest you have to make a few educated guesses and then search within the specific collection. I think this is a major drawback in the accessibility of these collections." At the same time, just offering search and browse options is not enough. The key issue is how to make the search and browse functions more effective. Subjects needed the search function, but they also concerned the overwhelming results generated by their searches. Just as subject 13 discussed his own experience in searching for his own question, "Unfortunately, I found browsing the collection to be much more effective than searching the collection. For example, I searched using the term "e.e. cummings." I received 1000 results, but none could be clearly identified in the listing of results as having anything to do with the poet." Irrelevant results is another concern, subject 17 identified the problem, "While it is extremely important that the search function searches across collections, the amount of material from which the search is drawing is so large that the results that come up can be irrelevant and contain false drops."

5.3.1.4. General help versus specific help. Learning is another type of information seeking strategies applied in the information retrieval process. Users have to learn if they encounter problems in their information-seeking process. AM's help features were rated pretty high ($x = 4$), but UWDCs' unavailability of explicit help leads to its low rating ($x = 2.1$). Subjects were pleased with the comprehensiveness of help mechanism in AM. Subject 1 well summarized these help features, "Help features are quite extensive. Help is divided into the categories "How to View", "Search Help", "FAQ" and "Contact Us". Answers to basically all possible technical problems are given, search strategies are detailed and if one has questions, it was possible to write and email, call or chat with a librarian." Subject 7 added, "Other noteworthy features of the American Memory are the help features that aim to give a Technical Overview of the website and provide information and tips about the formats that are used, and how to obtain the applications that are necessary to fully access all elements of the collections, especially the content that is in multi-media formats, such as video."

The problem with help features in AM and UWDCs is different. In UWDCs, there was no general explicit help available, but help does exist in specific collections. According to subject 2, "UWDCs does not have a generalized help feature; it appears that the user must be in a specific collection and then use that collection's help feature, of which the quality varies." In AM, there are help features available in the main page, but users cannot access them from each specific collection. Subject 9 complained, "Help features are sufficient but they are often not on the page you would need them to be." Moreover, both digital libraries do not have standard help in specific collections, and they are not identifiable easily. "Help files usually did exist, which were specifically devoted to each collection. However, often, they were not evident, nor necessarily easily accessible," subject 7 commented.

5.3.1.5. View options and relevancy/usefulness evaluation. Interestingly, view and output options were rated the highest among other criteria within usability in AM ($x = 4$) as well as in UWDCs ($x = 3.8$). Subjects were satisfied with the following aspects of viewing and output: (1) help on how to view. Subjects need to access information about how to view materials with different formats. Subject 2 praised AM for its help on viewing, "American Memory has a prominently displayed button to the help page, which is divided into four parts, concerning viewing/accessing parts of collections, search help, frequently asked questions, and contact information (which is also available as its own button near the top of every page, next to the help button)." (2) high resolution. Subjects in general desired for high resolution items. Subject 1 discussed her example in UWDCs, "This collection is also a good example for discussing view and output. The original pages have been scanned and give an accurate impression of these survey records. Thanks to the high quality/resolution of the scanned images they can also be printed out. However, there is no fancy feature like zooming available." Some subjects liked the zoom feature. Other preferred multiple forms to view items. (3) multiple view options. Subject 19 applauded the different ways to view the items in AM, "Besides the search and browse, I would like to say the view options are very good. For example, all the items have both list view and gallery view; photographs have both thumbnail view and high resolution view. It successfully avoids zoom which often time frustrates people."

One major problem of view feature occurred in both digital libraries is small fonts of some of the items. "...unconscionably small type once you browse the collections...Regarding correction of the type style and/or font issue, to create greater readability...I can't believe they haven't had complaints," subject 8 criticized. One related issue of view is how to effective evaluate the relevance/usefulness of the results. Overwhelming results discussed above require digital libraries to support users to effectively evaluate the results. However, none of the two digital libraries actually assist users in doing so.

5.3.2. Collection quality

Collections of digital libraries reflect the content of the digital libraries, and they are the core of the digital libraries. Collections quality in both digital libraries (AM, $x = 4.6$; UWDCs, $x = 3.8$) were rated higher than usability of the two digital libraries. Subjects were impressed with the collections of AM. Subject 4 spoke highly of the AM collections, "The collection is superb. Once again, for someone with an interest in the history of our nation, this is a treasure of knowledge to be tapped. In general, the content is phenomenal and is the strongest feature of the AM." One weakness of AM comparing with UWDCs is its description of each collection. Subject 17 pointed out, "The descriptions of individual items were often quite brief and did not give much background information about them."

