How Usable are Operational Digital Libraries – A Usability Evaluation of System Interactions

Xiangmin Zhang
Rutgers University
4 Huntington Str.
New Brunswick, NJ 08901
xiangminz@gmail.com

Jingjing Liu
Rutgers University
4 Huntington Str.
New Brunswick, NJ 08901
jingjing@scils.rutgers.edu

Yuelin Li
University of Southern
Mississippi
118 College Dr. #5146
Hattiesburg, MS 39406
Yuelin.Li@usm.edu

Ying Zhang
University of California at
Irvine
Irvine, CA 92697-8100
yingz@uci.edu

ABSTRACT

This paper reports a usability evaluation of three operational digital libraries (DLs): the ACM DL, the IEEE Computer Society DL, and the IEEE Xplore DL. An experiment was conducted in a usability lab and 35 participants completed the assigned tasks. The results demonstrate that all three DLs have more or less usability problems by various measures. Searching in Xplore by inexperienced users was problematic, and browsing in IEEE CS was extremely difficult for all users. Interaction design features that caused these usability problems were identified and discussed. The study implies there is still large room for operational DLs to improve in order to provide more satisfactory services.

Categories and Subject Descriptors

H.5.2 User Interfaces --*User-centered design*

General Terms

Design; *Human Factors; Experimentation*

Keywords

Digital libraries; interaction design; usability testing

1. INTRODUCTION

Digital libraries (DLs) have become important sources for people to fulfill their information needs. In this paper, we report the results of a usability study on three operational DLs: the Association for Computing Machinery digital library (ACM), the Institute of Electric and Electronic Engineering Computer Society digital library (IEEE CS), and the Institute of Electric and Electronic Engineering Xplore digital library (Xplore). The three DLs are all well known, serving millions of users worldwide. Although their collection foci differ to some extent, they all cover the fields of engineering and computer science. The three DLs

represent different styles of interface design. Figures 1, 2, and 3 are screenshots of their homepages.
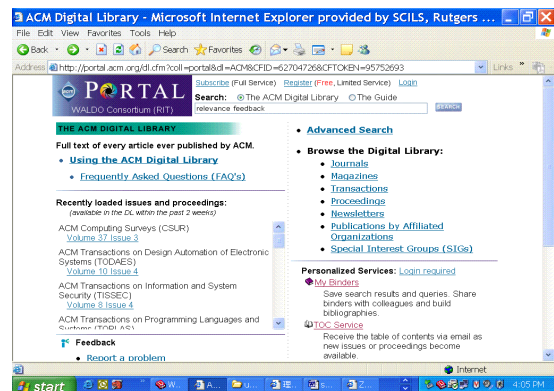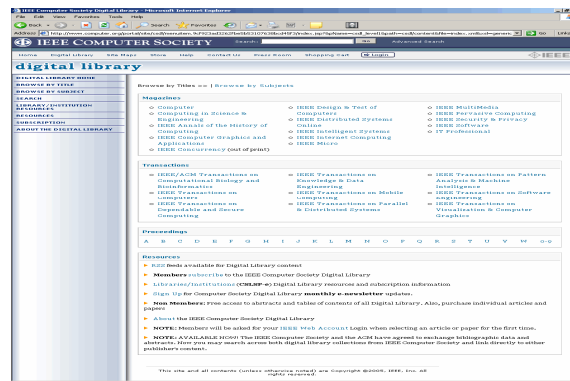

Figure 1. The ACM DL homepage

Figure 2. The IEEE CS DL homepage[1]

Usability of a system is critical for its success. Evaluation of the system's usability identifies problems and provides feedback for future system design. Among the many dimensions that DLs may be evaluated [26, 33], this study focused on the usability of the interaction design. We looked particularly at searching and browsing, two interactions users frequently engage in with DLs [36]. Designs for both searching and browsing have drawn plenty of research interest, but so far little in-depth knowledge has been provided on how designs relate to user performance in operational DLs. We were interested in finding how well

---

[1] The study was completed in 2006. The IEEE CS DL homepage has changed since our study.

these three DLs serve the users, what usability problems exist, and how they compare to each other in usability.
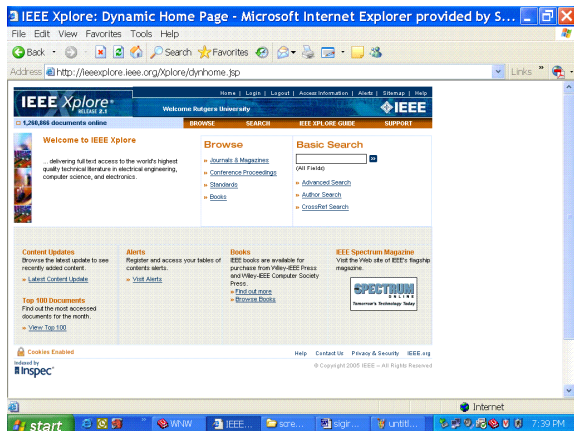

Figure 3. The Xplore DL homepage

2. RELATED WORK

A number of major usability issues with search systems has been identified, including zero-hit results, difficulties in formulating queries (including using special syntax such as Boolean operators), and feelings of being lost or overwhelmed [12, 29]. While some of these may be due to the internal functions of the system, many of the problems are caused by poor interaction designs for search systems.

