

29 September 2004

Evaluation of digital libraries: An overview

Tefko Saracevic
School of Communication, Information and Library Studies
Rutgers University
New Brunswick, NJ USA
tefko@scils.rutgers.edu

“Evaluating digital libraries is a bit like judging how successful is a marriage” (Marchionini, 2000)

Abstract

The purpose of this report is to provide an overview of works on digital library evaluation that included data. The approach was first to analyze and then isolate (or even deconstruct) some 80 evaluation studies along the lines of: 1. construct that were evaluated; 2. context in which evaluation were conducted; 3. criteria that were chosen as a basis for evaluation, and 4. methods that were used. For each of these lines of analysis approaches taken in evaluation studies are identified, enumerated, and approximately ranked: a list of constructs evaluated as to entities or processes is given; context or approaches taken in evaluation are enumerated; numerous criteria used as a base of evaluation are classified; and finally the methodologies used are identified, listed and discussed. Evaluation of digital libraries is not widely practiced. The corpus in this analysis represents the majority of efforts in digital library evaluation that contain data. Conclusions, among others, speculate as to the reasons for a relatively low presence of evaluation in digital library research and practice.

1. Introduction

Digital libraries have a short history. While a discussion about digital libraries, under various names, started in the 1960s, it wasn't till the start or even mid of 1990s that the research, development and practice related to digital libraries really took off. The growth was phenomenal. In little over a decade thousands of digital libraries in a variety of forms were built globally and are functioning operationally, with more to come. Libweb, (a directory of libraries on the web maintained at University of California, Berkeley), currently lists over 7100 pages from libraries in over 115 countries. Hundreds of research projects were devoted to great many aspects of digital libraries in great many countries, and more are reported each year. Everything about digital libraries is explosive, except one thing: evaluation.

The literature on digital library evaluation can be divided in two distinct types:

1. meta or “about” literature: works that suggest evaluation concepts, models, approaches, methodologies or discuss evaluation, but do not contain data
2. object or “on” literature: works that report on actual evaluation and contain data; even data reporting is of two kinds: hard data or soft (sort-of) data (impressions).

The objective of this report is to synthesize the literature on digital library evaluation that reports on actual evaluation and contains data, i.e. the object or “on” literature. The literature about evaluation is **not** covered here, even though many of these works are highly significant for framing and guiding evaluation.

2. Corpus

The literature that reports evaluation (with data) is not large, because there are not that many evaluation efforts around. The proceedings of the two major conferences on digital libraries, the European Conference on Digital Libraries (ECDL) and the Joint ACM/IEEE Conference on Digital Libraries (JCDL), year in year out contain no more than 5% or so of papers or posters that have data pertaining to evaluation of whatever aspect of digital libraries. The overall research literature on digital libraries has even lesser percent devoted to evaluation containing data.

The meta or “about” literature is larger than the object or “on” literature, i.e. up to now there are more works that discuss evaluation than those that report evaluation. A historical parallel could be made: actual evaluation of information retrieval (IR) systems started in the early 1960s or a bit before; for more than a decade around that time there were great many more papers that talked about IR evaluation than reported results of evaluation.

The bibliography here contains 80 items, including a book with several reports. These were selected from a much larger corpus of digital library literature, arrived at after exhaustive searches of many sources and from experience. The criteria for selection were: 1. the work should directly address a digital library or a process related to digital libraries and 2. the work should contain data in whatever form. This does not represent all of literature on digital library evaluation, but I believe that it is not far from the total corpus of reports on evaluation efforts. I estimate that the total corpus is not exceeding 100 reports; this includes a number of duplicate reports from the same project.

A major difficulty was establishing boundaries: what to include/exclude as to the first criterion? What falls under a “digital library?” Or even more difficult: What falls under a digital library process? For instance, where does IR evaluation stop and digital library evaluation start? If we broaden the criteria for processes to include evaluations that have implications for digital libraries but do not address digital libraries directly, e.g. many TREC reports or human information behavior studies, then the number of reports covered would be much larger.

3. Approach

The overview of evaluation reports is structured along the following aspects:

1. **Construct** for evaluation. *What was evaluated? What was actually meant by a “digital library”? What elements (components, parts, processes...) were involved in evaluation?*
2. **Context** of evaluation - selection of a goal, framework, viewpoint or level(s) of evaluation. *What was the basic approach or perspective? What was the level of evaluation? What was the objective(s)?*
3. **Criteria** reflecting performance as related to selected objectives. *What parameters of performance were concentrate on? What dimension or characteristic were evaluate?*
4. **Methodology** for doing evaluation. *What measures and measuring instruments were used? What samples? What procedures were used for data collection? For data analysis?*

4. Constructs

Constructs here refer to items under evaluation: What was evaluated? Two broad and distinct constructs can be recognized as a subject of evaluation:

1. a specific digital library as an entity
2. a specific process that is related to digital libraries but is not a part of any given entity.