5.3.2.1. Individual collection versus cohesiveness of all the collections. The quality of collection was rated high in both digital libraries (AM, $x = 4.6$; UWDCs, $x = 3.8$). The quality of collections needs to be evaluated from different aspects. Subject 6 well put it, "The Library of Congress American Memory Project, with its more than 100 collections and nine million items, is an exemplary digital library, with a clear and consistent focus, unique collections and services, and careful organization." Another reason for the high quality is the collections of digital libraries are own by the physical libraries. Subject 1 commented on UWDL, "The quality of the items of libraries that actually possess their own digitized items is in general high." However, one issue related to digital libraries is the cohesiveness of all the collections. Subject 17 criticized UWDL, "The collections are on a range of random topics, with little cohesiveness between collections." Subject 13 further commented,

"While each of the individual collections has a worth of its own, and many of the collections are very good indeed, the entirety, the overarching University of Wisconsin Digital Collections is horribly crippled by the varying quality and usefulness of the individual collections and by a lack of a unifying search functions and help function." That raises the issue of how to create digital libraries for diverse users with diverse topics.

5.3.2.2. Scope versus completeness. Scope of a digital library in general is presented in "About", which helps users make a decision whether what they look for are actually included in the digital library. Comparatively speaking, scope (AM $x = 4.4$; UWDCs $x = 3.2$) and completeness (AM $x = 3.9$; UWDCs $x = 3.4$) were rated lower than other criteria within the collection category in both digital libraries. Subjects were questioned the specifications of scope presented in the digital libraries. For example, subject 2 pointed out, "The scope of UWDCs, although explained in 'about', is not very intuitive. The collections are designed to support curriculum or research interests, and therefore there does not seem to be any sort of overarching theme other than relevance to the statewide UW community." Simultaneously, they were amazed by the scope of AMDL collections in its huge size. Subject 1 explained, "I browsed the *Abraham Lincoln Papers* and was amazed by the scope of this collection, the view and output. All his correspondence had been scanned which amounted to 122,000 pictures and the digital images had been made from microfilm which was an additional challenge." The scope of AMDL is further validated by a variety of topics, and their organization. Here are the related justifications from two subjects. "The categories cover a wide scope of American history from presidents and their portraits to women's history and suffrage," subject 4 illustrated the topics. "The classification system defines the scope so the digital library is broken down into more manageable chunks, without sacrificing usability," subject 17 added.

The scope of a digital library does not guarantee the completeness of the collections; instead, it is normally used to judge the completeness of the collections of the digital library. "American Memory, on the other hand, is enormous group of collections, but this is explained even in the title itself. I think almost any type of material can belong to the 'American Experience' and while there are of course limits, it seems like their collections house just about anything feasible," subject 2 described it. Even though AM has huge items in collections, it does not indicate completeness. However, subjects evaluated the completeness of collections more based on the name recognition. Here is a typical comment from subject 13, "It is by no means complete, or even impressively comprehensive, but because it is a collection of the Library of Congress, there is every reason to assume the collection will grow in scope and completeness."

5.3.2.3. Authority versus accuracy. Authority and accuracy are the two criteria highly correlated to each other. The ratings for authority and accuracy are similar in AM ($x = 4.5$ and $x = 4.4$) as well as UWDCs ($x = 3.9$ and $x = 3.9$). Authority was judged by the following aspects: (1) the organizations/sponsors' reputation. Both Library of Congress and University of Wisconsin System are highly reputable. Just as subject 1 stated, "Taking into account that the *Library of Congress* is considered the national library of the United States and the biggest library in the world the national digital library program is well equipped to meet its collection scope. This is why I have given this library a rating of "5" in terms of authority." Subject 7 talked about the UWDCs, "From what I saw, the accuracy and authority of the content was quite good as most of it was generated as the result of various projects involving academic scholars." (2) copyright information. The offering of copyright information is another way of demonstrating authority. Subject 15 confirmed it, "The authority of the page is comprehensive and can be viewed by clicking on the Legal icon on the bottom of the page where copyright is listed as well as privacy issues and a mission statement." (3) clear presentation of the authority behind the collections. Authority is also confirmed if each individual collection is clearly presented with information regarding who/which organization is behind the collection. Subject 12 explained, "The authority behind the presentation of the collections is clear as is the authority behind individual collections. This in turn provides considerable assurance as to the quality and accuracy of the content of the collections."

To users, accuracy is determined by the authority. Just as Subject 4 put it, "I can only assume that the accuracy is exceptional, judging by the authority of the Library of Congress." As stated above, for users, the reputable organization and scholars who involve in the development of digital libraries guarantee the accuracy of the collections. In addition, checking for error is effective approach for assessing accuracy. Subject 7 justified his rating with example, "It turns out that the translations had been removed, because errors had been found.

This reflects well on the American Memory project's commitment to accuracy." Finally, referring to other collection is also considered an effective approach for accuracy. Subject 1 stated her reason, "The references to other collections also indicate to me that accuracy and completeness are very important to most collections of this library."

5.3.3. Service quality

Service is an essential component for a traditional library. In digital library environments, it is important to examine what users need and desire for services. Two related issues emerged: (1) mission of digital libraries and their user communities, and (2) traditional versus new and unique services.

