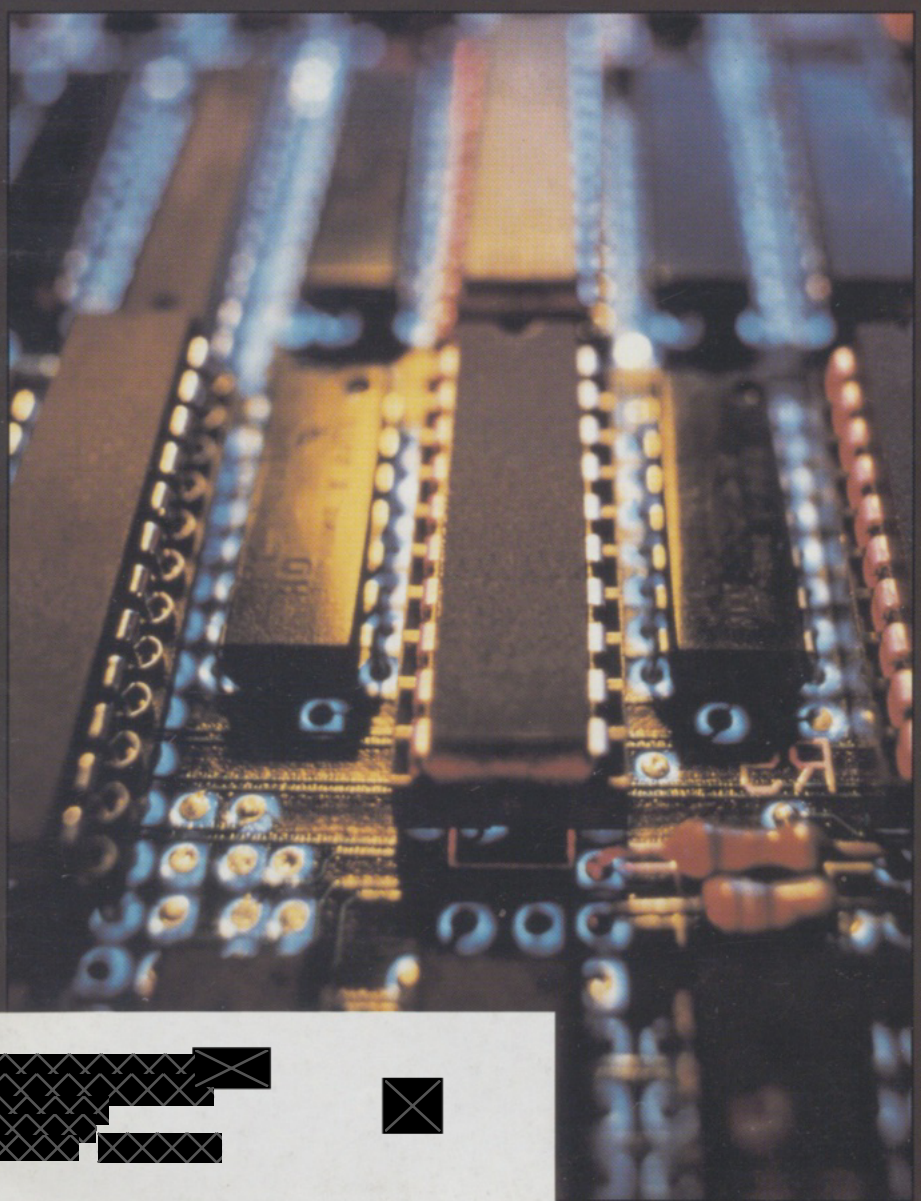
Information Technology AND Libraries

ISSN 0730-9295

Volume 20 Number 4

DECEMBER 2001



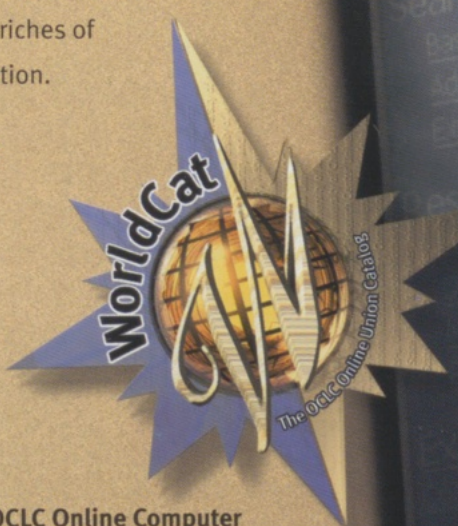
IN THIS ISSUE

- Library Web Proxy Use
- Personalized Information Environments

Local control global reach

The OCLC FirstSearch service delivers a world of information to your users, with a focus on your collection. Through unparalleled resources like WorldCat that are enhanced by searching tools and system features you can customize, FirstSearch makes finding information convenient for your users and affordable for your budget. Your users have access to the holdings of thousands of libraries worldwide, full text from more than 9,000 serials (including 3,400 electronic journals), and best of all, the riches of your library's collection.

www.oclc.org



OCLC Online Computer Library Center

OCLC—Furthering Access to the World's Information

Information Technology and Libraries (ISSN 0730-9295) is published quarterly in March, June, September, and December by the American Library Association, 50 E. Huron St., Chicago, IL 60611. It is the official publication of the Library and Information Technology Association, a division of the American Library Association. SUBSCRIPTIONS: Subscription price, $25, is included in membership dues. Nonmembers may subscribe for $50 per year in the U.S.; $55 in Canada, Mexico, Spain, and other PUAS countries; $60 in other foreign countries. Send subscription requests to *Information Technology and Libraries*, Subscription Department, American Library Association, 50 E. Huron St., Chicago, IL 60611; 800-545-2433; fax: (312) 944-2641; subscription@ala.org. Single copies, $15. Periodical-class postage paid at Chicago, Illinois, and at additional mailing offices. POSTMASTER: Send address changes to *Information Technology and Libraries*, 50 E. Huron St., Chicago, IL 60611.

Information Technology and Libraries publishes material related to all aspects of libraries and information technology, including digital libraries, metadata, authorization and authentication, electronic journals and electronic publishing, telecommunications, distributed systems and networks, computer security and intellectual property rights, technical standards, geographic information systems, desktop applications, online catalogs and bibliographic systems, optical information systems, software engineering, universal access to technology, futuristic forecasting, library consortia, vendor relations, and technology and the arts. *ITAL* welcomes unsolicited manuscripts. Submissions should follow the guidelines stated under "Instructions to Authors" in the June 2001 issue or on the LITA Web site at www.lita.org/ital/infoauth.htm. Manuscripts of feature articles, communications, and tutorials should be addressed to Dan Marmion, 221 Hesburgh Library, The University of Notre Dame, Notre Dame, IN 46556-5629. Books for review should be addressed to Tom Zillner, WILS, 464 Memorial Library, 728 State St., Madison, WI 53706. Software for review should be addressed to Andy Boze, Hesburgh Library, The University of Notre Dame, Notre Dame, IN 46556-5629.

As a matter of policy, *ITAL*, as the scholarly organ of LITA, does not review LITA publications.

PRODUCTION: ALA Production Services (Troy D. Linker, Kevin Heubusch; Ellie Barta-Moran, Angela Hanshaw, Kristen McKulski, and Karen Sheets), American Library Association, 50 E. Huron St., Chicago, IL 60611.

Publication of material in *Information Technology and Libraries* does not constitute official endorsement by LITA or the ALA.

Abstracted in *Computer & Information Systems, Computing Reviews, Information Science Abstracts, Library & Information Science Abstracts, Referativnyi Zhurnal, Nauchnaya i Tekhnicheskaya Informatsiya, Otdyelnyi Vypusk,* and *Science Abstracts Publications.* Indexed in *CompuMath Citation Index, Computer Contents, Computer Literature Index, Current Contents/Health Services Administration, Current Contents/Social Behavioral Sciences, Current Index to Journals in Education, Education, Library Literature, Magazine Index, NewSearch,* and *Social Sciences Citation Index.* Microfilm copies available to subscribers from University Microfilms, Ann Arbor, Michigan.

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1992.∞

Information Technology AND Libraries

Volume 20, Number 4 December 2001 ISSN 0730-9295

Abstracts and contents pages of recent issues of *ITAL* can be found on the LITA Web site at www.lita.org/ital/index.htm.

171 Editorial: The Open Source Movement and Libraries
DAN MARMION

172 Library Web Proxy Use Survey Results
PETER E. MURRAY

179 eThemes: An Internet Instructional Resource Service
FENG-KWEI WANG AND JOHN WEDMAN

185 A Personalized Information Environment for Digital Libraries
CHAMPA JAYAWARDANA, K. PRIYANTHA HEWAGAMAGE, AND MASAHITO HIRAKAWA

COMMUNICATIONS

198 Weaving Multimedia into the Collection: One Library's Journey
GAIL GOLDERMAN

202 Creating an Instant Messaging Reference System
JODY CONDIT FAGAN AND MICHELE CALLOWAY

TUTORIAL

213 The Service of Server-Side Includes
MICHELLE MACH

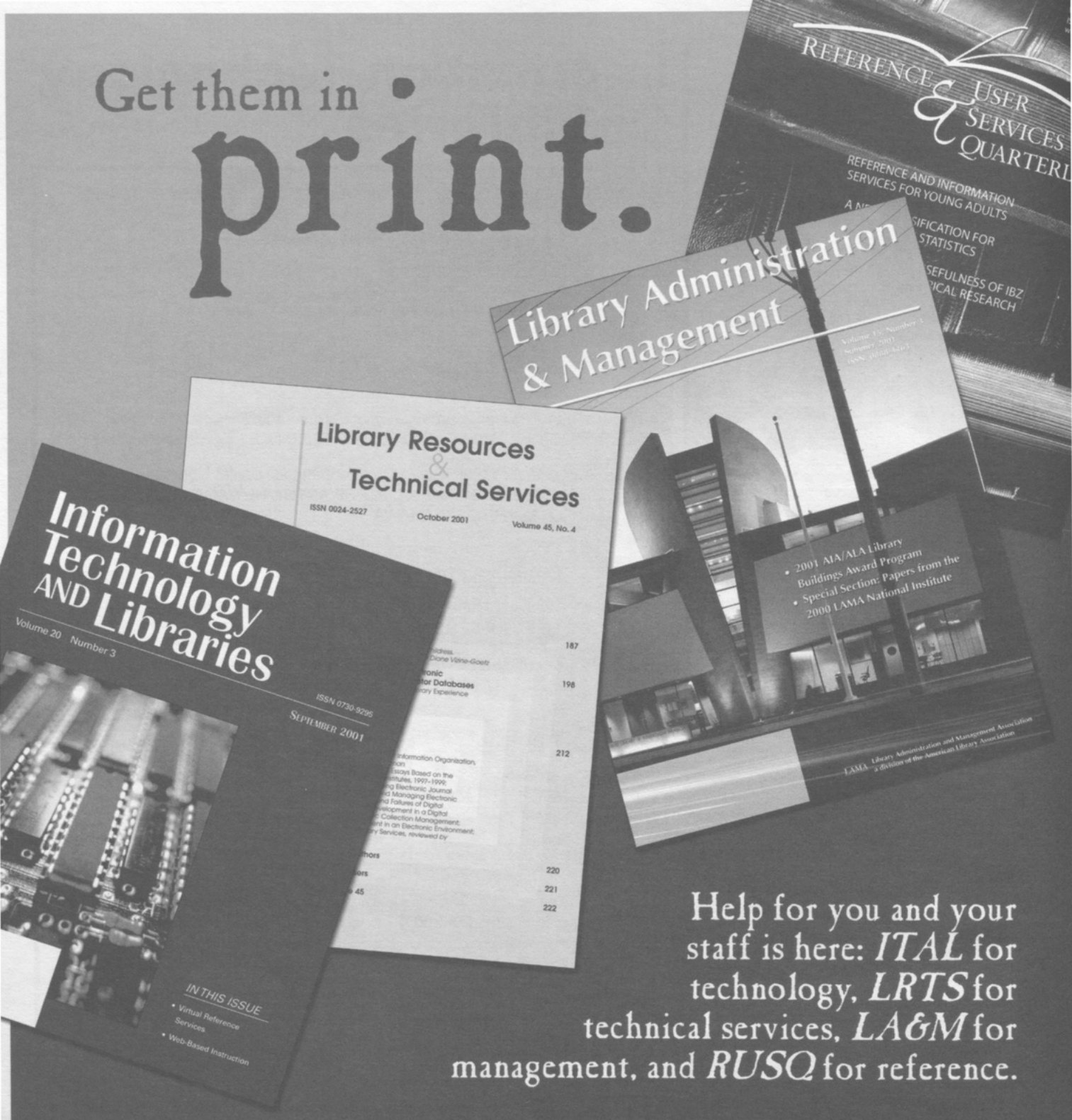
219 Index to Advertisers

220 Book Review

224 Index to Volume 20 (2001)

cover design Kevin Heubusch ● **interior design** Dianne M. Rooney

Get them in print.



Help for you and your staff is here: *ITAL* for technology, *LRTS* for technical services, *LA&M* for management, and *RUSQ* for reference.

To subscribe, call 1-800-545-2433, press 5.

ALCTS LAMA LITA Reference & User Services Association

Association for Library Collections and Development | Library Administration and Management Association
Library Information and Technology Association | Reference and User Services Association

Journal subscription is a member benefit. To join, call 1-800-545-2433, ext. 4299.

Editorial: The Open Source Movement and Libraries

Dan Marmion

I've been thinking about open source software[1] lately for a couple of reasons. One reason is that the March 2002 issue of *ITAL* will be a theme issue devoted to open source applications in libraries, guest edited by Jeremy Frumkin of the University of Arizona.

The other reason is a bit closer to home. I hired Eric Lease Morgan to be head of a new department at the Notre Dame Libraries called the Digital Access and Information Architecture Department (DAIAD, pronounced "die-ad," frontward or backward). Eric is something of a force in the open source movement and was the primary developer of the MyLibrary@NCState (http://my.lib.ncsu.edu) open source software created at North Carolina State University. (He is contributing an article to Jeremy's issue, incidentally.)

One of Eric's first responsibilities here at Notre Dame is to convert our library Web site to a database-driven technology and he wants to do it on an open source platform. My job is to decide whether to go along with that or to dictate a more conventional approach using something like Oracle and ColdFusion or Active Server Pages (ASP). Hence my immediate reason for devoting my time to the subject.

So what's the problem? Why not just do it? One reason is that despite a few well-publicized examples, there is not yet a whole lot of open source software available, and the movement is still fairly young and unproven. Another is that if we go this direction, we will have to commit some substantial human resources to do the required programming.

Does that mean we shouldn't do it? Why put my eggs in an unproven basket? Why commit valuable human resources to create more of this unproven software?

Maybe it isn't all that unproven. There are, as I mentioned, some fairly well known and widespread open source applications, such as the Apache Web server and the Linux operating system. My sense is that Linux is finding its way into more and more libraries these days, and that Apache is the leading Web server in colleges and universities. So to call those two unproven is really subject to debate. Likewise, open source products such as MySQL and Perl are being used in many applications.

While it's true that open source applications require programmers, the fact is that ColdFusion and ASP also require programming, and I would surely not want to run a critical library application based on Oracle without someone with extensive Oracle knowledge and experience on hand. Thus the human resources argument is also less compelling, especially when you realize that not only do you have to invest dollars in people to make those commercial products works properly, you also have a substantial cost in licenses for them that you don't have with open source software.

Another thing about open source software, in theory at least, is that if somebody creates something that is truly good and makes it freely available, a community of users will emerge and collectively support, maintain, and improve upon it. That's an intriguing, almost '60s concept that is also unproven. Certainly for Apache and Linux it seems to be working. On the other hand, a brilliant piece of work like jake (the Jointly Administered Knowledge Environment) is, ultimately, a failure, because even though the software is great, the user community did not keep current the data upon which jake depends.

One last point to make regarding open source applications is that there does seem to be a potential for something truly beneficial to libraries. MyLibrary@NCState is a good example. Many libraries do not have the resources to create software from scratch, but because there is no ongoing license fee for open source software, perhaps they can provide the resources needed to implement it at their library. There are other library-related open source projects underway, such as OpenBook, a library automation system developed by Technology Resource Foundation (www.trfoundation.org) for smaller public and school libraries.

On a different note, I call your attention to the lead article in this issue by Peter Murray, the first-ever winner of the LITA/Endeavor Student Writing Award. I was on the judging panel and can attest that we had many good papers submitted, so I'm pleased to know there are some burgeoning young writers out there. This is a new annual award—see www.lita.org/a&s/awards.htm#writing for more details.

Note

1. I assume most readers of this journal know what the phrase "open source" means, but just in case (and greatly simplified) it refers to software that is freely given away, along with source code, according to certain explicit distribution terms that allow it to be just as freely redistributed (although not necessarily for free). The concept actually goes back to the GNU Project (www.gnu.org/gnu/thegnuproject.html) that began in the mid-1980s. Back then the preferred term was "free software," meaning that users of the software were free to modify it and redistribute it. Most open source software is distributed under the terms of the GNU General Public License (GPL), ensuring that it retains this "free" status.

Dan Marmion (dmarmion@nd.edu) is Associate Director for Information Systems and Access at the University of Notre Dame Libraries, Notre Dame, Indiana.

Library Web Proxy Use Survey Results

Peter E. Murray

[Editor's note: The following article is winner of the first annual LITA/Endeavor Student Writing Award.]

Libraries face many policy and technological difficulties in providing remote access to databases, making effective use of Internet bandwidth, controlling where patrons browse on public computers, and gathering statistics on usage. Some libraries have chosen to employ proxy Web servers to solve these problems. This paper outlines the use of proxy Web servers by libraries to address these areas and documents survey results on their use in libraries.

In its most general definition, a proxy server is "[a]n intermediary server that accepts requests from clients and forwards them to other . . . servers."[1] In the general form of this definition, a proxy server may act as an intermediary for one of many Internet protocols (such as HTTP, FTP, Telnet, NNTP, and others). This paper focuses on the application of proxy Web servers in general, and specifically their use in library networks to solve library-specific problems.

There are four reasons a library may install a proxy Web server: to enable access to resources by patrons outside a library's network, to filter Web requests or responses on public stations, to conserve bandwidth and improve response time, and to gather statistics on Web usage. In order to identify why libraries had installed proxy servers and what proxy server software was used, a survey was prepared and distributed in November 2000. Additional questions were asked about the documentation provided to patrons about how to use a proxy server and any privacy statements about the disposition of proxy server log files. This paper will discuss the information gathered in the survey.

The survey was posted to the following electronic mailing lists: Web4Lib@sunsite.berkeley.edu, PACS-L@listserv.uh.edu, LITA-L@ala1.ala.org, teknoids@listserv.law.cornell.edu, SYSLIB-L@listserv.acsu.buffalo.edu, LIS-LINK@mailbase.ac.uk, and PUBLIB-NET@sunsite.berkeley.edu. E-mail messages were also sent to the attendees of the second and third LITA Regional Institutes on Web Proxy Servers and Authentication. A copy of the survey appears at the end of this paper in appendix A.

Between November 16 and December 22, 2000, eighty-five libraries responded to the call for participation in the survey. Respondents had the option to identify themselves and their institution; that information was used for follow-up information, much of which is discussed in this paper. A breakdown of library types is shown in table 1. Three of the responding libraries were not using proxy servers at this time.

Peter E. Murray (peter@pandc.org) is a graduate student in the MLS program, Simmons College, Boston.

Proxy for Remote Resource Access

By far the most frequent reason for libraries to use Web proxy servers is to enable off-network patrons to access vendor-provided resources. These resources are typically restricted to a particular institution's subscription by one of two methods: vendor-supplied username/password authentication or network address recognition. Although alternate methods exist for the purpose of authenticating access to resources (vendor-supplied scripts and referer-URL authentication, for example), these methods do not enjoy widespread implementation.

The problems of distributing vendor-supplied usernames and passwords to a community of users and keeping them secure are well known; such a method allows access to a resource from nearly any Web browser in the world. A single username and password supplied by the vendor can be distributed to individual patrons or posted on an internal Web site. In addition to the packet sniffing problems discussed in Cole, there is nothing inherent in this scheme that prevents the password from being given to patrons outside the institution's community.[2] The same problem exists for usernames and passwords distributed to individuals, although it is easier to identify abuses with a single user's password and cancel access for a username which has been compromised.

Alternatively, using network address recognition for authentication is very convenient for users on an institution's network because the only requirement for access to the resource is using a machine in the proper IP address range; no password is required. It is harder for unauthorized users to gain access to a resource because an unauthorized individual must be using a computer physically attached to the institution's network. This same physical requirement, though, prevents legitimate remote users from accessing the resource. Since resources are restricted to an institution's network addresses, placing a proxy server within that range of network addresses enables off-network users to appear to come from within the network to database vendors.

Seventy-two libraries responded that they use proxies for remote resource access; the breakdown of software packages is shown in table 2. EZproxy is the most popular, followed by Innovative Interfaces' proprietary Web Access Management. One library uses a combination of Apache and EZproxy and another uses Microsoft Proxy and Netscape Proxy. One library uses more than one proxy

server, but did not elaborate. In addition, a public library was using Web Access Management but is planning to install EZproxy; no proxy server is in place at this time.

Some proxy servers for remote resource access require users to configure their browsers to take advantage of the proxy service. Several libraries supplied URLs to documentation explaining this reconfiguration; these were the best in the author's opinion:

- Central Michigan University, http://ocls.cmich.edu/remoteindex.htm
- University of Waterloo, www.tug.uwaterloo.ca/proxy
- Tarleton State University, www.tarleton.edu/~library/proxy/instructions1.htm

An academic library noted that some patrons must be told to remove the proxy server setting automatically configured by a cable TV-based Internet service provider (ISP) before using the institution's proxy server.

One specialized type of proxy server that doesn't require the user to modify the browser configuration is a "rewriting" proxy Web server. Rewriting proxy servers transform the HTML pages from vendor databases such that URLs on the page are rewritten to point back to the proxy server. Several academic libraries have created their own rewriting proxy servers, often using existing free proxy servers as the basis. One example is the library at the University of Calgary as reported by Eric Tull. Another example, although not mentioned in the survey, is Brown University's implementation.[3] EZproxy is an example of a commercial rewriting proxy server.

One academic library makes its authenticating proxy server available to other campus departments besides the library, but at this time only the main library and the law school library are using the service. The same library is considering expanding the proxy server use beyond remote resource access to bandwidth conservation.

The library at SUNY–Oswego has set up its EZproxy server to allow access to IP-restricted resources on campus Web servers in addition to vendor provided databases. The types of resources made available in this fashion are campus network maps, faculty committee documents, and software for which a site-license for home access has been negotiated. The systems librarian and campus webmaster seek out other campus Web information to make available using this mechanism.

Proxy for Filtering

Since proxy servers are in the middle of the transaction between the client and the server, the proxy can examine the HTTP request from the client or response from the server. Based on programming or configuration parameters, the proxy can modify the client's request before delivering it to the server or modify the server's response before returning it to the client. Proxies performing this kind of action are referred to as "filtering" proxies. The modification can affect headers in the HTTP transaction or change the HTML files or other data returned by the Web server.

Eighteen libraries use Web proxies for filtering Internet stations. Software packages used are shown in table 3. There are many reasons for a library to install a filtering proxy. Most libraries use filtering proxies for "allow lists" that permit access to only specific Web sites and "deny lists" that prevent access to specified Web sites. Seven libraries use proxies for deny lists, three for allow lists, and seven use proxies in their libraries for both reasons.

Dan Lester from Boise State University (Idaho) included in his survey response details about how his library uses WinProxy to deny access to Web-based e-mail, gaming, and chat sites.[4] In addition, Lester edits a list of known Web sites with Web-based e-mail, gaming, and chat functions; libraries are encouraged to submit additions and corrections to the list.[5]

Table 1. Breakdown of Respondents to Survey

Library Type	Count
Academic	68
Public	10
Other/None	7

Table 2. Software Packages Used for Remote Resource Access

Software	Count
EZproxy	29
Web Access Management	14
Squid	9
Apache	6
Microsoft Proxy	4
Netscape Proxy	3
Homegrown	2
Delegate	1
Netware BorderManager	1
Other—More than one	3

Table 3. Software Packages Used for Filtering

Software	Count
Microsoft Proxy	5
Squid	3
WinProxy	3
Apache	1
Bess	1
Netscape Proxy	1
Novell BorderManager	1
WebManager	1
Other—More than one	2

A number of libraries are using plug-ins to Microsoft's Proxy Server to do various forms of filtering. One public library is using the CyberPatrol plug-in to filter content on library stations in addition to using a proxy server. Another library is using a plug-in called Websense to provide optional filtering of sexual materials from patrons. A multi-type library consortium is using the SmartFilter plug-in for Microsoft Proxy Server.

The University of Waterloo (Canada) Library forces public library stations to use a proxy server. A router between the public library network and the campus network restricts HTTP requests to just the proxy server (in addition to other network restrictions). Stations must therefore use the proxy server to access Web resources. The proxy server includes allow/deny directives denying access to Web-based e-mail services.

A public library uses a proxy to filter advertisements out of responses sent back from servers. Other uses for filtering proxy servers not reported in survey responses are to scan files for viruses before they are received, or prevent certain file types (movies, audio files, executables, and others) from being downloaded.

Proxy for Bandwidth Conservation

Bandwidth conservation is typically the reason that organizations other than libraries install a proxy server. The goals are twofold: to reduce the amount of traffic crossing an Internet connection, and to reduce the amount of time it takes for a Web browser to receive content. A caching proxy server does this by storing Web requests and responses for use by subsequent requests. The caching proxy is a server on the local network and browsers are configured to contact the proxy server for every Web request. A browser's first request for an entity (HTML page, graphic, and so on) may take slightly longer to be received because of the added processing required by the proxy server. However, subsequent requests for the same entity by the same browser or other browsers using the same caching proxy will be served faster because the proxy server on the local network can respond more quickly to subsequent requests without traversing the Internet connection.

Nineteen libraries use proxy servers for bandwidth conservation; the proxy servers used by libraries are listed in table 4.

In a related response, one library reported that it employs a proxy server to reduce the load on an old, proprietary Web server that cannot be replaced for several months. Because requests come through the proxy server first, the proxy server rather than the old Web server can handle requests for static content such as graphics and HTML files that don't regularly change. This implementation of a proxy server is called a "reverse" proxy server.[6]

A number of the responses to this question did not specify traditional software proxies, but rather interception proxies. An interception proxy requires no changes to Web clients; it operates instead at a network infrastructure level. Network routers and switches redirect HTTP requests to the interception proxy transparently where the proxy returns the response out of its cache or contacts the Web server for response on behalf of the client. Of the responses to this question, CacheFlow 5000, Cisco's cache engines, Cobalt, Novell BorderManager, and Novell Internet Router are interception proxies. (Cobalt and Novell Internet Router can also be noninterception, traditional proxies.)

Table 4. Software Packages Used for Bandwidth Conservation

Software	Count
Microsoft Proxy	6
Squid	3
Netscape Proxy	2
WinProxy	2
CacheFlow 5000	1
Cisco cache engine	1
Cobalt	1
Novell BorderManager	1
Novell Internet Router	1
Other—More than one	1

Proxy for Gathering Statistics

Another side effect of the interaction among the client, proxy, and Web server is that the proxy server will contain log entries for all of the accesses by the client. By configuring OPAC stations to use a proxy server for requests to vendor databases, a library can get a rough gauge of database usage by examining the log files of the proxy server. Thirteen libraries use proxy servers to gather statistics on Web requests; the proxy servers used are listed in table 5.

The same proxy server used for filtering or bandwidth conservation can be used for gathering statistics. The log files of a proxy server for remote resource access, when correlated with information about what percentage of resource accesses was aided by that proxy server, can also be used to report on resource access.

In the survey, libraries were asked to identify what applications they used to create statistical reports from the proxy server logs. The applications listed were (each application was mentioned once unless otherwise noted):

- WebTrends (4)
- Analog (2)
- Program developed in-house (2)
- HttpAnalyze
- Websense
- Software built into the Novell Internet Caching System proxy
- Excel manipulation
- MS Access for most; WebTrends LogAnalyzer for some

One academic library uses a homegrown counter on links to databases; tracking the number of times the link is accessed gives the library an idea of how often databases are used. Another academic library periodically uses Squid (a full-featured proxy cache) for in-house activity views. Only a limited, random number of sessions are examined.

Table 5. Software Packages Used for Statistics

Software	Count
Microsoft Proxy	5
EZproxy	2
Squid	2
Netscape Proxy	1
Novell Internet Router	1
WinProxy	1
Other—More than one	1

The survey response for a multi-type library consortium included a comment that the institution's proxy Web server does not collect statistics. There are special programs running on the system to delete personal data collected by the proxy server that cannot be disabled in order to protect users' privacy.