Theng [31] found that the feeling of being "lost in hyperspace" was a prevalent problem in DLs. When interacting with DLs, users experienced different forms and degrees of feelings of being lost. They sometimes could not identify where they were, sometimes could not return to the previously visited information, and sometimes could not go to the information they believed existing.

Shiri and Molberg [27] investigated 33 digital collections in Canada. The authors detected many problems in the simple search interaction designs. Different from the advanced search mode, which usually allows the user to structure queries using multiple criteria such as title, author, year, keyword, etc., the simple search mode restricts the user to keyword searches, typically using title and/or author [18]. As Shiri & Molberg [27] found, the simple search approach hindered the user to enter more precise and complicated search queries and caused confusion whether it was searching the institution's digital collection or searching its web site. The user usually also did not know in which field the system was searching.

While the above usability evaluations were conducted for established DLs, efforts were also made during the development process of DLs. Hill et al. [14] describes user evaluations for an on-going DL project: the Alexandria Digital Library. During the system development period, three different user interfaces were developed and tested by user groups. User feedback was collected through various formal and informal approaches and the results were fed back into the design and implementation cycle.

Some studies focused specifically on the usability issues of browsing designs. Paynter et al. [23] argued that metadata information, such as authors, titles, and keywords, organized and represented properly, could be used for browsing a collection effectively. Chang [2] found that both information organization and representation largely influenced "the extent to which people engage in browsing and the effectiveness of that browsing" (p.47). According to the author, browsing performance relies not only on the user's awareness of what and where the destination is, but also on his awareness of how to get there. Inappropriate access point to resource is considered to be one of the problems for browsing design [1]. The issue of "how to get there" is closely related to how browsing interface is designed. Different designs may lead to different ways of reaching the desired information and may result in different user performance.

Several user studies have demonstrated that some browsing approaches may be superior to others for users to locate the needed information. In a fine art image retrieval setting, Yee et al. [35] demonstrated some advantages of faceted metadata display over a standard image retrieval/browsing interface. The results of their usability test with 32 art history students showed that overall the participants preferred more faceted metadata display, and evaluated it as more informative, flexible, and easier to use. Similarly, in a video DL browsing setting, Zhang and Marchionini [37] demonstrated that their faceted browsing interface was superior to traditional form-fill in video DL in terms of being more effective, efficient and satisfactory for data exploration tasks. Although these studies are about image and video retrieval, their proposed browsing approaches are not necessarily limited to the image and video collections, but have potentials to be used in other digital libraries with other types of collections.

Based upon their examination on users' attitudes towards various interface features provided by existing Web directories, Chen, Magoulas, and Dimakopoulos [3] proposed a flexible interface design guide. They found that users' cognitive styles had impacts on their reactions to the organization of subject categories. Whereas one group of users preferred alphabetical display, the other group of users seemed to favor the ordered display. Accordingly, they suggested a flexible interface (e.g. switching visual cues, offering successive options) to accommodate the preferences of users with different cognitive styles.

Chimera and Shneiderman [4] evaluated three interface types, i.e. expandable, expandable/contractible, and multi-pane types, for browsing large hierarchical tables of contents. They found that the latter two information display approaches yielded more efficient performance for the given tasks. Believing that multifaceted hierarchical display was superior to linear display, Dakka, Ipeirotis, and Wood [7] designed an automatically generated multifaceted interface. This new approach proved to be efficient in providing a fast access to information.

As can be seen, most of previous studies on browsing proposed new approaches. However, few of them were concerned with evaluation of browsing designs in established operational DLs by real end users. Testing of the browse features implemented in operational DLs should be helpful to detect design problems and to provide feedback to browsing interface designs.

Usability issues affect users' attitude towards DLs. Hong et al. [16] and Thong, Hong, & Tam [32] investigated user acceptance of DLs. They found that potential users of DLs may not use the developed DLs though millions of dollars were invested. Their studies revealed that the perceived ease of use was one of the determinants of the user acceptance of DLs.

Tsakonas and Papatheodorou [33] explored usefulness and usability of electronic information service system, such as digital libraries, e-journal platforms, portals, and e-prints. The results indicated that user interaction was affected by both content and system characteristics. Also, ease of use and navigation are the most influential attributes of a system. They [34] further identified that easiness of use, learnability, and familiarization with DL functionalities dramatically affected users' interaction and satisfaction with DLs.

Several studies were concerned with operational DLs and aimed to help improve current DLs' design. Hartson, Shivakumar, & Pérez-Quinones [11] performed a usability inspection on the Networked Computer Science Technical Reference Library (NCSTRL). They revealed some problems with consistency, feedback, wording, and layout and graphic design. Based on the results, the authors suggested reconsidering the interaction design of searching and browsing functions in DLs. The limitation of the inspection was that it was performed by usability specialists, rather than real users. Kengeri et al. [18] is among the few studies comparing user performance in multiple operational DLs. They looked at four DLs: ACM, IEEE CS, NCSTRL, and the Networked Digital Library of Theses and Dissertation (NDLTD). User performance in their study was measured by ease of use, the amount of search time, and the number of errors. This study used a small number of measures, and did not involve examining the browsing function. Chrzastowski and Scheeline [6] evaluated the web site of the Analytical Sciences Digital Library (ASDL) based on 10 students from two institutions. The study found that ASDL was not successful in helping students "suggest a site to ASDL" and "find class materials on gas chromatography." Zhang et al. [38] investigated the interaction design of search and browse function in ACM, IEEE CS, and IEEE Xplore. The results demonstrated significant differences in many aspects of the user interactions among the three DLs, and thus informed potentially future improvement.