5.3.3.1. Mission and targeted user community. Each digital library needs to clearly define its mission and targeted user communities. The results show that AM offers clear mission statement ($x = 4.6$) and its targeted user communities ($x = 4.3$). Subject 11 described the relationship between the two, "The Mission statement is accessed by the About button on the homepage. It defines the user community." Furthermore, the mission statement is not just for one or several collections. "The mission of this project is clearly defined and cohesively links together the different collections," subject 17 added. That is the problem for UWDCs' mission statement ($x = 3.2$). Subject 17 also commented UWDCs, "There were missions for the individual collections, but none for the whole digital library." Subject 10 complained, "The mission statement is vague and general. The statement mentions (is it a mission statement per se?)"scholars at large," but does not really say what they hope to add, exactly to the invisible college. They have a K-12 collection, but this statement does not address an attempt to reach out to educators and children." It is a challenge for digital libraries that have a variety of collections representing different themes to define their mission statements. Comparing with mission statement problem, the developers of UWDCs did a good job in defining user communities. According to subject 18, "This collection did score higher in user community because there was a pint made about UW faculty, staff and students being important consumers of the collections along with "citizens of the state, and scholars at large."

5.3.3.2. Traditional services versus unique services. Users care unique services of digital libraries as well as traditional libraries. Subjects of this study rated the two digital libraries higher in unique services (AM $x = 4.1$, UWDCs $x = 3.1$) than traditional services (AM $x = 3.7$, UWDCs $x = 2.9$). AM did a better job in offering traditional library service than UWDCs mainly because users are able to interact with librarians as they normally do in traditional libraries. Subject 15 discussed her feeling about AM services, "This digital library has a strong traditional library service feel because of its interactive contact information such as Ask a Librarian." "I also gave the digital library high ratings for their service and feedback features. For traditional services, American Memory offers email and live chat with a librarian," subject 18 agreed. Simultaneously, UWDCs was criticized for its lacking of traditional library services. Subject 1 listed the problems of UWDCs in offering traditional service, "However, I think that customer satisfaction and traditional library services is less important for this library. There were in general never telephone numbers available. This is also true for contact and help features. . . .the possibility to contact someone if one needed more help was only given when one clicked on 'Technical Assistance'." Subject 7 concluded, "However, traditional library services were practically nonexistent. One could not expect reference support." Subject 9 even checked individual collection, "None of the collections I came across purported to offer any type of standard library services."

Two specific unique features were praised for the two digital libraries: involving users and providing services to specific user groups. UWDCs' "Submit an Idea" provides an opportunity for users to be involved in the digital collection development. Subject 15 discussed her pick on the unique service, "I think a unique service that the UWDCs gives is the *Submit an Idea* page. I think it is great to get patrons involved in ideas of the collections because they may have unique historical collections that they could contribute." AMDL's learning page offers services to teachers. "One unique service this collection offers is a page especially for teachers to help them use American Memory as a classroom teaching tool."

5.3.4. System performance

The objective of digital libraries is to facilitate users finding the information that they need. Therefore, digital libraries, are also IR systems that should allow users to search for information. Again AM was outper-

formed by UWDCs in efficiency and effectiveness ($x = 3.9$ versus $x = 2.9$) as well as relevance ($x = 3.8$ versus $x = 3.3$). Again it was caused by the unavailability of UWDCs' overall search function.

5.3.4.1. Searching across collections versus searching a specific collection. No doubt, searching is needed in digital libraries for efficiency and effectiveness. The question is: which one is more effective: searching across collections or searching a specific collection? Subject 1 justified her low rating of UWDCs, "Because of the non-existing search options I rated the effectiveness and precision of the system performance "3". Subject 17 argued, "Retrieval performance could be better with a specific collection. Since each collection must be searched separately, it is difficult even to evaluate." However, it is inefficient unless the searcher has a clear idea of what to look for and which collection to use." It seems that AM serves as a role model, "Besides search all collections, it also has search selected collections feature which can make the search more efficient and effective, subject 19 recommended. Of course, the availability of searching functions are not enough, and efficiency also requires good help features as suggested by subject 1, "As already mentioned, one can search very precisely and into the individual collections, help features and search explanations are available in abundance. This shows that effectiveness and precision are very important for this library."

5.3.4.2. Browsing and precision. Users prefer precision or recall depending on their tasks. In this study, their search tasks mainly require precision. However, the digital libraries still have problems in their search mechanisms. On the one hand, UWDCs does not have an overall search function. On the other, AM's search mechanism does not yield high precision. "The retrieval system is good for broad, general searches, but did not have as high a rate of precision for specific searches," subject 18 offered her opinion. Searching is not always leading to high precision. According to subject 12, "As it turns out, at least in this case, strictly browsing increases precision, while searching all collections in another case increases recall." At the same time, search functions in specific collection of UWDCs cannot always generate relevant results; however the browsing functions are well designed. Subject 4 discussed the outcomes of the two types of functions, "The relevance was questionable at times with keyword searches, though the browsing was effective as long as you knew exactly what you were looking for and had previous knowledge of the subject."

5.3.5. User satisfaction

User satisfaction is essential for the success of a digital library. Multiple channels are needed for developers of digital libraries to solicit feedback from their targeted users.