4.1 Entities

The first construct encompasses three kinds of entities:

1. Evaluation of digital libraries that have been specifically constructed as digital libraries in R&D projects. Examples are
 - Perseus – classics literature and objects. Perseus is the digital library that has been evaluated more than any other and over a longer period of time. It is highly successful. As far as I can determine Marchionini & Crane (1994) [57] were the first ever to report on an evaluation of a digital library. Other reports are [56] [57] [58] [80]
 - Water in the Earth System (WES) part of Digital Library for Earth System Education (DLESE). WES collection scope is water and users are primarily high school educators [52]
 - Envision –computer science literature [30]
 - ADEPT – Alexandria Digital Earth Prototype – geo-referenced resources for undergraduate education [17] [18] [37]
 - DeLiver – journal articles from scientific and technical journals [7] [8] [9]
 - National Gallery of the Spoken Word (NSGW) [73]
 - MICeval – Moving Images Collection evaluation - online catalog of moving images held by a variety of organizations [72]
 - Making of America prototype – 19th century journals and monographs [53]
2. Evaluation of some aspect of operational digital libraries. Examples are:
 - New Zealand Digital Library (NZDL) focusing on Computer Science Technical Reports [45]
 - ARTEMIS digital library – science materials for students grade 6 to 12 [1]
 - Internet Public Library – digital reference questions [21]
 - Comprehensive Access to Printed Materials (CAMP), Johns Hopkins University – browsing of off-site collection [23]
 - UK National Electronic Library for Health (NeLH) – focused on a large hospital [2]
 - Mann Library Gateway, Cornell University – web access interface [67]
3. Evaluation of multiple digital libraries. Examples are:
 - Project SOUP, Cornell University – six digital collections in museums and libraries [43]
 - Middlesex University – six general digital libraries accessing journals and articles [12]

Glaringly missing from this list are evaluations of operational digital libraries in institutions, i.e. as instituted in academic libraries, museums, national libraries, public libraries and the like. Considerable amounts of statistics are being collected about these libraries, but as yet they have not been a subject of evaluation. As to evaluation, institutional digital libraries are a terra incognita. Yet, there are thousands of them. Similarly missing from evaluation are commercial products for digital libraries. A number of companies are offering digital library software and services; they have not been a subject of evaluation, other than commercial comparisons.

4.2 Processes

A variety of processes related to digital libraries have been evaluated without a reference to a specific library. They are difficult to generalize. Here is a relatively arbitrary classification:

1. Evaluation of various representations for use in digital libraries, e.g.:
 - noun phrasing in medical digital libraries [75]
 - context-based representation for medical literature [68]
 - extraction of keyphrases in digital libraries [46]
2. Evaluation of various tools, e.g. evaluation of:
 - generation of links [59]
 - searching of digital video [78]
 - user-centered interface [4]
 - image retrieval method [35]
 - load balancing on servers [47]
3. Evaluation of various services, e.g.
 - recommender system for digital library [39]
 - digital reference service [77]
4. Evaluation of an evaluation scheme [31]
5. Studies of user behavior in connection with digital libraries or services, e.g.:
 - usage patterns of service logs [15]
 - perception of quality in digital libraries by educators [74]
 - work patterns of experts for development of tools for a digital library [33]
 - user preferences in searching full text databases [66]
 - information seeking in hypermedia digital libraries [70]

Inclusion of the last group of studies, i.e. studies that involve users and use raises a question of the border of digital library evaluation studies. To what extent are user studies also evaluation studies? To what extent are studies of specific user behavior in digital libraries also evaluation studies? Or studies on how people use systems? Clearly, some of these studies are directly related to evaluation (e.g. when examining difficulties or barriers), but other more general user studies are evaluation by implication only.

5. Context

By context is meant the general framework of evaluation, including approach, orientation, level and objectives. Digital libraries are complex social, institutional and technical systems. No evaluation can possibly address all of these aspects together. Thus, different approaches have been used for different evaluation goals.

I am concentrating here on identifying different approaches applied and presenting them in descending order of use in a number of studies:

- **Systems-centered approach:** most prevalent. Involves study of some aspect of performance. Includes assessing effectiveness and/or efficiency of some feature or some specific design or some technological component. Applied in a number of studies with results that may inform specific choices in design or operations.
- **Human-centered approach:** also widely applied. Involves study of behavior in respect to given information needs, such as information seeking, browsing, searching or performance in completion of given tasks, either predetermined or observed in natural settings. Used in a number of studies that illuminated human behavior, requirements, needs, or difficulties encountered. Provides implications for design, but indirectly rather than directly.