Other Results

Respondents were asked if the library publishes a privacy policy regarding the use of proxy server log files. One library includes a statement regarding data collection and use on the proxy server login page:

> This information is collected under the Freedom of Information and Protection of Privacy Act. It is required to verify the identification of the researcher and to authorize access to the database. If you have any questions about the collection or use of this information, please contact the Public Services Systems Librarian.[7]

Boise State University used proxy logs in the arrest of a patron who was viewing child pornography. Four days of logs were given to law enforcement personnel. There was no issue of needing a court order to get the data as it was the library that filed the complaint.

An academic library is installing a Virtual Private Network (VPN) for off-campus clients on DSL and cable modem connections to access resources restricted by IP address. VPNs extend the institution's IP addresses to machines outside the local area network by tunneling network traffic through the general Internet. As such, VPNs work at a network infrastructure layer below that of a Web proxy server, but can accomplish the same result as a Web proxy server for remote resource access.

Conclusions

It comes as no surprise that proxy servers are most often used for remote resource access. Attendees at the four LITA Proxy Web Servers and Authentication workshops stated that learning about remote resource access is their primary reason for attending. In addition, the most common reason attendees have installed proxy servers prior to attending the workshop is to provide remote resource access.

One of the surprising outcomes from the survey is the use of interception proxies by libraries and institutions. Almost one quarter of the responses to the "Proxies for Bandwidth Conservation" question came from libraries using interception proxies. Although public libraries only made up 10 percent of the survey responses, two of four

interception proxy installations are in public libraries. The great benefit of this type of proxy server is that it performs the proxying function without requiring modifications to the browser configuration. Consequently, an institution can use one of these proxy servers for filtering, bandwidth conservation, or statistics without a visible impact on the user. Interception proxies cannot be used to enable remote access to databases.

Interception proxies have caused problems for libraries in the past, however, especially when installed by ISPs. The interception proxy changes the IP address of the client making the request to the IP address of the interception proxy. As a result, the database vendor detects the resource request as coming from an IP address outside the range of the institution's IP addresses, and the database vendor will deny access. The interception proxy can typically be bypassed for specified Web servers, but the library must submit a list of database vendor server addresses to the ISP for inclusion in the interception proxy's exception table.

Another surprising result was the lack of privacy statements for the log files of proxy servers. The proxy server's log files are particularly sensitive because the proxy will record all accesses by a client. It is possible to reconstruct the actions and perhaps even the individual searches of a user by analyzing the log files. As users become more sensitive of their personal information being misused in e-commerce transactions, patrons may begin to question the security of their information in the library.

View of the Future

Proxy Web servers are beginning to gain acceptance in library networks. Although the proxy Web function was included in the first Web server and specified in the first version of the HTTP protocol specification, widespread use of proxy Web servers for library issues is only now being seen. Some libraries have used proxy Web servers to solve remote access problems, but there are other ways they can be exploited. What institution doesn't want to offer access to subscription databases to patrons without regard to where the patron is physically located? Or fulfill a policy directive to control the types of material accessed at all or a subset of public workstations? Or extend the life of an expensive connection to the Internet or a network connection between branches by reducing repetitive network traffic? Or be assured that the money spent for subscription databases is effectively used? Proxy servers, ranging from freely available software packages to vendor-supplied turnkey systems, can solve the technical and policy requirements of libraries.

Although proxy Web servers provide a means to address a number of important library issues, in time one hopes that better alternatives will evolve to meet some library needs. Remote resource access is the most popular use of proxies in libraries today, but it represents a cumbersome and inefficient way to solve the remote resource access problem. These proxies can be complicated to set up, both for the user and the library, and cause content for the remote resource user to cross an institution's Internet connection twice. Proxies for statistics give the library just a crude measurement of the use of databases, representing the number of actual HTTP transactions to the database vendor and not the number of searches or records retrieved by patron search sessions. The adoption of standards for interinstitution access control and for the gathering of common statistics should reduce the reliance of proxy servers for these uses. Proxy servers, however, will likely remain a useful way to reduce bandwidth consumption and implement filtering requirements for some time to come.

References

1. Ari Luotonen, *Web Proxy Servers* (Upper Saddle River, N.J.: Prentice Hall, 1998): 4.
2. Timothy W. Cole, "Using Bluestorm for Web User Authentication and Access Control of Library Resources," *Library Hi Tech* 14, no. 1–2 (1997): 62–63.
3. Richard Goerwitz, Pass-through Proxying as a Solution to the Off-Campus Web-Access Problem. Accessed Mar. 28, 2001, www.brown.edu/Facilities/CIS/Network_Services/libproxy.
4. Dan Lester, Blocking Chat, E-Mail, and Game Playing at Library Internet Workstations. Accessed Mar. 28, 2001, www.riverofdata.com/tools/blocking.htm.
5. Dan Lester, Proxy Server Blacklist for Chat, Web E-Mail, and Game Playing Sites. Accessed Mar. 28, 2001, www.riverofdata.com/tools/blacklist.htm.
6. I. Cooper, I. Melve, and G. Tomlinson, Internet Web Replication and Caching Taxonomy. Accessed Mar. 28, 2001, www.rfc-editor.org/rfc/rfc3040.txt.
7. Eric Tull to Peter Murray, personal communication, Nov. 16, 2000.

Appendix A. Web Proxy Use Survey

Introduction

The purpose of this survey is to gather information about the use of Web proxy servers in libraries. Responses to this survey may be used in future presentations and publications. This survey is being conducted in conjunction with the LITA Regional Institute "Proxy Web Servers and Authentication."

The survey consists of a group of questions regarding your library's use or planned use of proxy servers to solve one or more of these problems:

1. bandwidth conservation;
2. gathering statistics;
3. filtering; and
4. remote resource access.

For each area, you will be asked which proxy server you are using or plan to use to solve a particular problem along with followup questions specific to each problem. You can skip an entire major section if it does not apply to your institution.

Proxies for Bandwidth Conservation

Proxies for bandwidth conservation are used to reduce latency (the average time it takes for Web pages to display due to network delays) and network traffic on your network segments. If you do not use a proxy server for bandwidth conservation, you can skip to the next section.

What proxy server does your institution use for bandwidth conservation?
___ Apache
___ Squid
___ Delegate
___ Microsoft Proxy Server
___ Netscape Proxy Server
___ WinProxy
___ WebManager (Sagebrush)
___ Homegrown software
___ Other (please specify): _____
___ More than one software package
___ No response

Proxies for Statistics

By forcing all Web requests for Web resources through a proxy server, an institution can use the log files from the proxy server to gather statistics on what resources are used. If you do not use a proxy server for statistics, you can skip to the next section.

What proxy server does your institution use for statistics?
___ Apache
___ Squid
___ Delegate
___ Microsoft Proxy Server
___ Netscape Proxy Server
___ WinProxy
___ Obvia
___ WebManager (Sagebrush)
___ Homegrown software
___ Other (please specify): _____
___ More than one software package
___ No response

What statistics program do you use to process the log files? _____

Proxies for Filtering

Some proxy servers can be configured to allow access to only specific Web sites (also known as "allow lists") or deny access for specified Web sites (also known as "deny lists"). If you do not use a proxy server for filtering, you can skip to the next section.

What proxy server does your institution use for filtering?
___ Apache
___ Squid
___ Delegate
___ Microsoft Proxy Server
___ Netscape Proxy Server
___ WinProxy
___ WebManager (Sagebrush)
___ Homegrown software
___ Other (please specify): _____
___ More than one software package
___ No response

For what purpose do you use a filtering proxy server?
___ Allow lists
___ Deny lists
___ Both
___ Neither
___ No response

Does your institution use proxy servers for other types of filtering (such as removing cookies, blocking advertisements, and virus scanning). If so, please describe: _____
_____
_____
_____

Proxies for Remote Resource Access

Some proxy servers can be used to provide remote access to vendor databases from computers outside your institution's network. If you do not use a proxy server for remote resource access, you can skip to the next section.

What proxy server does your institution use for remote resource access?
___ Apache
___ Squid
___ Delegate
___ Microsoft Proxy Server
___ Netscape Proxy Server
___ WinProxy
___ Ezproxy
___ Obvia
___ WebManager (Sagebrush)
___ Remote Patron Authentication from epixtech
___ Web Access Management (WAM) from Innovative Interfaces
___ Homegrown software
___ Other (please specify): _____
___ More than one software package
___ No response

If your proxy server requires users to make changes to their browser configuration, do you provide instructions on your Web site?
___ Yes; please list URL: _____
___ No
___ No response

Other Uses of Proxy Servers

Is your library using proxy servers for another reason? If so, please describe: _____
_____

Does your library publish a statement regarding the type and amount of information collected by the proxy server, and the use and disposition of proxy server log files?
___ Yes; please list URL if published on your Web site:
_____
___ No
___ No response

Demographics

Type of library
___ Academic
___ Public
___ School
___ Corporate
___ Special
___ Other (please specify): _____
___ No response

Where did you hear about this survey?
___ Posting on Web4Lib
___ Posting on PACS-L
___ Posting on LITA-L
___ Posting on Teknoids
___ Posting on SYSLIB-L
___ Posting on LIS-LINK
___ Posting on PUBLIB-NET
___ E-mail received because I attended a LITA Regional Institute
___ Word-of-mouth
___ Nick Moore's column in *Library Review*
___ Other (please specify): _____
___ No response

Institution name: _____

Your name: _____

Your e-mail address: _____

Your name, institution, and e-mail address are optional. This information will be used for followup to survey answers, and will not be published or disclosed to third parties without your consent.

eThemes: An Internet Instructional Resource Service

Feng-Kwei Wang
and John Wedman

Many studies indicate that computer technology has the potential of changing teachers' pedagogical practices. This paper describes a major initiative to support teachers in integrating Internet resources into the instructional process while shifting their instruction to a more constructivist approach. Referred to as eThemes, this service accepts requests from teachers, finds Web sites that meet the requirements specified in the requests, and creates an archive of quality Internet resources for easy access and searching. It minimizes teachers' resource-seeking time and maximizes their resource-using time in their instruction to enhance teaching practice and student performance.

> . . . the real promise of technology in education lies in its potential to facilitate fundamental, qualitative changes in the nature of teaching and learning.[1]

Is the "real promise" of technology in education being fulfilled? The literature indicates that computer technology has the potential of changing teachers' pedagogical practices.[2] However, such changes are not automatic; teachers must be supported as they transform their teaching to take advantage of computer technology and digital information resources.

Supporting teachers in the process of changing their instructional practice is a multifaceted effort. Performance expectations must be communicated; teaching and learning resources must be available; rewards and incentives must be provided; and knowledge and skills must be developed.[3] This paper describes a major initiative to support teachers in integrating Internet resources into the instructional process while shifting their instruction to a more constructivist approach.[4] The article begins with the background of the project followed by the challenges being addressed; specific attention is given to improving resource availability while minimizing the demand on teachers' time. Evaluative data regarding the project are included, providing a context for future efforts to facilitate change in teacher practices.

Background

Like educators across the nation, Missouri teachers are working to change their instructional practices to incorporate technology in their teaching, the goal being to improve student performance as measured by state performance assessments. In 1997, MOREnet, the primary Internet Service Provider (ISP) for K–12, higher education, and public libraries in Missouri, launched the Multimedia Interactive Networked Technologies (MINTs) project. Consisting of selected classrooms in six schools in the greater St. Louis area, MINTs was designed to raise student performance and change teacher practice through the infusion of technology into the classroom. By placing state-of-the-art technology in the hands of teachers and students and eliminating the technology barriers traditionally experienced by schools, this project encouraged a new way of teaching—a way that engaged students by making resources available in a learning environment that fostered cooperation, collaboration, problem solving, and higher order thinking skills.

The accomplishments of the MINTs project received national recognition; the project was awarded the 1999 Computerworld Award from the Smithsonian Institution. At the state level, the project was considered so successful that it prompted the Missouri Commissioner of Education to launch a statewide initiative to change the way Missouri was thinking about educating its K–12 students. This new initiative was dubbed eMINTS—Enhancing Missouri's Instructional Networked Teaching Strategies. In the fall of 2000, the eMINTS project involved 188 third and fourth grade teachers in eighty-eight school districts throughout Missouri. Each teacher's classroom had a teacher workstation with desktop video, a SmartBoard and projection system, student workstations with a 2:1 student-to-computer ratio, and a fast Internet connection. As described below, this significant investment in hardware and networking was necessary, but not sufficient, to achieve the dual goals of changing teacher practices and improving student performance.

eMINTS Challenges

Like any large-scale, systemic change effort, eMINTS was confronted with many interrelated challenges, including:

1. communicating clear expectations and feedback to the teachers,
2. providing responsive technical support,
3. developing teachers' knowledge and skills, and
4. making high-quality teaching and learning resources accessible.

A brief overview of how these challenges are being addressed is offered in the following paragraphs, as is an in-depth look at the complexities of providing high-quality

Feng-Kwei Wang (wangfeng@missouri.edu) is Assistant Professor and **John Wedman** (wedmanj@missouri.edu) is Director of the School of Information Science and Learning Technologies, University of Missouri–Columbia.

teaching and learning resources to eMINTS teachers and their students.

The eMINTS teachers are expected to use technology to support teaching and learning consistent with a constructivist framework. The project adopted *Learning with Technology: A Constructivist Approach* as a means of setting and communicating performance expectations for teachers.[5] Online discussions, periodic training sessions, and individual consultation are used to help communicate the expectations.

Technical support for the project is available at two levels. At the state level, MOREnet (www.more.net) provides the Internet service to the participating school districts. At the local level, the school district's technical personnel are available to resolve technical problems in the classrooms. MOREnet also provides technical training to the school district technical staff, with special attention given to the technology used in the eMINTS classrooms.

EMINTS teachers' knowledge and skills (technical and pedagogical) are developed through a range of activities consisting of over one hundred hours of professional development per year. This training is supplemented with one-on-one consulting provided by MOREnet staff distributed across the state. Additionally, informal, teacher-to-teacher training is common throughout the eMINTS project.

The technology resources available to eMINTS teachers included software, hardware, and networking. A word processor, spreadsheet, presentation tool, browser, and keyboarding tutorial was installed on modern workstations (two students per computer). The software was selected to decrease the likelihood that drill and practice applications would dominate. EMINTS networking consists of a T-1 line to each participating school district, with the district being responsible for the wide-area and local-area networking to the building and classrooms.

One of the lessons learned from the MINTS project was that teachers were quick to integrate Internet resources in their instruction, primarily in the form of Web sites related to teaching units. However, we also learned the technology infrastructure for tapping into Internet resources was inadequate in terms of sustaining changes in teacher practice. While the teachers valued Internet resources for supporting teaching and learning, three resource-access barriers were identified: (1) limited search skills on the part of teachers; (2) lack of teachers' time to find and maintain resources (such as "fixing" broken resource links); and (3) minimal resource sharing among eMINTS teachers.

These three barriers are not unique to the eMINTS project. Balas and Snow pointed out that without mastering several Internet search techniques, information seekers may get unexpected or inconsistent search results.[6] The complexity of the resource finding task is exacerbated by the fact that Web content increases by more than 3.2 million new pages and more than 715,000 images every twenty-four hours.[7] To illustrate this point, a keyword search of "Missouri History" using Yahoo! returned nineteen matched categories and 204 matched sites in early November 2000. Four weeks later, the same search returned twenty matched categories and 222 matched sites. Information overload and content reliability on the Internet become untamed problems for educational professionals who utilize Internet resources.[8] In fact, according to an Internet user survey conducted by the Georgia Institute of Technology, Internet users ranked broken links as the second most frequently cited problem on the Internet.[9] The critical point here is that maintaining a current list of high-quality Internet resources is a difficult and time-consuming task, which most teachers are too busy to carry out.[10]

Even if all eMINTS teachers had the required search skills and the time to apply them, the practice of individual teachers finding and maintaining an inventory of Internet resources is grossly inefficient. For example, in the eMINTS project, there are approximately ninety-five third-grade teachers, all of whom use the same state-mandated standards (www.dese.state.mo.us/standards) to guide their instruction. A common component of the third grade curriculum is Missouri history. Little is gained by having each third grade teacher individually find, review, and catalog Internet resources for Missouri history.

The skill, time, and resource-sharing barriers described above were identified through focus group interviews with MINTS and eMINTS teachers and resulted in the generation of a solution considered to be quickly scalable to other grade levels and transportable to other states.

The Solution—Internet Instructional Resource Service

The need to minimize teachers' resource-seeking time and to maximize their resource-using time gave rise to a new venture—the Internet instructional resource service. Referred to as "eThemes," this venture is described below from service, process, and technical perspectives.

eThemes Service Perspective

As a service, eThemes accepts Internet resource requests from eMINTS teachers. Graduate assistants with library science and educational technology backgrounds serve as Internet resource scouts for the eMINTS teachers. The scouts find Web sites that meet the requirements specified

in the request, and distribute their search results in a way that is accessible to all teachers (including teachers not involved in the eMINTS project). Since its launch in late 1999, the eThemes service has fulfilled nearly two hundred resource requests, cross-referencing approximately one thousand Web sites appropriate for third and fourth grade learners to the Missouri student performance standards. The resource scouts are trained to use many Web search engines and online databases, thus increasing the likelihood that helpful resources are found. A "link checker" is used daily to find broken links, enabling the resource scouts to isolate and resolve broken link problems before they appear in the classroom. Periodically, the existing resources are reviewed and updated with the most current Internet Web sites available.

eThemes Process Perspective

As a process, eThemes is a means of finding, organizing, and managing Internet resources for teaching. As depicted in figure 1, the eThemes process starts with a request by an eMINTS teacher. Once the resource request is received, a resource scout is assigned to search Web sites that match the criteria of the request. Then, an eThemes resource record that contains the information of matched Internet resources is created and published in the eThemes database. Meanwhile, an e-mail notification is sent to the requester who then reviews the resources found by the scout and potentially provides feedback for the resource scout to consider in revising the search. Once a resource record is published, other teachers can access and search the database for resource records that can be used in their instruction.

eThemes Technical Perspective

As a database, eThemes is a Web-based system that uses Oracle database software to:

1. collect and catalog resource requests from eMINTS teachers;
2. organize and display the Web sites found by the resource scouts that match the resource request; and
3. provide efficient search processes so that all teachers can quickly locate high-quality Internet resources linked to the Missouri student standards.

These three functions are supported by the three primary modules in the eThemes database—Request Fulfillment, Resource Search, and Resource Administration (see figure 2 for eThemes' main menu).

The Request Fulfillment module keeps track of the requests that eMINTS teachers make for resource scouting services. A request record contains information about the content and format of the Web sites the requester wants a scout to find. Figure 3 shows a request example. The output of a request is a resource record (simply called "resource" in eThemes). An eThemes resource is a collection of Web sites and information describing the scope of the collection. Figure 4 shows a resource example.

The Resource Search module allows a teacher to use keywords, names, grade levels, and Missouri education standards to help teachers quickly find Internet resources they want. The search results can be sorted by

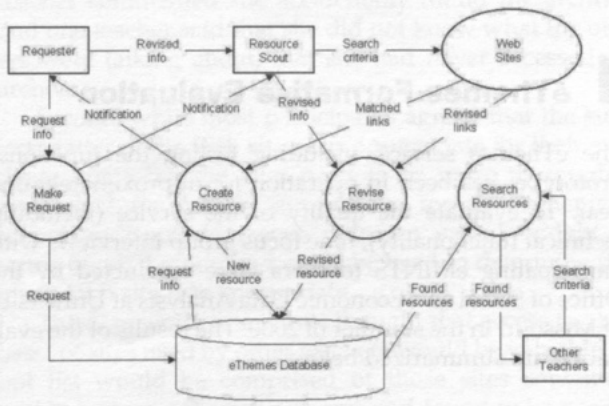


Figure 1. Process Flow of the eThemes Service

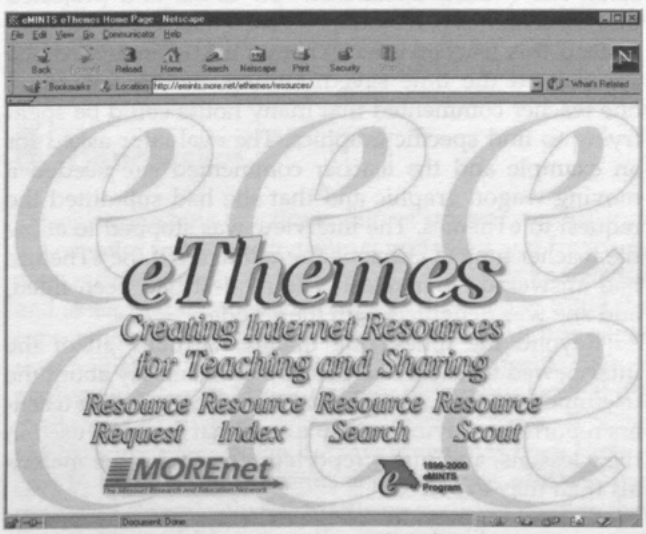


Figure 2. eThemes Main Menu

their relevance to the search query or by the alphabetic order of the resource names. Figure 5 shows the search menu.

The Resource Administration module is the place where the service team manages their services. The administration module contains the functions for a scout to create and revise an eThemes resource based on a request. The module tells which requests are new, which requests are assigned to whom, and which requests are fulfilled. In addition, it has a link checker that checks broken links on a regular basis with an automatic mechanism to inform the service team and the link users. Also, a monthly usage report is generated for the management to track the eThemes services. Figure 6 shows the administrative menu.

eThemes Formative Evaluation

The eThemes service, including testing the functional prototype, has been in operation for approximately one year. To evaluate the quality of the service (including technical functionality), nine focus group interviews with participating eMINTS teachers were conducted by the Office of Social and Economic Data Analysis at University of Missouri in the summer of 2000. The results of the evaluation are summarized below.

Value of the eThemes Service

The most common benefit attributed to eThemes was that the service was a timesaver for teachers. One teacher indicated she needed information on China and projected that the search "... would have taken me days or weeks." Instead this teacher used eThemes and commented she appreciated the time saved. During another interview, one teacher commented that many hours could be spent trying to find specific graphics. The evaluator asked for an example and the teacher commented she needed a moving wagon graphic and that she had submitted the request to eThemes. The interview was stopped to allow the teacher time to check her e-mail to see if the eThemes had answered her request. The request had been filled, and she was delighted with the results.

Beyond the timesaving quality, virtually all of the interviewed teachers had positive things to say about the usefulness of the materials that they received. The teachers reported receiving multiple sites that could be used in their lessons, and many reported receiving more materials than they expected.

Central to the value of the eThemes service were the activities of the resource scouts. Many teachers relayed how cordial the scouts were. Several teachers were impressed by the responsiveness of the scouts and

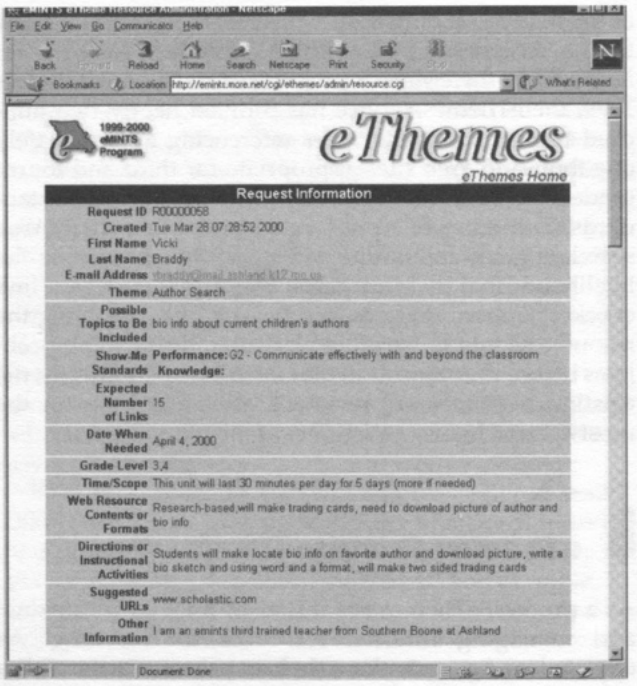


Figure 3. A Request Example

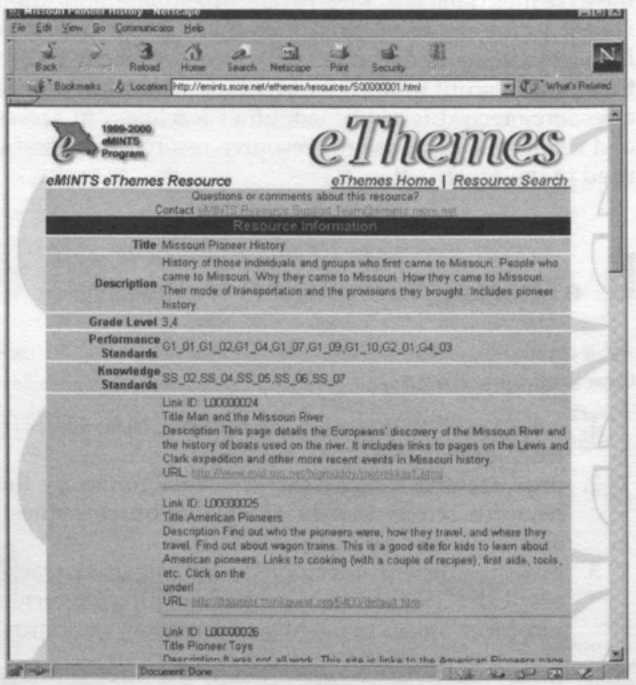


Figure 4. A Resource Example

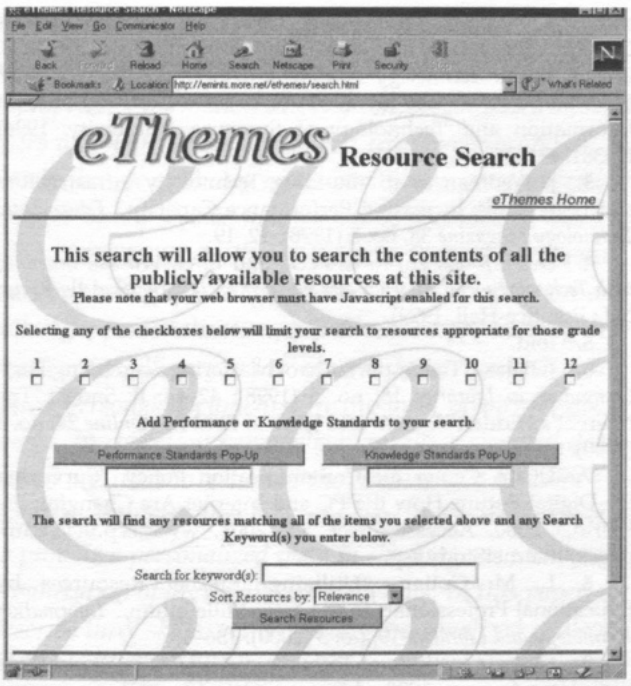


Figure 5. The Resource Search Menu

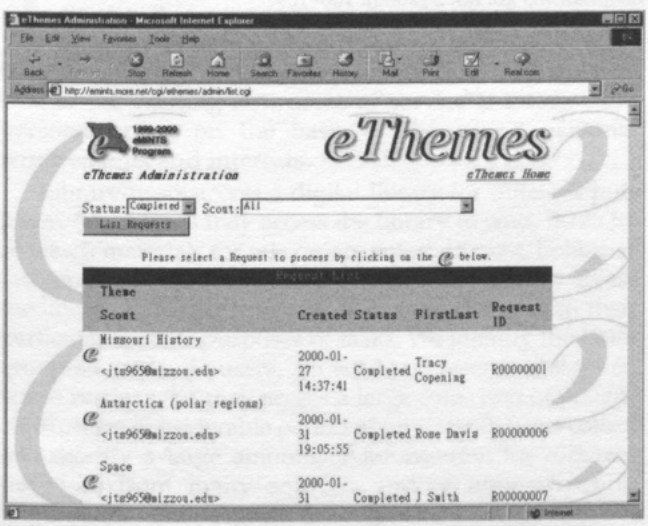


Figure 6. eThemes Administration Menu

appreciated their efforts to provide useful and appropriate Internet resources for their students. By and large the teachers were satisfied with the turnaround time for resources. The following comment is typical, "Anytime I've requested anything from them, they've been fast." The scouts are perceived as an important resource for teachers, especially as teachers develop new lessons using Internet resources.