Users' perception and evaluation criteria also gained attention. Xie [36] found that more than half of the participants discussed both usability in general and interface usability as key evaluation criteria. In particular, search and browse functions were ranked as the second most essential evaluation criterion, following content in general. Kani-Zabihi, Ghinea, and Chen [17] surveyed users' suggestions for digital libraries. With respect to usability, users expected that a DL should be easy to learn and reliable in providing search results. In summary, efforts have been made on usability evaluation of DLs, but few empirical studies have been conducted to investigate the usability issues of operational DLs.

3. METHODOLOGY

Usability testing normally is conducted on a single system, against certain usability guidelines or benchmarks. In our study, because we focused on multiple DLs (ACM, IEEE CS, and IEEE Xplore), and there were established usability standards for the interactions in DLs, we explored usability issues first by comparing user performance data across different DLs. The comparisons were based on the assumption that if there were different user performances between the DLs evaluated, there would be usability problems with one or more DLs involved (This does not mean there would be no problem if the user performance is at the same level on all the DLs). We then relied on other methods, such as user feedback, or heuristic evaluation, to further identify usability problems. Following this approach, we designed the study as described in this section.

3.1. Participants
Thirty-six participants were recruited from a large research university: 12 undergraduate engineering or computer science (UE) students, 12 graduate engineering or computer science (GE) students, and 12 masters of library and information science graduate (LIS) students. They were considered as the end users of the DLs investigated. Among the 36 participants recruited, one LIS participant dropped off from the study. The valid number of participants was therefore 35, of whom 15 were female and 20 were male.

The participants as a whole were a very computer literate group. When asked in the background questionnaire about their level of expertise with computers, over 94% of participants rated themselves above the medium level of computer expertise (4 in a 7-point scale), and nearly 50% considered themselves as experts (ratings 6 or 7). They were also experienced users of search engines: about 83% considered themselves very experienced with searching on the Internet.

In contrast to their experience with internet search engines, the majority (23 out of 35, 66%) of the participants did not have any experience with ACM and Xplore. An even higher percentage of participants, 80% (28 out of 35), never used IEEE CS. In general the participants were new or inexperienced users of the three DLs.

3.2. Tasks
To explore the interaction designs in the three DLs for searching and browsing, two information exploration tasks, one searching and one browsing, were designed for the study.

The search task was a typical topic search task. It required the participants to locate relevant information about "protecting the online repository from fraudulent activities by watermarking". The participants were allowed to use any search method.

The browsing task required the participants to use "Browse" feature (including the "Search within" function associated with "Browse" feature) in the three DLs to locate the specified source and relevant papers on a specific topic. The participants were asked to save the relevant results on the experiment computer. This task was featured as requiring certain efforts to explore the information space and meanwhile providing a general browsing goal to guide the browsing interactions. For IEEE CS and Xplore, the users were asked to browse the proceedings of ITCC: International conference on Information Technology (2004): Coding and Computing, and locate two papers about data streaming, and save their abstracts. For ACM which does not have ITCC, Annual Symposium on Computational Geometry (SCG) (2004) was used.

Considering that a task for usability test of a DL could not be too simple or too complex, as Notess, Kouper, and Swan [22] pointed out, these two tasks were selected from a list of candidate tasks provided by the IEEE Xplore library reference services. The candidate tasks were tested in the three DLs by the researchers of this study in order to identify the appropriate ones that could lead the participants to experience the different aspects of interaction designs. A pilot study was also conducted. It indicated only one search and one browse task could be completed by the participants, due to the time constraint.

3.3. Usability Measures
Both objective and subjective measures were employed in this study. Some of the measures were particularly chosen for this study, and some others were widely used measures in usability testing of interactive information retrieval systems.

3.3.1. Objective measures
Number of queries issued: The number of queries a participant issued to obtain the relevant results set.

Amount of search time: The amount of time a participant spent in order to get the satisfactory result set. It started when a participant issued the first query and ended after he/she clicked the mouse to copy the search results.

Amount of browsing time: The amount of time the participants spent for the browsing task. It started when the participant moved the mouse to locate the entry point for browsing and ended after the participant clicked the mouse to copy the results.

Number of search steps: The number of steps the participants made before they finally obtained the relevant results (i.e. clicking the mouse to copy the relevant results), including choosing a search method, inputting a query, changing search fields, clicking the button for submitting a query (or using the "Enter" key), and clicking "Back" and "Forward" in the browser to navigate between pages.

Number of browsing steps: The number of moves the participants made before they obtained the desired results (i.e., clicked the mouse to copy their results). The steps included selecting the category for browsing, clicking links, entering a query (if the participant used the embedded search feature), issuing the query, etc.

Number of zero-hit ("no results") pages returned: The number of times the system failed to find any document that would match the user's query.

Number of user errors included:

- Clicking on a link leading to wrong web pages.
- Errors in search query and the system sent error messages after searching, such as "You have entered an invalid search."
- Actions causing system process errors such as "Error occur when processing request," or "Enter key is disabled" when the user tried to use the enter key on the keyboard instead of clicking the mouse.