- **Usability-centered approach:** Involves assessment of different features, particularly in respect to portals, by users. It is a bridge between systems- and human-centered approaches. Used in several studies with mixed, or self-evident results.
- **Ethnographic approach:** Involves study of life-ways, culture and customs in a digital library environment. Also involves study of impact of a digital library on a given community. Applied successfully in a few studies, with illuminating results, particularly as to impact.
- **Anthropological approach:** Involves study of different stakeholders or communities and their cultures in relation to a given digital library. Applied in one study with interesting results illuminating barriers between stakeholder communities.
- **Sociological approach:** Involves assessment of situated action or user communities in social setting of a digital library. Applied in one study with disappointing results.
- **Economic approach:** Involves study of costs, cost benefits, economic values and impacts. Strangely, it was applied at the outset of digital library history (e.g. project PEAK) but now the approach is not really present at all.

The levels of evaluations varied from micro-evaluation of given features to macro evaluation of the impact of a digital library on a field. An example of first level is evaluation of different speeds for recognition of image surrogates [78] and the latter level is the evaluation of the impact of Perseus on education in the field of classics, and the field itself [58].

Every evaluation is also temporal. Some evaluations had a very pronounced temporal component, particularly in relation to evaluation of given technological features. With technological advances such components become obsolete fast, thus evaluations become obsolete as well. For instance, evaluations of access speeds and problems became irrelevant with broadband access. But some evaluations have also been longitudinal in respect to time. The best, if not the only, example of a longitudinal study was the evaluation of Perseus.

6. Criteria

Criteria refer to chosen standard(s) to judge things by. Criteria are then used to develop measures. (To define the differences by examples: *time* is a criterion, *minute* is a measure, and *watch* is a measuring instrument; *relevance* is a criterion, *precision and recall* are measures, and *human relevance judgment* is a measuring instrument). The importance of criteria follows from this truism: there can be no evaluation without explicitly or implicitly having some criterion or criteria first.

Since 1950's evaluation of IR systems uses relevance as the basic criterion for evaluation. Libraries use a variety of (more or less) standardized criteria for evaluation of components, such as a collection, or services, such as reference. Digital library efforts have not as yet developed anything similar as to evaluation criteria. There is nothing like relevance to be a basic criterion, there are no more or less standardized criteria for digital library evaluation. Several efforts that are devoted to developing digital library metrics have not produced, as yet, generalizable and accepted metrics, some of which may be used for evaluation. Thus, evaluators have chosen their own evaluation criteria as they went along. As a result, criteria for digital library evaluation fluctuate widely from effort to effort.

A summary of most often used criteria follows.

6.1 Usability

Usability has been used widely in digital library evaluation, but there is no uniform definition of what does it cover in digital library context. Usability is a very general criterion that covers a lot of ground and includes many specific criteria – it is a meta term. ISO defines usability “as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.” This was pretty much the umbrella under which usability was used in digital library evaluation.

Here is a list of specific usability criteria applied in various studies:

- **Content** (of a portal or site)
 - accessibly, availability
 - clarity (as presented)
 - complexity (organization, structure)
 - informativeness
 - transparency
 - understanding, effort to understand
 - adequacy
 - coverage, overlap,
 - quality, accuracy
 - validity, reliability
 - authority
- **Process** – carrying out tasks such as search, navigate, browse, find, evaluate or obtain a resource etc.
 - learnability to carry out
 - effort/time to carry out
 - convenience, ease of use
 - lostness (confusion)
 - support for carrying out
 - completion (achievement of task)
 - interpretation difficulty
 - sureness in results
 - error rate
- **Format**
 - attractiveness
 - sustaining efforts
 - consistency
 - representation of labels (how well are concepts represented?)
 - communicativeness of messages
- **Overall assessment**
 - satisfaction
 - success
 - relevance, usefulness of results
 - impact, value
 - quality of experience
 - barriers, irritability
 - preferences
 - learning

6.2 Systems features

As digital libraries are systems, many traditional systems evaluation criteria were used. Some pertain to performance of technology others to performance of given processes or algorithms using technology. .

- **Technology performance**
 - response time
 - processing time, speed
 - capacity, load
- **Process/algorithm performance**
 - relevance (of obtained results)
 - clustering
 - similarity
 - functionality
 - flexibility
 - comparison with human performance
 - error rate
 - optimization
 - logical decisions
 - path length
 - clickthroughs
 - retrieval time
- **Overall system**
 - maintainability
 - scalability
 - interoperability
 - sharability
 - costs

6.3 Usage

A number of studies concentrated on use of a digital library, considering this as evaluation. Usage in itself is a criterion involving studies of:

- usage patterns
- use of materials
- usage statistics
- who uses what, when
- for what reasons/decisions

6.4 Ethnographic and other criteria

Several ethnographic and anthropological studies used criteria derived from those fields. Among others, these involve determination of conceptions and misconceptions in and between different groups, their practices and environments, their culture, their language and frames of reference, priorities, learning, impacts and similar aspects that provide comparative evaluation in a broader context of a digital library.