Possible Areas for Improvement

In addition to underscoring the valuable service provided by eThemes, the formative evaluation served to point out areas needing improvement. First, the teachers' opinions about access to the eThemes archive vary widely. Some teachers thought that the archive is easy to understand and work through. These teachers went to the Search module first because chances are that whatever they were looking for had already been collected. However, another teacher commented she accidentally found the archive, and one teacher said that she did not know what the others were talking about, that she had never accessed an archive.

Second, while most participants agreed that the subject matter of the Web sites was appropriate for their students, some felt the sites need to be more thoroughly researched for content appropriate for an elementary school audience. However, opinions varied widely in terms of what guidelines could be used to determine if a given Web site was appropriate.

Third, several participants thought that access to "hot lists" of sites used by other teachers would be valuable. A hot list would be comprised of those sites contained within a given eThemes resource and found to be particularly valuable by teachers who used them in teaching. Although eThemes was referred to as an excellent resource, the teachers said they still had to figure out the links and see what was appropriate. If another teacher used the site, however, they felt more assured that the site was classroom ready.

Conclusion

Changing teaching practice to impact student performance is a complex task involving new expectations and feedback, appropriate rewards and incentives, up-to-date knowledge and skills, and the tools and resources needed to fulfill the expectations. The eThemes service described in this paper rests within the context of a comprehensive effort to change teaching practice in Missouri. Early indications are that the challenge is at least as great as anticipated at the onset of the eMINTS initiative. While the investment has been significant in terms of money, time, and resources, eMINTS has generated several valuable resources, such as eThemes, that are readily scalable across the entire K–12 curriculum in Missouri, and with slight modification to all states that have identified student performance standards.

Missouri is not alone in its efforts to provide ready access to Internet resources. For example, the California Instructional Technology Clearinghouse (http://clearinghouse.k12.ca.us) parallels eThemes in some respects. The same can be said for TrackStar (http://trackstar.hprtec.org) supported by the High Plains Regional Consortium for Educational Technology. What differentiates eThemes from others is the recognition that Internet resource access involves much more than connectivity, a computer, a browser, and a database of Website information. eThemes is an integrated, responsive set of services and support that meets the needs of teachers using Internet resources. Indeed, eThemes appears to be about building relationships of credibility between the scouts and the teachers while providing a service that supports changes in teaching practice and improvements in student performance.

References

1. President's Committee of Advisors on Science and Technology Panel on Educational Technology, *Report to the President on the Use of Technology to Strengthen K–12 Education in the United States* (Washington, D.C.: U.S. Government Printing Office, 1997).
2. Henry J. Becker and Jason Ravitz, "The Influence of Computer and Internet Use on Teachers' Pedagogical Practices and Perceptions," *Journal of Research on Computing in Education* 31, no. 4 (summer 1999): 356–84; J. H. Sandholtz, C. Ringstaff, and D. C. Dwyer, *Teaching with Technology: Creating Student-Centered Classrooms* (New York: Teachers College Pr., 1997); B. Wilson et al. *Technology Making a Difference: The Peakview Elementary School Study* (Syracuse, N.Y.: ERIC Clearinghouse on Information and Technology at Syracuse University, 1994) ED381149.
3. J. Wedman et al. "Building Technology Infrastructure and Enterprises: Increasing Performance Capacity," *Educational Technology Magazine* 38, no. 5 (1998): 12–19.
4. D. H. Jonassen, K. L. Peck, and B. G. Wilson, *Learning with Technology: A Constructivist Perspective* (Upper Saddle River, N.J.: Prentice-Hall, 1999).
5. Ibid.
6. J. Balas, "The Importance of Mastering Search Engines," *Computers in Libraries* 18, no. 5 (1998): 42–44; B. Snow, "The Internet's Hidden Content and How to Find It," *Online* 24, no. 3 (2000): 61–66.
7. UCLA Center for Communication Policy, "Surveying the Digital Future: How the PC and Internet Are Changing the World," 2000. Accessed Oct. 13, 2001, www.ccp.ucla.edu/pages/InternetStudy.asp.
8. L. M. Golian, "Utilizing Internet Resources by Educational Professionals in the New Millennium," *Information Technology and Libraries* 19, no. 3 (2000): 136–43.
9. Georgia Institute of Technology, GVU's 9th World Wide Web User Survey, 1998. Accessed Oct. 13, 2001, www.gvu.gatech.edu/user_surveys/survey-1998-04/reports.
10. T. J. Waldhart, J. B. Miller, and L. M. Chan, "Provision of Local Assisted Access to Selected Internet Information Resources by ARL Academic Libraries," *The Journal of Academic Librarianship* 26, no. 2 (2000): 100–109.

A Personalized Information Environment for Digital Libraries

Champa Jayawardana,
K. Priyantha Hewagamage,
and Masahito Hirakawa

A digital library provides one of the most important information environments in which to retrieve and refer to appropriate information directly online. Different library users will have different personal requirements and interests in the use of library materials. Hence, personalization is an essential service that should be provided to users to allow them to create their own personalized information environments. Two stages, material personalization and collection personalization, are discussed. In material personalization, library users can create customized views, called personal documents, in the form of virtual documents. In collection personalization, these personal documents are used to generate the profiles that describe the user's working context and interests. The main purpose is to provide a personalized view in the usage as well as the organization of digital library materials. In this paper, we describe three skills, active reading, personalized retrieval, and personalized filtering, which are considered essential to construct a personalized information environment in the digital library.

To make a digital library attractive and to encourage its use, it is necessary for the library environment to provide new and value-added services beyond the conventional distribution and access mechanisms. One of the exciting new areas of service is information personalization on the basis of the user's personal requirements and interests.

Library users access a digital library for different purposes. Some users may access the library to read, listen to, or watch materials for relaxation, entertainment, hobby, or just on their routine visits to the library. Others may access the library to gather information in accordance with their particular specific purposes or tasks. We identify the latter group as academic users. An academic user is always an active reader who may access a large collection of materials during a considerable period of time.[1] Such users collect and modify a large amount of information for different purposes from many sources, adding annotations to improve its value. Again, these activities are carried out according to users' personal requirements of information usage in the digital library. Hence, in the library environment, users expect more than being able to filter, retrieve, and refer to library materials, and they prefer to use them as their own personal sources of information.

It is therefore necessary that the digital library environment should support personalization services. Such services must provide facilities to satisfy the user's personal requirements and interests for the digital library. In our opinion, personalization is not a linear facility that goes from the library to its subscribers/users, because individual requirements and interests change from time to time even with respect to a particular single user. Hence, the success of personalization depends on the way those requirements and interests are captured implicitly in a cyclic process between the library and the user. We classify personalization into two stages: material personalization and collection personalization.[2] In material personalization, facilities are provided for users to acquire library materials according to their individual requirements. Specifically, they allow users to carry out active reading on library materials. Collection personalization, on the other hand, captures the user's working context and interest in order to provide a personalized view of the organization of library materials.

Currently, in the majority of studies with respect to personalization in digital libraries, the main issue addressed corresponds to information-filtering based on the user's interest profile, which is a part of collection personalization.[3] At the same time, there is little study with respect to material personalization. More importantly, these two have been discussed separately although they benefit each other. In this paper, we present in detail both material personalization and collection personalization while showing how the former benefits the latter. These two schemes provide the foundation to develop integrated tools for the success of a personalized information environment in a digital library.

In section 2, we first explain about the expectations of a large information space for personalized information environments. Material personalization is described in detail with the personal document model in section 3. In section 4, the details of collection personalization are given. This section also describes how to generate the profiles of personal documents and the user's interest profile to support the collection personalization. A prototype system that has been developed is presented in section 5. The related work is summarized in section 6 and the conclusion is given in section 7.

Personalized Information Environments

Bush identified the necessity of a personalized information environment and felt that current mechanisms for

Champa Jayawardana (champa@isl.hiroshima-u.ac.jp) is a Ph.D. Candidate, **K. Priyantha Hewagamage** (hewa@isl.hiroshima-u.ac.jp) is the Research Fellow, and **Masahito Hirakawa** (hirakawa@isl.hiroshima-u-ac.jp) is the Director of the Information Systems Lab, Faculty of Engineering, Hiroshima University.

dealing with information were wholly inadequate given the volume of work being produced.[4] He then presented his vision of the Memex, a tool that would allow its user to note, bookmark, and organize information in whatever fashion made most sense to that user. Later this work gave motivation and direction to many research studies in information storage and retrieval.

The personalized information environment is a framework of integrated sets of tools that allow users to create their own view of a large diverse information space through a highly customizable access. The main issues for a personalized information environment are customizability of information resources, and effective and efficient search of them. Public information spaces are not developed for specific user-interests and requirements, but are organized to meet the needs of a general user group using published information about sources. Hence, access capability to information is often poor and the awareness of the existence of specific data is becoming increasingly difficult. What is needed is a set of tools that enable users to create a personal collection of information resources based on their interests and working requirements. In other words, these tools should provide facilities to personalize content, organization, and finding of information resources.

In our vision, a personalized information environment should provide a user-centered and user-organized information space. When users select particular information sources, they may prefer to modify the appearance of the content based on their personal preferences, and to use them in a way that satisfies their individual requirements. Hence, they will be able to carry out activities such as active reading.[5] On the other hand, the organization of this information space must be helpful for users with respect to what they have been doing. Then the retrieval and filtering of the information resources become more effective while providing a personalized view of the information space.

In contrast to the typical search of multiple information sources, where the search engine controls which resources or categories are explored, a personalized information environment places the control in the user's hands. Although the main searching facility would be centralized in a server, many controls are supposed to be decentralized at client sides. It would then be possible to search a subset of items that are already known to the user as well as to direct queries to subcategories identified in the information space. In addition, these controls may modify the user's queries to improve their effectiveness, send them to each of the engines, collect the results, and present them to the user in an intelligent manner. At the same time, the information space would be diverse with a continuous flow of new materials. Hence, the filtering which must be available to maintain a better awareness also plays an important role in providing a personalized view of the information space.

In this paper, we present a user-centered framework called the personalized information environment for digital libraries based on the two ideas of personalization called material personalization and collection personalization. Figure 1 shows a conceptual view of the proposed framework.

Tools that are developed to maintain the personalized information environment are incorporated in a single framework, which makes it possible to enhance the productivity of customizability, retrieval, and filtering. In this paper, we show how an active reading tool developed in material personalization improves the retrieving and filtering facilities in collection personalization.

Material Personalization

When digital library users find useful library materials, they regard them as their own personal information resources. The task of customizing digital library materials according to personal requirements is called material personalization. These personal requirements include annotating, modifying, formatting, and integrating the collected information. By providing facilities for material personalization, the digital library enables its users to carry out active reading on library materials.

Active Reading

Reading frequently involves not just looking at words on a page, but also underlining, highlighting, and making

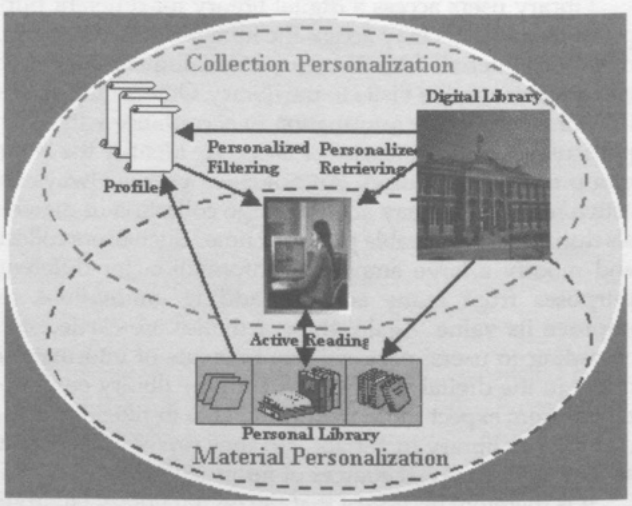


Figure 1. Overview of the Personalized Information Environment for a Digital Library

notes either in the text or in a separate book. This combination of reading with critical thinking and learning is a common behavior of academic users. It is generally referred to as active reading. [6]

Although the way users interact with digital libraries differs from their interaction with conventional libraries, the primary requirements and ultimate objectives of the data gathering process remain the same. In a conventional library, the materials collected from the library are public items that should be treated with care and returned without any marks on them. In such a situation, people use a notebook to keep a record of things, making notes of important parts from several materials. If something is very useful for their studies, users may keep a copy of it for future reference in their personal libraries.

The personal library contains such personal documents as notebooks, articles, and books belonging to the user. Later, the user can access personal library materials for different purposes. When reading the materials from the personal library, the user can modify them by adding annotations to improve their value, underline or highlight some important sentences or phrases, and organize them according to personal requirements. Those activities are considered as the user's active reading activities.

In order to provide similar functionality for the active reading process in a digital library environment, we carried out an investigation by interviewing the actual library users in our university. Based on the gathered results, we determined that the following information activities should be supported for material personalization. Studies done by Nancy et al., Marchionini, and Dillon also verify these points.[7]

- Library users collect information from the digital library by writing down pertinent sections in their notebooks and copying important articles and books
- Library users maintain their personal libraries to store the collected information
- Library users organize the structure of the materials in their personal library (collected information)
- Library users customize the contents of library materials by annotating, formatting, modifying, integrating, summarizing, collecting, and term-defining

Design for the User's Interaction in Material Personalization

To achieve information activities described in the previous section, we designed a model for interaction while combining both the digital library and the user's personal library.

Figure 2 describes how the user-interaction is provided for material personalization. The digital library is a collection of digital objects (referred to as library books or materials) organized in a computer system to provide library services. We consider these digital objects to be the same as published materials, which have some copyright requirements.[8] Since the library possesses the ownership of these materials, it is a responsibility of library service providers to protect them from illegal usage whenever a third party borrows them. A personal document, on the other hand, belongs to a subscriber of the library who can edit its contents as desired, for example, by adding annotations, by virtually deleting (hiding), and by changing its appearance. Hence, a personal document is proposed as a virtual document for the library user.[9]

In this interaction model, the user first selects a required library book from the digital library and refers it. At that time, the user can create or open a personal document with a specific purpose. When some important parts are found in the library book, the user copies them to the personal document. Such important parts are called "booksegments." The process of copying to a virtual document is implemented by shallow copy, which is the key technique used to link parts from library books to personal documents.[10] Basically, in shallow copy, only a pointer to the copied segment is stored and no physical copy of material is made. Thus, this shallow copy technique safeguards the copyrights and ownership requirements of digital library materials.

The documents in the personal library are categorized into three groupings according to their virtual contents. They are notebooks, personalized articles, and personalized books. A notebook is the simplest form of a personal document and allows the user to integrate different segments of library materials while also adding annotations. The personalized articles and personalized books are to provide personalized views of the library materials. A personalized article corresponds to a logical part of library material such as an article from a magazine or a

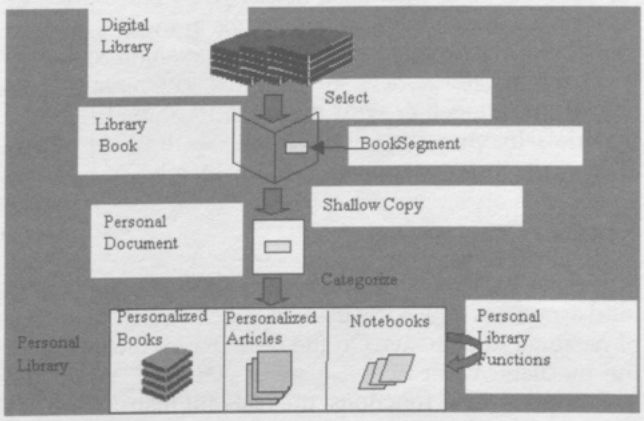


Figure 2. Interaction for Material Personalization

chapter of a book. A personalized book, on the other hand, corresponds to the whole of library material. The content development of those personal documents is carried out by using functions assigned from the personal document model.

Personal Document Model

The personal document model is designed to develop the system for material personalization in a digital library. It is actually an extended version of the notebook model.[11] The architecture of the personal document model integrates documents in the personal library with those in the digital library, supporting the development of the interface for material personalization of a digital library. It mainly consists of the document architecture, and functions for the content management in personal documents.

Document Architecture

Figure 3 illustrates the document architecture of the personal document model and it describes how a personal document exists as a personalized view of library materials and their relationships in a virtual authoring process. A class hierarchy is used to describe all documents together with their constituents. Attributes are used to describe specific properties of documents. In this architecture, classes are defined considering subscribers' data-gathering processes in a real-world library environment.

DocObject class is the based class for the whole class hierarchy and maintains the primitive attributes for all subdocument classes. LibDoc class corresponds to library materials in the digital library with properties protecting their copyrights. A library document is also considered as a combination of text and image objects, which are instantiated under TextObject and ImageObject classes. On the other hand, PersonalDoc class defines in general the logical elements for the personal library. They correspond to personal documents described in the previous section. Thus, three subclasses, Notebook, Personalized Book, and Personalized Article, are declared to maintain their classification together with their relationships to library materials. In other words, objects in these three subclasses are specific instances of personal documents.

Functions

Functions in the personal document model define behavioral aspects in which the user interacts with the contents of personal documents. On the other hand, they provide the mechanism for active reading activities. When the user applies these functions, the relevant instructions are generated for the content management of personal documents. Many of these functions appear as normal editing functions found in the authoring of the document. But some of them are declared considering specific features of personal documents and we briefly describe them below.

Shallow Copying

Shallow copy is the main technique for the creation of personal documents as a kind of virtual documents.[12] When data are copied from one library document to a personal document, only the addresses of the source data are persistent inside the target. Hence it requires fetching the actual content from the original source whenever the document is opened for browsing or editing. This technique supports both the text and image data irrespective of their data formats. The addresses of segments are determined after projecting them to a virtual coordinates system. Java virtual machine provides facilities for the practical implementation.[13] This function also adheres to the concept of transculation advocated by Ted Nelson. [14]

Annotating

This function inserts the library user's own narrations or comments into personal documents. They are probably based on the reader's own thoughts and become new artifacts on top of the published materials.

Virtual Deleting

This enables the user to hide some contents shown in personal documents. Probably such contents may represent minutiae or things the reader is not interested in.

Summarizing

The summarizing function allows retrieval of the summary of a document based on a specific set of rules that the user may have designated. It can also be used to interactively redefine the summary of a document by selecting parts such as headings or highlighted sentences in a personal document.

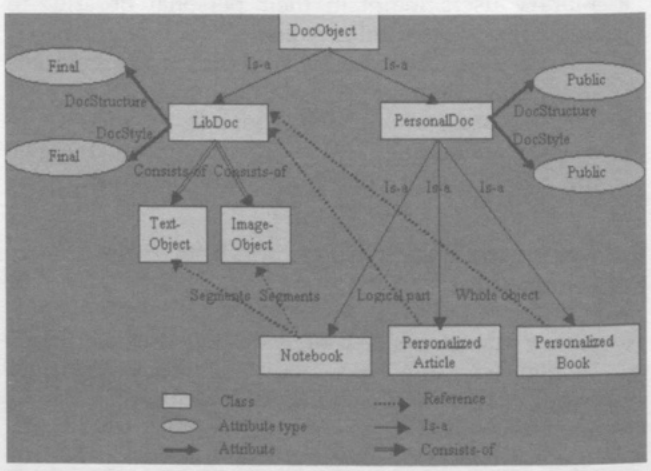


Figure 3. Document Architecture of the Personal Document Model

Collecting

The collecting function supports the creation of a new personal document by integrating existing personal documents in the personal library. Such a new document would probably represent an abstract discussion about those existing documents or particular user-defined hierarchy for the document organization.

Term-Defining

This function extends the reader's vocabulary and inserts or redefines terms in the personal dictionary of the personal library. The personal dictionary is a special personal document used to represent selected dictionaries in the digital library.

Collection Personalization

The main emphasis of collection personalization is to provide skills which can be used to define a personalized view of the organization of the digital library. The user becomes able to easily interact with, obtain, and receive information sources in the digital library. As a result, it would reduce the problem of information overload that is encountered in a dynamic and heterogeneous information environment. The success of this collection personalization mainly depends on the way in which the user's information needs are captured and maintained during the usage of the digital library.

Traditionally, personalization is addressed to customize the retrieving or filtering of information from dynamic sources with respect to the user's interest. However, collection personalization is based not only on the user's interest but also on the user's working context. The user's working context describes the user's current information requirements whenever the digital library is accessed. Generally, people access the library with different purposes that necessitate some information sources that would be different from their current personal interests. For example, a graduate student may want to refer to certain studies in image processing as a part of his or her course work though it would not be his or her personal interest. But the system should be able to personalize based on such short-term personal requirements during the usage of the library. Hence, in this research, collection personalization is designed to provide skills such as personalized retrieving and personalized filtering based on both the user's interest and the user's working context.

In this section, we are presenting how information needs with respect to personal interest and working context are captured and maintained using the personal documents introduced in the material personalization stage. As we stated earlier, the personal documents provide the mechanism to personalize the library materials depending on the user's personal requirements. Therefore, the user's interest can be identified by considering the contents of all personal documents, and the current working context can be identified by considering the contents of the currently opened personal documents. Hence, by analyzing the content of a personal document, personal interest and working context can be captured. It is understood that there are two types of knowledge—content-based knowledge and context-based knowledge—in the personal document. A separate profile is maintained for each personal document with both content-based knowledge and context-based knowledge. Using these profiles, an interest profile is generated. Figure 4 illustrates how these profiles of personal documents and the user's interest profile are related in the personalized retrieving and filtering processes.

A profile is maintained for each personal document and an interest profile is generated based on all those profiles. Personalized retrieving and personalized filtering are the main processes in collection personalization. All profiles of the personal documents, the interest profile, and the history file (which contains the user's access information) provide the required information for the success of the personalized retrieving and filtering processes. More details are described in the next subsections.

Profiles of Personal Documents

As described in the section on material personalization, the user creates a personal document to interact with and manage the information found in library materials. Some personal documents, specifically notebooks, correspond to many library materials and they contain selected segments. Others probably are based on a single library

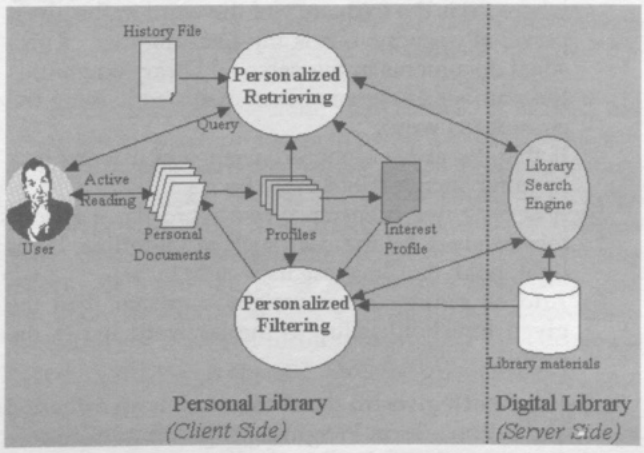


Figure 4. Information Flow in the Collection Personalization

source with a personalized view. The user generally creates a personal document with a specific purpose in mind while engaging in academic work or persuing a personal interest. Hence, a personal document can be considered as one that describes a particular context during the usage of library materials. The content of a personal document indicates what has resulted due to the user's interaction with materials. At the same time, we use a profile for each personal document from the beginning of its creation to maintain both content- and context-based knowledge.

Content-based knowledge, which is one of two categories of information to be contained in the profile of a personal document, contains a set of keywords and a set of phrases determined by analyzing the content of the personal document. The actual contents of each shallow segment and the user's comments in a personal document are sampled to determine these keywords. The phrases are determined by analyzing such structures as headings, subheadings and section headings, as well as the formatted text phrases (highlighted, underline, bold, etc.). These phrases provide different semantics than the set of keywords of the relevant personal document.

Steps to determine keywords in a personal document:

1. Create an index of words other than stop words in the personal document having the following structure: <Word>, {addresses of its occurrences in the personal document}.
2. If a word in the index can be directly identified in the personal dictionary, give an initial weight of w_1 to it.
3. Increase the weight of a word by adding its frequency weight (w_2) that is determined based on the total number of personal documents

$$W2 = f(i)\log\left(\frac{n}{df(i)}\right)$$

where *f(i)* is the frequency of the word in the given personal document, *n* is the total number of personal documents in the personal library and *df(i)* is the number of personal documents containing the considered word.
4. If there is at least one occurrence of a word in a heading, increase its weight by w_3.
5. If at least one occurrence of a word has a special format given by the user such as underline, highlight, bold, etc., increase its weight by w_4.
6. Add all words whose weight is greater than the given threshold value to the keyword list in the profile.

The approach given in the above steps is an extended version of Salton's Term Weighting Algorithm.[15] The personal dictionary that is built in the material personalization is used as an important resource to determine possible keywords in the personal documents. If the particular word's frequency is high and it is repeated in many documents, its weight is decreased using a logarithm value of *n* over *df(i)*. The position and user-given formatting information are also used to identify possible keywords. The values for w_1, w_3, and w_4 are allowed to modify at the personal documents level but they all are given as initial default values determined at the experimental stage. If the keyword list of a personal document gives an incorrect or deviated keywords list, the user can change those weights and the cutoff threshold value. Such modifications are considered as a kind of user's feedback.

On the other hand, the contextual representation of a personal document in its profile has two subsections. The first one describes the corresponding context of a personal document in the digital library. These are identified based on the original sources of each segment in the personal document (the most relevant context) and based on the hyperlinks in the content of a personal document. These pieces of contextual information are used by personalized retrieving and filtering processes to properly identify the clusters of library materials in the digital library. The details are described below.

The second subsection of contextual representation in a profile of a personal document gives the categories of library materials in which the user is interested. The digital library, just like conventional libraries, provides the catalog information. This catalog information can be easily used to identify the relevant categories to which the personal document is related. The library materials, to which the personal document has links, are used to identify the corresponding categories by searching the digital library catalog. Figure 5 shows a sample profile that would be generated based on the personal document called "Java Programming."

Interest Profile

The interest profile describes the degree of user interest and the preferred library sources. Once the profiles for all personal documents are built, the user's interest profile is constructed by amalgamating all these profiles. It is periodically updated to maintain the current status. The interest profile also contains two sections to describe the content- and context-wise interests.

We first explain a procedure to get the content-based interests. Keywords and phrases are again used to describe these interests.

Steps to select keywords for the user's interest profile:

1. Construct an index of all possible keywords using profiles of personal documents.
2. Calculate weights for each keyword as follows: Weight = (Frequency of occurrence of the keyword) * (Average weight of the keyword).

3. If the weight of a keyword is greater than a given threshold value, it is selected as a keyword to include in the interest profile.

Steps to determine phrases for interest profile:

1. Create an index of phrases considering all personal documents.
2. Calculate the weight of a phrase by adding the weights of keywords in the phrase (use weights given in the interest profile keywords).
3. Multiply weight of a phrase by its frequency of occurrence.
4. If the accumulated weight of the phrase is greater than the given threshold value, it is selected to include in the interest profile.

Meanwhile, the contextual interest is determined by analyzing the information sources (library materials) consulted by the user. One such interest describes actual library materials and the other one indicates the categories of library materials.

Steps to identify interested library materials:

1. Create an index of library materials using all profiles of personal documents.
2. Calculate a weight for each library material as follows: Weight = (Frequency of occurrence of the keyword) * (Average weight of the keyword).
3. If the weight of a material is greater than the specified threshold value, it is selected as a library material the user is interested in.

Steps to identify interested categories of library materials:

1. Create an index of categories for all library materials using all profiles of personal documents.
2. Calculate the weight for each category as follows: Total number of library materials accessed in a category/Total number of library materials consulted.
3. If the weight for a category is greater than the given threshold value, it is selected as an area the user is interested in.

Personalized Retrieving

The process of information retrieving can be divided into three subcategories: text representation, retrieval (comparison) techniques, and the acquisition of the user's needs.[16] The text representation corresponds to indexing, clustering, and reorganization mechanisms in the domain of data. The retrieval techniques are based on comparison models. The three main approaches are the Boolean, vector space, and probabilistic retrieval models.[17] The acquisition of user-needs focuses on how a proper query can be formed to extract the correct information. The latter category plays

```
// Profile of Personal Document – Java Programming

// Content-based Knowledge of Personal Document
//Keywords List
   Java (0.7), Compiler (.45), Interpreter (0.5),
   Javac (0.44), Programming (0.6),.....
//Phrases List
   "Java Programming Language" (1.3), "Java Virtual Machine" (0.7),
   "Java Interpreter" (1.2), "Java Compiler" (1.59), .....

//Context-based Knowledge of Personal Document
// Direct Context based on Shallow Segments
   http://www.isl.hiroshima-u.ac.jp/~DL/JavaTutorial/getstarted/
   http://www.isl.hiroshima-u.ac.jp/~DL/JavaProgramming/Introduction/
   http://www.isl.hiroshima-u.ac.jp/~DL/JavaTutorial/getstarted/Beans/
   .....
//Indirect Context based on the hyperlinks
   http://www.isl.hiroshima-u.ac.jp/~DL/JavaTutorial/Introduction/1.3
   http://java.sun.com/
  .....
//Relevant Categories identified using catalog information
 Programming, Java .....
```

Figure 5. A Sample Profile of a Personal Document

an equally important role in the success of information retrieval and our work in this paper is also based on it.