These measures have been frequently used in usability testing and evaluations of interactive information retrieval systems. For each of these measures, the higher the value, the more difficult the system is to use.

3.3.2. Subjective measures
Subjective measures were also used to test the systems' usability. Post-task questionnaires included some rating statements which elicited the perceived ease of use, satisfaction, etc. Perceived ease of use was initially constructed by Davis [8], and was validated and empirically tested by Doll, Hendrickson, & Deng [9]. In this study, the statements for perceived ease of use were based on Thong, Hong, & Tam [32], but were tailored to specific search and browsing tasks in this study. User's satisfaction was elicited by two statements developed based on Chin, Diehl, and Norman [5]. One was on the satisfaction with the search results and the other one was on the satisfaction with the overall system feature.

Participants rated these statements on a 7-point Likert scale based on how closely they describe their experience, where "1" was for the least in agreement and "7" for the most. In general, "1" was the most negative and "7" was the most positive. One exception is for the statement "I exerted great effort to complete this task," "1" for the least effort (most positive) and "7" the most effort (most negative).

3.4. Experiment Design
The experiment involved three DLs and two information seeking tasks. The study employed a within-subject, Latin-square design to balance the system and task orders [30]. For 36 participants, there are 12 orders of searching and browsing tasks combined with the system order. The orders were assigned to the participants and every 3 among the total of 36 recruited participants followed the same order.

3.5. Experiment Procedures
The participants were invited individually to an on-campus usability lab to complete the experiment. Upon arrival, the

participant was first asked to fill out a consent form and a background questionnaire. A brief instruction session then followed to inform the participant of the tasks that s/he was to complete. The participant was then asked to perform the search and browsing tasks in an assigned task and system order. After each task (either the search or the browsing task) was completed in one DL, a post-task questionnaire was administered. All the participants were asked to 'think-aloud' during the experiment. A hardcopy of detailed instructions was also given to the participant. The entire session was recorded by the Morae usability testing software. The whole experiment last 2 and a half hours.

3.6. Data Analysis

Because multiple variables (different measures and three DLs) were involved and we wanted to test the differences on all measures across the three DLs, the standard ANOVA and Chi-Square tests were used as the major statistical procedures for data analyses. Our hypothesis for these tests was there would be no differences between/among the three DLs on the usability measures. To minimize the error effect of individual differences among the participants on the accuracy of the experimental results, when all participants were treated as one group to compare different DLs, we used One Factor Repeated Measures ANOVA [15], assuming spherical form of the variance-covariance matrix of the dependent variable. When this assumption for the tests did not hold, the Greenhouse-Geisser criterion was chosen for adjustment, which is especially suitable for small sample sizes [15]. However, when the three groups of participants were compared, since different participants in the three groups, regular one-way ANOVA was used [15].

4. RESULTS

4.1. Objective Measures

Table 1 presents the results of objective measures.

Number of queries
It was found that in Xplore, the participants issued significantly more search queries than in the other two DLs.

Task completion time
For the search task, the participants did not show significant differences in completion time on each of the three DLs. For the browse task, they spent significantly more time in IEEE CS than in the other two DLs.

Task steps
For the search task, it was found that the participants took more search steps in Xplore. For the browse task, they took significantly less steps in ACM than in the other two DLs.

Zero-hit returns
Data showed that Xplore returned significantly more zero-hit pages than the other two DLs. This could possibly explain, at least to some extent, why the participants had to make more steps and issue more queries to complete the search task in Xplore.

Table 1. Objective measures for usability

Measures	ACM	IEEE CS	Xplore	F value
# of queries issued	3.32	2.60	6.14	11.11*
Search time (Seconds)	271.54	240.80	327.91	1.81
Browse time (Seconds)	283.63	595.37	349.34	20.65*
# of search steps	11.51	12.09	22.97	9.999*
# of browsing steps	13.29	36.80	29.83	9.44*
# of "zero-hit" returns	0.69	0.49	3.26	15.61*
# of search errors*	3	7	17	NA
# of browsing errors	1	6	2	NA

* p<.01

User errors
The participants encountered various errors in all three DLs. A Chi-square test found that the participants had significantly more errors in performing the search task with Xplore than with the other two DLs. For the browsing task, meanwhile, no differences were detected in the number of errors among the three DLs.

4.2. Subjective Measures (Search Interaction)

Table 2 lists users' perceptions on their search interactions with the three DLs.

Table 2. Subjective measures of searching interactions

Statements	Mean user ratings		
	ACM	IEEE CS	Xplore
Easy to get started (searching)	5.46	5.66	4.97
Search result lists easy to read	5.23	5.43	5.40
Made great effort to accomplish the task	4.09	3.94	4.49
Satisfaction with final search results	4.77	4.77	4.71
Overall ease of search	5.37	5.26	4.83
Satisfaction with overall search feature	5.06	4.89	4.97

Perceived ease of search
Although not statistically significant, descriptively, the ratings for Xplore were lower than that for ACM and for IEEE CS in terms of the ease of search. The ratings for Xplore fell about a half point below the other two systems on three statements: "Easy to get started (searching)," "Made great effort to accomplish the task," and "Overall ease of search." The lower ratings for Xplore were consistent with the results of the search effort the users exerted, that they issued more queries, made more search steps, and obtained more zero-hit returns in Xplore. The ease of reading the search results from the three DLs was about the same cross the systems.