The most illuminating studies using these criteria and methods addressed impact of a digital library in a given field (such as the already mentioned evaluation of Perseus) or in a given environment (such as a hospital).

7. Methodology

The range of methods used in digital library evaluations is wide. It would be hard to find a scholarly evaluation method that was not used. Here is one: historical method was not used so far..

A number of evaluation studies involved several methods. Here is an inventory in approximate order of number of application:

- surveys, including direct questionnaires and online surveys
- structured interviews
- focus groups
- observations
- task accomplishment
- think aloud
- case studies
- transaction log analyses
- experimentation
- records analysis
- usage analysis
- documents, meeting, communication analysis (anthropology)
- economic analysis

Digital libraries are complex entities, as mentioned. Thus, many methods are appropriate. Each method has, by definition, certain strengths and weaknesses. There is no one “best method” and there never will be one. Still in order to be able to derive some generalizations there will have to be some agreement and standardization of criteria and methods that are appropriate for given evaluation goals and contexts.

8. The versus hypothesis

In this report the emphasis was on synthesis of the constructs, context, criteria and methodologies found in studies evaluating digital libraries. Findings were not synthesized. Anyhow, generalizations of findings are hard to come by, if not even impossible, because of the disparity and great variety of all the other aspects in these studies. However, one finding can be generalized. Expressed in a number of ways this general finding stands out:

A number of studies reported various versions of the same result:

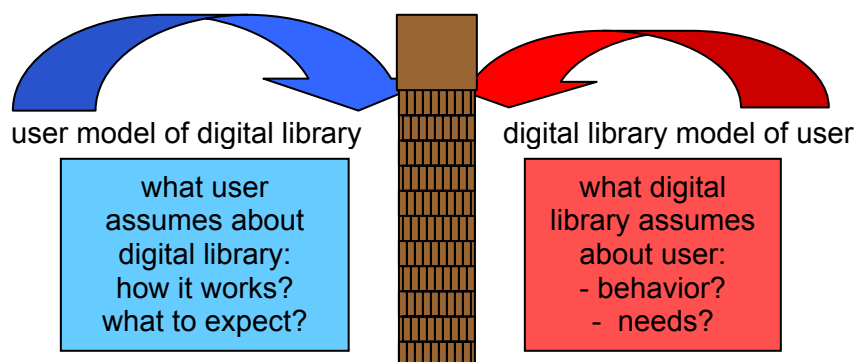
users have many difficulties with digital libraries, such as:

- they usually do not fully understand them
- they hold different conception of a digital library from operators or designers
- they lack familiarity with the range of capabilities, content and interactions provided by a digital library.
- they often engage in blind alley interactions.

Perceptions of users and perceptions of designers and operators of a digital library are generally not very close. To use an analogy from a title of a popular book: Users are from Venus and digital libraries are from Mars. Questions about the relation can be raised:

is it users AND digital libraries? or users VERSUS digital libraries? Is there are firewall between digital libraries and their users as depicted in Figure 1?

Figure 1. Assumptions of users and digital libraries about each other



This leads to the versus hypothesis:

in use, more often than not, digital library users and digital libraries are in an adversarial position.

The versus hypothesis does not apportion blame. It does not say that digital libraries are poorly designed, nor that that users are poorly prepared. It just says that the adversarial relation exist and as all hypotheses that it should be confirmed. A confirmation may find factors that may lead to lessening of the adversity. But then on the other hand, an adversarial relation may be a natural order of things.

9. Conclusions

This report provides an overview of works on digital library evaluation that included data. I tried to first analyze and then isolate (or even deconstruct) some 80 evaluation studies along the lines of: 1. construct that were evaluated; 2. context in which evaluation were conducted; 3. criteria that were chosen as a basis for evaluation, and 4. methods that were used.

Evaluation is not a wide or even growing activity in digital libraries. As a matter of fact, evaluation is more conspicuous by its absence or minimal presence in vast majority of works on digital libraries, in both research and practice; evaluation seems to be an exception rather than a rule. This is in stark opposition to research in IR, where evaluation is an absolute rule.

There have been many works, including my own, suggesting digital library evaluation concepts, approaches, and models. It seems that they had little or no visible impact on those actually doing evaluation. It seems that evaluation theorists and evaluation practitioners do not communicate well, at least not in ways that are visible.