5.3.5.1. Feedback mechanism. Users would like to have multiple avenues to send their feedback to the designers/developers of digital libraries. That is why AM ($x = 4$) was rated much higher than UWDCs ($x = 3$) in that aspect. Subject 1 compared the two digital libraries, "There is also an option for user feedback. Under the category 'Contact' is a section called 'Comment on the American Memory Site' and 'Report an Error'. However [UWDCs], besides the few options of writing emails there was no feedback possibility for users." Subject 3 agreed, "Another extremely good mark [for AM] was gained in the User Feedback category. Feedback was solicited in four distinct formats: General Comments, Ask a Librarian, Chat with a Librarian, and Technical Problems." Subject 17 further illustrated the problem of UWDCs, "The overall UWDCs only allows one conduit for feedback, which is an email form that requires a name and email address, while asking non-mandatory questions about university affiliation."

Contact information is highly correlated to user feedback, and it is one of the main channels for feedback. However, it is not enough just to offer contact information. It also needs to be design in a standard and consistent way, and it also needs to be easy to access. Subject 9 complained, "...every page had contact information. The location of the contact information differed across the board; sometimes it was on the about page, while other times it was located on the help page. It was often hidden, though which was a deterrent." Moreover, "There is no encouragement to send comments or suggestions," subject 18 added.

5.3.5.2. User satisfaction and its assessment. The development of digital libraries is to satisfy user need. User satisfaction is determined by all the criteria discussed above: interface usability, collection quality, service variety, system performance, and user feedback. That is why subjects were more strictly in assigning their ratings

for the user satisfaction of AM ($x = 3.9$) and UWDCs ($x = 2.9$). Accordingly, the ratings of user satisfaction is lower than most of the other criteria for both of the digital libraries. At the same time, it also reflects their overall evaluation of digital libraries based on their evaluation of each aspect of the digital libraries. While it is easy to rate AM user satisfaction, it is difficult to evaluate UWDCs user satisfaction because "the user satisfaction is directly related to the ease of searching and the completeness of the collections. As these functions vary from collection to collection, it is hard to generalize (S11)."

6. Discussion

Borgman et al. (2000) suggest that technical complexity, variety of content, uses and users, and lack of evaluation methods contribute to the problem of DL evaluation. Citing Marchionini's (2000) metaphor considering evaluation of digital libraries as evaluation of a marriage, Saracevic (2004) well illustrates the complexity of DL evaluation. However, the only people who can best evaluate a marriage are the people who are involved in the marriage. That also applies to the digital library evaluation. Xie (2006) pointed out that users who use digital libraries, researchers who work on theoretical framework of DL evaluation and its criteria, and practitioners who conduct actual evaluation studies emphasize on different aspects digital library evaluation and its criteria. Moreover, this study shows that digital library evaluation, even just from user side, is not simple. In this section, the questions regarding user evaluation raised in Introduction and Literature Review are further discussed. No doubt, digital library use affects its evaluation. Moreover, there are discrepancies between the perceived importance of DL evaluation criteria and the actual evaluations of digital libraries. Finally, users are not same, nor are their evaluations of digital libraries.

6.1. DL use versus DL evaluation

Evaluation itself is a form of sense-making of the evaluators, and it is also situated (Van House et al., 1996). Digital library use and digital library evaluation is interrelated to each other. Users' evaluation of digital libraries is largely based on their use of these digital libraries. To be specific, digital library use affects digital library evaluation in two folds. First, the problems users encountered in their use of digital libraries lead to the negative evaluation of digital libraries. Second, the availability of new features or design sets up a higher standard for digital library evaluation.

User-oriented system design requires the design of digital libraries to adapt to the users. However, in reality, users have to adapt to the digital libraries. In that sense, the design of digital libraries affects how users interact with digital libraries. Further, that also influences how they evaluate digital libraries. The results show that when there is no overall search function available, users have to browse to find information. That is why subjects of this study browsed more in UWDCs than in AM because they had to browse in different collections in order to find the right collection before they can search for relevant information. Just as subject 3 explained, "I did not see any search features at the homepage of UWDCs so I browsed the collections titles for possibilities. This was frustrating." This is one of the main reasons that usability of UWDCs was rated lower than AM. For users, it is not enough to just have search and browse functions. They also have to be effective. For browsing, their use of AM, which consists of multiple access points for browsing, establishes a higher standard for assessing digital libraries browsing functions. In users' evaluation of digital libraries, they were no longer satisfied with the availability of the browsing function. Instead, users desired multiple ways of browsing from subject, format, time, or geography to any other access points that users might look for information. For searching, users do not want to deal with overwhelming and unrelated results. High precision is what they what, that is particularly true in searching for information in digital libraries with large collections. Their negative evaluation of AM's performance is derived from their problems in achieving high precision in their use of the digital library.

More important, users do encounter problems in their interactions with digital libraries. If there is no help available, users have to apply trial and error to solve their problems themselves, such as in UWDCs. They need help, but they do not want to ask for help. Not only their current use of digital libraries affect their help use, but also their experience in using help in other types of IR systems affects their perceptions of help features in digital libraries (Xie & Cool, 2006). If the help features are not context-sensitive and intuitive, they do not

use them. The subjects of this study complained that Help features are sufficient but they are often on the page you would need them to be. In addition, general help does not help specific problems. Help needs to be designed to answer specific questions. Without using the digital libraries, they would not have specific suggestions for DL evaluation.