Satisfaction
Despite the difficulties the participants had with Xplore during searching, user satisfaction with the search results did not show differences for the three DLs, indicating that the participants were able to find the satisfactory results from each of the DLs. This result seemed inconsistent with the results of objective measures, in which poorer performance

was demonstrated for Xplore. This inconsistent is discussed in later section of this paper.

4.3. Subjective Measures (Browsing Interaction)

Table 3 shows user ratings regarding browsing interactions.

Table 3. Subjective measures on browsing interactions

Statements	Mean user ratings			F value
	ACM	IEEE CS	Xplore	
Easy to get started on the browsing task	5.91	5.06	5.71	3.61*
Result lists easy to read	5.66	4.34	5.66	7.982 **
Links easy to follow	5.71	5.06	5.63	4.36*
Made great effort to accomplish the task	3.14	5.66	3.77	19.90 **
Satisfaction with the browsing result	5.34	2.31	4.77	37.26 **
Overall easy to browse	5.63	3.29	5.37	24.70 **
Overall satisfaction with the browse feature of the system	5.77	3.57	5.31	20.88 **

* $p<.05$, ** $p<.01$

Perceived ease of browsing

For all statements, user ratings for IEEE CS were the lowest among the three DLs. Statistical significances were detected on ratings for all statements. Compared with ACM and Xplore, IEEE CS was rated significantly worse in "Easy to get started on browsing". "Links easy to follow" reflected the ease of navigation in the digital collection. IEEE CS was rated significantly lower than ACM. "Made great effort to accomplish the task" was a reversely rated statement: the higher the rating, the more effort the user spent. IEEE CS was rated significantly higher than both ACM and Xplore For "Overall easy to browse," user rating for IEEE CS was about two points lower than the other two systems, and the differences were again significant. Last but not least, for the ease of reading the result list, IEEE CS was found to be significantly more difficult to read than the other two DLs.

Satisfaction

Given that IEEE CS had poor ratings with all statements, it is not surprising that the users' ratings on the two satisfaction statements for IEEE CS were significantly lower than the other two DLs. For satisfaction with the browsing results, the rating for IEEE CS was significantly lower than that for ACM and for Xplore. For "Overall satisfaction with the browse feature," the users' satisfaction with the IEEE CS was again significantly lower than the other two DLs. No statistically significant differences were found between Xplore and ACM.

4.4. Usability Problems

Based mainly on the differences described in previous sections, major usability problems were observed for each of the three DLs. Some problems are common in nature and some problems are unique with specific DLs. These problems are discussed in more detail in the following subsections for each DL. Although many of the problems are trivial, altogether they could make an impact on users' experience.

It should be noted that because the study focused on the interaction design issues, we did not conduct a thorough usability inspection of the whole system (for each of the three DLs). The problems were identified through user interactions. There may be other usability violations that are still uncovered by this study.

Xplore

The results demonstrate that in terms of the usability of search interaction designs, Xplore was significantly weaker than the other two DLs. The participants spent more efforts and made more mistakes. This difficulty of use was related to the system's frequent zero-hits return problem, which was almost solely with Xplore.

Although the zero-hit problem has been documented [12, 20], the reason why it occurs has not been well explored. An examination of the interface designs of the three DLs found that the problem appeared to be related to the Xplore's design that the fielded search was set as the default method, and "AND" as the default Boolean operator. The fielded search method strictly limits searching to the narrower metadata database. When searching metadata, if users use long queries, they may obtain zero-hit results if the query phrases or terms are not available as a valid phrase in the metadata. For example, one participant (GE6) commented: *"It is funny for IEEE Xplore basic search…I gave three keywords, it does not give me anything. It is strange, and not very friendly."* This may be able to explain the result regarding the query length, that the participants submitted significantly shorter queries in Xplore. This finding is consistent with Shiri and Molberg's [27] findings about the weakness of basic search in digital collections, that it cannot well support more precise and complicated search queries. Resnick & Vaughan [25] suggest that if no results are found, the system should provide suggestions for improving the query.

The difficulty of searching was reflected by lower ratings on the corresponding perceived ease of use statements for Xplore, although the differences were not statistically significant. However, the satisfaction of overall search feature for Xplore was rated about the same as the other two systems, despite the difficulties the users actually had when searching Xplore. The reason may be that this is a very general statement. The impression of other parts of the interface design such as the presentation design may have effects. It could be the case that the overall impression of the interface design compensated the inferior search interaction design of Xplore for the users' overall satisfaction ratings.

In addition to the search usability problems, Xplore's browse design also had problems. The major problem is that there is no acronym or complete title for conference proceedings for

"Browse alphabetically." The list of proceeding titles was ordered based on the title keywords, rather than the original titles. All the titles were presented in such a way that the keywords were placed at the beginning of the entry and other complementary words such as "conference" or "proceeding" was placed following the keywords. For example, the proceeding "International Conference on Information Technology (2004): Coding and Computing" was transformed in the list as "Information Technology: Coding and Computing (ITCC), International Conference on." This order contradicts with the instruction displayed on the screen for this "Browse alphabetically" feature, which reads: "Select a letter to view titles beginning with that letter," indicating an alphabetical order of the titles. It also contradicts with many users' expectations. Participants were looking for either the whole title or the acronym "ITCC" in the list but failed to find it. An example of the participants' comments on "Browse alphabetically" was: "where is ITCC? What kind of order (the proceeding list) follow?" (GE9).