I can speculate as to why evaluation of digital libraries is not more widespread and embraced:

- **Complexity:** Digital libraries are highly complex, they are much more than technological systems alone; evaluation of complex systems is very hard; we are just learning how to do this job and have a lot more to learn. In other words, we as yet do not know how to evaluate and we are experimenting with doing it in many different ways.
- **Premature:** Even though they are exploding and are widespread, it may be too early in the evolution of digital libraries for evaluation. At this stage, evaluation on a more organized and wider scale may be premature.

- **Interest:** There is no interest in evaluation. Those that do or research digital libraries are interested in doing, building, implementing, breaking new paths, operating ... evaluation is of little or no interest, plus there is no time to do it.
- **Funding:** There are inadequate or no funds for evaluation. Evaluation is time consuming, expensive and requires commitment – all these are in short supply. Grants have minimal or no funds allocated for evaluation. Granting agencies, while professing evaluation, are not allocating programs and budgets for evaluation. If there were funds there would be evaluation. With no funds there is no evaluation.
- **Culture:** evaluation is not a part of the culture in research and operations of digital libraries. It is below the cultural radar. A stepchild. Plus many communities with very different cultures are involved in digital libraries. This particularly pertains to differences between technical and humanists cultures: language and frames of reference, priorities and understandings are different; communication is hard and at times impossible. Under these circumstances evaluation means very different things to different constituencies.
- **Cynical:** who wants to know or demonstrate actual performance? Are there any emperor clothes around? Evaluation may be subconsciously or consciously suppressed.

The ultimate evaluation of digital libraries will revolve around assessing transformation of their context – determining possible enhancing changes in institutions, learning, scholarly publishing, disciplines, small worlds and ultimately society due to digital libraries.

References: Studies reporting data on digital library evaluation

Compiled by Ying Zhang and Tefko Saracevic

Annotated and expanded version at http://www.scils.rutgers.edu/~miceval/research/DL_eval.html

1. Abbas, J., Norris, C. & Soloway, E. (2002). Middle school children's use of the ARTEMIS digital library. Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries, 98-105.
2. Adams A, & Blandford, A. (2001). Digital libraries in a clinical setting: friend or foe? Research and advanced technology for digital libraries Proceedings of the 5th European Conference on Digital Libraries, ECDL'01, 214-224.
3. Bainbridge, D., Dewsnip, M. and Witten, I.H. (in print) Searching digital music libraries. Information Processing & Management. Retrieved Sept. 23, 2004 from <http://www.sciencedirect.com/>
4. Baldonado, M.Q.W. (2000). A user-centered interface for information exploration in a heterogeneous digital library. Journal of American Society for Information Science, 51(3): 297-310.
5. Bergmark, D., Lagoze, C., & Sbityakov, A. (2002). Focused crawls, tunneling, and digital libraries. , Research and advanced technology for digital libraries Proceedings of the 6th European Conference, ECDL'02, 91-106.
6. Bertot, J.C., McClure, C.R. (2003). Outcome assessment in the networked environment: research questions, issues, considerations, and moving forward. Library Trends, 51(4): 590-513..
7. Bishop, A.P. (1998). Measuring access, use, and success in digital libraries. The Journal of Electronic Publishing, 3 (December 1998), Retrieved on September 25, 2004 from <http://www.press.umich.edu/jep/04-02/bishop.html>
8. Bishop, A.P. (1999). Making digital libraries go: comparing use across genres. Proceedings of the Fifth ACM Conference on Digital Libraries, 94-103.
9. Bishop, A. P. (1999). Document structure and digital libraries: How researchers mobilize information in journal articles. Information Processing & Management, 35 (3) 255-279.
10. Bishop, A.P., Neumann, L.J., Star, S.L. & Merke, C. et al. (2000). Digital libraries: situating use in changing information infrastructure. Journal of American Society for Information Science, 51 (4): 394-413.
11. Bishop, A.P., Van House, A.A., & Buttenfield, B.P.(Ed.) (2003). Digital Library Use: Social Practice in Design and Evaluation. Massachusetts, Cambridge: The MIT Press.
12. Blandford, A. & Buchanan, G. (2002). Workshop report: Usability of Digital Libraries @ JCDL'02, ACM SIGIR Forum, 36(2): 83-89.
13. Blandford, A., Keith, S., Connell, I. and Edwards, H. (2004) Analytical Usability Evaluation for Digital Libraries: A Case Study. Proceedings of the 2004 joint ACM/IEEE conference on Digital libraries. 240 – 241.
14. Blocks, D., Binding, C., & Cunliffe, D. et al. (2002). Qualitative evaluation of thesaurus-based retrieval.: ECDL'2002 Proceedings of the Second European Conference, ,. 346-361.
15. Bollen, J. & Luce, R. (2002). Evaluation of digital library impact and user communities by analysis of usage patterns. D-Lib Magazine, 8 (6).