In order to ensure the success of a digital library, the provision of information retrieval must be a priority. A library user encountering a particular problem (an information requirement) specifies it as a query using the retrieval tool given to search the library materials. Generally, such a query is a collection of words with Boolean conditions or a phrase with some semantic meaning. Queries are passed to a library search engine without any consideration of the user's current information needs. Hence, the retrieved results cannot fulfill the user needs. Query reformulation is required to solve these problems.[18] At the same time, the interface for the retrieved results is poorly organized, making it a bit difficult or time-consuming to identify appropriate library materials.

In the proposed system, the retrieval is carried out using the information found in profiles of personal documents, the user's interest profile in the personal library, and the interaction history file, which describe the usage of library materials. It is basically done at two stages. First, when the user specifies a query to retrieve some information, the query is modified interactively and internally to satisfy the user's current information need. Secondly, the retrieved result is formed by categorizing it based on the user's specific declared preferences.

Personalized Filtering

Information filtering systems can help users by eliminating irrelevant information. They are akin to mediators

between sources of information and end-users. Although there are many similarities between information retrieval and filtering, a notable distinction is that an information filtering system must remember the user and individualize its performance for that particular user's personal requirements and interests.[19] Specifically, it should motivate the users in their information-seeking by providing appropriate information at the right time. Hence, the user's long term interests and short-term requirements (working contexts) should be incorporated into a personalized filtering process.

Generally, there are three stages in the filtering process: (1) obtaining the user's interests and requirements, (2) identifying the most suitable information sources, and (3) presenting those results to the user at the right time in a user-friendly way. Personal documents and their profiles work as the basic foundation for these three stages in filtering. In the proposed system, filtering is carried out based on each profile as an automated background process whenever the profile is updated significantly, and also periodically as specified by the user. Keywords, phrases, and contextual information in profiles are used to obtain the user's interests and requirements.

The filtering process is performed in two ways based on the contextual information given in profiles. The first way of filtering is carried out considering the categories of library materials. All library materials in specified categories, except those directly referred to the personal document, are accessed to identify the relevant library materials. For this purpose, it is required to calculate their relevancy factors.

Let F_d be the relevancy factor of a library material for a given personal document.

$$F_d = \frac{\sum_{i=1}^{n_1} x_i w_i + \sum_{i=1}^{n_2} x_i w_i}{\sum_{i=1}^{n_1} w_i + \sum_{i=1}^{n_2} w_i}$$

Where n_1 and n_2 are the number of keywords and phrases in the profile, respectively, X_i is the frequency of a keyword or a phrase in the given document, and is the weight for a keyword or phrase found in its profile.

Library materials whose relevancy factors are greater than the given threshold are selected as filtered documents identified inside the context of a personal document. The above formula for the calculation of the relevancy factor is based on the theory presented in Salton.[20]

The second way of filtering is done without considering the categories of library materials. In this case, the library materials contain at least 50 percent of keywords or phrases and are retrieved by using library search engine. Then their relevancy factors are calculated to select the most suitable materials for a given threshold value.

The filtered library materials are recorded in the profiles themselves and presented to the user at appropriate times. For example, when a personal document is opened, new filtered information segments in its profile are added to the document.

The Prototype System

We developed a prototype system to investigate and demonstrate the mechanism discussed in the previous sections. It provides the software engineering approach that could be used to establish a personalized library service. The architecture of the prototype system is given in figure 6, and the system is presented as an integrated interface of the digital library and the user's personal library. We use a document base of published materials on a Web server as the library, and the prototype system is developed as a client component while maintaining a low coupling between them. This prototype runs on top of the Java Virtual Machine and the user interacts with the system through two components called DL Browser and Personal Document Editor. Facilities for material personalization are provided in the Personal Document Editor, which also provides an interface to the user's personal library. DL Browser facilitates navigation in the library just like a normal browser and consists of tools for personalized retrieval and personalized filtering, which are provided in collection personalization. In addition, an agent called Address Accuracy Detector is used to maintain the personal documents in the personal library. It detects the changes in the shallow-copied segments and updates them if possible.

DL Browser

The DL Browser is similar to a normal Web browser in its ability to navigate through library materials once a user logs into the digital library. Its starting page is called Public BookShelf, which is the main gate of the digital library. It provides facilities to navigate the collection of library materials under different classifications (alphabetical, author, publication date, and catalog entries). Links to library services and other notices are also included in this page. The user can also access the main library search engine directly from this page.

Whenever material is selected for reading, the user can apply shallow copy to it to personalize its content. As we described in the section on material personalization, when the user selects a segment of library material, its reference addresses are determined and can be pasted on the personal documents. When an article or whole book is selected, the relevant entries are updated in the Personal BookShelf, which is the main page of the Personal Document Editor.

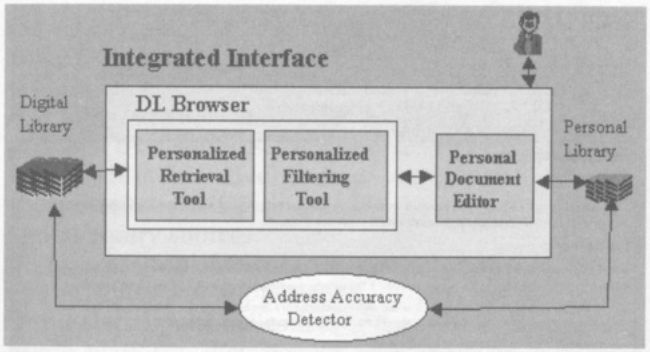


Figure 6. Architecture of the Prototype System

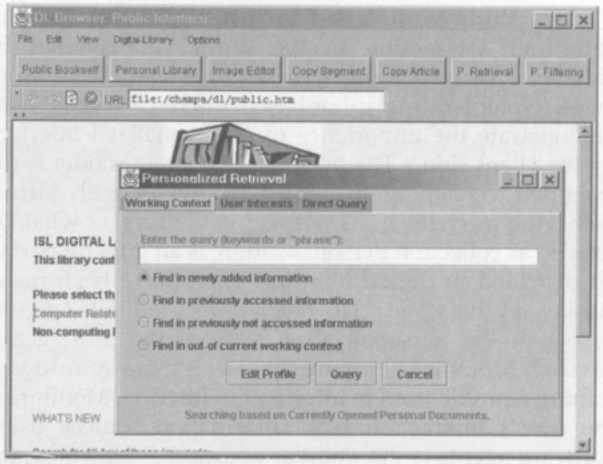


Figure 7. Personalized Retrieving in DL Browser

Figure 7 shows the screen shot of DL Browser and the subwindow of a personalized retrieval tool. When the user presses the P.Retrieval button on its menu bar, the sub-window of the personalized retrieval tool appears allowing the user to specify a query and different options. Generally, it retrieves the documents based on the user's working context (default option), which is described by the currently opened personal documents. Queries can be given based on the user's interest profile or as direct queries by suspending personalized features. Additional options are provided to classify the output results list.

The personalized filtering tool, which is another main tool in the DL Browser, filters the information based on the user's interest profile as well as the user's current working context. The filtered library materials based on the user's working context are listed at the end of the personal documents. Also, by pressing the P.Filtering button on the toolbar, the user can see the filtered materials based on the user's interest profile. The user can also modify the options for personalized retrieval and filtering going through the Options menu in the DL Browser. It allows the user to modify relevant weights and threshold values for the user's interest profile and to manually edit the current interest profile.

Personal Document Editor

The Personal Document Editor provides facilities to create, open, and maintain the personal documents. Its main page, Personal BookShelf, is a personalized interface for the digital library and integrates it with the user's personal library. The user can interactively modify the content and appearance of Personal BookShelf and it also lists the personal documents in the personal library classified under notebooks, personalized articles, and personalized books.

The Personal Document Editor provides the relevant functions for the content management of personal documents. Whenever the user wants to perform the shallow copy on opened materials in the DL Browser, the relevant segment should be selected using the mouse just like a normal copy function. If the segment consists of an image object, it should be opened in a separate window mapping it into virtual coordinates. Then the virtual coordinates are used to determine the addresses of image segments.[21] The user can insert shallow-copied segments into any location inside the personal document by placing the cursor and pressing the relevant button.

Figure 8 shows a notebook in the Personal Document Editor. This notebook shows the book segments copied from many library materials and the user's annotations/comments. All contents have been reformatted and their appearance is different from the original. Each segment acknowledges original segments by placing corresponding references to it. At the end of the personal documents is a list of the relevant library materials filtered for this personal document using the filtering tool. If the user is unsatisfied with the filtered items, he or she can review the weights and thresholds assigned to the opened personal document by going through the Options menu in the DL browser.

Figure 9 displays a personalized article with its original article opened in the back window. The front window shows the personalized article as a user-defined view of original library material. The user has reformatted the original article and added annotations.

Related Work

The work in this paper presents our ongoing project on developing a personalized library environment in the digital library.

This study was started to investigate how human-computer interaction in the digital library can be enhanced by providing a better information environment. Notebook metaphor-based interface mechanisms demonstrate the importance of a personalized interface on the client side.[22] The personal document model is the extended version of the notebook model with virtual authoring mechanism.[23] Information filtering, or what we refer to as collection personalization, is an integral part to be provided in digital libraries. There have been many studies on this issue.[24] However, trials are limited to one aspect of the personalization. One of the most active research topics is to capture the user's interest for determining a profile used in information filtering. Monitoring the user's interaction with information sources using software agents is the popular approach.[25] However, the personalization is not merely collection personalization according to the user's interests. Facilities provided for material personalization enhance the user's active reading activities, and the shallow copy technique can protect both copyright requirements and ownerships of library materials.[26] The personalized information environment for the digital library, which we are presenting in this paper, illustrates how material personalization is useful in enhancing software development requirements for collection personalization.

The model in the study by Joaquim et al. is close to our personal document model, in which two entities (information tokens and information artifacts) are described on the level of library materials.[27] Information tokens represent particular information contents (what we described as segments), and information artifacts are views created recursively over information tokens by copying, structuring, combining, annotating, and so on. The personal document model offers a practical and technical framework for the specification of personalized digital materials.

A conceptual structure for the personalized information environment of the digital library is proposed in a study by French and Viles.[28] It addresses the customizability of resources based on the palette of resources. A system called InfoBeans, given in Bauer and Dengler, can be used to personalize a process of gathering and integrating information sources.[29] A framework for analyzing the role of personalization in the design of any user experience is presented in Kramer, Noronha, and Vergo.[30] It describes a user-centered design (UCD) process to develop the object-oriented interfaces for e-commerce applications. According to Carroll and Rosson, and Quiroga and Mostafa, the explicit way for capturing the interests is not always successful since the user is not willing to spend a considerable amount of time and effort to maintain the functionality of the system.[31] Using implicit methods, the user's changing interests can be captured without interfering with the user. In our proposed framework, material personalization can be used to capture interest profiles implicitly.

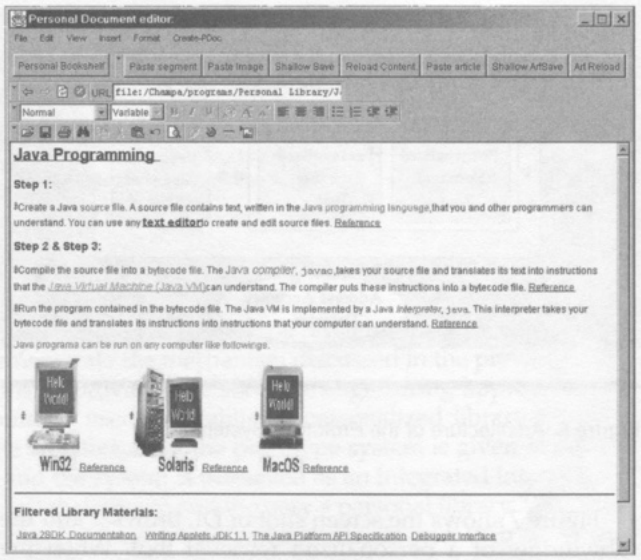


Figure 8. Notebook

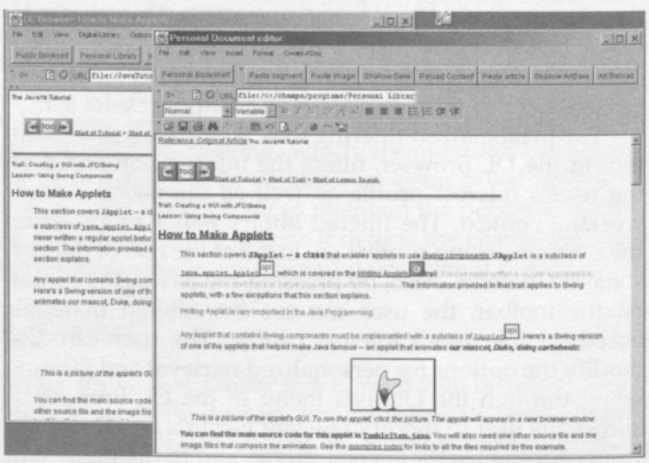


Figure 9. Personalized Article

Conclusion

In this paper we introduced the framework called a personalized information environment in a digital library

that is accessible through a network. Tools in a personalized information environment, which are developed as client-side components, can be used to personalize access to library materials and those tools benefit each other to maintain their services. We presented the mechanism for three such tools: active reading, personalized retrieving, and personalized filtering under the personalization of digital library sources.

In our vision of the personalization of digital libraries, two schemes are supported: collection personalization and material personalization. If they are addressed on a single platform, they benefit each other by creating a cycle of enhanced interaction. Collection personalization, which includes personalized information retrieval and filtering, is generally considered as the mechanism that reduces the information overload. Material personalization can be used to personalize the content of library materials, thus improving their usage.

In material personalization, facilities are provided for users to consume library materials according to their individual requirements and interests, and specifically allowing them to carry out active reading of library materials. The personal document model proposed in this paper is a successful design for the material personalization of digital libraries. It describes how a personal document exists as a personalized view of library materials and functions for active reading. We applied the shallow copy technique for construction of personal documents to safeguard the copyright requirements of original sources and also to protect ownership of library materials. In shallow copy, when the data segments of published materials are copied to the personal documents, only the addresses of these segments are stored on the documents.

In collection personalization, retrieval and filtering are personalized by connecting those processes with personal documents. Personalized retrieval and filtering are based on what the user is generally interested in or wants in a particular working context. Hence, the main issue is to capture these working contexts and interests in an implicit way as much as possible through monitoring the user's interaction. In this paper we demonstrated how the personal documents of material personalization could be used to determine the working context and user-interests in a personalized information environment. The user creates a personal document with a specific objective in the usage of library materials as described in material personalization. Hence, one or more personal documents that are currently opened describe the user's working context at a particular stage. Content- and context-based knowledge are determined by analyzing the personal documents and are represented in the corresponding profiles. The profiles of all personal documents in the personal library are used to determine the user's interests. A term-weighting schema is considered to generate those profiles. The filtered library materials are presented to the user by appending relevant links of them to corresponding personal documents. The user can modify the filtering by modifying weights or editing the profiles.

In this paper we also briefly presented the prototype system which we are developing to demonstrate the personalization of digital library. The work discussed in this paper contributes to application development of personalized information environments in the digital library. In the future, we are planning to investigate other issues in personalization of the digital library such as situated personalization and personal documents-based information-sharing for collaborative activities among digital library users.

References and Notes

1. B. N. Schilit, G. Golovchinsky, and M. N. Price, "Beyond Paper: Supporting Active Reading with Free-Form Digital Ink Annotations," in *Proceedings of CHI '98, April 18–23, 1998* (New York: ACM Pr., 1998), 249–56. Accessed Mar. 1, 2001, www.fxpal.xerox.com/abstracts/sch98a.htm. M. J. Adler and C. Van Doren, *How to Read a Book* (New York: Simon and Schuster, 1940).

2. C. Jayawardana et al., "Interface for the Personalization of the Digital Library," in *Proceedings of International Symposium on Digital Libraries ISDL '99, Sept. 1999* (University of Library and Information Science, Tsukuba, Ibaraki, Japan, 1999), 155–58.

3. W. Lam et al., "Detection of Shifts in User Interests for Personalized Information Filtering," in Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1996. Accessed Mar. 1, 2001, http://dev.acm.org/pubs/citations/proceedings/ir/243199/p317-lam. J. C. French, A. L. Powell, and W. R. Creighton, "Efficient Searching in Distributed Digital Libraries," in *Proceedings of 3d ACM International Conference on Digital Libraries, June 1998* (New York: ACM Pr., 1998), 283–84. Accessed Mar. 1, 2001, http://citeseer.nj.nec.com/french98efficient.html. B. Sheth, "A Learning Approach to Personalized Information Filtering," master's thesis, MIT, 1994. Accessed Mar. 1, 2001, http://citeseer.nj.nec.com/sheth94learning.html. B. Sheth and P. Maes, "Evolving Agents for Personalized Information Filtering," in *Proceedings of the 9th IEEE Conference on Artificial Intelligence for Applications, 1993* (Los Alamitos, Calif.: IEEE Computer Society Press, 1993) 345–52.

4. V. Bush, "As We May Think," *Atlantic Monthly* 176, no. 1 (1945): 101–08 (republished in *ACM Interactions* 3, no. 2 [Mar. 1996]: 35–46).

5. C. Marshall, "Annotation from Paper Books to the Digital Library," in *Proceedings of the ACM International Conference on Digital Libraries, 1997* (New York: ACM Pr., 1999), 131–40; J. M. Carroll and M. B. Rosson, "The Paradox of the Active User," in *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction* (Boston: MIT Pr., 1987).

6. Schilit, Golovchinsky, and Price, "Beyond Paper"; Adler and VanDoren, *How to Read a Book.*

7. A. Nancy et al., "User-Centered Iterative Design for Digital Libraries," D-Lib Magazine, Feb. 1996. Accessed Mar. 1,

2001, www.dlib.org/dlib/february96/02vanhouse.html. G. Marchionini, "User-Centered Methods for Library Interface Design," Univ. of Illinois, 37th Allerton, 1995. Accessed Mar. 1, 2001, http://edfu.lis.uiuc.edu/allerton/95. A. Dillon, "What is the Shape of Information? Human Factors in the Development and Use of Digital Libraries," Indiana Univ., 37th Allerton, 1995. Accessed Mar. 1, 2001, http://edfu.lis.uiuc.edu/allerton/95/s4/dillon.html.

8. S. Brin, J. Davis, and G. Hector, "Copy Detection Mechanisms for Digital Documents," in *Proceedings of ACM SIGMOD Conference, 1995* (New York: ACM Pr., 1995), 398–409. Accessed Mar. 1, 2001, http://citeseer.nj.nec.com/brin95copy.html; "Copyright in the Library," copyright course. Accessed Mar. 1, 2001, www.utsystem.edu/OGC/IntellectualProperty/l-diglib.htm.

9. C. Jayawardana, K. P. Hewagamage, and M. Hirakawa, "Virtual Authoring Based on the Shallow Copy Technique for a Collection of Digital Documents," in *Proceedings of the International Symposium on Multimedia Software Engineering, Dec. 2000* (Los Alamitos, Calif.: IEEE Computer Society Pr., 2000), 77–84.

10. Ibid; C. Jayawardana and M. Hirakawa, "Interface Mechanism to Personalize the Digital Library," in *Transactions of Information Processing Society of Japan*, vol. 41, no. 10 (Oct. 2000): 2863–2372.

11. T. Watanabe, C. Jayawardana, and T. Ushiama, "Personal Interface Mechanism on Digital Library," in *Proceedings of IEEE ADL '98, 1998* (Los Alamitos, Calif.: IEEE Computer Society Pr., 1998), 76–85; C. Jayawardana, T. Ushiama, and T. Watanabe, "Personal Interface Adaptable to Digital Library," in *Proceedings of International Conference on Virtual Systems and Multimedia (VSMM) '96* (Gifu, Japan: Gifu Research and Development Foundation, 1996), 491–96.

12. Jayawardana, Hewagamage, and Hirakawa, "Virtual Authoring Based on the Shallow Copy Technique for a Collection of Digital Documents."

13. Java 2D API Specification. Accessed Mar 1, 2001, http://java.sun.com/products/jdk/1.2/docs/guide/2d/index.html.

14. T. H. Nelson, "Literary Machines" (Swarthmore, Penn.: self-published, 1983).

15. G. Salton and C. Buckley, "Term Weighting Approaches in Automatic Text Retrieval," Technical Report [87–881], Cornell Univ., Nov. 1987.

16. N. Belkin and W. B. Croft, "Information Filtering and Information Retrieval: Two Sides of the Same Coin?" in *Communications of the ACM* 35, no. 12 (Dec. 1992): 29–37.

17. G. Salton, E. Fox, and H. Wu, "Extended Boolean Information Retrieval," in *Communications of the ACM* 26, no. 11 (Nov. 1983): 1022–36; G. Salton, *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer* (Reading, Mass.: Addison-Wesley, 1989); H. R. Turtle and W. B. Croft, "A Comparison of Text Retrieval Models," *Computer Journal* 35 no. 3 (1992): 279–90.

18. N. J. Belkin, "Helping People Find What They Don't Know," in *Communications of the ACM* 43, no. 8 (Aug. 2000:) 58–61.

19. Belkin and Croft, "Information Filtering and Information Retrieval."

20. Salton, *Automatic Text Processing.*

21. Jayawardana, Hewagamage, and Hirakawa, "Virtual Authoring Based on the Shallow Copy Technique for a Collection of Digital Documents."

22. Watanabe, Jayawardana, and Ushiama, "Personal Interface Mechanism on Digital Library"; Jayawardana, Ushiama, and Watanabe, "Personal Interface Adaptable to Digital Library."

23. Jayawardana, Hewagamage, and Hirakawa, "Virtual Authoring Based on the Shallow Copy Technique."

24. Bush, "As We May Think"; Schilit, Golovchinsky, and Price, "Beyond Paper"; Adler and VanDoren, *How to Read a Book;* Nancy et al., "User-Centered Iterative Design for Digital Libraries"; Marchionini, "User-Centered Methods for Library Interface Design;" Dillon, "What is the Shape of Information."

25. A. G. Moukas, "Amalthaea: Information Discovery and Filtering Using a Multiagent Evolving Ecosystem," *International Journal of Applied Artificial Intelligence* 11, no. 5 (Apr. 1997): 437–57.

26. Jayawardana and Hirakawa, "Interface Mechanism to Personalize the Digital Library"; Brin, Davis, and Hector, "Copy Detection Mechanisms for Digital Documents"; "Copyright in the Library."

27. W. S. Joachim et al., "Linguistic and Architectural Requirements for Personalized Digital Libraries," *International Journal of Digital Libraries* 1, no. 1 (Oct. 1997): 89–104.

28. J. C. French and C. L. Viles, "Personalized Information Environment," D-Lib Magazine, June 1999. Accessed March 1, 2001, www.dlib.org/dlib/june99/french/06french.html.

29. M. Bauer and D. Dengler, "InfoBeans: Configuration of Personalized Information Assistants," in *Proceedings of International Conference on Intelligent User Interfaces, 1999* (New York: ACM Pr., 1999), 153–56.

30. J. Kramer, S. Noronha, and J. Vergo, "A User-Centered Design Approach to Personalization," in *Communications of the ACM* 43, no. 8 (Aug. 2000): 45–48.

31. Carroll and Rosson, "The Paradox of the Active User"; L. M. Quiroga and J. Mostafa, "Empirical Evaluation of Explicit versus Implicit Acquisition of User Profiles in Information Filtering Systems," in *Proceedings of the 4th ACM International Conference on Digital Libraries, Aug. 1999* (New York: ACM Pr., 1999), 238–39.

Weaving Multimedia into the Collection: One Library's Journey

Gail Golderman

For well over fifteen years, Schaffer Library at Union College in Schenectady, New York, has experimented with a variety of methods to best manage acquiring and granting access to multimedia material. This article conveys the journey one library has taken in order to evolve to the present system of full integration of nonprint processing with print processing within the cataloging department. This has included reorganizations of the cataloging department, creation of detailed documentation, maintenance of a regularly updated Web site for immediate access to relevant links and procedures, and most significantly, providing staff members with essential skills and equipment to better enable them to have a working familiarity with contemporary information formats and standards.

Literature suggests that nonprint media has occupied an enhanced role in library collections for quite some time, particularly with the adoption of AACR2 for all holdings.[1] For well over fifteen years, Schaffer Library has experimented with a variety of methods to best manage acquiring and granting access to multimedia material. The experimental phases necessary for conceptual and hopefully actual realization and fulfillment have steadily evolved to the present system of full integration of nonprint processing with print processing within the cataloging department. This has involved the perception that appropriate changes needed to occur, including two reorganizations of the cataloging department within the past six years, creation of detailed documentation, maintenance of a regularly updated Web site for immediate access to relevant links and procedures, and most significantly, providing staff members with essential skills and equipment to better enable them to have a working familiarity with contemporary information formats and standards. All staff in technical services now have Internet-based access to OCLC from their desktops and first-hand experience with cataloging nonprint items such as videos, audio CDs, CD-ROMs, and resources with multiple formats. The end result is that presently all acquisitions are now administered more efficiently as a benefit not only to the daily workflow in the department, but to the library as a whole and the entire Union College campus community.

One of the first steps taken to address integrating multimedia into the collection was a proposal submitted to the library director in 1985 by the automation committee to purchase audio CDs and to discuss a future CD collection development plan. The proposal requested funds for a minimum initial collection of thirty-five classical and fifteen jazz and popular music titles. This proposal was accepted along with the request to purchase a portable CD player that would be made available for in-house use only, similar to other equipment kept under supervision of the circulation desk. Money was allocated and procedures were established for cataloging the CDs utilizing a local cutter system. The cataloging and processing task was assigned to one individual in the cataloging department and so began the journey.

It soon became apparent that wider library involvement was called for, which led to the formation of a multimedia committee in the late 1980s. This committee's charge was to create policy for an eclectic mix of nonbook resources that had begun to accumulate within the library. As this material gathered, several pressing issues emerged not only for technical services, but for both collection development and public services as well. A major concern was where material would be housed. A renovation and new building project was in the preliminary planning stages and serious space limitations were a reality for the existing collection. Should this new material coexist in the stacks with monographs or be housed in a separate location? There already existed the precedent for separation in Schaffer Library with the existing phonograph records and the newly acquired audio CDs. Should the material circulate? Should the material have different borrowing terms than the regular circulating collection? Should the material have Library of Congress (LC) classification? Should specified hardware and software be purchased for access to certain in-house resources? The list of questions grew. Most importantly, responsibility for training, creating processing procedures, and cataloging this new assortment of material had to be assigned. Lists were made, policy was decided on, changed, and changed again. The multimedia committee became less involved as the cataloging department took action, but issues continued to arise between departmental sections and committees relating to circulation, storage, preservation, and security.

Literature on managing multimedia within a main circulating collection illustrates that we were not alone with our questions. Weihs recommended intershelving of nonprint resources in the main library shelves to promote easy browsing and retrieval by patrons.[2] Other libraries grappled with concerns of circulation, access, and processing as well. The debate continued, evident at the American Library Association (ALA) Midwinter Meeting, where Errickson discussed the advantages and disadvantages of storing disks and CD-ROMs with or without the accompanying book, and

Gail Golderman (goldermg@union.edu) is Electronic Media Librarian, Schaffer Library, Union College, Schenectady, New York.

Seaman and Carter provide insight on the University of Colorado library's policy to separate accompanying materials from books.[3] Changing technology will continue to foster debate for our profession as noted by the European Library Automation Group (ELAG) 1999 Library Systems Seminar on Managing Multimedia Collections.[4] Additionally, a recent study by the Texas A & M University Libraries provided a set of recommendations for access to material with accompanying CD-ROMs, which resulted in a revision of current library policy of housing material separately.[5]

When I first joined the cataloging department we were comfortable with processing audio CDs and books with accompanying software, which at that time primarily consisted of 5¼-inch floppy disks. Our procedure was to separate the workflow from regular new book cataloging and with books with accompanying software, send the material to the systems librarian, who would review the book and create a back-up copy onto a 3½-inch disk. The book and back-up disk would be sent back to the cataloging department, where the designated staff member would catalog and process both the software and new book. The back-up disk and the book would then be available to circulate together, although the software was stored behind the circulation desk. Soon computers with 5¼-inch drives became obsolete and books began to be accompanied solely by 3½-inch disks. We still chose to create a back-up disk in the event of corruption or loss and the copy would circulate. In the case of audio CDs, the staff member would wait for a book truck to accumulate and then designate time to assign cutter numbers, locate OCLC records, catalog, and process.

