responding rise in scholarly journal prices. NESLI neither encourages nor hinders changes in scholarly communication and therefore the question of restructuring the scholarly communication process remains.[20]

References and Notes

1. Barbara McFadden and Arnold Hirshon, "Hanging Together to Avoid Hanging Separately: Opportunities for Academic Libraries and Consortia," *Information Technology and Libraries* 17, no. 1 (March 1998): 36. See also International Coalition of Library Consortia, "Statement of Current Perspective and Preferred Practices for the Selection and Purchase of Electronic Information," *Information Technology and Libraries* 17, no. 1 (March 1998): 45.
2. Martin S. White, "From PSLI to NESLI: Site Licensing for Electronic Journals," *New Review of Academic Librarianship* 3, (1997): 139–50. See also CHEST. *CHEST: Software, Data, and Information for Education* (1996).
3. Thomas J. DeLoughry, "Library Consortia Save Members Money on Electronic Materials," *The Chronicle of Higher Education* (Feb. 9, 1996): A21.
4. Information Services Subcommittee, "Principles for the Delivery of Content." Accessed Nov. 17, 1999, www.jisc.ac.uk/ pub97/nl_97.html#issc.
5. Joint Funding Council's Libraries Review Group. *The Follett Report.* (Dec. 1993): Accessed Nov. 20, 1999, www.niss.ac.uk/education/hefc/follett/report/.
6. John Kirriemuir, "Background of the eLib programme." Accessed Nov. 21, 1999, www.ukoln.ac.uk/services.elib/background/history.html.
7. PSLI Evaluation Team, "UK Pilot Site License Initiative: A Progress Report," *Serials* 10, no. 1 (1997): 17–20.
8. White, "From PSLI to NESLI," 149.
9. Tony Kidd, "Electronic Journals: Their Introduction and Exploitation in Academic Libraries in the UK," *Serials Review* 24, no. 1 (1998): 7–14.
10. Jill Taylor Roe, "United We Save, Divided We Spend: Current Purchasing Trends in Serials Acquisitions in the UK Academic Sector," *Serials Review* 24, no. 1 (1998): 3–6.
11. PSLI Evaluation Team, "UK Pilot Site License Initiative," 17–20.
12. Beverly Friedgood, "The UK National Site Licensing Initiative," *Serials* 11, no. 1 (1998): 37–39.
13. University of Manchester and Swets & Zeitlinger, *NESLI: National Electronic Site License Initiative* (1999). Accessed Nov. 21, 1999, www.nesli.ac.uk/.
14. NESLI Brochure, "Further Information for Librarians." Accessed Nov. 21, 1999, www.nesli.ac.uk/nesli-librarians-leaflet.html.
15. A copy of the model site license is available on the NESLI Web site. Accessed Nov. 22, 1999, www.nesli.ac.uk/ModelLicense8.html.
16. Albert Prior, "NESLI Progress through Collaboration," *Learned Publishing* 12, no. 1 (1999).
17. Science Direct. Accessed Nov. 24, 1999, www.sciencedirect.com.
18. Declan Butler, "The Writing is on the Web for Science Journals in Print," *Nature* 397, (Jan. 21 1998).
19. The Journal Access Core Collection Request for Proposal. Accessed Nov. 22, 1999, www.calstate.edu/tier3/cs+p/rfp_ifb/980160/980160.pdf.
20. Frederick J. Friend, "UK Pilot Site License Initiative: Is it Guiding Libraries Away from Disaster on the Rocks of Price Rises?" *Serials* 9, no. 2 (1996): 129–33.

A Low-Cost Library Database Solution

Mark England, Lura Joseph, and Nem W. Schlecht

Two locally created databases are made available to the world via the Web using an inexpensive but highly functional search engine created in-house. The technology consists of a microcomputer running UNIX to serve relational databases. CGI forms created using the programming language Perl offer flexible interface designs for database users and database maintainers.

Many libraries maintain indexes to local collections or resources and create databases or bibliographies concerning subjects of local or regional interest. These local resource indexes are of great value to researchers.