16. Bonn, M. S., Lougee, W. P., Mackie-Mason, J. K. & Riveros, J. F. (1999). A Report on the PEAK Experiment. Context and Design. D-Lib Magazine, 5 (6).
17. Borgman, C.L., Gilliland-Swetland, A.J. (2000). Evaluating digital libraries for teaching and learning in undergraduate education: a case study of the Alexandria Digital Earth Prototype (ADEPT). Library Trends, 49 (2): 228-250.
18. Borgman, C.L., Leazer, G.H., et al. (2001). Iterative design and evaluation of a geographic digital library for university students: a case study of the Alexandria Digital Earth Prototype (ADEPT). Research and advanced technology for digital libraries s Proceedings of the Fifth European Conference, ECDL'01,390-401.
19. Borgman, C.L, et al. (2004). How geography professors select materials for classroom lectures: implications for the design of digital libraries. Proceedings of the 2004 joint ACM/IEEE conference on Digital libraries, 179 - 185
20. Bosman, F.J.M., Bruza, P.D., & van de Weide, Th.P. et al. (1998). Documentation, cataloging, and query by navigation: A practical and sound approach. Research and . Research and advanced technology for digital libraries Proceedings of the Second European, ECDL'98, 459-478.
21. Carter, D., & Janes, J. (2001). Unobtrusive data analysis of digital reference questions and service at the Internet Public Library: and exploratory study. Library Trends, 49 (2): 251-265.
22. Champeny, L., Borgman, C.L., Leazer, G.H., Gilliland-Swetland, A. J., Kelli, A., Millwood, K.A., D'Avolio, L., Finley, J.R., and Smart, L.J. (2004). Developing a Digital Learning Environment: An Evaluation of Design and Implementation Processes. Proceedings of the 2004 joint ACM/IEEE conference on Digital libraries. 37 – 46.
23. Choudhury, S., Hobbs, B., & Lorie, M. (2002). A framework for evaluating digital library services. D-Lib Magazine, 8 (7/8).
24. Cornell University (1999). MESL technical report. Retrieved Sept. 24, 2004 from <http://cidc.library.cornell.edu/gateway.htm>
25. Covey, D.T. (2002). Usage and usability assessment: library practices and concerns. Washington, D.C.: Digital Library Federation Council on Library and Information Resources. Retrieved 4/12/2004 from <http://www.clir.org/pubs/reports/pub105/contents.html>
26. Cox, I.J., Miller, M.L., Minka, T.P., Papatomas, T.V. & Yianilos, P.N. (2000). The Bayesian image retrieval system, PicHunter: Theory, Implementation and Psychophysical experiments. IEEE Transactions on Image Processing, 9(1): 20-37.
27. Dillon, A. (1999?). Evaluating on time: a framework for the expert evaluation of digital interface usability. Retrieved 4/12/2004 from <http://www.ischool.utexas.edu/~adillon/publications/evaluating.html>
28. Dorward, J., Reinke, D. & Recker, M. (2002). An evaluation model for a digital library service tool. Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Library, 322-323
29. Entlich, R., Garson, L., Lesk, M., & Normore, L. et al. (1996). Testing a digital library: user response to the CORE project. Library Hi Tech, 14 (4): 99-118.
30. Fox, E.A., Hix, D., et al. (1993). Users, user interfaces, and objects: Envision, a digital library. Journal of the American Society for Information Science and Technology, 44 (8): 480-491.
31. Fuhr, N., Hansen, P., Mabe, M., Micsik, A., & Solvberg, I. (2001). Digital libraries: A generic classification and evaluation scheme. Research and Advanced Technology for Digital Libraries. 5th European Conference, ECDL'01. 187-199.
32. Fuhr N., Klas, C.P., Schaefer, A. & Mutschke, P. (2002). DAFFODIL: an integrated desktop for supporting high-level search activities in federated digital libraries. Research and advanced technology for digital libraries: Proceedings of 6th European conference, ECDL'02, 597-612.
33. Gorman, P. et al (2002). Following experts at work in their own information spaces: Using observational methods to develop tools for the digital library. Journal of the American Society for Information Science & Technology, 53 (14):1245-1250.
34. Greenberg, J., Bullard, K.A., & James, M.L. et al. (2002). Student comprehension of classification applications in a science education digital library. Research and Advanced Technology for Digital Libraries: Proceedings of the 6th European Conference, ECDL'02,. 560-567.
35. Han, J.W. & Guo, L. (2003). A shape-based image retrieval method using salient edges. Signal Processing: Image Communication, 18: 141-156.
36. Hartland-Fox, B., & Dalton, P. (2003). EVALUed-an evaluation model for e-library developments, Ariadne, 31. Retrieved June 1, 2003, from <http://www.ariadne.ac.uk/issue31/evalued/>.
37. Hill, L.L., Carver, L. et al. (2000). Alexandria Digital Library: user evaluation studies and system design. Journal of the American Society for Information Science and Technology, 51 (3): 246-259.
38. Hill, L.L., Dolin, R., et al. (1997). User evaluation: summary of the methodologies and results for the Alexander Digital Library, University of California at Santa Barbara. In C. Schwartz et. (Eds.) Proceedings of the American Society for Information Science (ASIS) Annual Meeting.,225-243, 369).
39. Huang, Z., Chung, W.Y., Ong, T.H. & Cheng, H.C. (2002). A graph-based recommendation system for digital library. Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries.,65-73.
40. Huxley, L. (2002). Renardus: following the Fox from project to service. Research and Advanced Technology for Digital Libraries: Proceedings of the Second European Conference, ECDL'02, 218-229.
41. Janssen, Olaf (2004) The European library user survey of Gabriel, Gateway to Europe's National Libraries. Retrieved September 2, 2004, from <http://www.bl.uk/gabriel/index.html>