The library began to make progress with cataloging additional formats. A collection of folk music on cassettes was added to the circulating collection. The library chose to integrate this material with the rest of the circulating collection. Videos were purchased for the first time on a small scale when faculty requested a title for class viewing. A separate fund was established for video purchases and we soon had a mini-collection, set out on shelves for browsing. The library also began to purchase audio CDs to replace the collection of aging phonograph records that had been part of the circulating collection. The audio CD procurement has evolved from that initial proposal to scheduled purchases several times per year with a separate fund from which material is acquired. Our CD collection, which is quite popular with students and faculty, has grown tremendously from that original fifty-CD purchase and includes a wide variety of artists and genres. The collection does circulate and is shelved separately for easy browsing.

Not much has changed and, at the same time, so much has changed in ten years regarding cataloging books with accompanying software and videos. We still perform the same original procedures to an extent, although our cataloging standards are more complete, and our processing methods have been greatly simplified. Of course, now CD-ROMs have become the media of choice for authors and publishers (with Web sites thrown in for errata and update information) even though floppy disks still accompany a small percentage of books. While we still make back-up disks, we do not burn CDs to maintain a safe copy and have not had problems with corruption or theft. We also fully catalog CD-ROM databases and selected government document CD-ROMs based on reference value and data type, as both circulating and noncirculating material. While the government document CD-ROMs are available for browsing, the remaining media that is available for checkout is held behind the circulation desk.

Within the past two years the library inherited responsibility for well over one thousand video titles, most of which were not integrated into our online catalog. We presently handle requests for new video and CD-ROM material the same as the acquisition of new books. Although we assign LC call numbers to our new video purchases, we have kept the local numbering classification originally assigned to the inherited videos and the collection is no longer browseable, but can be searched via the online Web catalog. Several new videos arrive on a weekly basis although the majority tends to be processed as "rush" material for a particular faculty's classroom or reserve-room use. While the multimedia committee has long-since dissolved, we still contend with collection, housing, database display, bibliographic standards, circulation, software, and hardware issues. It of course becomes more complicated with each new format we introduce into the collection, which more often than not makes past decisions obsolete or incorrect.

The latest challenge has been DVDs, and in fact we recently received our first double-sided disc. This posed a question as how best to physically process the material, based on our current standards and procedures. Several initial concerns included whether we would circulate DVDs as we do videos, or designate equipment in the library for viewing. Our video policy had changed over the years to one that allows borrowing, at first to in-house use only, (except for faculty) and then recently we have reverted back to a circulating status for the entire college community. We had equipment already allocated for viewing videos in our new group study rooms and the library recently designated equipment for in-house DVD use as well.

An initial reorganization of the cataloging department occurred in 1995, with the addition of the electronic media librarian position. This position was primarily responsible for departmental processing of nonprint material and the inclusion of

electronic access to periodicals into the online catalog. At that time we thought it most efficient to designate one staff member to be the official nonprint assistant and to keep responsibilities separate from the other cataloging staff. This staff member would incorporate audio CDs, books with accompanying software, videos, or CD-ROMs into the daily workflow. With an unforeseen vacancy in that position, all cataloging staff had to lend a hand. Our written documentation was acceptable for audio CDs, books with software, and videos, and sketchy or nonexistent for the rest. Of course any documentation or procedure that did exist had been clearly understood by the assigned staff member who from the beginning kept the operation running quite smoothly. With the vacancy, we realized the information was totally unfamiliar to the rest of the cataloging department and needed to be updated and restructured not only to meet our current bibliographic standards, but to coexist with our ever-changing library-wide policies as well. It is possible that we never realized the volume of multimedia material we acquired on a regular basis, and not having complete and current documentation for all aspects of the routine was devastating to the department. This set us back quite a bit with backlogs and a great deal of scrambling to train not only ourselves, but to begin the creation of new and the updating of existing procedures.

Through all this development, what is radically different for our technical services section, and specifically our cataloging staff, is that we now routinely catalog a variety of multimedia material. Most importantly for the department and the daily workflow, this work is no longer isolated with one individual but is handled by everyone in the department. We have yet to mix different formats on a given daily work truck, but we are indeed moving in that direction. Our goal is to have books, videos, CD-ROMs, or material with multiple formats handled collectively on the cataloging side. We still will continue to separate formats for the student assistants to productively complete the physical processing. We have documented each step on our technical services Web site, which is updated regularly.[6] Each specified format has three associated components:

- cataloging procedures, which offer general statements about the particular format and relevant information to be utilized during the copy cataloging process with examples of complex situations;
- an editing guide, with notations for the fixed or variable MARC fields that are necessary to add, edit, or delete; and
- a set of detailed procedures with illustrations for physically processing the material that exists not only for our staff members, but also, more importantly, for our student assistants.

When initiating a new format, the entire cataloging staff uses the team approach to examine the procedures, ask questions, and offer suggestions. After the changes have been implemented and updated to the Web site, the new format is cataloged. Staff takes part in the actual labeling and other physical processing routines, so that all feel comfortable to train and assist students, to fill in during breaks without students, or to handle rush material.

For this latest stage to be implemented, we took a hard look at the department's needs and all the staff members' experiences and skills. We invested a great deal of time and energy on the second phase of the reorganization of the cataloging team. With proper documentation on hand, everyone accepted new challenges, shared responsibilities, and began to feel comfortable with cataloging a book truck of videos, or audio CDs, or books with accompanying material. Now that the routine is nearly complete, our next goal is to thoroughly review all formats, change any inconsistencies, simplify where possible, and it is hoped, continue to pat ourselves on the back when we realize how far we have come.

Having online documentation allows us to make changes in order for all to stay abreast with current standards, not only locally, but nationally as well. This process has been invaluable for the department, the library, and the campus community, with regards to handling the flow of all material in an effective, timely, and simplified manner. The journey continues as the technology changes and we await the next format to be introduced into our collection.

References

1. Sheila Intner, "Equality of Cataloging in the Age of AACR2," *American Libraries* 14, no. 2 (Feb. 1983): 102–3.

2. Jean Weihs, *The Integrated Library: Encouraging Access to Multimedia Materials* (Phoenix: Oryx Pr., 1991).

3. April Errickson, "Methods of Circulating Accompanying Materials," presented at ALA Midwinter Meeting, Library Administrative and Management Systems and Services Section, Washington, D.C., Feb. 16, 1997. Accessed Jan. 29, 2001, www.ala.org/lama/committees/sass/errickson.html. Scott Seaman and Nancy Carter, "Do Not Desensitize: Developing a Policy for Accompanying Materials at the University of Colorado," *Information Technology and Libraries* 16, no. 2 (June 1997): 86–92.

4. European Library Automation Group, "Managing Multimedia Collections," Twenty-third Library Systems Seminar, Bled, Slovania, Apr. 21–23 ,1999. Accessed Jan. 29, 2001, www.nuk.uni-lj.si/elag99.

5. Katherine Weimer et al., "Security and Access to CD-ROMs Accompanying Books: Data and Recommendations," *Library Resources & Technical Services* 44, no. 4 (Oct. 2000): 201–08.

6. Schaffer Library Technical Services Home Page, Union College, Schenectady, New York. Accessed Jan. 29, 2001, www.union.edu/PUBLIC/LIBRARY/TS.

Creating an Instant Messaging Reference System

Jody Condit Fagan and Michele Calloway

Libraries are expanding digital reference services to include chat and instant messaging. Various companies have developed systems for purchase that meet this need without requiring users to download new software or learn a new skill. These systems share many features in common, including transcript storage, sending URLs, and pushing Web pages. Several librarians at an Association of Research Libraries academic library wished to test instant messaging reference with the ability to control and experiment with different implementations of these features. This article describes the features of instant messaging reference systems and the in-house development of a system that incorporates them.

Online reference, specifically the flavor known as chat or instant messaging reference, is definitely a hot topic. At the American Library Association (ALA) 2001 Annual Conference, the topic cropped up in poster sessions, in programs, at the exhibits, and in planning committees.[1] There are many resources available for libraries interested in starting online reference services, including bibliographies, registries of services, and electronic discussion lists.[2]

Morris Library at Southern Illinois University–Carbondale has offered digital reference through e-mail and online Web form for about three years. Use has increased dramatically each year; in FY00, it received only 348 questions, while during FY01 it received 629 questions. Some of the increase in digital reference traffic is credited to the redesign of the library's home pages, in which the e-mail service was made more prominent. Also during the redesign, the library implemented a Web site search that categorizes results by resource type and ranks items by librarian-assigned metadata.[3] While the site search was a popular tool, library patrons used it ineffectively, with over half the searches getting no matches. Many were typing in specific topics as if they expected books and articles to magically appear. Their desire for reference assistance on the Web was an inspiration to expand services.

During a study of the site search in which transactions were logged, the number of searches approached the number of desk reference questions library wide.[4] This made it clear that some patrons wanted to get help from their workstations, and they wanted to use a tool that provided the advantages of site search. We hypothesized that the simple input box was the attraction of site search. It delivers instant results and is (seemingly) anonymous. Instant messaging services have these same advantages.

The usage statistics from libraries currently using instant messaging reference show patrons are beginning to show interest, but not at an overwhelming rate. Table 1 shows statistics for several libraries that were reported at the ALA 2001 Annual Conference poster sessions.[5]

We wanted a system that would take advantage of the success of site search, which we felt was due to its up-front implementation at the top of our Web pages, its speed, and the implicit promise of instant results. In a digital reference survey by Joseph Janes, only half of the one hundred fifty libraries surveyed linked to their digital reference (e-mail or Web form) from the library home page.[6] Ease of use and access were of top importance; this meant simplicity, high speed, and no technical difficulties.

What's Out There: Library Instant Messaging Reference Systems

After deciding to try out instant messaging reference, we needed to find out what software was available, what features were offered, and how well different systems met our requirements. Steve Coffman from Library Systems and Services (LSSI) says there are "fifty or more vendors developing software with varying degrees of sophistication at present."[7] Products to be evaluated for this study were selected by looking at the archives of the Dig-ref electronic discussion list and LiveRef: A Registry of Real-Time Digital Reference Services by Gerry McKiernan.[8] From these sources three products were selected for review: Live Assistance, LivePerson, and Virtual Reference Desk (VRD) (LSSI). Each of these companies' Web sites was visited and their chat demonstrations were evaluated.[9] Transcripts of the demonstrations, documentation available on the Web site, and e-mail from company representatives were used to gather information about these products. Although all three companies provided a live demonstration of what patrons experience, none provided a demonstration of the library side of the product. This made it difficult to determine if the various products would meet staff expectations.

Paul Constantine's presentation on Cornell University's LiveHelp Service at the 2000 Virtual Reference Desk Conference outlines several key points that libraries should look for when selecting a live reference service.[10] The following list was modified to use as a checklist for comparing the three software products:

1. Does not require special software or hardware for the user.
2. Offers ability to push patrons' browser to a desired page.
3. Offers ability to send active embedded URLs.

Jody Condit Fagan (jfagan@lib.siu.edu) is a Social Sciences Librarian, and **Michele Calloway** (mcallowa@lib.siu.edu) is an Undergraduate Librarian, Morris Library, Southern Illinois University–Carbondale.

Table 1. Use of Real-Time Reference at Selected Institutions

Institution	Software	Hours	Usage	Usage per Hour of Service
University of Florida	eShare's Net Agent	M–F 10 A.M.–5 P.M.	375 questions in 3 months	5.3
Golden Gateway Library Network*	LSSI	unknown; 6 hours	5–6 questions in 6 hour shift	1
Carnegie Mellon University	live person	M-F 1–5 P.M.	60–80 questions per month	3–4
North Carolina State University	LSSI	M-F 9 A.M.–5 P.M.	100 questions per month	less than 1
Alliance Library System**	LSSI	24x7	474 transactions in approx. 3 months	less than 1 every four hours

* Includes 22 libraries and 120 staff members participating in online reference.
** Includes eight institutions.

4. Offers ability to survey user.
5. Supports multiple browsers.
6. Logs transcripts for future reference.
7. Provides searchable transcripts.
8. Offers ability to create preformatted answers for frequently asked questions.
9. Identifies patrons by IP address.
10. Offers fast response time.
11. Minmizes technical difficulties.
12. Is reasonably priced.

Summary and Evaluation of Library Chat System Features

Installation of Software

None of the products required installation of software for the user or for the library. Although many free instant messaging and chat programs exist, most require installation, which is why systems are being developed specifically for Web customer support and chat reference.

Push the Patrons' Browser

All three provided the ability to push patrons' browsers to Web sites but implemented this feature in different ways. Live Assistance opened a new window on top of the chat window, LivePerson opened a new window behind the chat window with both pages visible, and VRD opened the new window in a frame next to the chat window. LivePerson's chat session crashed twice when a Web site was pushed.

Active Embedded URLs

VRD did not support sending active embedded URLs in the chat box but when a patron exits the system a summary list is provided of hyperlinks to sites visited in the session. The links are also active in the transcripts that are sent via e-mail to patrons. Both LivePerson and Live Assistance make URLs active hyperlinks in the chat windows. Live Assistance also allows patrons to send URLs as active hyperlinks.

Surveying the Patron

Live Assistance provided a password-protected administrator account that allows the library to create and edit surveys. LivePerson offered surveys in their Corporate edition. VRD did not have a survey feature at the time of writing; however, company representatives say it is in development.[11]

Supports Multiple Browsers

Live Assistance documentation claims that their product works with all standard browsers version 3.x or better. VRD's online documentation could not be viewed with Netscape 4.7. When viewed through Internet Explorer, the VRD documentation recommended Internet Explorer or Netscape 4.0 or greater. According to customer service representatives, LivePerson Pro supports all browsers.

Logs Transcripts for Future Reference and Searchable Transcripts

Live Assistance creates a database of transcripts that can be searched and allows librarians to assign categories to each transaction for report purposes (such as circulation and reference). Librarians can also type additional notes for follow-up using a post chat wrap-up tool. Live Assistance does not automatically make transcripts available to patrons, but the librarian can e-mail a transcript after the chat session is over.

LivePerson logs transcripts and creates patron histories. These patron histories identify repeat visitors with cookies. Since LivePerson Pro does not automatically make transcripts available and they do not require patrons to

log in with an e-mail address, the librarian would have to ask the patron for their e-mail during the chat session and cut and paste the transcript into an e-mail message. It is possible to search transcripts by date, librarian, for the past week, or for a given month.

VRD logs transcripts but they are not searchable. Patrons are automatically sent a summary of their chat session as they close the chat and also via e-mail. In one demonstration, the e-mail of the transcript failed to arrive, and the customer service associate noted she had not received a transcript of the conversation in question either. When the patron closes the chat, a window opens, listing all of the Web sites or documents that were sent by the librarian.

Ability to Create Preformatted Answers for Frequently Asked Questions

This feature was available with all three products. Since only the patron side was seen during demonstrations, the various implementations of this feature are unknown.

Identify Patron by IP Address

Live Assistance identified the IP address of the patron for the librarian to view during the chat. In Live Assistance, librarians can also view the patron data collected prior to the start of the chat and update the patron data during the chat. LivePerson uses IP addresses and cookies to identify a repeat customer and build a user history. If cookies are not enabled, the history feature will not be available to the librarian. VRD does not track by IP address. Some libraries don't see tracking as a positive feature due to privacy concerns, particularly when an outside company is providing the service.[12]

Fast Response Time

All three systems were reasonably fast once the chat session began. Some of VRD's more advanced delivery features such as slide show and escort took some time, and it seemed to take a long time before an operator replied to the initial inquiry during a demonstration. They do send several messages indicating that someone knew a user was waiting and that "a senior librarian would answer shortly." These response times may not replicate the real world where patrons or librarians may be using various connection speeds.

No Technical Difficulties

All three systems were easy to use and had a simple patron interface. Live Assistance and VRD required the patron to log on with a name and e-mail address while LivePerson provided just an input box. One of the chat sessions with VRD had some system slowdown problems, and the e-mail transcript of one of the conversations never arrived. In two separate demonstrations with LivePerson, the chat session crashed when the company representative pushed a page.

Price

LivePerson Pro Edition was the least expensive of the three tested. The Pro edition offered the basic chat features for $89 per month. LivePerson Corporate Edition allows surveying and more reporting features for $350 per month. LivePerson does not provide on-site training, free or otherwise, but provides live help through their home page, and one-on-one usage instruction from account managers. Live Assistance was the next most expensive at $150 per month. They also charge a one-time fee of $500 for installation and over-the-phone training. The most expensive by far was VRD at $6,000 per year ($500 per month). VRD also charges a one-time fee for setup and training of $8,000. Two-day on-site training sessions are conducted by experienced online reference librarians. This initial training is supplemented with regularly scheduled online updates.

Summary of Features and Product Comparisons

A summary of the above features and products is provided in table 2. All three products provided the same core capabilities. VRD provides more features for sending information to patrons but at $14,000 was deemed too expensive, particularly because potential usage is unknown. Also, VRD did not provide some of the features we were looking for. The VRD system does offer nice features and has the advantage of specializing in library services.

Live Assistance had a more reasonable price and offered most of the features that we thought were important. They also had ample documentation on their Web site and very helpful customer service representatives who were knowledgeable about the needs of library customers.

LivePerson Pro was the least expensive product and offered the basic services we wanted but unlike the other two providers, LivePerson did not seem interested in the special needs of the library customer.

Our Instant Messaging Reference System

Although Live Assistance would have been our choice of the three systems we evaluated, we decided to create our own instant messaging system incorporating the features on the

checklist. This would allow us to customize the way it would be presented, how quickly it responded, how it operated, and what kind of data it stored. We wanted to provide patrons with a service similar in appearance to our site search—but with librarians answering the queries rather than a search engine. Creating our own system would also allow us to protect the privacy of transcripts and patron data and develop customized reports. Also, our library was not in a position to purchase a system at this time. The administration was supportive of our ideas, but not ready to commit to a new service. External funding would have taken time to acquire, and we wanted to get started right away.

Morris Library is not the first to use a homegrown system. Marc Meola and Sam Stormont created a chat program called TalkBack, described by Meola as more of an "Internet paging" system using pop-up windows than a true chat.[13] In their test from fall of 1998, librarians at the physical reference desk additionally answered questions through TalkBack. In six weeks (seventy-six hours per week), they received eighty-six questions through TalkBack. Meola and Stormont further refined their program and have an updated program called TalkNow, which uses PHP and MySQL.[14] Comparing the features of their system and the three systems above helped us in the development process of our own system.

Our system was designed to meet the requirements of the preceeding checklist; figure 1 provides a brief summary, while a more detailed description of how the system works follows.

The Morris Library Online Reference system uses a Web browser for both staff and patron clients, and was tested with success in both Internet Explorer (5+) and Netscape Navigator (4.0+ but not Netscape 6). Both staff and patron clients are extremely responsive, with nearly instant response when both parties are directly connected to the Internet, and only a second or so delay when one party is on a 28K dial-up connection. Since the only requirement for the system is a Web browser, library staff and patrons can be in virtually any location. A test between the library, located in southern Illinois, and Cleveland, Ohio, resulted in no significant time lag.

Table 2. Comparison of Product Features

Features	Live Assistance	Live Person	Virtual Reference
No special software or hardware	Yes	Yes	Yes
Push the patrons browser	Yes	Yes	Yes*
Send active embedded URLs	Yes	Yes	No
Ability to survey user	Yes	No **	No
Supports multiple browsers	Yes	Yes	Yes
Creates transcripts	Yes	Yes	Yes
Searchable transcripts	Yes	Yes	No
Preformatted answers for FAQs	Yes	Yes	Yes
Identify patron by IP address	Yes	Yes	No
Fast response time	Yes	Yes†	Yes
Technical difficulties	No	Yes	Yes
Training/setup fee	$500	$0	$8,000††
Subscription price	$150/month	$89/month	$6,000/yr

* VRD also offers delivery options of escort, slideshow, and capture.
** LivePerson Corporate edition provides surveys for $350 per month.
† LivePerson was fast, but experienced technical difficulties when pages were pushed.
†† Two-day, on-site training.

When a librarian is online, the library's home page (see figure 2) and the footer on every subsequent page have an input box and an invitation to patrons to enter a question if they need help, thus providing access to our instant messaging reference service in a similar manner to the site search. After clicking the submit button, a patron is connected with the service.

Patron Interface

The patron window is split into two frames (see figure 3). The top frame contains brief instructions, an input box with a submit button, and a disconnect button. The bottom frame contains either information about the connection or a conversation.

After a staff member responds, the bottom frame patron window changes from "attempting to connect" to a display of the conversation so far: the patron's first submission and the staff member's first response. The top frame changes to include a graphic showing that the patron is online, and also displays the screen name of the staff member at the other end of the connection.

Once the patron is connected, the conversation continues as long as both parties are interested. The patron can disconnect at any time with the Disconnect button, and is then taken to a screen that offers the option to reconnect, close the window, or display a transcript of the conversation. This screen is also where a brief Web

1. Does not require special software or hardware for the user. Requires only a Web browser on both ends, similar to all three systems.
2. Offers ability to push the patrons' browser to a desired page. Opens new pages in the background window, but leaves the focus on the online reference window.
3. Offers ability to send active embedded URLs. URLs appear as active hyperlinks to both staff and patrons. Only staff can send active URLs
4. Offers ability to survey user. Survey pops up upon disconnecting; questions can be added, changed, or removed as staff desire. Surveys are linked to conversation transcripts and user information.
5. Supports multiple browsers. Tested with success in both Internet Explorer 5 and Netscape 4.0+.
6. Logs transcripts for future reference. Transcripts are logged and can be printed, saved, or emailed by both staff and patrons.
7. Provides searchable transcripts. Transcripts are stored in a database and can be searched by any field including staff member, IP address, keyword in full text.
8. Offers ability to create preformatted answers for frequently asked questions. Each staff member can create their own subdatabase of frequently asked questions. A pull down list facilitates answers requiring edits, while buttons send preformatted answers instantly.
9. Identifies patron by IP address. IP addresses are visible to staff during and after the conversations. A database also notes if the IP address is on or off campus, or is in the library.
10. Offers fast response time. Response time is a few seconds for all users, whether on-campus, out-of-network, or out-of-state.
11. Is free of technical difficulties. Our system has had no technical difficulties since implementation.
12. Is reasonably priced. Our programmer estimates development costs at approximately $2,000 based on his salary.

Figure 1. Features of Morris Library's Instant Messaging System

evaluation of the service is posted. In case patrons close the browser window or walk away from the workstation without disconnecting, the program periodically tests the patron's window to see if it is still active. If it is not, the staff member is notified that the patron's window probably has closed.

The patron transcript includes a full record of the conversation, including the time of each line of the conversation, the screen name of the staff member, and any URLs sent during the conversation. The patron may use the browser's print, e-mail, or save functions to preserve the transcript for future reference.

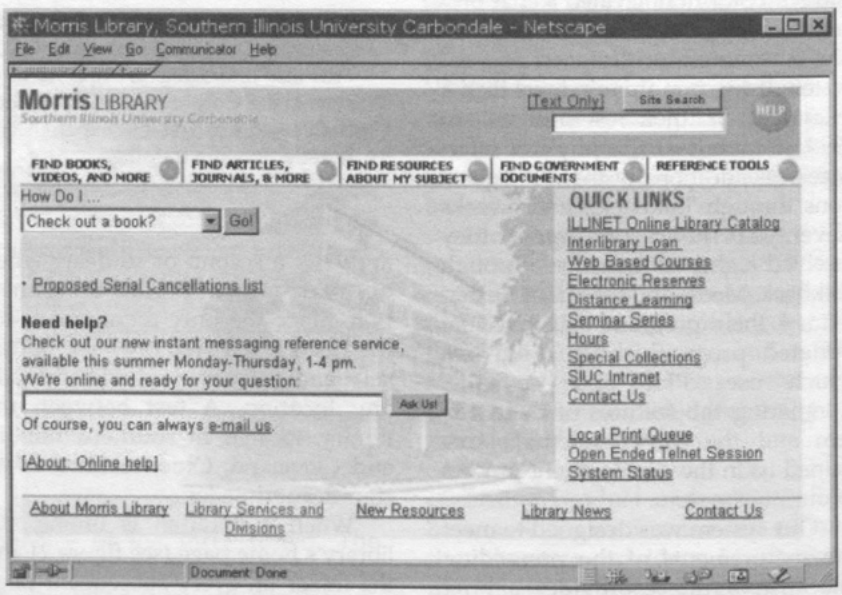


Figure 2. Morris Library Home Page

Staff Interface

The staff interface begins with an offline screen. The top frame of the screen offers three options to connect: "Willing to backup others," "Ready to answer questions," and "Able to take redirects only." The first option is equivalent to being on the front line of a reference desk, while the second is similar to being in a back office but monitoring desk traffic to help if things get busy. The third option was designed for nonreference staff to be available for specific types of questions—for example, a member of the circulation staff might be available for redirected conversations concerning renewals or overdue fine complaints.

The three connection options are followed by a password input and a Connect button. There is also a link to the maintenance pages. The bottom frame shows connection status (such as "You are offline") and conversations when they occur. When staff members choose a connection option and enter their passwords successfully, the bottom frame changes to

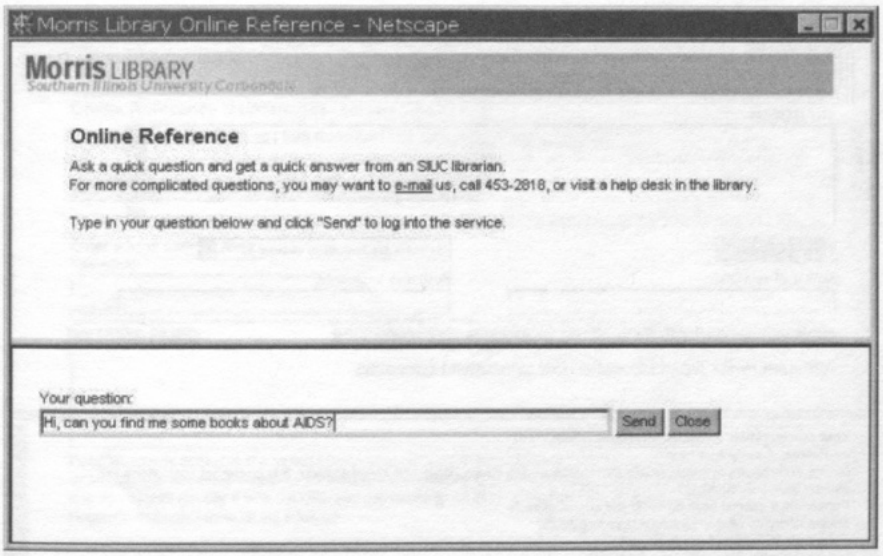


Figure 3. The Patron Interface

"Waiting . . . ." When a patron connects to the service, the staff screen top frame changes to an input workspace. The bottom frame shows the conversation and information about the patron: IP address, Web browser, and the time the conversation began (see figure 4). The lines of the conversation are listed in reverse chronological order, so that the most recent remark in the conversations is listed first. This line is also in bold text. Each line is labeled "Patron" or by the staff members' nicknames.

The top frame of the connected staff window identifies the name under which the staff member has logged in and has various input boxes for sending messages in different ways.

The largest text area is for free-text input, and is immediately followed by the Send button, making it easy to select after entering a comment. Other input boxes allow staff to select a preformatted answer, send a URL as a hyperlink the patron can click on, or enter a URL to which the user's browser is pushed. If a preformatted answer is selected, the text appears in the main input box, giving the staff person time to edit any details before sending the message. The row of buttons following the input boxes are staff-specific and are similar to the preformatted answers except they send their associated messages instantly. Their intention was for the most common of comments, for example, "Welcome to online reference," or "Hang on, I'm looking for the answer." The Warn buttons send warning messages of increasing severity to ask if the patron is still there and interested in continuing the conversation, while the Kill button disconnects the patron.

If other staff members are logged in, their names appear at the top as options (radio buttons) to send messages to. The system defaults to "patron," however, as that is who the staff member is supposed to be communicating with when connected. If staff send each other messages, these lines only appear in the staff interface and are not displayed in any transcripts. A redirect feature is also available when multiple staff are logged in, allowing staff to pass entire conversations between each other. This was designed to facilitate referrals.