The Web provides an inexpensive means for broadly disseminating these indexes. For example, Kilcullen has described a nonsearchable, Web-based newspaper index that uses Microsoft Access 97.[1] Jacso has written about the use of Java applets to publish small directories and bibliographies.[2] Sturr has discussed the use of WAIS software to provide searchable online indexes.[3] Many of the Web-based local databases and search interfaces currently used by libraries may:

- have problems with functionality;
- lack provisions for efficient searching;
- be based on unreliable software;
- be based on software and hardware that is expensive to purchase or implement;
- be difficult for patrons to use; and
- be difficult for staff to maintain.

After trying several alternatives, staff members at the North Dakota State University Libraries have implemented an inexpensive but highly functional and reliable solution. We are now providing searchable indexes on the Web using a microcomputer running UNIX to serve relational databases. CGI forms created at the North Dakota State University Libraries using the programming language Perl offer flexible interface designs for database users and database maintainers. This article describes how we have imple-

Mark England (england@badlands.nodak.edu) is Assistant Director, **Lura Joseph** (ljoseph@badlands.nodak.edu) is Physical Sciences Librarian, and **Nem W. Schlecht** (schlecht@plains.nodak.edu) is a Systems Administrator at the North Dakota State University Libraries, Fargo, North Dakota.

mented this technology to distribute two local databases to the world via the Web. It is hoped that recounting our experiences will facilitate other such projects.

Creating the Databases

The two databases that we selected to use as demonstrations of this technology are a community newspaper index and a bibliography of publications related to North Dakota geology.

The Forum Index

The *Fargo Forum* is a daily newspaper published in Fargo, North Dakota. It began publication in 1879 and is the paper of record for North Dakota. For many years, the North Dakota State University Libraries have maintained an index to the *Forum*. Beginning with the selective indexing of notable events and editions, we started offering full-text indexing of the entire paper in 1996. Until early in the 1980s, all indexing was done manually and preserved on cards or paper. Then for several years, indexing was done on one of the university's mainframe computers. Starting in 1987, microcomputers were used to compile the index, first using DBASE and then using Pro-Cite as the database management software. Printed copies of the database were sold annually to subscribing libraries and businesses. Starting in the summer of 1996, the library made arrangements with the publisher of the paper to acquire digital copy of the text of each newspaper.

In early 1997, the NDSU Libraries began a project to place all of our *Forum* indexes on the Web. DBASE, Pro-Cite, WordPerfect, or Microsoft Access computer files existed for the newspaper index from 1879 to 1975, 1988, and from 1990 to 1996. All other data was unavailable or unreadable. Printed indexes from 1976 to 1987 and 1989 were scanned using a Hewlett Packard 4C scanner fitted with a page feeder. Optical character recognition was accomplished using the software OmniPage Pro. Once experience was gained with scanner and software settings, the scanning went very quickly with very few errors appearing in the data. Various members of the library staff volunteered to check and edit the data, and the digitizing of approximately 1,500 pages was completed in about three weeks.

All data were checked and normalized using Microsoft's Excel spreadsheet software and then saved as tab-delimited text. Programmer's File Editor was used to do the final text editing. Because of variations in the completeness of the indexing, three separate relational database tables were created: one each for the years 1879–1975, 1976–1996, and 1996–the present.

The Collective Bibliography of North Dakota Geology

In 1996 a project was initiated to combine three bibliographies of North Dakota geology and to make the final product searchable and browsable on the Web. All three of the original print bibliographies were published by the North Dakota Geological Survey. Scott published the first bibliography as a thesis. It is a bibliography of all then-known North Dakota geological literature published between 1805 and 1960, and most entries are annotated.[4] The second print bibliography, also by Scott, focuses on North Dakota geological literature published in the years 1960 through 1979, and also includes some material omitted in the first bibliography.[5] Most entries in the second bibliography include annotations in the form of keywords or keyword phrases. The third bibliography covers the years 1980 through 1993, and is not annotated.[6] All three bibliographies are indexed. The third bibliography was available in digital format, whereas the first two were in print format only.

Library staff members began rekeying the two print bibliographies using Microsoft Word. The remaining pages were digitally scanned using a new Hewlett Packard 4C scanner and the optical character recognition software OmniPage Pro. There were many errors in the resulting text. Different font sizes in the original documents may have contributed to optical recognition errors. Editing of the scanned pages was nearly as time consuming and tedious as rekeying the documents. The Microsoft Word documents were saved as text files and combined as a single text file. Programmer's File Editor was used as a final editor to remove any line breaks or other undesirable formatting.