IEEE CS
The results highlighted an inferior user browsing performance with IEEE CS in almost all the measures used. The participants experienced more steps, spent longer time, and made more mistakes during the process of accomplishing their browsing task. The perceived ease of browsing and satisfaction were also the poorest.

The major usability problem with IEEE CS could be attributed to the lack of the "Search within" feature in the browse interface. Among the three systems, both ACM and Xplore allowed the users to search within a proceeding, which had the advantage that the users did not need to browse through all issues of the proceeding or all documents within an issue to find desired papers. Using search to quickly locate related items within a particular proceeding apparently saved users' effort and time.

Unfortunately, IEEE CS did not provide this function. The participants had to manually follow each issue/year of the proceeding, which required more time and steps. Egan et al. [9] and Hertzum and Frøkjær [12] found that users favored the combination of browsing and search function. Our results confirmed that this feature on the browsing interface of DLs was highly preferred by the user. In addition, the systems with a combination of browsing and searching may lead to better user performance.

The problems were reflected in participants' perceived ease of use and satisfaction ratings. The ratings showed IEEE CS was difficult to use and the participants were less satisfied with the IEEE CS's browse design. It is interesting to find that when users had difficulties in browsing, the users' satisfaction with the results decreased significantly. This seems to be different from the search interaction. Although the users had problems when searching in Xplore, they were still as satisfied with the results as with the results from the other two systems. This is certainly an interesting issue worth further investigation in the future.

For the usability of the search interface, both objective measures and subjective ratings for IEEE CS were not the worst. However, they were not the best either. Particularly in the perceived ease of use and satisfaction measures, the ratings are far away from the top ratings (7 on the scale), which indicate that even in the search interface arena, IEEE CS also could be improved.

ACM
Comparing the three DLs in terms of their usability, Xplore had major search interaction design problems. IEEE CS, on the other hand, had browse design problems that led to poor user performance. Unlike these two DLs, ACM did not show major problems. In all areas measured, it seldom was the worst among the three DLs. However, this does not mean ACM has no usability problems. While the objective measures could only reveal obvious problems, not how best a design is, the perceived ease of use and satisfaction ratings can indicate the degree of the usability level. In the case of subjective ratings for the searching task, even though some of the ratings for Xplore were low, the ratings for the other two systems were not high either: neither of them was above 6 on a 7-point scale. The highest was 5.77, below 6 out of the 7-point scale.

The browsing case had the similar results. IEEE CS was rated significantly lower than the other two systems for all statements. However, the highest rating for the other two systems was 5.91 for ACM's "Easy to get started." The highest overall satisfaction rating was about 5.77 (for ACM), again below 6 out of the 7-point scale. Apparently there is still large room for ACM to improve.

4.5. Suggestions for Future Design
Interaction design experts have proposed various types of principles for design of friendly user interfaces, for example, the principles of *match between the system and the real world*, *consistency and standards*, etc., proposed by Preece, Rogers, & Sharp [24]. These principles are usually general guidelines, and it is normal that a system interface has more or less violations to such guidelines even though the design engineers are aware of the principles. Therefore, it is necessary to offer detailed and specific suggestions that are of practical importance. Our inspection of the three DLs detected quite some usability problems, which varied from search to browse interface to other categories such as result display, labeling, and so on. Based on the usability problems that were found in the three DLs, some suggestions for future design of more friendly interfaces could be offered, as shown in Table 4.

Table 4. Suggestions for future design

Category	Problem	System	Suggestion
Search interface/facility	No "abstract" as a search field in the advanced search mode	IEEE CS	In addition to title, author, date, etc., also provide "abstract" as a search field in the advanced search mode
	Missing "Basic search" in the pull-down search menu	Xplore	In the pull-down menu in all search pages, include all types of search modes for the ease of accessing and choosing any search mode in any page
	Default search field set as fielded search, sometimes causing zero-hit result in case of long search query	Xplore	Clarify or provide informative feedback message with zero-hit return when use fielded search as default, or use full-text search as default
	Problematic "Modify search" box containing both the user entered search terms and the system generated code: "<in> (*search field*)."	Xplore	Be friendly in the display of "modify search" box, for example, keep only the user entered terms
Browse interface/facility	Poor organization order in "Browse alphabetically" which uses the alphabetical order of title keywords rather than full or acronyms of the conference	Xplore	Use the alphabetic order of either the full name or the acronyms of the conference, and explicitly clarify this to the users
	No "search within" proceedings function	IEEE CS	Provide a "search within" proceedings function
	Confusing "Advanced search" in "Search within" (a specific proceeding) in the browse function, which directed the users to "advanced *search*" rather than advanced "*search within*" the proceeding	ACM	Avoid misleading, unclear, or confusing search mode labeling
Result display	No "abstract" presented in the result page	IEEE CS	Provide an abstract in the result display for the user to have a quick overview of the document
	No "abstract" link for some documents in search results	ACM	Provide an abstract link for all documents in result display
	Search terms not highlighted in the results	ACM	Highlight search terms in the result display
Labeling	Inconsistency in labeling, e.g., the use of two labels: basic search and simple search, to refer to the same thing; the use of "browse by keyword" and "search within this proceeding" which are both hard to understand	IEEE CS, Xplore	Keep the label use consistent, simple, and easy to understand
Error message	Uninformative (generic) error messages	Xplore	Be specific in error messages descriptions
Homepage relocating	Lack of "Home" button/link, or confusing one which directed the users to the homepage of the ACM or IEEE CS rather than the DLs	ACM, IEEE CS	Provide access to "Home" in all pages for the ease of going back to the DL's homepage

Some problems detected in this study are not unique. Other types of interactive systems, such as search engines, databases, or library catalogs, may have the same or similar problems in their interface design. The suggestions listed in Table 4 are proposed to solve the specific problems of the inspected three DLs. Meanwhile, they could serve as helpful guidelines to interface design of other systems.