Searching and browsing are the most used search strategies. The design of digital libraries focuses on these two types of features. However, users have to evaluate results, especially when the results are overwhelming and contain many false drops. Therefore, efficiency of system performance and precision and recall were rated lower in both digital libraries. Moreover, the selected digital libraries did not offer much assistance in facilitating users to evaluate results. Subjects of this study had to go over the results one by one to find the right information. That greatly reduces the efficiency of the information retrieval process.

It is not the more the better for collections of digital libraries. For developers of digital libraries, it is also important to consider the cohesiveness between collections. All individual collections have to be interrelated under one theme. Otherwise users are confused, and they are unable to determine whether a digital library contains items that they need. In this study, users had to go through each digital collection of UWDCs to find the collection that might have the needed information. Subjects' experience in accessing UWDCs raised a question. The question is how to build collections for a digital library that serves a variety of uses with diverse need, such as UWDCs. One possible solution is to create multiple digital libraries for each of the theme with multiple collections.

In Web environments, users have to be concerned with authority and accuracy of presented information. There is no standard form for authority and accuracy statements or validation. Subjects of this study evaluated authority via checking the organizations/sponsors' reputation, the presentation of the authority behind the collection, and copyright information. At the same time, accuracy assessment was limited by checking authority, announcements regarding removing documents with errors, and links to other collections. In that sense, errors still cannot be avoided before the launching of digital libraries. An error checking mechanism needs to be created before and after the launching of digital libraries, so accuracy of information can be checked, reported, and corrected. Users should be informed about the checking mechanism to have more confidence on authority and accuracy of information contained in digital libraries. By understanding how users verify authority and accuracy of information in their use of digital libraries, DL evaluation offers suggestions for the design of digital libraries to incorporate authority and accuracy validation mechanisms.

6.2. Perception of DL evaluation criteria versus actual DL evaluation

One interesting finding of this study is that subjects have higher standard in discussing the DL evaluation criteria. They actually lower their standard in evaluating digital libraries. In other words, the discussion of DL evaluation criteria reflects their desired digital libraries, and their evaluation of actual digital libraries reflects their expectation of digital libraries. They did not apply the same standard in evaluating the digital libraries as they discussed in evaluation criteria. In that sense, subjects compromised their evaluation criteria in assessing the actual digital libraries. For example, in the discussion of usability, they emphasized that usability needs to consider users with different skills and knowledge structure. In particular, they suggested that interface design should not be too difficult for novice users and not too simplistic for expert users. In actual evaluation of usability of digital libraries, they more stressed the simplicity of interfaces and attractive interface, but offered no discussion of usability for expert users. In discussing accessibility in perceived criteria, they especially took account of people with special needs, but in actual evaluation, they only thought of regular users.

At the same time, their evaluations of digital libraries are also inspired by the digital libraries that they evaluate. They identified some specific issues in their evaluation that they did not discuss in their perceived DL evaluation criteria. For example, the issue of general help versus specific help emerges in the evaluation of the two digital libraries. General help is offered in AM, but users cannot access to these general help from some of the specific collections. At the same time, no general help is offered in UWDCs, it does contain help in specific collections. Subjects apparently had problems in help mechanism of both digital libraries. Another issue is the cohesiveness of all the collections. Digital collections of UWDCs fit to target users, but the coverage of these collections is so broad that they have little cohesiveness. While subjects focused on collection quality in discussing perceived DL evaluation criteria, they concentrated more on the cohesiveness issue that

they had to deal with in evaluating actual digital libraries. These issues are related to specific digital libraries, and they are easily missed in general DL evaluation criteria discussion.

Furthermore, they also had a more in-depth discussion of the specifics of evaluation criteria than they discussed in the perceived evaluation criteria. For example, in discussing perceived DL evaluation criteria regarding navigation, search and browsing, they focused on the general design principles. However, in their actual evaluation, they covered not only searching and browsing, but more specific and efficient searching and browsing options. Moreover, the integration of searching and browsing options together. In addition, in discussing user feedback criteria, subjects only suggested to open communication channels between developers and users of digital libraries. In evaluating the feedback mechanism offered by two digital libraries, subjects discussed a variety of feedback channels offered in AM, such as "Comment on the American Memory site," "Report an error," "general comments," "Ask a librarian," "Chat with a librarian," and "Technical problems."

6.3. User preference, experience and knowledge structure and DL evaluation

Users are not same. Their evaluation of digital libraries is affected by their preferences, experiences and knowledge structure. Different users have different preferences. For example, while some subjects liked the simplicity of the digital library interface, others preferred more attractive interface with illustrations or pictures. Same applies to the design of interface for individual collection and multiple collections. While the majority of subjects loved the standard interfaces for all the collections for easy access, some of them desired the design to show uniqueness of each individual collection so that users can quickly attract to and understand the collection.