42. Jewell, T. D. (1998). The ARL "investment in electronic resources" study: Final report to the council on library and information resources. Retrieved April 13, 2003, from <http://www.arl.org/stats/specproj/jewell.html>
43. Jones, M.L.W., Gay, G.K. & Rieger, R.H. (1999). Project Soup: comparing evaluations of digital collection efforts. *D-Lib Magazine*, 5 (11).
44. Jones, G.J.F., & Lam-Adesina, A.M. (2002). An investigation of mixed-media information retrieval. *Research and Advanced Technology for Digital Libraries: Proceedings of the Second European Conference, ECDL'02*, 463-478.
45. Jones, S., Cunningham, S.J. et al (2000). A transaction log analysis of a digital library. *International Journal of Digital Libraries*, 3:152-169.
46. Jones, S. & Paynter, G.W. (2002). Automatic extraction of document keyphrases for use in digital libraries: evaluation and application. *Journal of the American Society for Information Science & Technology*, 53 (8): 653- 677.
47. Kapidakis, S., Terzis, S., Sairameshi, J., & Nikolaou, C. et al. (1998). A management architecture for measuring and monitoring the behavior of digital libraries. *Research and Advanced Technology for Digital Libraries: Proceedings of Second European conference, ECDL '98, Heraklion, Crete, Cyprus, September 21-23, 1998*, 1513: 95-114.
48. Kassim, A.R.C., and Kochtanek, T.R. (2003). Designing, implementing, and evaluating an educational digital library resource. *Online Information Review*, 27 (30),160-168.
49. Kenney, A.R., Sharpe, L.H., & Berger, B. (1998). Illustrated book study: digital conversion requirements of printed illustration. *Research and Advanced Technology for Digital Libraries: Proceedings of the Second European Conference, ECDL'98, September 21-23, 1998, Heraklion, Crete, Greece*. 279-293.
50. Kengeri, R.; Seals, C.D.; Harley, H.D.; Reddy, H.P.; Fox, E.A., (1999). Usability study of digital libraries: ACM, IEEE-CS, NCSTRL, NDLTD. *International Journal on Digital Libraries*,.2, (2-3) 157-69
51. Khoo, C.S.G., Poo, D.C.C., & Toh, T.K. et al. (1998). E-referencer: a prototype expert system Web interface to online catalog. *Research and Advanced Technology for Digital Libraries: Proceedings of the Second European Conference, ECDL'98*,. 316-333.
52. Khoo, M. (2001). Ethnography, evaluation, and design as integrated strategies: a case study from WES. *ECDL'01; Proceedings of the 5th European Conference on Digital Libraries*, 263-274.
53. Kilker, J. & Gay. G. (1998) The social construction of a digital library: A case study examining implications for evaluation. *Information Technology and Libraries*, 17 (2), 60-70
54. Kwak, B.H., Jun, W., Gruenwald, L. & Hong, S.K. (2002). A study on the evaluation model for university libraries in digital environments. *Research and Advanced Technology for Digital Libraries: Proceedings of the 6th European Conference, ECDL'02*, 204-217.
55. Mackie-Mason, J.K., Riveros, J. F., Bonn, M. S., & Lougee, W.P. (1999). A Report on the PEAK Experiment. *Usage and economic behavior. D-Lib Magazine*, 5 (7/8).
56. Marchionini, G. & Crane, G. (1994). Evaluating hypermedia and learning: Methods and results from the Perseus Project. *ACM Transactions on Information Systems*, 12 (1), 5-34.
57. Marchionini, G. (2001). Evaluating digital libraries: a longitudinal & multifaceted view. *Library Trends*, 49 (2): 304-333.
58. Marchionini, G., Plaisant, C., & Komlodi, A. (2003). The people in digital libraries: multifaceted approaches to assessing needs and impact. In: Bishop, A., van House, N. A., Battenfield, B.P. (Eds.) *Digital Library Use Social Practice in Design and Evaluation*. MIT Press, 119-160.
59. Mellucci, M. (2004). Making digital libraries effective: Automatic generation of links for similarity search across hyper-textbooks. *Journal of the American Society for Information Science and Technology*, 55 (5) 414-430.
60. Melucci, M. & Orio, N. (2004) Combining melody processing and information retrieval techniques: Methodology, evaluation, and system implementation, *Journal of the American Society for Information Science and Technology*, 55 (12) 1058-1066)
61. Meyyappan, N. Foo, Schubert, and Chowdhury, G.G. (2004). Design and evaluation of a task-based digital library for the academic community. *Journal of Documentation*, 60 (4), 449-475.
62. Monopoli, M., Nicholas, D. Georgiou, P. and 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: new information perspectives*, 54 (2), 103-117.
63. Nicholas, D., Huntington, P. & Watkinson, A. (2003). Digital journals, Big Deals and online searching behaviour: a pilot study. *Aslib Proceedings: new information perspectives*. 55 (1): 84-109.
64. Orio, N. (2002). Alignment of performance with scores aimed at content-based music access and retrieval. *Research and advanced technology for digital libraries: Proceedings of 6th European conference, ECDL'02*, 479-492.
65. Paliouras, G., Papatheodorou, C., & Karkaletsis, V. et al. (1998). Learning user communities for improving the services of information providers. *Research and Advanced Technology for Digital Libraries: Proceedings of the Second European Conference, ECDL'98*, 316-333.
66. Park, S. (2000). Usability, user preferences, effectiveness, and user behaviors when searching individual and integrated full-text databases: implications for digital libraries. *Journal of the American Society for Information Science and Technology*, 51(5): 456-468.
67. Payette, S.D. & Rieger, O.Y. (1998). Supporting scholarly inquiry: Incorporating users in the design of digital library. *Journal of Academic Librarianship*, 24 (2): 121-129