5. DISCUSSION AND CONCLUSIONS

One aspect of the study that warrants further discussion is the differences between/among different types of participants. In general, the system differences were dominant in this usability test. Significant differences among different types of participants, such as in terms of academic background, were found on only two occasions: one was satisfaction with

IEEE CS between the LIS and the UE participants: The LIS users were significantly more unsatisfied with their search results and the search feature of IEEE CS than the UE participants were. The second was the perceived ease of use and satisfaction of browsing design with ACM, for which the LIS and the GE participants were more positive than the UE participants.

The reason that the participants did not exhibit significant differences on all the three systems may be that the difference of the participants' academic background was not a factor as important as the interface design of the DLs. For the search case, the users had the worst performance in Xplore. IEEE CS seemed fine in objective measures. For the browse case, ACM was actually the best among the three DLs. These indicated that when the interface design was particularly poor, different types of users would have similar

experience. This was supported by our observation that the subjects who had prior experiences with the DLs tested actually would make the same mistakes or be frustrated by the system as the inexperienced subjects did. Only a few subjects reported they had some experience with one or more of the DLs in the study. Our observation indicated that they in general experienced, more or less, the same usability problems as those inexperienced subjects did. The implication of the results seem to be that, when designing interfaces, efforts first need to be made on a better design for all users.

Our results reveal various usability problems with the three DLs. The design deficiencies leading to the usability problems were further identified. We hope that these problematic designs could be addressed and corrected in the future versions of the systems, so that the user experience with the systems can be enhanced.

Like any other studies, this research had its limitations. First, for the convenience reason, the user population was limited to novice or infrequent users of the DLs investigated. Therefore, the results may not be applicable to all users, particularly to expert users of these DLs. Second, limited by the experiment time, there were only one search task and one browsing task. The variety of tasks in real life situations was not really present in the study, which may limit the generalization of the findings. In future studies, a larger sample with different types of users and different types of searching and browsing tasks can be used for experiments, so that the results from the current study can be verified. Finally, limited by available resources, we were unable to provide alternative designs based on our findings. Hope this could be done in the future.

Despite these limitations, our results were consistent with some previous studies [23, 15, 1, 2, 3], and the suggestions we made should be applicable to other interactive information retrieval systems

6. ACKNOWLEDGEMENT

7. REFERENCES

[1] Belkin, N.J., Marchetti, P.G., Cool, C. 1993. Braque: design of an interface to support user interaction in information retrieval. Information Processing & Management. 29(3), 325-344.

[2] Chang, S. L. 1995. Toward a multidimensional framework for understanding browsing. Unpublished doctoral dissertation, Rutgers, The State University of New Jersey, New Brunswick, NJ.

[3] Chen, S. Y., Magoulas, G. D. , Dimakopoulos, D. 2005. A flexible interface design for Web directories to accommodate different cognitive styles. Journal of the American Society for Information Science and Technology. 56(1), 70-83.

[4] Chimera, R., Shneiderman, B. 1994. An exploratory evaluation of three interfaces for browsing large hierarchical tables of contents. ACM Transactions on Information Systems (TOIS). 12(4), 383-406.

[5] Chin, J., Diehl, V. A., & Norman, K. L. 1988. Development of an instrument measuring user satisfaction of the human-computer interface. Proceedings of the SIGCHI conference on Human factors in computing systems (pp.213-218). Washington, D.C., May 15 - 19, 1988.

[6] Chrzastowski, T. E., & Scheeline, A. 2006. ASDL: The Analytical Sciences Digital Library taking the next steps. Science & Technology Libraries, 26(3/4), 79-94.

[7] Dakka, W., Ipeirotis, P. G., Wood, K. R. 2005. Automatic construction of multifaceted browsing interfaces, in: Chowdhury, A., Fuhr, N., Ronthaler, M., Schek, H.-J., Teiken, W. (Eds.), Proceedings of the 14th ACM international conference on Information and knowledge management CIKM '05. ACM Press, New York, pp. 768-775.

[8] Davis, F.D. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13, 3, 319–340.

[9] Doll, W.J.; Hendrickson, A.; and Deng, X. 1998. Using Davis's perceived usefulness and ease of use instruments for decision making: A confirmatory and multigroup invariance analysis. Decision Sciences, 29, 4, 839–869.

[10] Egan, D.E., Remde, J.R., Landauer, T.K. 1989. Behavioral evaluation and analysis of a hypertext browser, in: Bice, K., Lewis, C. (Eds.), Proceedings of CHI'89 Conference on Human Factors in Computing Systems. ACM Press, New York, pp.205-210.