Users' past experience of other IR systems influences their use of digital libraries. For example, users are normally not satisfied with the usefulness of help mechanisms. Their perception of help features and their experience in the past affect their perception of the importance and usefulness of help features in digital libraries. Help features were rated the lowest among other features for usability. Only 2 of the 19 subjects each actually used help feature once even through all subjects encountered problems in their information retrieval process based on their diaries. This also applies to digital library services. They have been enjoying the traditional library services for a long time. They expect digital libraries to offer similar services as physical libraries, such as digital reference service. However, they only identified the unique services provided by the existing digital libraries, such as AM learning page, but did not suggest any other unique services limited by their experience with the development of digital libraries.

User knowledge also has an effect on their evaluation of digital libraries. In discussing their perceived importance of DL evaluation criteria, subjects emphasized the importance to consider both types of users with and without knowledge in designing search and browsing features. Majority of the subjects are novice users of the two digital libraries, and that is probably the reason that in actual evaluation they more focus on ease of use rather than user control of the interfaces. Event though the subjects of this study have basic knowledge of digital libraries, their knowledge is still limited. When they suggest new criteria or features, they are largely based on their own experiences. The previous discussion of their lack of ability to suggest additional unique services of digital libraries is a good example.

7. Conclusion

Digital Library evaluation is not an easy task. The value of digital libraries needs be judged by the users of digital libraries. Subjects of this study not only rated the importance of the DL evaluation criteria developed by another group of users in a previous study but also evaluated the two digital libraries by applying these criteria. The results of this study yielded some interesting findings. The use of digital libraries shows that the design of digital libraries affects how users interact with them. In particular, the availability or unavailability or the actual design of features suggest or guide users how to use a digital library. The ratings of DL evaluation criteria generate the most important criteria for users. Surprisingly, comparing with the previous study, usability and system performance instead of usability and collection quality were regarded as the highest among all the criteria because they trust the developers of the two selected digital libraries for their content.

The findings also identified some problems or dilemmas challenged the development and design of current digital libraries. Some of these dilemmas are caused by individual preference, experience or knowledge.

One of the main contributions of this study is the identifications of relationships between perceived importance of DL evaluation criteria and actual evaluation of digital libraries and the relationships between the use of digital libraries and the evaluation of digital libraries. The findings of this study show that users' use of digital libraries, their perceived DL evaluation criteria, and their preference, experience, and knowledge structure co-determine their evaluation of digital libraries. While users' proposed DL evaluation criteria represent their desired images of digital libraries, their actual evaluation reflects their expectation of digital libraries. At the same time, users' evaluation of digital libraries also enhances their perceived DL evaluation criteria and associated variables. The problems users encounter in their DL use process normally lead to disapproving evaluation of digital libraries, their experience of using new features of one digital library raises the bar of their evaluation for other digital libraries. Of course, users' experience, preference, and knowledge structure also affect their evaluation of digital libraries, in particular the design of interfaces and the offered services.

This study offers the following suggestions for digital library research and development:

In order to incorporate users' perspectives into the development of digital libraries we need to integrate users' perceived importance of DL evaluation criteria, their use of digital libraries, and their evaluation of digital libraries as well as their preferences, experience, and knowledge structures. Since users' DL evaluation is co-determined by users' perceived important DL evaluation criteria, their actual use of digital libraries, and their preferences, experience and knowledge structures, just focusing on one aspect cannot portray a complete picture of user evaluation of digital libraries.

There are similarities and differences in terms of DL evaluation criteria proposed by users, researchers, and professionals. The major categories of evaluation criteria applied in this study such as interface usability, collection quality, service quality, system performance, and user satisfaction have been identified by users and researchers and applied in actual DL evaluation. However, while users focus more on the usefulness of digital libraries from their own perspective instead of from researchers' and professionals' perspectives, they care less about cost, treatment, preservation, social impact of digital libraries, and so on. Moreover, users' DL evaluation criteria are limited by the existing digital libraries and their design. This is why users are more concerned with whether or not some of the features are available and less concerned with feature effectiveness. Also, users emphasize accuracy and authority rather than completeness and currency of the collection. They also suggest some specific variables for evaluation that researchers have ignored, such as unique services offered by digital libraries. In a word, we need to integrate perspectives from users, researchers, and professionals in terms of DL evaluation criteria in order to develop better digital libraries.

The design of digital libraries has to take consideration of users' preference, experience, and knowledge structure. It seems impossible to design a one-size-fits-all digital library to satisfy all types of user needs. The findings of this study identified some dilemmas, such as simplistic versus attractive interfaces, default versus customized interfaces, general help versus specific help, individual collection versus cohesiveness of entire collection, traditional services versus unique services, searching across collections versus searching specific collections, etc. These dilemmas are caused by users' diverse preferences, experiences, and knowledge structures. Further research in digital libraries needs to investigate how to solve these dilemmas.