68. Purcell, G.P., Rennels, G.D. & Shortliffe, E.H. (1997). Development and evaluation of a context-based document representation for searching the medical literature. *International Journal on Digital Libraries*, 1 (3): 288-296.
69. Rui, Y., Gupta, A., & Acero, A (2000). Automatically extracting highlights for TV baseball programs. *ACM Multimedia 2000*, 105-115.
70. Salamapasis, M., Tait, J. & Bloor, C. (1998). Evaluation of information-seeking performance in hypermedia digital libraries. *Interacting With Computers*, 10: 269-284.
71. Sanderson, M. & Crestani, F. (1998). Mixing and merging for spoken document retrieval. *Research and Advanced Technology for Digital Libraries: Proceedings of the Second European Conference, ECDL'98, September 21-23, 1998, Heraklion, Crete, Greece*. 397-407.
72. Saracevic, T., Zhang, Y., Li, Y. & Jeng, J. (2004). Moving images collection evaluation: Final report. School of Communication, Information and Library Studies, Rutgers University. Retrieved Sept. 23, 2004 from <http://www.scils.rutgers.edu/~miceval/>
73. Seadle, M. (2001). Project ethnography: an anthropological approach to assessing digital library services. *Library Trends*, 49 (2): 370-385.
74. Sumner, T.; Khoo, M.; Recker, M.; Marlino, M (2003). Understanding educator perceptions of "quality" in digital libraries. *Proceedings of the Third ACM/IEEE-CS Joint Conference on Digital Libraries*, 269-279.
75. Tolle, K.M., & Chen, H. (2000). Comparing noun phrasing techniques for use with medical digital library tools. *Journal of the American Society for Information Science*, 51 (4): 352-370.
76. Wesson, J., & Greunen, D.V. (2002). Visualization of usability data: measuring task efficiency. *Proceedings of the Conference of South African Institute of Computer Scientist and Information Technologists, SAICSIT 2002*, 11-18.
77. White, M.D. (2001). Digital reference services: framework for analysis and evaluation. *Library & Information Science Research*, 23 (3): 211-231.
78. Wildemuth, B.M. et al (2003). How fast is fast? Evaluating fast forward surrogates for digital video. *Proceedings of the Third ACM/IEEE-CS Joint Conference on Digital Libraries*, 221-230.
79. Wilson, R., Landoni, M., & Gibb, F. (2002). Guidelines for designing electronic books. *Research and Advanced Technology for Digital Libraries: Proceedings of the 6th European Conference, ECDL'02, September 16-18, 2002, Paris, France*, 47-60.
80. 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*, 53 (14): 1210-1223.