[11] Hartson, H. R., Shivakumar, P., & Perez-Quinones, M. A. 2004. Usability inspection of digital libraries: a case study. International Journal of Digital Library, 4, 108-123.

[12] Hearst, M., Elliot, A., English, J., Sinha, R., Swearingen, K., & Yee, K.P. 2002. Finding the flow in Website search. Communications of the ACM, 45(9):42-49.

[13] Hertzum, M., Frøkjær, E. 1996. Browsing and querying in online documentation: a study of user interfaces and the interaction process. ACM Transactions on Computer-Human Interaction (TOCHI). 3(2), 136-161.

[14] Hill, L.L., Carver, L., Larsgaard, M., Dolin, R., Smith, T. R., Frew, J., & Rae, M.-A. 2000. Alexandria Digital Library: User Evaluation Studies and System Design. Journal of the American Society for Information Science. 51(3):246–259.

[15] Hinton, P. R., Brownlow, C., McMurray, I., & Cozens, B. 2004. SPSS Explained. New York, NY: Routledge, Taylor & Francis Group.

[16] Hong, W., Thong, J. Y.L., Wong, W.-M., & Tam, K.-Y. 2001-2002. Determinants of user acceptance of digital libraries: An empirical examination of individual differences and system characteristics. Journal of Management Information Systems, 18(3), 97-124.

[17] Kani-Zabihi, E., Ghinea, G., Chen, S. Y. 2006. Digital libraries: what do users want? Online Information Review, 30(4), 395-412.

[18] Kengeri, R., Seals, C. D., Harley, H. D., Reddy, H. R., & Fox, E. A. 1999. Usability study of digital libraries: ACM, IEEE-

CS, NCSTRL, NDLTD. International Journal of Digital Library, 2, 157-169.

[19] Lesk, M. 2005. Understanding Digital Libraries, 2nd ed. San Francisco, CA: Morgan Kaufmann Publishers.

[20]Marchionini, G. 1995. Information Seeking in Electronic Environments. Cambridge University Press, United Kingdom.

[21] Marchionini, G., Norman, K., & Morrell, K. 1992. Final project report: Computer interface design for intermediate results for Grateful Med. Report to the National Library of Medicine. Bethesda, MD: NLM.

[22]Notess, M., Kouper, I., & Swan, M. B. 2005. Designing effective tasks for digital library user tests: lessons learned. OCLC Systems & Services, 21(4), 300-310.

[23]Paynter, G. W., Witten, I. H., Cunningham, S. J., Buchanan, G. 2000. Scalable browsing for large collections: a case study, in: Nürnberg, P. J., Hicks, D. L., Furuta, R. (Eds.), Proceedings of the fifth ACM Conference on Digital libraries. ACM Press, New York, pp.215-223.

[24] Preece, J., Rogers, Y., & Sharp, H. 2007. Interaction design: beyond human-computer interaction. New York: John & Wiley Sons, Inc.

[25]Resnick, M.L. & Vaughan, M.W. 2006. Best practices and future visions for search user interfaces. Journal of the American Society for Information Science & Technology, 57(6), 781-787.

[26]Saracevic, T. 2000. Digital library evaluation: towards an evolution of concepts. Library Trends, 49(3): 350-369.

[27]Shiri, A. and Molberg, K. 2005. Interfaces to knowledge organization systems in Canadian digital library collections. Online information Review, 29(6), 604-620.

[28] Shneiderman, B., & Plaisant C. 2005. Designing the user interface. Boston, MA: Pearson Education, Inc.

[29] Shneiderman, B., Byrd, D., & Croft, B. 1997. Clarifying search: A user-interface framework for text searches. D-Lib Magazine, January 1997. Retrieved January 20, 2008, from http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html.

[30]Tague-Sutcliffe, J. 1992. The pragmatics of information retrieval experimentation, revisited. Information Processing & Management, 28(4), 467-490.

[31] Theng, Y. L. 1999. "Lostness" and digital libraries. Proceedings of the fourth ACM conference on Digital libraries, 250-251.

[32] Thong, J. Y. L., Hong, W., & 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.

[33] Tsakonas, G., & Papatheodorou, C. 2006. Analysing and evaluating usefulness and usability in electronic information services. Journal of Information Science, 32 (5), 400-419.

[34] Tsakonas, G., & Papatheodorou, C. 2008. Exploring usefulness and usability in the evaluation of open access digital libraries. Information Processing and Management, 44, 1234-1250.

[35] Yee, K.P., Swearingen, K., Li, K., Hearst, M. 2003. Searching and organizing: Faceted metadata for image search and browsing, in: Cockton, G., Korhonen, P. (Eds.), Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM Press, New York, pp. 401-408.

[36] Xie, H. 2006. Evaluation of digital libraries: Criteria and problems from users' perspectives. Library & Information Science Research, 28, 433-452.

[37] Zhang, J., Marchionini, G. 2005. User interfaces for integrating web-based information: Evaluation and evolution of a browse and search interface: relation browser, in: Delcambre, L. M. L., Giuliano, G. (Eds.), Proceedings of the 2005 National Conference on Digital Government Research. Digital Government Research Center, California, pp.179-188.

[38] Zhang, X., Li, Y., Liu, J., & Zhang, Y. 2008. Effects of interaction design in digital libraries on user interactions. Journal of Documentation, 64(3), 438-463.