Staff Maintenance

The staff maintenance pages have five screens: View Summary, which shows current settings (see figure 5); Edit Buttons and Edit Answers, which allow staff members to add or change their preformatted buttons and answers on an individual basis (see figure 6); Edit Nicknames, which allows users to add or edit nicknames and passwords; and View Conversations, which allows for viewing or simple searching of past conversations (see figure 7). Staff can see others' settings and conversations by selecting a name from a pull-down menu before they enter the maintenance section. Changes, updates, and deletions may be made while the staff person is logged into the service.

How Does It Work?

Online reference is mounted on a UNIX machine (Sun/Solaris) using the Apache Web server and uses Perl, MySQL, JavaScript, and HTML. All four languages are free and ample documentation is available on the Web.[15] Perl has excellent process, file, and text manipulation facilities, particularly with Web applications. This made it a natural choice for the backbone, as described below. MySQL, a database management system, is easy to install and can store and retrieve large amounts of data very quickly. Anyone who can write SQL queries can query it. JavaScript and HTML, are necessary for the manipulation of Web browsers, which serve as the client.

In summary, the operation of this instant messaging system can be thought of as a database that contains the full text of a conversation. By submitting Web forms, the staff member and the patron update the same database table. Updated results are displayed in the bottom frame of each interface, so that each sees the current

conversation. The mechanisms that coordinate these actions are written in Perl, the information is stored in MySQL, and the display is presented with both JavaScript and HTML.

Perl

Perl is the actual program which sends the MySQL queries, JavaScript scripting, and HTML tags to the correct place at the correct time. The MySQL queries, the JavaScript scripts, and the HTML tags all reside in the Perl code.

The program itself resides in three files written in Perl: chat_patron, chat_staff, and chat_maint. The first two are the actual instant messaging programs, while chat_maint provides an interface for database management. All three Perl scripts use the Perl DBI module, which contains the necessary functions to interact with the MySQL database.[16]

Chat_patron and chat_staff directly communicate with each other very little. Instead, each program takes the input provided by the staff or patron and saves it into a database. The other program then reads and displays the conversation from the database. Because of the speed of MySQL, this querying process can take less than a second. The only variables directly passed between the two programs are the session ID and the recipient, which link the clients to the correct database records. When a conversation is redirected from one staff member to another, only the recipient variable is changed. In this way, a new staff member receiving a redirected conversation gets all the information the previous staff member had—they literally pick up the conversation where the last person left off.

Although the patron interface looks quite simple, the underlying procedures are quite complex. Chat_patron generally alternates between two procedures. The first, bottom_connect, contains the intricacies of creating a file for the two programs to share, properly identifying which one it is and who the recipients are, and updating the status of staff

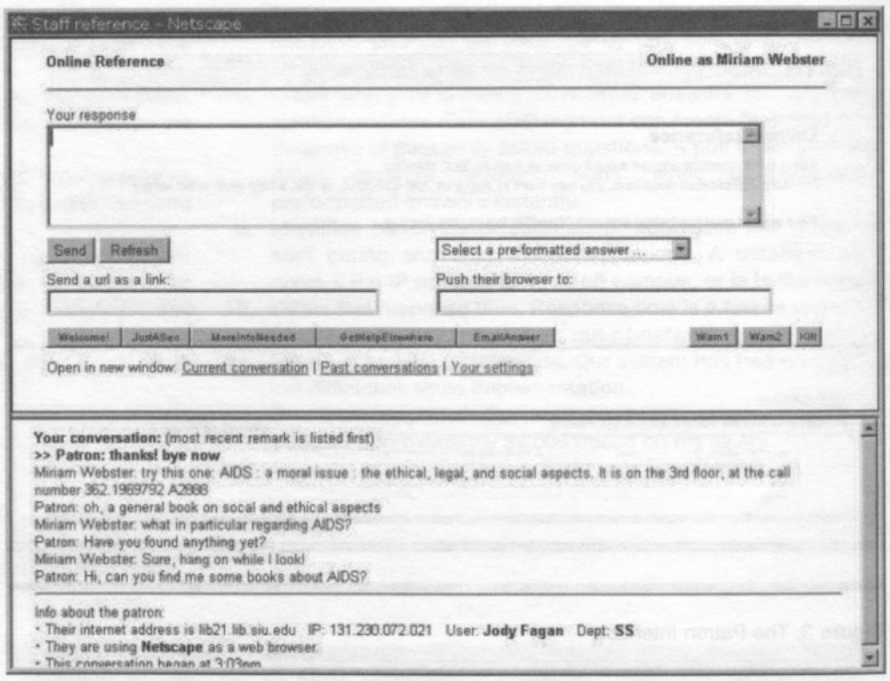


Figure 4. The Staff Interface

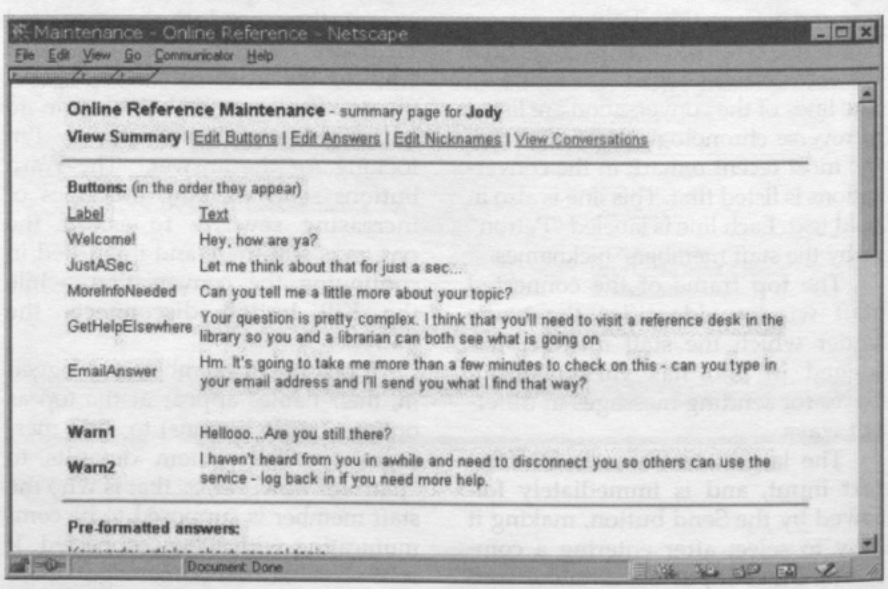


Figure 5. Staff Maintenance, Main Screen

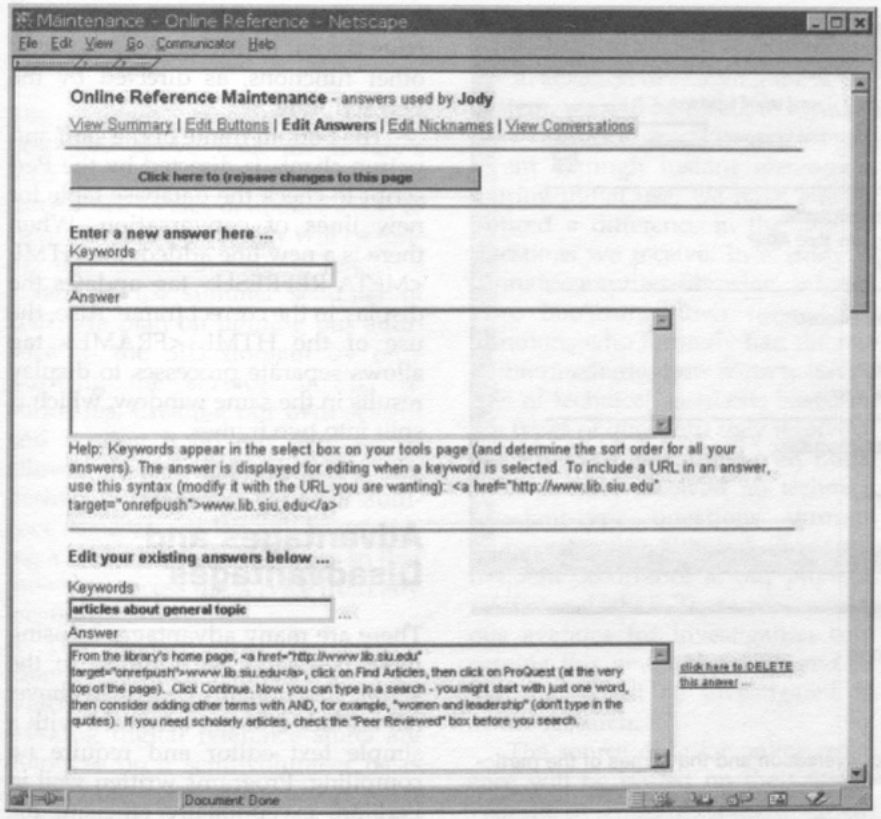


Figure 6. Staff Maintenance, Preformatted Answers Screen

and patrons connecting and disconnecting. The second, do_work, checks to see if anyone is trying to talk to someone else and lets the involved parties know they are connected.

Chat_staff is very similar to chat_patron, in that it is reading from the MySQL database and posting the results of a query to the bottom frame of the staff window. However, the staff client also checks to see if the patron process is still alive—(about every thirty seconds) in case the patron has closed their window. The staff side is also watching for a "warn" file, which would indicate that a message from another staff member has arrived, or a warning that a patron is waiting to get on the system. The warn file reloads the staff interface to include the new information.

The two programs share many features involved in providing clear conversations; one example is that actions such as pushing the patron's browser or sending a URL as a hyperlink are sent as lines of the conversation—but the actual characters are replaced with a message that makes sense to a librarian: instead of

> <script>open("http://url2goto.com","onrefpush")</script>,

the conversation would display

> [opened page: http://url2goto.com].

Chat_maint contains the insert, append, and query statements necessary to update, edit, and display the database of conversations, staff members, and preformatted answers. The Perl script allows users to interact with the MySQL database through Web forms.

MySQL

MySQL is an open source relational database management system. It is freely available for download on the Internet, is easy to install on several platforms and operating systems, and supports several client programs and programming interfaces, including the Perl DBI module.

The database is structured into seven tables, which are listed in figure 8.

SQL queries within the Perl scripts query the database for several functions:

1. Create a staff member's customized workspace, including preformatted answers and buttons.
2. Update and display the current conversations to both staff members and patrons.
3. Show transcripts of conversations to both staff members and patrons.
4. Display the current settings of staff members' profiles and allow additions, editions, and deletions through a Web form.

JavaScript

Functionality provided by JavaScript includes most windows-based management tasks in both staff and patron Web browsers. One of the most important things JavaScript takes care of is demanding focus from the correct window, frame, and input box. Focusing simply means keeping the most important window at the front of the screen, preventing the online reference window from getting buried under other browser windows that might be open. The Perl scripting initiates the focusing action at the correct time, but a JavaScript method

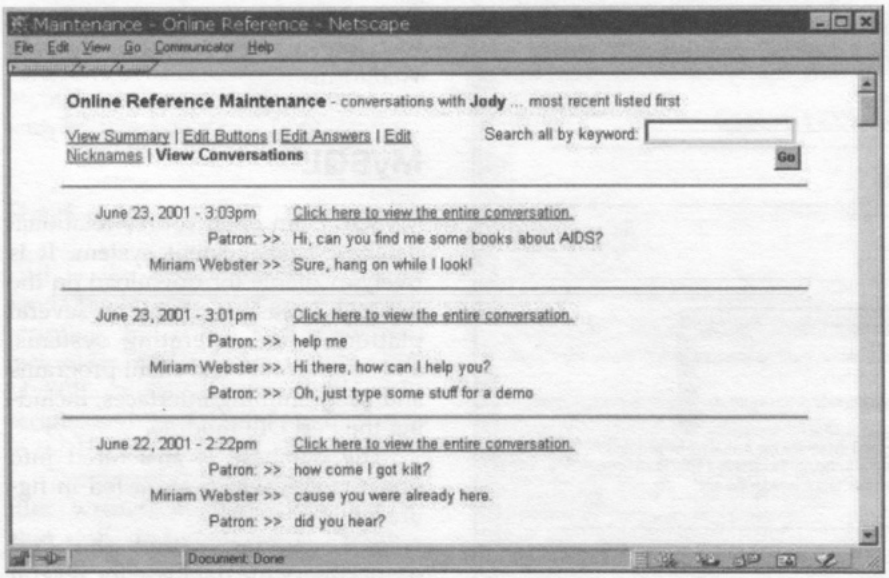


Figure 7. Staff Maintenance, Viewing Conversations

Conv_content: contains the actual lines of the conversation and the names of the participants, the time and date, and the session ID.

Conv_headers: contains patron information such as IP address, type of browser used, which staff member answered the call, and the time and date, and the session ID number.

Staff_answers: contains each staff members' pre-formatted answers, sort keywords for each answer, and the associated staff ID.

Staff_buttons: contains the button labels, the text each button delivers, the associated staff ID, and a numerical field by which the buttons are sorted.

Staff_names: contains only the staff ID number and the associated staff person's name.

Staff_nicknames: contains the nicknames, the staff ID number, and the password associated with the nickname.

Patron_response: contains the data from the patron evaluation form.

Figure 8. Descriptions of Database Tables

accomplishes it. Under direction from Perl, JavaScript methods also open and close windows, creating them to be the correct size and with the appropriate scrollbars and names.

JavaScript also plays a role in managing the various HTML forms. In the staff interface, for example, JavaScript transfers the text of the preformatted answers to an input box where they can be edited. JavaScript also confirms actions with dialog boxes, for example, deleting records from database tables.

HTML

HTML is used to provide a graphical interface for the online reference system. In addition to fonts, colors, and page layout, HTML performs several other functions, as directed by the Perl scripting.

The bottom frame of the staff and patron clients is directed by the Perl script to check the database table for new lines of conversation. When there is a new line added, the HTML <META REFRESH> tag updates the display in the correct frame. Also, the use of the HTML <FRAME> tag allows separate processes to display results in the same window, which is split into two frames.

Advantages and Disadvantages

There are many advantages to using these languages in addition to the simple functionality described above. All are free and can be edited with a simple text editor and require no compiling. Programs written well in Perl are exceptionally efficient; the same is true of MySQL.

The only disadvantage of using Perl is that programming knowledge is required to install and configure the system. As a database system goes, MySQL is standard and therefore would present no problem to anyone familiar with writing SQL queries. However, anyone familiar with the other three can make changes within the Perl script. For example, someone who knew HTML could enter the Perl code and change such things as colors, fonts, and words.

Almost all of the disadvantages in using these languages to create an instant messaging system resulted from the system's reliance on Web browsers as clients. For example, it is challenging to write JavaScript that is acceptable to both Netscape Navigator and Internet Explorer. Both browsers have internal rules for handling JavaScript commands, and use slightly different flavors of HTML. This made it difficult to design a system that worked identically in both languages.

Testing the New System

The instant messaging system described above took the library's Web programmer the equivalent of three work weeks, or about $2,000, to complete. Morris Library will be testing this instant messaging reference system in the summer semester of 2001. We plan on limiting our audience to the SIU domain by only showing the service to SIU IP addresses. Creating our own system and hosting it on our own server allows us to authenticate by any desired IP range and limit our audience as necessary. We will be providing a URL to distance learners so that they may access the service from any location.

The service will be staffed Monday through Thursday, 1–4 P.M., which are our peak hours of use. Staff working digital reference shifts are doing so on a semi-volunteer basis and do not necessarily receive relief from other duties. In general, staff are taking one three-hour shift per week. Staff participants meet regularly to evaluate experiences and suggest improvements. As a result, we may change the way we offer access to the service, the wording of links to the service, staffing, or other options as we respond to patron input.

If our system offers good service to patrons and it is something the library is interested in pursuing, policy issues of greater depth will be examined. These would include who we will offer service to, how we will staff the service, and issues of reference quality and evaluation. We are continually reexamining our system requirements to meet our needs.

Future Exploration

By creating our own system, we will be able to tweak it to accommodate our use. We are able to quickly insert and edit survey questions, search our transcripts, and manage the data with great flexibility.

In addition to studying the actual system, we will be trying to evaluate how reference practices may be different through instant messaging. During initial use, we have already noticed a difference in the type of questions we receive. In a study of human-computer-librarian interaction, SooYoung Rieh found that librarians who formerly had the role of intermediaries now seem to have a role of technical assistants based on the types of questions they receive at the reference desk. Based on initial use, we have received no technical assistant-type questions through online reference, although they are a frequent occurrence at our physical reference desks.[17] There are numerous avenues for investigation concerning this new technology and its use that could be investigated in future research.

The source code for online reference will be posted on the OSS4lib Web site (www.oss4lib.org) for download, however no technical support is available due to lack of resources. Librarians are encouraged to become involved in the development of reference software, whether it is through suggestions to companies who make such software or by creating in-house systems.

Acknowledgement

The authors thank Keith VanCleave, one of the "fretful minority" and the creator of our online reference system.

References and Notes

1. Karen Ciccone, "Interactive Online Reference at NCSU Libraries," (poster session presented at ALA Annual Conference, June 2001, San Francisco, Calif., June 2001; Matthew Marsteller and Paul Neuhaus, "The Chat Reference Service Experience at Carnegie Mellon University," (poster session presented at ALA Annual Conference, June 2001, San Francisco, Calif.); Susan Holmer and Janie B. Silveria, "Live in Five (Minutes!): Implementing a Virtual Reference Service in a Multitype Library Consortium," (poster session presented at ALA Annual Conference, June 2001, San Francisco, Calif.); Jana Ronan, Mimi Pappas, and Colleen Seale, "RefXpress, Part II: Implementing a Chat Reference Service Using eShare's NetAgent," (poster session presented at ALA Annual Conference, June 2001, San Francisco, Calif.).

2. Bernie Sloan, Digital Reference Services: A Bibliography. Accessed July 10, 2001, www.lis.uiuc.edu/~b-sloan/digiref.html. Several registries of real-time reference services divided by type of library are provided by Gerry McKiernan at LiveRef: A Registry of Real-Time Digital Reference Services. Accessed July 10, 2001, www.public.iastate.edu/~cyberstacks/ LiveRef.htm. University of Florida provides a similar listing at A Survey of Online Interactive Reference Services. Accessed July 10, 2001, http://Web.uflib.ufl.edu/hss/ref/chat/cc3.html. For the most up-to-date issues in virtual reference, the reader is pointed to either of two active electronic discussion lists, livereference@yahoo.com andDIG_REF @ELECTRONIC DISCUSSIONLIST.SYR.EDU.

3. Jody Condit Fagan, "Use of an Academic Library Web Site Search Engine," *Reference & User Services Quarterly* , forthcoming.

4. The total number of all types of questions for one week at all reference desks was 2,644; the total number of site searches in a week was 2,053.

5. Poster sessions presented at ALA Annual Conference, June 2001, San Francisco, Calif.

6. Joseph Janes, David Carter, and Patricia Memmott, "Digital Reference Services in Academic Libraries," *Reference & User Services Quarterly* 39, no. 2 (winter 1999): 145–50.

7. Steve Coffman, "Distance Education and Reference: Where Are We Headed?" *Computers in Libraries* 21, no. 4 (Apr. 2001): 20–25.

8. DIG_REF electronic discussion list; Gerry McKiernan, LiveRef: A Registry of Real-Time Reference Services. Accessed July 10, 2001, www.public.iastate.edu/~cyberstacks/LiveRef.htm.

9. Welcome to LiveAssistance, the Web Call Center and Online Customer Support Service. Accessed June 28, 2001,

www.liveassistance.com. LivePerson: Give Your Site a Pulse. Accessed June 28, 2001, www.liveperson.com. Virtual Reference Desk. Accessed June 28, 2001, www.virtualreference.net/virtual.

10. Paul Constantine, "Cornell University's LiveHelp Service," presentation at Virtual Reference Desk Conference, Oct. 2000, Seattle, Wash. Accessed June 2001, http://vrd.org/conferences/VRD2000/proceedings.

11. Steve Coffman, E-mail to LIBREF-L on June 12, 2001. LSSI Introduces New Privacy Protections for Virtual Reference, Libref-l@listservkent.edu.

12. Ibid.

13. Marc Meola and Sam Stormont, "Real-Time Reference Service for the Remote User: From the Telephone and E-mail to Internet Chat, Instant Messaging, and Collaborative Software," *The Reference Librarian* 67/68 (1999): 35.

14. Sam Stormont and Marc Meola. Interactive Reference Project. Accessed June 28, 2001, www.library.temple.edu/ref/interactref.html.

15. Although many sites which document these languages exist, the authors refer to Perl Mongers. Accessed July 15, 2001, www.perl.org. MySQL. Accessed July 15, 2001, www.MySQL.com. JavaScript.com—The Definitive JavaScript Resource: JavaScript Tutorials and Free Java Scripts. Accessed July 15, 2001, www.javascript.com. HTML Help from the Web Design Group. Accessed July 15, 2001, www.htmlhelp.com.

16. Perl.com: A Short Guide to DBI [Oct. 22, 1999]. Accessed July 15, 2001, www.perl.com/pub/a/1999/10/DBI.html. Download Perl/DBI downloads. Accessed July 15, 2001, www.MySQL.com/downloads/api-dbi.html.

17. Soo Young Rieh, "User Computer Librarian Interaction in End-User Online Searching," in *Proceedings of the National Online Meeting* (New York: Information Today, 1998), 321.

Tutorial

The Service of Server-Side Includes

Michelle Mach

The use of server-side includes (SSI) simplifies Web site maintenance by generating Web pages on-the-fly, rather than creating static HTML pages. This article provides examples and instructions for common uses of SSIs, including displaying environmental variables, inserting file information, and using the advanced capabilities of extended SSI (XSSI). Potential concerns including security, WYSIWYG (what-you-see-is-what-you-get) editors, filename extensions, and server load are discussed, as are the advantages and disadvantages of SSI over more well-known tools like CGI, JavaScript, or Active Server Pages (ASP).

Many librarians are turning to dynamically generated Web pages to more effectively manage the growing numbers of library Web pages. Creating Web pages on-the-fly minimizes the amount of HTML coding required by staff, encourages HTML-resistant staff to participate in Web site maintenance and development, and allows libraries to offer more customization to their Web site visitors. It also allows libraries to respond more quickly to needed design or content changes. As Atelman notes, "To move forward, libraries must stop thinking of their Web sites as collections of HTML pages and view them as dynamic resources for information and services that patrons will use in highly individualized ways."[1]

Different levels of dynamically generated pages are obtainable, depending upon the level of skill, staffing, money, and training available at the library. At one end are the fully customized databases that use PHP, ColdFusion, or other tools to generate entire Web pages or sites based entirely upon user requests or queries. At the other end are tools like ASP, JavaScript, CGI, or SSIs that generate parts of Web pages or add limited customization features. For many libraries, especially those with medium-size Web sites and limited technical staff, this second group of tools is worth investigating.

In particular, SSIs exhibit several advantages over other, similar dynamic techniques. Introduced in the early 1990s by the National Center for Supercomputing Applicatons (NCSA), SSI support comes standard with most servers. ASPs, on the other hand, work only on Internet information service (IIS) servers. Unlike CGI scripts or PHP programming, implementation of SSIs is fairly quick and does not require special programming skills. In addition, pages containing SSIs have a normal URL for users to remember and cite, rather than the lengthy search query string URL that database-generated pages create. SSIs also do not depend upon users to turn on browser features like JavaScript, nor do they rely on certain browser versions to work like cascading style sheets (CSS). While not as customizable or full-featured as other solutions, SSI-generated Web pages offer an attractive solution for libraries seeking limited customization incorporated in a fairly quick, simple way.

Basically, SSIs embed special commands into an HTML document that tells the server to perform specific actions when a user requests the page. The server then creates the Web page on-the-fly by merging files or inserting requested information. From the Web administrator's perspective, the maintenance is lessened considerably, since a single change on a single file will affect all other files pointing to it. Users viewing the Web page will be unable to tell that SSIs have been used (see figure 1). Even if users view the source code, the page will display only the generated HTML code, not the SSIs (see figure 2). Only the Web administrator or author with Web server access sees the code with the SSIs (see figure 3).

Setup

The use of SSIs requires two steps to set up the Web server. A general description of the steps used for the Apache server running under UNIX follows. Filenames and directory paths will vary depending upon the system version and configuration. For Netscape or IIS servers, see the server documentation for setup.

First, find the options directive in the directory section of the global access configuration file. On newer systems, this file is named http.conf, while older versions may use access.conf. In the directory section, one of two lines are required to indicate that either all type of include files are permissible (Options Includes), or to exclude the executable types of includes (Options IncludesNoExec). For example:

```
<directory /usr/local/http/docs>
Options Includes
</directory>
```

Some Web administrators may prefer the second type (Options IncludesNoExec) for security or server-load reasons.

In the second step, tell the server which files should be parsed by indicating the file extension. Depending upon the system, this information may be included in the same http.conf file or in the srm.conf file. In the file, insert the following line:

```
AddHandler server-parsed.shtml

AddType text/x-server-parsed-html.
shtml
```

This tells the Web server to treat all files ending in .shtml as if they contained SSIs. Some Web administrators prefer this extension, as it limits the number of files processed by

Michelle Mach (mmach@manta.colostate.edu) is an Assistant Professor and Web Librarian, Colorado State University Libraries, Fort Collins.

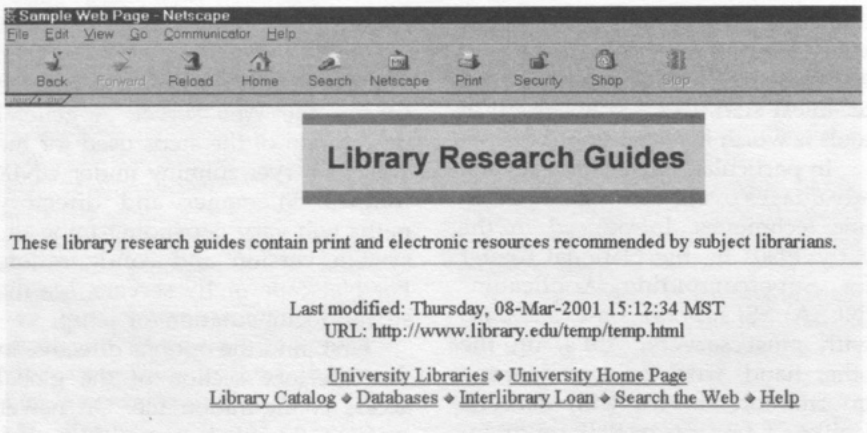


Figure 1. Web Browser View of Web Page Using SSIs

```
<HTML>
<HEAD><TITLE>Sample Web Page</TITLE>
</HEAD>
<BODY BGCOLOR="#FFFFF2">
<CENTER>
<IMG SRC="head.gif" alt="Library Research Guides">
</CENTER>
<P>These library research guides contain print and electronic resources recommended
by subject librarians.
<HR>
<P ALIGN="CENTER">Last modified:  Thursday, 08-Mar-2001 15:12:34 MST<BR>
URL: http://www.library.edu/temp/temp.html
<P ALIGN="CENTER"><A HREF="http://www.library.edu/">University Libraries</A>
<IMG SRC="http://www.library.edu/images/sdi_gr.gif" width="11" height="11" alt="">
<A HREF="http://www.university.edu">University Home Page</A><br>
<A HREF="http://catalog.library.edu/">Library Catalog</A>
<img src="http://www.library.edu/images/sdi_gr.gif" width="11" height="11" alt="">
<A HREF="http://www.library.edu/databases/">Databases</A>
<img src="http://www.library.edu/images/sdi_gr.gif" width="11" height="11" alt="">
<a href="http://www.library.edu/ill/">Interlibrary Loan</A>
<IMG SRC="http://www.library.edu/images/sdi_gr.gif" width="11" height="11" alt="">
<a href="http://www.library.edu/search.html">Search the Web</A>
<img src="http://www.library.edu/images/sdi_gr.gif" width="11" height="11" alt="">
<A HREF="http://www.library.edu/help.shtml">Help</A>
</BODY>
</HTML>
```

Figure 2. HTML Source Code Containing SSIs, as Viewed by the User

the server. However, renaming hundreds or thousands of files and links with a different extension is not a viable option for many libraries. Unless the Web site is extraordinarily popular, most Web servers should be able to handle the extra load of parsing all files and simply using the .html extension already in place. This is especially true if the Options IncludeNoExec setting, which disables executable scripts or programs, is in place. The .htm or .html extensions may be added to the file like this:

```
AddHandler server-parsed .shtml
.htm .html

AddType text/x-server-parsed-
html .shtml .htm .html
```

An alternative setup method uses the XBitHack directive to parse HTML files based on file permissions.[2] This method, which also allows you to automatically keep the .html file extension, works by setting the user's execute bit in the file permissions. This means that the file would need to be set to read, write, and execute for the user and read only for others. With this method, security becomes an issue if executable files are housed in the same directory as the files with SSIs.