Each record was edited to occupy one line, and each field was delimited by two asterisks. Asterisks were used because there were many occurrences of commas, semicolons, and other symbols that would have made it difficult to parse any other way. Because italics were removed by converting to a text file, some errors were made in parsing. In retrospect, parsing should have been done before the document was saved as a text file.

Punctuation between fields was removed because the database would be converted to a large table. It would have been better to leave the punctuation intact, since it cannot easily be put back in for the output to be presented in bibliographic form. The alphabetical additions to publication dates (e.g. Baker, 1966a) were left intact to aid in hand-cutting and pasting index terms into the records at a later date.

Initially, the resulting document was converted to a Microsoft Access file so that it would be in a table format. However, many of the fields

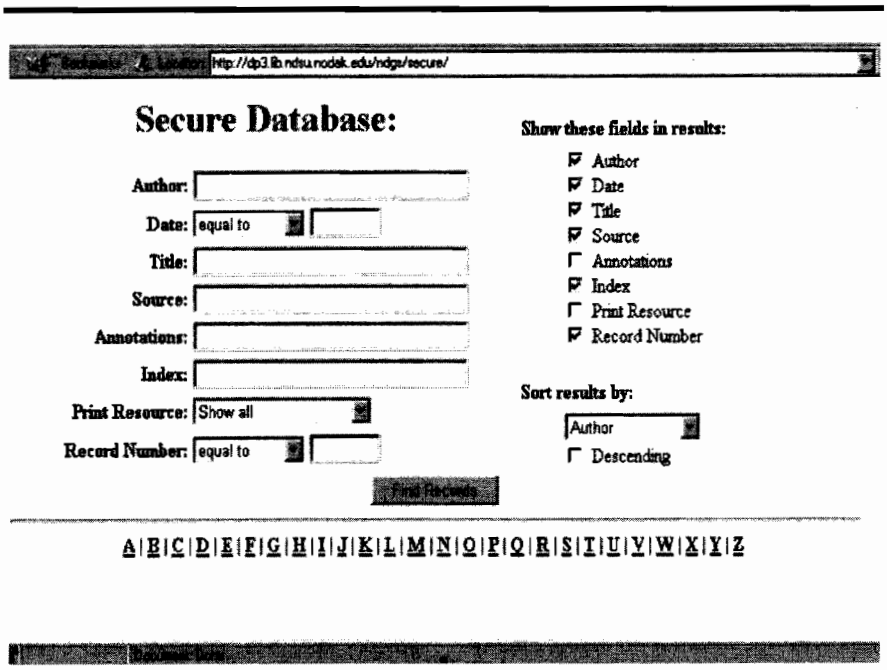

Figure 1: Secure Database Editing Interface

were well over the 256 character limit of individual fields. To solve this problem, the data were imported into a relational database called MySQL, which allows large data fields called "blobs." Running under UNIX, MySQL is very flexible and powerful.

Database and Search Engine Design

We examined the features and capabilities of various online bibliographies and indexes when deciding on our search interfaces and search engine designs. We wanted our databases to be both searchable and browsable and, in the case of the Collective Bibliography of North Dakota Geology, we wanted to provide the option of receiving search results accurately in a specific bibliographic format. We wanted both simple and advanced search capabilities, including the ability to do highly sophisticated Boolean searching. Finally, we wanted to provide those maintaining the databases with the ability to easily add, delete, and change records from within simple forms on the Web and immediately see the results of this editing.

MySQL uses a Perl interface, DBI (Database Independent Interface), which makes accessing the database simple from a Perl script. Essentially, a SQL statement is generated, based on data from an HTML form. This SQL statement is then run against the MySQL database, returning matching rows that the same script can handle and display as needed. All of the dynamically generated pages in this database are created this way. Using both MySQL and Perl provided a nice, elegant way to integrate database functionality with the Web.

The databases were installed on a server and made available via the Web. It soon became apparent that there were problems with large numbers of returns. Depending upon the client machine's hardware configuration, browsers could lock up the machine. While an efficient search should not result in such a large number of hits, we decided to limit returns to reduce this problem. Following suggestions from users, various search tips were added, and some search interface terminology was changed.