No doubt, this study has its own limitations. First, the convenience sample and the sample size do not represent a variety of users of digital libraries even though they are the real users of these digital libraries and their uses were recorded in real settings. The results of the analysis of the nineteen graduate students cannot be generalized to other types of users. However, the findings offer insightful information regarding users' evaluation of digital libraries for further research. In addition, this study is the extension of the author's previous study (Xie, 2006), in which 46 students participated. Second, although the assigned tasks include subjects' own questions, the assigned tasks still mainly focus on finding information from the digital libraries. Other types of task, such as service related tasks, also need to be explored. Third, although the diaries enable users to record their actual use of digital libraries, users might not be able to record all the activities they engaged in the user-system interactions. The diaries concentrate mostly on users' searching and browsing behaviors. A transaction log of the interaction process or interview could help to collect more complete data regarding different types of information-seeking behaviors/strategies.

Further research needs to involve real users with real problems/tasks to represent a variety of users of digital libraries. Simultaneously, more data collection methods, such as log analysis and interview need to be employed to collect data. Finally, DL evaluation needs to involve users, but more importantly, it needs to consider all the factors that might affect their evaluation of digital libraries.

Acknowledgement

References

Association of Research Libraries. (1995). Definition and purposes of a digital library. Retrieved July 17, 2002, from: http://www.arl.org/sunsite/definition.html.

Atilgan, D., & Bayram, O. (2006). An evaluation of faculty use of the digital library at Ankara University, Turkey. *Journal of Academic Librarianship, 32*, 86–93.

Bertot, J. C., Snead, J. T., Jaeger, P. T., & McClure, C. R. (2006). Functionality, usability and accessibility: Iterative user-centered assessment strategies for digital libraries. *Performance Measurement and Metrics, 7*(1), 17–28.

Bishop, A. P., Neumann, L. J., Star, S. L., Merkel, C., Ignacio, E., & Sandusky, R. J. (2000). Digital libraries: Situating use in changing information infrastructure. *Journal of the American Society for Information Science, 51*, 394–413.

Blandford, A., & Buchanan, G. (2002). Workshop report: Usability of digital libraries @JCDL'02. *SIGIR Forum, 36*(2), 83–89.

Blandford, A. & Buchanan, G. (2003). Usability of digital libraries: A source of creative tensions with technical developments. *TCDL Bulletin*. Retrieved April 12, 2007 from: http://www.ieee-tcdl.org/Bulletin/v1n1/blandford/blandford.html.

Bollen, J., & Luce, R. (2002). Evaluation of digital library impact and user communities by analysis of usage patterns. *D-Lib Magazine, 8*(5), Retrieved March 15, 2005 from http://www.dlib.org/dlib/june02/bollen/06bollen.html .

Bollen, J., Luce, R., Vemulapalli, S. S., & Xu, W. (2003). Usage analysis for the identification of Research trends in digital libraries. *D-Lib Magazine, 9*(5), Retrieved March 15, 2005 from http://www.dlib.org/dlib/may03/bollen/05bollen.html .

Borgamn, C. L., & Larsen, R. (2003). ECDL 2003 Workshop report: Digital library evaluation - metrics, testbeds and processes. *D-Lib Magazine*, 9(9). Retrieved May 8, 2007 from http://www.dlib.org/dlib/september03/09inbrief.html#BORGMAN%20.

Borgman, C. L. (1999). What are digital libraries? Competing visions. *Information Processing and Management, 35*, 227–243.

Borgman, C. L., Leazer, G. H., Gilliland-Swetland, A. J., & Gazan, R. (2000). Evaluating digital libraries for teaching and learning in undergraduate education: A case study of the Alexandria digital earth prototype (ADEPT). *Library Trends, 49*, 228–250.

Budhu, M., & Coleman, A. (2002). The design and evaluation of interactivities in a digital library. *D-Lib Magazine, 8*(11), Retrieved April 12, 2006 from http://www.dlib.org/dlib/november02/coleman/11coleman.html.

Buttenfield, B. (1999). Usability evaluation of digital libraries. *Science and Technology Libraries, 17*(3/4), 39–50.

Carter, D. S., & Janes, J. (2000). Unobtrusive data analysis of digital reference questions and service at the Internet Public Library: An exploratory study. *Library Trends, 49*, 251–265.

Cherry, J. M., & Duff, W. M. (2002). Studying digital library users over time: A follow-up survey of early Canadiana online. *Information Research, 7*(2), Retrieved March 12, 2005, from http://informationr.net/ir/7-2/paper123.html.

Choudhury, G. S., Hobbs, B., Lorie, M., & Flores, N. E. (2002). A Framework for evaluating digital library services. *D-Lib Magazine, 8*(7/8), Retrieved March 8, 2007 http://www.dlib.org/dlib/july02/choudhury/07choudhury.html.

Chowdhury, G. G., & Chowdhury, S. (2003). *Introduction to digital libraries*. London: Facet Publisher.

Dillon, A. (1999). Evaluating on TIME: A framework for the expert evaluation of digital library interface usability. *International Journal on Digital Libraries, 2*(2/3), 170–177.

Fox, E. A., Akscyn, R. M., Furuta, R. K., & Leggett, J. J. (1995). Digital libraries. *Communications of the ACM, 38*(4), 23–28.

Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., et al. (2007). Evaluation of digital libraries. *International Journal on Digital Libraries, 8*(1), 21–38.