Even if the Web administrator does not allow changes on the server settings, individual users may still be allowed to run SSIs in their own directories. To do this, create a file named .htaccess and insert the Options Includes, AddType, and AddHandler statements mentioned in the preceeding example.

Examples

SSIs have three common uses. First, SSIs may be used to echo variables, meaning to display information that the computer receives about the user. This information could include the user's Web browser type or the date and time the user viewed the page. Second, SSIs may embed information like graphics, text, or HTML from another file or directory. The third use, executing an external script, will be discussed in the security section of this article.

SSI statements look similar to HTML comments. They begin with

```
<HTML>
<HEAD><TITLE>Sample Web
Page</TITLE>
</HEAD>
<!--#include
virtual="/research/header.htm" -->
<P>These library research guides contain
print and electronic resources
recommended by subject librarians.
<hr>
<P ALIGN="CENTER">
Last modified: <!--#echo
var="LAST_MODIFIED" --><br>
<!--#include file="footer.html" -->
</BODY>
</HTML>
```

Figure 3. HTML Code Containing SSIs, as Viewed by the Web Administrator or Author

<!—# and end with —>. There is no space after the # and the command, but at least one space before the closing tag. No hard returns are permitted inside the comment tags. The basic general syntax for a server-side include is:

```
<!—#command tag1=value
tag2=value —>
```

The server searches for the <!—# sequence and then replaces it through the —> with the include information. The include information will be inserted wherever the SSIs exist, so add in whatever extra text, spacing, or HTML formatting needed.

To see the list of all the environmental variables available on a specific computer, use the Printenv command. For example:

```
<pre>
<!—#printenv —>
</pre>
```

The <pre> tags help formatting. The first part of the output listed in capitals is the variable; the second part is its value for that particular computer (see figure 4). Lists of the SSIs commands and descriptions are available on the NCSA and W3C Web sites.[3]

Date and Time

While a standard criterion for Web site quality, the most recent date and time of an update on a Web site requires constant attention to maintain by hand. This SSI automatically places this information on the page:

```
Last updated: <!—#echo var=
"LAST_MODIFIED"—>
```

In addition to LAST_MODIFIED, echo includes a number of different variables like DATE_LOCAL. This second variable, DATE_LOCAL can be useful for Web pages where the user may need to be reminded of the current date and time. Calendars, instruction session schedules, hours pages, or document delivery services that maintain a specific time schedule are all good candidates for the DATE_LOCAL variable. For example:

```
Library Workshops are held every
Friday at 2 P.M.

Today is: <!—#echo var="DATE_
LOCAL" —>
```

or

```
If a book is requested from storage
before noon, Monday–Thursday,
then it will be available for pickup
after 3 P.M. that day. Your book was
requested at <!—#echo var=
"DATE_LOCAL"—>
```

Be careful not to mix up the DATE_LOCAL and LAST_MODIFIED variables. "It is easy to mistakenly (or intentionally) use the DATE_LOCAL SSIs after a statement such as 'Page last updated on,' which then makes it look as if that page is always updated every day," comments Notess.[4]

URL

Displaying the URL on a Web page offers a nice service for users who print pages for later reference, especially if the pages have long URLs that are likely to be cut off. This requires cobbling together several echo variables to form a full URL, including the domain (HTTP_HOST) and the directory path and filename (DOCUMENT_URI). Notice that http:// is not included in the HTTP_HOST variable and is simply typed into the HTML code.

```
URL: http://<!—#echo
var="HTTP_HOST" —><!—#echo
var="DOCUMENT_URI" —>
```

The user will see:

```
URL: http://www.library.edu/
research/bananas.html
```

File Size

In a similar manner, indicate file size for files that users must download, like a browser plug-in needed for a library tutorial or a program to automatically configure a proxy server. Users accessing the library site over a slow modem will especially appreciate this information. For example:

```
<a href="proxy.exe">Download
Proxy Configuration
Program</a> (Filesize: <!—#fsize
file="proxy.exe" —>)
```

will be viewed as:

```
Download Proxy Configuration
Program (File size 44 K)
```

Headers and Footers

In many libraries, standard page elements like headers and footers are subject to sudden change by outside parties such as the university administration or the board of directors. Making a change such as adding a link to the university home page or changing the color of the navigation bar becomes much simpler and faster with SSIs in place because a change in a single file will affect all other files pointing to it. Easing the management of standard features also encourages more staff to participate in Web development. Most medium and large libraries have gone beyond the "webmaster" model where one person is responsible for every change on the

```
DOCUMENT_ROOT=/libraryweb/public_html
HTTP_ACCEPT=image/gif, image/x-xbitmap, image/jpeg
HTTP_ACCEPT_CHARSET=iso-8859-1,*,utf-8
HTTP_ACCEPT_ENCODING=gzip
HTTP_ACCEPT_LANGUAGE=en
HTTP_CONNECTION=Keep-Alive
HTTP_HOST=www.library.edu
HTTP_PRAGMA=no-cache
HTTP_REFERER=http://www.library.edu/temp/
HTTP_USER_AGENT=Mozilla/4.73 [en] (WinNT; U)
PATH=/usr/local/bin:/bin:/usr/bin:/etc:/usr/bin/X11:/usr/local/X11/bin:/usr/local/mysql/bin:
REMOTE_ADDR=123.45.678.910
REMOTE_HOST=lib123-456.library.edu
SCRIPT_FILENAME=/libraryweb/public_html/temp/temp2.html
SERVER_ADMIN=sysadmin@library.edu
SERVER_NAME=library.edu
SERVER_SOFTWARE=Apache/1.3.4 (Unix)
GATEWAY_INTERFACE=CGI/1.1
SERVER_PROTOCOL=HTTP/1.0
REQUEST_METHOD=GET
DATE_LOCAL=Thursday, 08-Mar-2001 12:22:59 MST
DATE_GMT=Thursday, 08-Mar-2001 19:22:59 GMT
LAST_MODIFIED=Wednesday, 07-Mar-2001 10:22:56 MST
DOCUMENT_URI=/temp/temp2.html
USER_NAME=mmach
DOCUMENT_NAME=temp2.html
```

Figure 4. Sample Output of the Printenv Command

site.[5] In theory, having many people edit Web pages lessens the workload. However, the use of multiple Web authors of varying skill levels invariably means that some standard features will get mangled in the process and will need to be redone. The use of SSIs simplifies the template, making it easier for staff to understand and edit.[6] If standard headers and footers are currently embedded into all library Web pages, a simple search-and-replace script may be used to replace bits of HTML code with the SSIs.

To include a standard footer, create an HTML file with the footer information, which may include other include tags. If the files are in the same directory, only the filename is needed in the tag like this:

```
<!—#include file="footer.html"—>
```

If the files are in different directories, then indicate the subdirectory like this:

```
<!—#include virtual="/subdir/
page.html"—>
```

When pointing to other files, be careful not to end up with a document with multiple HTML, HEAD, or BODY tags. In addition, be certain of the filename and path. If the server cannot find the correct file or path, the user will see the error message "an error occurred while processing this directive" generated as part of the Web page, rather than the more meaningful information in the header or footer.

Non-HTML Files

In addition to headers and footers, SSIs also work for non-HTML files. This enables every person who is capable of typing a few lines a potential Web-page contributor. As Notess writes, "By using includes, a simple text file can contain the content, and the people with no HTML experience can be given access to change that text."[7] A staff Web page, for example, may include a single paragraph of information about each staff member. Using SSIs, staff members could edit their individual text file, rather than the HTML document. For example:

```
<!—#include file="linda.txt"—>
```

When including fragments, you may want to use a non-HTML extension, so that the file is not indexed as a separate Web page either internally or externally. HTML tags may still be included, even if the file extension is not .htm or .html.

Smart Links

Although as librarians we might like to think otherwise, most users are probably not visiting our pages in an orderly linear fashion. Sometimes it may be desirable to allow users to return to their previous screen, whatever it might be. Two examples are within a Web virtual tour or within the library catalog. While this can be done in JavaScript, use of SSIs means that users do not need to have JavaScript turned on. A smart link looks like:

```
<a href="<!—#echo var="HTTP_
REFERER"—>">Return to
Previous Page</a>
```

Multiple Web Authors and WYSIWYG Editing Tools

If multiple staff members edit Web pages containing SSIs, two issues must be considered. First, if staff save files from the Web browser for editing, rather than directly accessing the

Web server directories, the saved file will not contain SSIs, only the generated code. If direct access to the Web server directories is not possible, then a workaround like a local FTP mirror site will be required in order to preserve the SSIs.

The second consideration is how popular WYSIWYG editors such as Microsoft FrontPage, Adobe GoLive, and Macromedia Dreamweaver treat SSIs. The good news is that no matter which editor is used, Web authors will be able to edit the SSIs in the HTML code view, assuming the file was copied from the server. However, beginning Web authors, especially those who use only the WYSIWYG editing screen, may require some initial training to avoid several common pitfalls.

First, Web authors must recognize how SSI content is displayed in their WYSIWYG editing screen, so that they do not accidentally alter or delete this content. Newer Web editors like Adobe GoLive 5.0 display placeholder graphics to indicate SSIs in the WYSIWYG editing screen, while older editors like FrontPage 98 do not. Second, Web authors who use the preview function available in most WYSIWYG editors may be dismayed to discover that the SSIs will not translate on their local machine. Previewing a Web page without crucial elements like background color, header, or footer for the first time can be disconcerting, especially to visually oriented designers (see figure 5). Third, Web authors will need a firm grasp on constructing valid SSI statements to include other files, since the WYSIWYG editors do not generally check the validity of these statements. Just as beginning Web authors may include local file information when creating links to Web pages or graphics, they may also create faulty include statements like this:

```
<!—#include virtual="file:///
C|temp/spiders.html"—>.
```

New Web authors may also be tempted to insert non-Web documents into an include statement, such as a Microsoft Word or Adobe Acrobat document. For example, a WYSIWYG editor may permit the statement:

```
<!—#include virtual="word.
doc"—>,
```

but when the computer attempts to translate this file, meaningless characters like ÐÏà¡± will be inserted into the document.

Currently, Dreamweaver 4.0 offers the most support for SSIs with an "Insert Server-Side Include" menu option for local files and SSI rendering for files within the same directory. Figure 6 demonstrates how the information in footer.html is displayed in the Dreamweaver 4 editing screen, while header.html is not. Nontranslated SSIs like the environmental variables used to display the last modified date and URL are shown by placeholder graphics. For new Web authors, the rendering of footer.html appears misleading, as the included file may not be edited from that screen, unlike other parts of the page. Instead, the Web author must open the included file separately in order to edit it.

Although Dreamweaver offers the most support for SSIs, other WYSIWYG editors like Adobe GoLive 5.0 and FrontPage 2000 offer a similar feature called components. Like SSIs, components simplify Web maintenance by housing common information in a single file, such as a page header. However, there are two major differences. The first difference is that unlike SSIs, Web pages using components are not generated on-the-fly. Instead, the WYSIWYG editor uses a search-and-replace function to hard code the information onto each Web page. This means that if a common file like a page header is changed, then all files pointing to it must be reloaded on the server. Second, editors such as Adobe GoLive or FrontPage recognize and update these components by inserting a proprietary code in the HTML. This means that if a library decides to change Web editors, then none of the component code will work. SSIs, in contrast, will work no matter which editor a library uses.

XSSI

XSSI, available from Apache, allows some advanced dynamic features, such as hiding or displaying links depending upon IP address or browser type, generating random images, or using hit counters. Basically, XSSI works by allowing the use of variables in commands and by allowing conditionals. To set variables, the basic string is:

```
<!—#set var="color"value="blue"
—>.
```

To set conditionals (if, else, elif, endif), the basic syntax is:

```
<!—#if expr="first" —>do this first
task

<!—#elif expr="second —>do this
other task

<!—#else —>do something else

<!—#endif —>
```

As the example demonstrates, conditionals make it possible to send users to various pages or display different information depending upon their particular needs. This feature is extremely valuable for libraries, particularly for its distance users. For example, users outside of the university IP domain could first be sent to a proxy server setup page, rather than immediately to the listing of databases on-campus folks would see. Users with a noncompatible browser for the proxy server could automatically view an explanatory message, rather than expecting the user to find the help page himself (figure 7). Instead of frustrating users with a "forbidden" message, links to restricted sites could be hidden from external users. For example, to hide the link to staff_intranet.html from users outside the IP address 1.2.3, use:

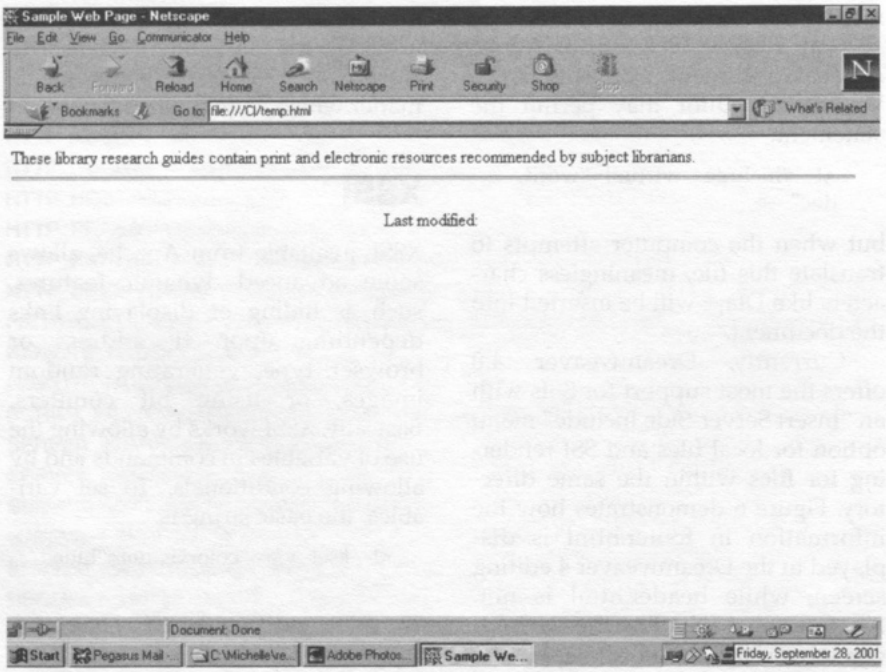


Figure 5. Preview in Web Browser of Local Copy of Web Page with SSIs

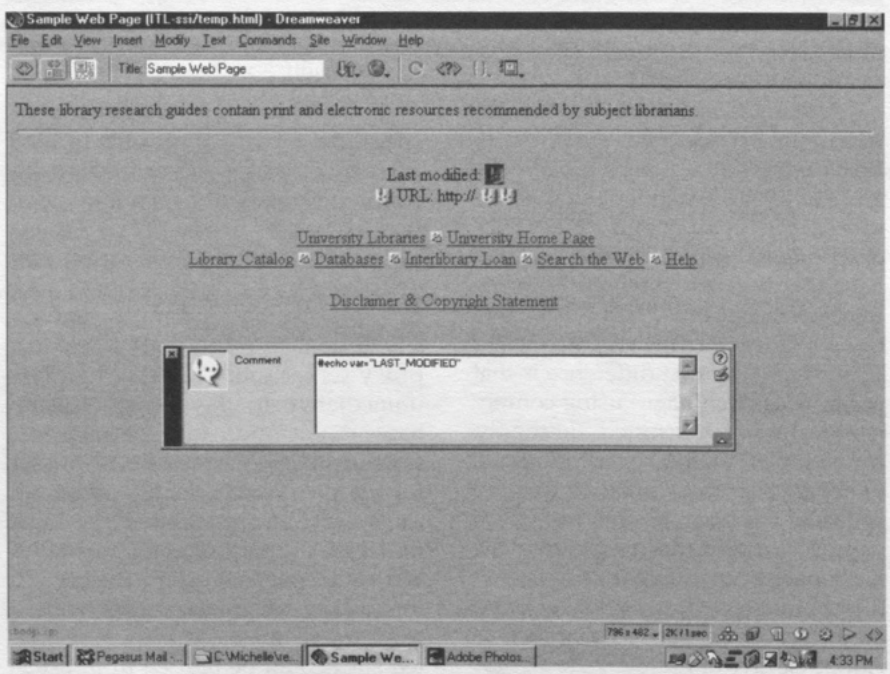


Figure 6. Dreamweaver 4.0 Editing Screen Showing Web Page with SSIs

```
<!—#if expr="$REMOTE_ADDR = /^1.2.3./"—> <a href="staff_intranet.html">Library Staff Intranet</a><!—#endif—>
```

Security and Other Issues

In addition to the server-load issue mentioned earlier, SSIs also raise some security concerns. The list of potential security issues includes crashing Web servers, killing other users' processes, and sending e-mail.[8] Many of these risks are triggered by external programs using the #Exec command in combination with either CMD or CGI to launch a program. For example, to run the script called sample.cgi, use:

```
<!—#exec cgi="sample.cgi"—>
```

On the positive side, using SSIs in combination with a CGI script allows the Web administrator to offer more sophisticated customization than is possible with SSIs alone. However, running scripts opens the server to possible harm, causing many Web administrators to use the Options IncludesNoExec setting. Many articles offer some steps to minimize the security risks, including:

- Disable includes in the CGI-bin directory or other directories with executable files.
- Run the Web server as the user nobody, not root. Root has unlimited access to the system.
- Keep server software up-to-date and use the IncludesNoExec option to disable scripts.
- Use the Virtual command, rather than Exec.[9]

Conclusion

The use of SSIs is a simple, relatively quick way to add much-needed manageability to growing library Web sites. It is also an effective way to involve all staff, even those with no

```
<!--#if expr="${HTTP_USER_AGENT} ~=
/MSIE5.5/" -->

<!--#set var="browser" value="IE5" -->
<!--#include file="ie5_warning.txt" -->

<!--#elif expr="${HTTP_USER_AGENT} ~=
/MozillaV4/" -->

<!--#set var="browser"
value="netscape4" -->
<!--#include file="netscape4_warning.txt"
-->

<!--#else -->

<!--#set var="browser" value="working" -
->
<!--#include file="proxy_setup.txt" -->

<!--#endif -->
```

Figure 7. Example of XSSI Code

HTML experience, in the development of a library Web site. While not as full-featured as a database driven site, use of SSIs does allow the library to add limited customization to the site. Most importantly, it also frees up the Web administrator's time from the more routine maintenance, affording more time for all those other Web projects at the library.

References and Notes

1. Kristin Antelman, "Getting Out of the HTML Business: The Database-Driven Web Site Solution," *Information Technology and Libraries* 18, no. 4 (Dec. 1999): 177.
2. For more information about the XBitHack method, see: "How to Enable Server Side Includes on Your Web Server." Accessed Jan. 19, 2001, http://bignosebird.com/sdocs/enable.html.
3. NCSA HTTPd Tutorial: Server Side Includes (Sept. 28, 1995). Accessed Aug. 9, 2000, http://hoohoo.ncsa.uiuc.edu/docs/tutorials/includes.html. W3C Server Side Include Commands. Accessed Jan. 19, 2001, www.w3.org/Jigsaw/Doc/User/SSIs.html.
4. Greg Notess, "Server Side Includes for Site Management," *Online* 24, no. 4 (July 2000): 79.
5. Antelman, "Getting Out of the HTML Business," 176.
6. For a comparison of various templating solutions including SSIs, see: Darlene Fichter, "Building Web Sites Using Dynamic Templating," *Online* 24, no. 6 (Nov./Dec. 2000): 84–88.
7. Notess, "Server Side Includes for Site Management," 80.
8. Jared Karro and Jie Wang, "Protecting Web Servers from Security Holes in Server-Side Includes," in Proceedings of the 14th Annual Computer Security Applications Conference (Los Alamitos, Calif.: IEEE Computer Society Pr., 1998): 103–11.
9. Art Sackett, SSIs for the Rest of Us. Accessed July 12, 2000, http://artsackett.com/grey_papers/ssi/rest_of_us.html. Chuck Musciano, "Inject Compelling and Up-to-the-Minute Data into Your Web Site," Sun World, Dec. 1996. Accessed Oct. 5, 2000, www.sunworld.com/swol-12-1996/swol-12-webmaster.html. Karro and Wang, "Protecting Web Servers from Security Holes in Server-Side Includes," 109. Notess, "Server Side Includes for Site Management," 80.

Index to Advertisers

ALA Editions	228	Library Technologies	223
ALCTS/LAMA/LITA/ RUSA	170	LITA	212, cover 3
Axonix	cover 4	OCLC	cover 2
infoUSA	201	Pacific Data Conversion	197

Book Review

Karen Coyle

Digital Copyright

Protecting Intellectual Property on the Internet

by Jessica Litman. Amherst: Prometheus Books, 2001. 208p. $25 (ISBN 1-573-92889-5).

If you're like most of us, you have some idea of what the Digital Millennium Copyright Act (DMCA) is about but feel you should know more. And if you're like most of us, you're not about to read a whole book on the topic in order to gain (or attempt to gain) more understanding of the topic.

Jessica Litman offers us an informative, intelligent, and even amusing way to further our knowledge of the DMCA and other copyright issues through her book *Digital Copyright*. To begin with, this is not a law book, although it is about law. *Digital Copyright* is instead a social history of copyright law. It is not about the law per se but about how the technology developments of the twentieth century changed how copyright law is crafted in the United States and who reaps the benefits.

Although few of us will ever read the thirty thousand words of the DMCA (and we probably wouldn't learn much from it if we did), Litman's book provides the reader with a clear picture of why the DMCA matters by presenting a picture of how this law evolved, who it favors, and why it is generally bad for the rest of us.

Litman begins chapter 3 with the sentence: "If history bores you, you should skip this chapter." Don't. This is the most important chapter in the book. It explains how our copyright law went from being a piece of legislation (like other laws) to its current state as a negotiated compromise between major commercial stakeholders. It began early in the twentieth century when the legislature and the U.S. Copyright Office decided that although the copyright law clearly needed updating, the issues were too complex for the legislators to understand. The U.S. Copyright Office therefore drafted the proposed legislation in a series of meetings with representative members of industries with an interest in copyright issues. This technique led to a revision of the U.S. Copyright Act in 1909 and has been the method of revising copyright law ever since.

Admittedly, it does make some sense that stakeholders would be included in the making of laws that affect them and their business, and presumably this is not uncommon in our law-making today. (That's what lobbyists are all about, after all.) The consequences of having the stakeholders dominate the process, however, are quite negative for all who were not part of the discussion. Not included in the process are representatives of future or emerging technologies, minority and noncommercial interests and, of course, members of the public.

Each iteration of copyright law in the twentieth century addresses particular technologies and their commercial interests. The 1909 copyright law mainly addressed issues relating to composers and music publishers. Not included in that process were members of the piano roll and "talking machine" (phonograph) industries, whose representatives were not present at the discussions. The law made the unlicensed manufacture of piano rolls and phonograph records illegal. The 1912 law addressed the right of movie companies to make motion pictures based on books, arising from a lawsuit over a movie version of *Ben Hur*. Movie companies participated to negotiate a law that limited their exposure in copyright infringement actions. But of course radio was not included.

Technology progressed and copyright discussions continued. From 1912 to 1976, numerous attempts were made to pass a law that would cover all of the new technologies: commercial radio, talking movies, commercial television. Negotiations fell apart before bills could be drafted or during the actual legislative process when new stakeholders got wind of the changes afoot and complained to their representatives. It was a chaotic tug-of-war of competing interests, each with their specific needs. Some of these needs introduced new concepts into copyright law: where once there had been one single copyright for a work, new technologies allowed copyright holders to license diverse uses for the same work, such as a music recording, radio play, and use in a film.

The 1976 copyright law, which was a major revision, was the result of decades of committee meetings, negotiations, drafting, and compromises. The final steps to this law began with a six-year study and five years of conferences that produced bill text. Another eleven years passed as Congress and the "interested parties" met to compromise on specific issues and add new language to the bill. The resulting law was a preagreed upon statement of copyright that satisfied those who had participated in its making.

This doesn't mean that everyone was covered. "Just as there had been no commercial broadcaster to invite to the conferences in 1905, there were no videocassette manufacturers, direct satellite broadcasters, digital audio technicians, personal computer users, motion picture colorizers, online database subscribers, or Internet service providers to invite in 1960" (51). It took decades of work to craft the law and it was out of date before Congress voted on it.

Even more important, however, was the nature of the law that came out of such a process. Copyright law that is crafted in this manner tends to address very specific industries, technologies, and situations. These specific legal points do not lend themselves to application to situations that weren't included in the negotiations; they don't generalize well. As an example, application of

the 1976 copyright law to computer technologies has led to some very unfortunate court rulings, at least some of which were used to bolster the early process that resulted in the DMCA. One of these rulings stated that every use of a computer file, such as turning on the computer and launching the operating system or opening a document to read it on the screen, made a copy of the software or file and therefore was potentially an infringement of copyright. In the story of the DMCA, it just goes downhill from there.

Most of the book, and decidedly the most engaging part of it, revolves around the development of the law that became the Digital Millennium Copyright Act and its aftermath. Litman introduces the Internet into the copyright picture and prefaces the remaining two-thirds of her book with a pithy one-line wrap-up of the process that led to the DMCA's implementation: "What happened next was a failure of imagination" (89). The members of the Clinton administration charged with looking into the potential effects of the Internet on copyright, along with the usual bevy of industry representatives, showed so little understanding of the new technology that it would have been amusing if it hadn't led to such a tragically mistaken law.

As we know now, the recent changes to the copyright law have greatly enhanced the rights of copyright holders and have consequently greatly reduced the rights of the reading public. The justification for this shoring up of the law was good old American patriotism: according to the Working Group on Intellectual Property, chaired by the Clinton administration Patent Commissioner Bruce Lehman from 1994 to 1996, nothing short of the entire future of the American economy was at stake. New protections for copyright in the networked economy were needed to protect American profitability in that environment. There would be no investment in the National Information Infrastructure and no content to sell both at home and abroad if adequate protection was not provided.

For those of us who were actively working on the Internet in those years, it is astonishing to hear that the Net was considered by Washington insiders to be ". . . a collection of empty pipes, waiting to be filled with content" (93). The entire movement to reform our copyright law ignored the fact that we already had a vital electronic information infrastructure bursting with content, albeit primarily non-commercial content. Yet it wasn't difficult to sell this vision to industry, the legislature, and the few members of the press who got wind of the Working Group's activities, most likely because most people had very little, if any, experience with the Internet itself. If they had, they dismissed it as being "academic" and therefore not relevant to the vital commercial world that the National Information Infrastructure promised to be.

The entire DMCA was based on a vision of a market that didn't exist in a world that was destined not to be. Not only that, the resulting bill was so long that it was subsequently published in book form. The original bill was about three thousand words long. As it went through committees it grew to four thousand, then twelve thousand, and finally passed at a whopping thirty thousand words. It covered not only the digital millennium but such special interests as boat hull designs. And no one, not even the most ardent of copyright attorneys, understands it all. As Litman says: "Our current copyright statute has more than merely a provision or two or three or ten that don't make a lot of sense; it's chock-full of them" (114).

This incomprehensibility of the copyright law has a great effect on the ability and willingness of the public to obey it. A theme that runs throughout Litman's book is that no law is effective if the people whose behavior it is supposed to control cannot understand it, or find it so absurd they are not willing to believe in it. Laws that are not obeyed are ineffective, as we saw with the federally imposed fifty-five-mile-per-hour speed limit that was routinely ignored even by those who consider themselves generally law-abiding. Rather than embarrass itself further, the government rescinded the law. The DMCA already has examples of public disbelief, the foremost being the development of the DeCSS program, software that allows DVD owners to play DVDs on Linux machines even though the DVD industry hasn't provided drivers for that platform. To many individuals it is an affront to be told that you can't choose the hardware platform in which to view a movie you legally own. It's also rather absurd that the industry is making a big deal about the decryption of their copy protection when the details of the copy protection scheme are at this point public knowledge, having been published on the Internet, and simple enough that it was first broken by a fifteen-year-old. Litman doesn't come right out and say that the DMCA is doomed to failure, but she offers some good reasons not to bet on its success.

Arguably the most significant development in the DMCA is that for the first time in the history of our copyright law we have specific provisions aimed at the individual. Previous incarnations of copyright law were focused on competing commercial interests; individual non-commercial actions, although theoretically infringing, were not considered sufficiently threatening to warrant a response. As Litman says, "As a comprehensive strategy, litigation works best against commercial actors" (167). The anti-circumvention language of the law makes breaking through technological protection, such as the CSS encryption of DVDs, a criminal act. This might be an effective strategy unless a large portion of the public engages in such acts, in which case an

industry is faced with having to treat its customer base as criminals. Even those who haven't passed Marketing 101 can see that taking legal action against the bulk of your customers is probably not good for an industry's bottom line.

You can add to this the complete failure of some major industries to understand the new technology, most notably the music recording industry. The Recording Industry Association of America (RIAA) spent years trying to outlaw the sale of devices that play MP3 files, while the total number of available files in that format grew into the millions. Meanwhile, the RIAA offered no alternative format to MP3. While talking the moral high ground, the recording industry continued to sell CDs at inflated prices (which became even more obvious when writable CDs came to the consumer market and people learned that even they could purchase the blank CDs for around fifty cents each). And although Metallica sided with the recording industry (and alienated many of its fans), other artists spoke out against the industry with stories of exploitation and near starvation, even while they were making number one hits for record companies. Napster is the evidence that many, many individuals don't believe what the record industry is saying, even though they care about its product. "If forty million people refuse to obey a law, then what the law says doesn't matter. It may be that people flout it because they're natural lawbreakers, or it may be . . . that they don't comply because it doesn't make sense to them. Whatever the reason, the law is not going to work well in the real world" (169).

The DMCA is a huge, incomprehensible law that entirely failed to take into account the hundreds of millions of consumers of intellectual property. It has placed the intellectual property industries in direct opposition to the public, the very public that buys its goods. When you think about it in these terms, it's a pretty amazing mess that they have gotten themselves into.

I first encountered Litman through her testimony at hearings related to the Green Paper on Intellectual Property that was the draft report of Lehman's Working Group. The transcripts of the hearings were published on the Department of Commerce Web site (presumably, though, without counting as "content" by the Working Group's own definition). After slogging through hundreds of pages of some of the most toadying lawyer-ese, Litman's testimony leapt from the page like some minor miracle of truth and justice. It was Litman who pointed out during those hearings that, as formulated in the Green Paper, copyright law was moving from control over copying to control over *access*. For the first time the law sought to control the acts of reading, listening, and viewing of copyrighted materials. She said then (and reprises this theme in her book): "What I think is needed now is for someone to act as the copyright lawyer for the public. To examine these proposals as one would if the public had retained one as its copyright lawyer and said: here is a proposal—is this in my interest? . . . I believe that the public's copyright lawyer would see an amendment expressly privileging individuals using their computers for ordinary reading, viewing, or listening to authorized copies of copyright work."[1]

Unfortunately, the public's copyright lawyer did not appear and the DMCA was crafted without such representation. We now have a law that actually encourages limitations on access and use of copyrighted materials. The CSS program that protects DVDs does not prevent copying of the DVD content, it only prevents unlicensed access, even though the DVD may itself be a legally owned copy. The digital rights management systems that are being developed for electronic books will have the same effect: they will regulate access and use, not copying. The intellectual property industries seem to be bent on developing some of the most user-hostile controls over their products that the world has seen. Where can we possibly go from here?

The last chapter of Litman's book opens with the sentence: "As of this writing, the future is murky." I wanted the book to have a neat ending, one that wraps up the whole question and makes me feel better about things. Of course, it's difficult writing a book on such a volatile topic, and Litman can't be faulted because she can't predict where we are going with digital copyright—no one else knows either. We can only hope that she comes back to us from time to time to continue her analysis. When she does, she'll be speaking or writing to an audience that has a better understanding of the current situation. Although we may not know where copyright is going, thanks to Litman we can understand how we got to where we are today.

Reference

1. United States Department of Commerce U.S. Patent and Trademark Office, National Information Infrastructure Task Force Working Group on Intellectual Property, Public Hearing on Intellectual Property Issues Involved in the National Information Infrastructure (Sept. 22, 1994).

Information Technology and Libraries
Index to Volume 20 (2001)

Compiled by Charles Anderson

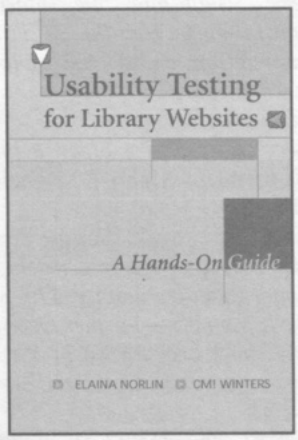
Academic Libraries.
E-Books and New Library Service Models (Communication), Jantz, Ronald, 104–13.
Searching for the Library: University Home Page Design and Missing Links, Astroff, Roberta, 93–99.
Three Iterations of an Academic Library Website, Vaughan, Jason, 81–92.
Use of CD-ROM Databases in Nigerian Academic Libraries (Tutorial), Oduwole, A. A., 43–47.
Virtual Reference at North Carolina State: The First One Hundred Days, Boyer, Joshua, 122–28.
Allen, Bryce L., *Boolean Browsing in an Information System: An Experimental Test*, 12–20.
Astroff, Roberta. *Searching for the Library: University Home Page Design and Missing Links*, 93–99.
Ault, Meredith, and Viggiano, Rachel. *Online Library Instruction for Online Students*, 135–38.
Bibliographic Instruction.
Online Library Instruction for Online Students, Viggiano, Rachel, and Ault, Meredith, 135–38.
Speakeasy Studio and Café: Information Literacy, Web-Based Library Instruction, and Technology, Jacobs, Mark, 66–71.
Boolean Browsing in an Information System: An Experimental Test, Allen, Bryce L., 12–20.
Boyer, Joshua. *Virtual Reference at North Carolina State: The First One Hundred Days*, 122–28.
Boze, Andy. *GIMP* (Software Review), 165.
———. *HotDog Professional 6.2* (Software Review), 51–54.
———. *IrfanView 3.36* (Software Review), 162–63.
———. *NoteTab Pro 4.83* (Software Review), 54–55.
———. *PrintKey 2000 v5.10* (Software Review), 164–65.
———. *Satori PhotoXL 2.29* (Software Review), 163–64.
———. *SuperToolz Plug-ins* (Software Review), 53–54.
———. *TrayMenu 2.01* (Software Review), 55–56.
———. *WinGrab 1.40* (Software Review), 164–65.
Bregman, Alvan, and Winnie S. Chan. *Customization of Library Service in a Cross-Platform Programming Environment*, 21–28.
Calloway, Michele, and Fagan, Jody Condit. *Creating an Instant Messaging Reference System*, 202–12.
Catalog Maintenance.
Customization of Library Service in a Cross-Platform Programming Environment, Bregman, Alvan, and Winnie S. Chan, 21–28.
CD-ROM Databases.
Use of CD-ROM Databases in Nigerian Academic Libraries (Tutorial), Oduwole, A. A., 43–47.
Chan, Winnie S., and Alvan Bregman. *Customization of Library Service in a Cross-Platform Programming Environment*, 21–28.
Chat Reference. *See* Virtual Reference Services.
Ciccone, Karen. *Virtual Reference: Today and Tomorrow* (Editorial), 120–21.
Coffman, Steve. *We'll Take It from Here: Further Developments We'd Like to See in Virtual Reference Services*, 149–53.
Collaborative Digital Reference Service (CDRS).
From Sshh to Search Engine: Reference.net on the Web, Kresh, Diane Nester, 139–42.
Collection Development.
Golderman, Gail. *Weaving Multimedia into the Collection: One Library's Journey* (Communication), 198–200.
Computer-Based Training.
Using Macromedia Authorware for Web-Based Instruction (Communication), Cullen, Kevin F., 154–58.
Copyright.
Digital Copyright: Protecting Intellectual Property on the Internet (Book Review), by Litman, Jessica, 220–22.
Cornish, Alan. *Data Retrieval from Library Systems Based on the Open Systems Model* (Communication), 100–103.
Coyle, Karen. *Digital Copyright: Protecting Intellectual Property on the Internet* (Book Review), by Litman, Jessica, 220–22.
Creating an Instant Messaging Reference System, Fagan, Jody Condit, and Calloway, Michele, 202–12.
Cross-Platform Programming.
Customization of Library Service in a Cross-Platform Programming Environment, Bregman, Alvan, and Winnie S. Chan, 21–28.
Cullen, Kevin F. *Using Macromedia Authorware for Web-Based Instruction* (Communication), 154–58.
Customization of Library Service in a Cross-Platform Programming Environment, Bregman, Alvan, and Winnie S. Chan, 21–28.
Databases.
Data Retrieval from Library Systems Based on the Open Systems Model (Communication), Cornish, Alan, 100–103.
Designing Web Usability (Book Review), by Nielsen, Jakob, 159–61.
Digital Copyright: Protecting Intellectual Property on the Internet (Book Review), by Litman, Jessica, 220–22.
Digital Libraries.
A Personalized Information Environment for Digital Libraries, Jayawardana, Champa, Hewagamage, K. Priyantha, and Hirakawa, Masahito, 185–96.
Digital Millenium Copyright Act.
Digital Copyright: Protecting Intellectual Property on the Internet (Book Review), by Litman, Jessica, 220–22.
Digital Reference Services. *See also* Virtual Reference Services.
Creating an Instant Messaging Reference System, Fagan, Jody Condit, and Calloway, Michele, 202–12.
Going Where the Users Are: Live Digital Reference, Stormont, Sam, 129–34
Roles in Digital Reference, McClennen, Michael, and Memmott, Patricia, 143–48.

From Sshh to Search Engine: Reference.net on the Web, Kresh, Diane Nester, 139–42.
E-Books and New Library Service Models (Communication), Jantz, Ronald, 104–13.
E-Book Technology: Waiting for the "False Pretender," Sottong, Stephen, 72–80.
Electronic Resources—Costs.
Measure for Measure: Statistics about Statistics, Knight, Lorrie A., and Lyons-Mitchell, Kimberly A., 34–38.
eThemes: An Internet Instructional Resource Service, Wang, Feng-Kwei and Wedman, John, 179–84.
Fagan, Jody Condit, and Calloway, Michele. *Creating an Instant Messaging Reference System,* 202–12.
Fidishun, Dolores. *People Servers vs. Information Providers: The Impact of Service Orientation on Technology Training,* 29–33.
From Sshh to Search Engine: Reference.net on the Web, Kresh, Diane Nester, 139–42.
Fullerton, John Paul, and Wilson, Mary Dabney. *Improving Retrieval of Microform Set Records,* 60–65.
GIMP (Software Review), 165.
Going Where the Users Are: Live Digital Reference, Stormont, Sam, 129–34.
Golderman, Gail. *Weaving Multimedia into the Collection: One Library's Journey* (Communication), 198–200.
Gorman, Michael. *Human Values in a Technological Age,* 4–11.
Graphics Softare.
PrintKey 2000 v.5.10 (Software Review), 164–65.
WinGrab 1.40 (Software Review), 164–65.
Hewagamage, K. Priyantha, Jayawardana, Champa, and Hirakawa, Masahito. *A Personalized Information Environment for Digital Libraries,* 185–96.
HotDog Professional 6.2 (Software Review), 51–54.
The Humane Interface (Book Review), by Raskin, Jef, 48–50.
Human Values in a Technological Age, Gorman, Michael, 4–11.
Improving Retrieval of Microform Set Records, Wilson, Mary Dabney, and Fullerton, John Paul, 60–65
Information Retrieval.
Boolean Browsing in an Information System: An Experimental Test, Allen, Bryce L., 12–20.
A Personalized Information Environment for Digital Libraries, Jayawardana, Champa, Hewagamage, K. Priyantha, and Hirakawa, Masahito, 185–96.
Instructional Design.
eThemes: An Internet Instructional Resource Service, Wang, Feng-Kwei and Wedman, John, 179–84.
Intellectual Property.
Digital Copyright: Protecting Intellectual Property on the Internet (Book Review), by Litman, Jessica, 220–22.
Internet.
Digital Copyright: Protecting Intellectual Property on the Internet (Book Review), by Litman, Jessica, 220–22.
eThemes: An Internet Instructional Resource Service, Wang, Feng-Kwei and Wedman, John, 179–84.
Internet Access in Public Libraries: Results of an End User Targeted Pilot Study, 1997–2000 (Communication), Kibirige, Harry M., 113–15.
IrfanView 3.36 (Software Review), 162–63.
Jacobs, Mark. *Speakeasy Studio and Café: Information Literacy, Web-Based Library Instruction, and Technology,* 66–71.
Jantz, Ronald. *E-Books and New Library Service Models* (Communication), 104–13.
Jayawardana, Champa, Hewagamage, K. Priyantha, and Hirakawa, Masahito. *A Personalized Information Environment for Digital Libraries,* 185–96.
Kibirige, Harry M. *Internet Access in Public Libraries: Results of an End User Targeted Pilot Study, 1997–2000* (Communication), 113–15.
Knight, Lorrie A., and Lyons-Mitchell, Kimberly A. *Measure for Measure: Statistics about Statistics,* 34–38.
Kresh, Diane Nester. *From Sshh to Search Engine: Reference.net on the Web,* 139–42.
Laser Printing at Public Workstations: User Behaviors and Attitudes (Tutorial), Hart, Richard L., Olson, John A., and Simmonds, Patience L., 39–43.
Libraries and Technology.
Human Values in a Technological Age, Gorman, Michael, 4–11.
Library Instruction.
Speakeasy Studio and Café: Information Literacy, Web-Based Library Instruction, and Technology, Jacobs, Mark, 66–71.
Library Web Page Design: Are We Doing It Right? (Editorial), Marmion, Dan, 2–3.
Library Web Pages.
Library Web Page Design: Are We Doing It Right? (Editorial), Marmion, Dan, 2–3.
Library Web Proxy Use Survey Results, Murray, Peter E., 172–78.
Lyons-Mitchell, Kimberly A., and Knight, Lorrie A. *Measure for Measure: Statistics about Statistics,* 34–38.
Mach, Michelle. *The Service of Server-Side Includes* (Tutorial), 213–19.
Macromedia Authorware.
Using Macromedia Authorware for Web-Based Instruction (Communication), Cullen, Kevin F., 154–58.
Marmion, Dan.
Library Web Page Design: Are We Doing It Right? (Editorial), 2–3.
The Open Source Movement and Libraries, (Editorial), 171.
Why Is ITAL *Important?* (Editorial), 58–59.
McClennen, Michael, and Memmott, Patricia. *Roles in Digital Reference,* 143–48.
Measure for Measure: Statistics about Statistics, Knight, Lorrie A., and Lyons-Mitchell, Kimberly A., 34–38.

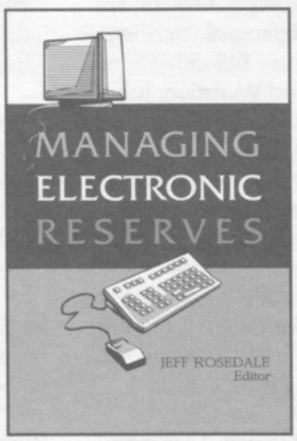
Memmott, Patricia, and Michael McClennen. *Roles in Digital Reference*, 143–48.
Microforms.
Improving Retrieval of Microform Set Records, Wilson, Mary Dabney, and Fullerton, John Paul, 60–65.
Multimedia.
Weaving Multimedia into the Collection: One Library's Journey (Communication), Golderman, Gail, 198–200.
Murray, Peter E. *Library Web Proxy Use Survey Results*, 172–78.
Nigeria.
Use of CD-ROM Databases in Nigerian Academic Libraries (Tutorial), Oduwole, A. A., 43–47.
NoteTab Pro 4.83 (Software Review), 54–55.
Oduwole, A. A. *Use of CD-ROM Databases in Nigerian Academic Libraries* (Tutorial), 43–47.
Olson, John A., Hart, Richard, and Simmonds, Patience L. *Laser Printing at Public Workstations: User Behaviors and Attitudes* (Tutorial), 39–43.
Online Library Instruction for Online Students, Viggiano, Rachel, and Ault, Meredith, 135–38.
Online Public Access Catalogs.
Boolean Browsing in an Information System: An Experimental Test, Allen, Bryce L., 12–20.
Laser Printing at Public Workstations: User Behaviors and Attitudes (Tutorial), Hart, Richard L., Olson, John A., and Simmonds, Patience L., 39–43.
Open Source Software.
The Open Source Movement and Libraries, (Editorial), Marmion, Dan, 171.
People Servers vs. Information Providers: The Impact of Service Orientation on Technology Training, Fidishun, Dolores, 29–33.
A Personalized Information Environment for Digital Libraries, Jayawardana, Champa, Hewagamage, K. Priyantha, and Hirakawa, Masahito, 185–96.
PrintKey 2000 v5.10 (Software Review), 164–65.
Privacy.
Human Values in a Technological Age, Gorman, Michael, 4–11.
Programming.
Customization of Library Service in a Cross-Platform Programming Environment, Bregman, Alvan, and Winnie S. Chan, 21–28.
Proxy Web Servers.
Library Web Proxy Use Survey Results, Murray, Peter E., 172–78.
Public Access Computer Systems.
Laser Printing at Public Workstations: User Behaviors and Attitudes (Tutorial), Hart, Richard L., Olson, John A., and Simmonds, Patience L., 39–43.
Library Web Proxy Use Survey Results, Murray, Peter E., 172–78.
Public Libraries.
Internet Access in Public Libraries: Results of an End User Targeted Pilot Study, 1997–2000 (Communication), Kibirige, Harry M., 113–15.
Public Services.
People Servers vs. Information Providers: The Impact of Service Orientation on Technology Training, Fidishun, Dolores, 29–33.
Roles in Digital Reference, McClennen, Michael, and Memmott, Patricia, 143–48.
Satori PhotoXL 2.29 (Software Review), 163–64.
Searching for the Library: University Home Page Design and Missing Links, Astroff, Roberta, 93–99.
Search Strategies.
Boolean Browsing in an Information System: An Experimental Test, Allen, Bryce L., 12–20.
The Service of Server-Side Includes (Tutorial), Mach, Michelle, 213–19.
Simmonds, Patience L., Olson, John A., and Hart, Richard. *Laser Printing at Public Workstations: User Behaviors and Attitudes* (Tutorial), 39–43.
Sottong, Stephen. *E-Book Technology: Waiting for the "False Pretender,"* 72–80.
Speakeasy Studio and Café: Information Literacy, Web-Based Library Instruction, and Technology, Jacobs, Mark, 66–71.
Staff Training.
People Servers vs. Information Providers: The Impact of Service Orientation on Technology Training, Fidishun, Dolores, 29–33.
Statistics.
Measure for Measure: Statistics about Statistics, Knight, Lorrie A., and Lyons-Mitchell, Kimberly A., 34–38.
Stormont, Sam. *Going Where the Users Are: Live Digital Reference*, 129–34.
Stress and Technology.
Human Values in a Technological Age, Gorman, Michael, 4–11.
SuperToolz Plug-ins (Software Review), 53–54.
Technology and Society.
Human Values in a Technological Age, Gorman, Michael, 4–11.
Technology—Life Cycle.
Book Technology: Waiting for the "False Pretender," Sottong, Stephen, 72–80.
Technology—Training.
People Servers vs. Information Providers: The Impact of Service Orientation on Technology Training, Fidishun, Dolores, 29–33.
Three Iterations of an Academic Library Web Site, Vaughan, Jason, 81–92.
TrayMenu 2.01 (Software Review), 55–56.
24/7 Reference Services. *See* Virtual Reference Services.
Use of CD-ROM Databases in Nigerian Academic Libraries (Tutorial), Oduwole, A. A., 43–47.
Use Studies.
Boolean Browsing in an Information System: An Experimental Test, Allen, Bryce L., 12–20.
Using Macromedia Authorware for Web-Based Instruction (Communication), Cullen, Kevin F. , 154–58.

USMARC.
Improving Retrieval of Microform Set Records, Wilson, Mary Dabney, and Fullerton, John Paul, 60–65.
Vaughan, Jason. *Three Iterations of an Academic Library Web Site*, 81–92.
Vendor Statistical Reporting.
Measure for Measure: Statistics about Statistics, Knight, Lorrie A., and Lyons-Mitchell, Kimberly A., 34–38.
Viggiano, Rachel, and Ault, Meredith. *Online Library Instruction for Online Students*, 135–38.
Virtual Reference: Today and Tomorrow (Editorial), Ciccone, Karen, 120–21.
Virtual Reference at North Carolina State: The First One Hundred Days, (Boyer, Joshua), 122–28.
Virtual Reference Services.
Creating an Instant Messaging Reference System, Fagan, Jody Condit, and Calloway, Michele, 202–212.
From Sshh to Search Engine: Reference.net on the Web, Kresh, Diane Nester, 139–42.
Going Where the Users Are: Live Digital Reference, Stormont, Sam, 129–34.
Online Library Instruction for Online Students, Viggiano, Rachel, and Ault, Meredith, 135–38.
Virtual Reference: Today and Tomorrow (Editorial), Ciccone, Karen, 120–21.
Virtual Reference at North Carolina State: The First One Hundred Days, (Boyer, Joshua), 122–28.
We'll Take It from Here: Further Developments We'd Like to See in Virtual Reference Services, Coffman, Steve, 149–53.
Wang, Feng-Kwei and Wedman, John. *eThemes: An Internet Instructional Resource Service*, 179–84.
Weaving Multimedia into the Collection: One Library's Journey (Communication), Golderman, Gail, 198–200.
Web.
Speakeasy Studio and Café: Information Literacy, Web-Based Library Instruction, and Technology, Jacobs, Mark, 66–71.
Web Pages—Design.
Searching for the Library: University Home Page Design and Missing Links, Astroff, Roberta, 93–99.
Web Pages—Usability.
Designing Web Usability (Book Review), by Nielsen, Jakob, 159–61.
Web Servers.
Library Web Proxy Use Survey Results, Murray, Peter E., 172–78.
Web Sites—Academic.
Three Iterations of an Academic Library Web Site, Vaughan, Jason, 81–92.
Web Sites—Education.
eThemes: An Internet Instructional Resource Service, Wang, Feng-Kwei and Wedman, John, 179–84.
Using Macromedia Authorware for Web-Based Instruction (Communication), Cullen, Kevin F., 154–58.
Speakeasy Studio and Café: Information Literacy, Web-Based Library Instruction, and Technology, Jacobs, Mark, 66–71.
Web Sites—Maintenance.
The Service of Server-Side Includes (Tutorial), Mach, Michelle., 213–19.
Wedman, John, and Wang, Feng-Kwei. *eThemes: An Internet Instructional Resource Service*, 179–84.
We'll Take It from Here: Further Developments We'd Like to See in Virtual Reference Services, Coffman, Steve, 149–53.
Why Is ITAL *Important?* (Editorial), Marmion, Dan, 58–59.
Wilson, Mary Dabney, and Fullerton, John Paul. *Improving Retrieval of Microform Set Records*, 60–65.
WinGrab 1.40 (Software Review), 164–65.
Winnie S. Chan, and Bregman, Alan. *Customization of Library Service in a Cross-Platform Programming Environment*, 21–28.
Zilner, Tom. *Designing Web Usability* (Book Review), by Nielsen, Jakob, 159–61.
———. *The Humane Interface* (Book Review), by Raskin, Jef, 48–50.

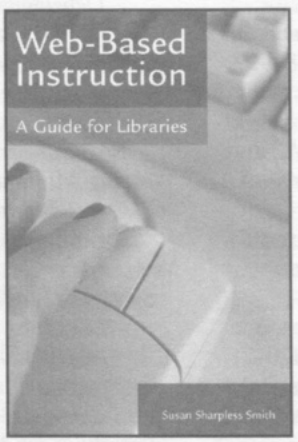
NEW Technology Titles from ALA Editions



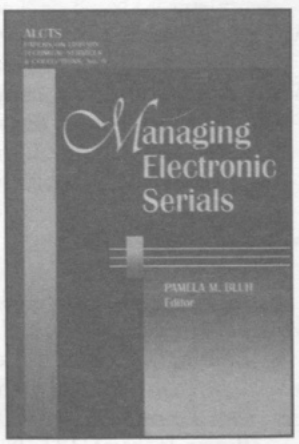
Price: $32.00
112 pages
6" X 9" • Softcover
ISBN: 0-8389-3511-7
ALA Order #: 3511-7-2262



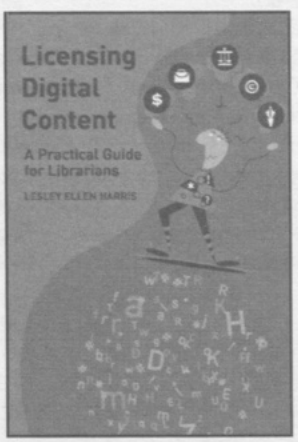
Price: $42.00
176 pages
6" X 9" • Softcover
ISBN: 0-8389-0812-8
ALA Order #: 0812-8-2262



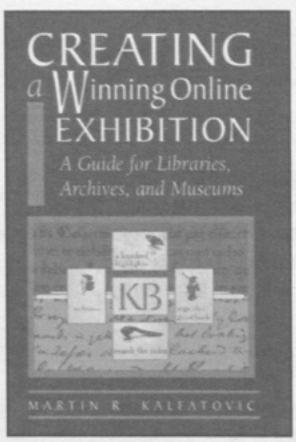
Price: $40.00
168 pages
7" X 10" • Softcover
ISBN: 0-8389-0805-5
ALA Order #: 0805-5-2262



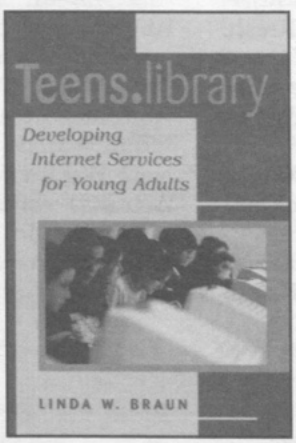
Price: $38.00
200 pages
6" X 9" • Softcover
ISBN: 0-8389-3510-9
ALA Order #: 3510-9-2262



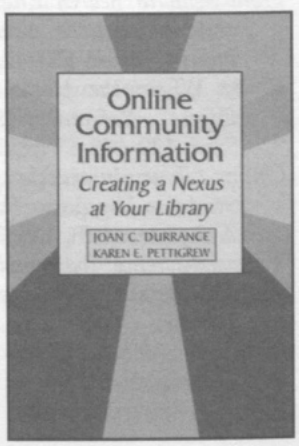
January 2002
Price: $45.00
144 pages
6" X 9" • Softcover
ISBN: 0-8389-0815-2
ALA Order #: 0815-2-2262



January 2002
Price: $40.00
136 pages
8.5" X 11" • Softcover
ISBN: 0-8389-0817-9
ALA Order #: 0817-9-2262



February 2002
Price: $28.00
96 pages
6" X 9" • Softcover
ISBN: 0-8389-0824-1
ALA Order #: 0817-9-2262



March 2002
Price: $42.00
208 pages
6" X 9" • Softcover
ISBN: 0-8389-0823-3
ALA Order #: 0823-3-2262

Visit the NEW "ALA Online Store" at

www.alastore.ala.org

(Or call us at 1-866-SHOP ALA)

ALA Editions

NETWORK CDs... WITHOUT CDs

SuperCD

AFFORDABLE AND FAST

Network your CD-ROMs today, but don't settle for yesterday's technology. Forget those big, slow and expensive CD towers. Using one high-speed CD drive and TurboCaching, SuperCD mirrors CD disks to a large hard drive. Super CD eliminates the need for multiple CD drives. It's super fast and super affordable, easy to use and easy to install.

Thousands of satisfied users appreciate:

- **Lowest cost per CD.**
- **Networks up to 512 CD titles.**
- **Serves 1,000 users.**
- **Fastest access times plus 66X transfer rates.**
- **Compact, fits on a shelf.**
- **Supports NT, Netware & Mac, simultaneously.**
- **Web Browser accessible.**
- **DVD compatible.**
- **Unburdens your LAN.**

BEST OF LAN TIMES 1996 SOFTWARE

Buy with confidence. Axonix backs its products with a FULL ONE YEAR WARRANTY. Plus, if you're not completely satisfied, you can RETURN THE PRODUCT WITHIN 30 DAYS for a complete refund. Axonix has a 18-year history and we understand the value of customer satisfaction.

Check our website for more SuperCD information.

For a free catalog, contact:
Axonix Corporation
844 South 200 East, Salt Lake City, UT 84111
800-866-9797 or sales@axonix.com

EDUCATION DISCOUNTS ARE AVAILABLE.

Tomorrow's Library TODAY

1.800 866.9797
www.axonix.com

AXONIX®
Share and Save