Fuhr, N., Hansen, P., Mabe, M., Micsik, A., & Sølvberg, I. (2001). Digital libraries: A generic classification and evaluation scheme. *Lecture Notes in Computer Science, 2163*, 187–199.

Heath, F., Kyrillidou, M., Webster, D., Choudhury, S., Hobbs, B., Lorie, M., et al. (2003). Emerging tools for evaluating digital library services: Conceptual adaptations of LibQUAL+ and CAPM. Article No. 170, 2003-06-09, *Journal of Digital Information 4*(2). Retrieved November 15, 2005, from http://jodi.tamu.edu/Articles/v04/i02/Heath/.

Hill, L. L., Carver, L., Larsgaard, M., Dolin, R., Smith, T. R., Frew, J., et al. (2000). Alexandria digital library: User evaluation studies and system design. *Journal of the American Society for Information Science, 51*, 246–259.

Jeng, J. (2005a). Usability assessment of academic digital libraries: Effectiveness, efficiency, satisfaction, and learnability. *LIBRI, 55*(2–3), 96–121.

Jeng, J. (2005b). What is usability in the context of the digital library and how can it be measured?. *Information Technology and Libraries 24*(2), 47–56.

Kassim, A. R. C., & Kochtanek, T. R. (2003). Designing, implementing, and evaluating an educational digital library resource. *Online Information Review, 27*(3), 160–168.

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

Lancaster, F. W. (1993). *If you want to evaluate your library* (2nd ed.). Urbana-Champaign, University of Illinois.

Lesk, M. (2005). *Understanding Digital Libraries* (2nd ed.). San Francisco: Morgan Kaufman.

Marchionini, G. (2000). Evaluation digital libraries: A longitudinal and multifaceted view. *Library Trends, 49*, 304–333.

Marchionini, G., Plaisant, C., & Komlodi, A. (1998). Interfaces and tools for the library of congress national digital library program. *Information Processing and Management, 34*(5), 535–555.

Meyyappan, N., Foo, S., & Chowdhury, G. G. (2004). Design and evaluation of a task-based digital library for the academic community. *Journal of Documentation, 60*(4), 449–475.

Monopoli, M., Nicholas, D., Georgiou, P., & Korfiati, M. (2002). A user-oriented evaluation of digital libraries: Case study: The "electronic journals" service of the library and information service of the University of Patras, Greece. *Aslib Proceedings, 54*(2), 103–117.

Nicholson, S. (2005). Digital library archaeology. A conceptual framework for understanding library use through artifact-based evaluation. *Library Quarterly, 75*(4), 496–520.

Nielsen, J. (1993). *Usability Engineering*. San Diego, CA: Academic Press.

Saracevic, T. (2000). Digital library evaluation: Toward an evolution of concepts. *Library Trends, 49*, 350–369.

Saracevic, T. (2004). Evaluation of digital libraries: An Overview. Presented at the DELOS Workshop on the Evaluation of Digital Libraries. Retrieved April 12, 2007 from http://www.scils.rutgers.edu/~tefko/DL_evaluation_Delos.pdf.

Saracevic, T., & Covi, L. (2000). Challenges for digital library evaluation. *Proceedings of the American Society for Information Science, 37*, 341–350.

Saracevic, T., & Kantor, P. (1997). Studying the value of library and information services. II Methodology and taxonomy. *Journal of the American Society for Information Science, 48*, 543–563.

Shneiderman, B. (1998). *Designing the user interface strategies for effective human-computer interaction* (3rd ed.). Reading, MA: Addison-Wesley.

Su, L. T. (1992). Evaluation measures for interactive information retrieval. *Information Processing and Management, 28*, 503–516.

Thong, J. Y. L., Hong, W. Y., & Tam, K. Y. (2002). Understanding user acceptance of digital libraries: What are the roles of interface characteristics, organizational context, and individual differences? *International Journal of Human-Computer Studies, 57*, 215–242.

Tsakonas, G., Kapidakis, S., Papatheodorou, C. (2004). Evaluation of user interaction in digital libraries. In Agosti, M., Fuhr, N. (Eds.), *Notes of the DELOS WP7 Workshop on the Evaluation of Digital Libraries*, Padua, Italy.

Van House, N. A., Butler, M. H., Ogle, V., & Schiff, L. (1996, February). User-centered iterative design for digital libraries: The Cypress experience. *D-Lib Magazine, 2*. Retrieved February 3, 2005, from http://www.dlib.org/dlib/february96/02vanhouse.html.

Xie, H. (2006). Evaluation of Digital Libraries: Criteria and problems from users' perspectives. *Library and information Science Research, 28*, 433–452.

Xie, H., & Cool, C. (2006). Toward a better understanding of help seeking behavior: An evaluation of help mechanisms in two IR systems. In Dillon, A., Grove, A. (Eds.), *Proceedings of the 69th ASIST Annual Meeting (Vol. 43)*. Retrieved February 1, 2007, from http://eprints.rclis.org/archive/00008279/01/Xie_Toward.pdf.

Yang, S. C. (2001). An interpretive and situated approach to an evaluation of Perseus digital libraries. *Journal of the American Society for Information Science and Technology, 52*, 1210–1223.