From a secure gateway, it is possible to call up different forms that allow individual records to be displayed, edited, and saved (see figure 1). New records are added by using a simple HTML form. It is also possible to bulk-load large numbers of records by using a special Perl program to load the data directly from a text file.

Advantages of the UNIX/MySQL Solution

After first using Glimpse, a popular Web search engine, under Linux, a free UNIX platform, and then Microsoft's Internet Information Server (IIS) software on a Windows NT platform to search the *Forum* newspaper index, we settled on using MySQL on a microcomputer running Linux and the Apache Web server. We found we could write Perl scripts that allowed users to make very sophisticated searches of the data from within very simple Web forms. MySQL is stable, reliable, free, and offers a high degree of functionality, flexibility, and efficiency. Apache is reliable, extendible, very fast, free, and offers tight control of data access.

Initially, each story received from the newspaper was maintained as a separate file on a microcomputer. By having the stories as separate files, it was easy to set up Glimpse as a searching tool for the articles. Although it did provide a nice preview of a workable system, Glimpse did not provide enough flexibility in how records were displayed, organized, or searched. It was not meant for managing data of this sort.

Windows NT, although a popular and successful IT solution, was

found to be somewhat cumbersome to implement and did not provide enough flexibility. The installation of these tools was easy, but it was difficult to obtain a high level of database and Web integration. Reliability and cost were also concerns.

We found that UNIX was more stable and practically eliminated any unavailability of the data. Perl, MySQL, and Apache were ultimately used to manage, store, and deliver the data. Although these products are available for Windows NT, their native platform is UNIX. By running these products on UNIX, we were able to take advantage of all the features offered by each of the products. We found that MySQL offered the flexibility and power to manage both sets of data efficiently. Also, to load the data into a relational database such as MySQL required the data to be normalized. Normalized data are data that are separated into logically separate components. To normalize data often takes some extra effort, as fields must be defined to contain certain types of data, but in the end the data is easier to manage and well organized. By having articles and bibliographies in a relational database, we are able to easily make updates, additions, and generate output or reports on the data in many different ways.

There are several Web servers available on the market today. However, Apache is often singled out as being the most popular server. Apache, like Perl and MySQL, is available free for all uses (educational and commercial). Using Apache and .htaccess control files, we are able to restrict access to administrative pages where data are added or modified. Many extensions for Apache are available to increase Web performance in different situations. For example, a module for Apache allows the Web server to execute Perl code within the server without the need to run the regular Perl interpreter.

Conclusion and Future Plans

Work is under way to refine and update *The Collective Bibliography of North Dakota Geology*. Because bibliography number three was not annotated, index terms are being added to facilitate searching and retrieval of citations. We have recently updated *The Collective Bibliography of North Dakota Geology* to include citations to publications through 1998, and we plan to update the database annually. Additionally, we receive monthly updates of *Forum* articles, which are added using a simple Perl script as soon as they are received. We have successfully implemented a number of other databases using these methods. We realize that this UNIX/MySQL solution is likely to be most helpful to other academic libraries: there are generally students and staff available on many campuses who are capable of programming in Perl and maintaining SQL databases on UNIX servers. Our Perl scripts are available at the URL ww.lib.ndsu.nodak.edu/lclds.

References and Notes

1. M. Kilcullen, "Publishing a Newspaper Index on the World Wide Web Using Microsoft Access 97," *The Indexer* 20, no. 4 (1997): 195–96.
2. P. Jacso, "Publishing Textual Databases on the Web," *Information Today* 15, no. 11 (1998): 33, 36
3. N.O. Sturr, "WAIS: An Internet Tool for Full-Text Indexing," *Computers in Libraries* 15 (June 1995): 52–54.
4. M.W. Scott, *Annotated Bibliography of the Geology of North Dakota 1806–1959* North Dakota Geological Survey Miscellaneous Series, no. 49. (Grand Forks, N.D.: North Dakota Geological Survey, 1972).
5. M. W. Scott, *Annotated Bibliography of the Geology of North Dakota 1960–1979* North Dakota Geological Survey Miscellaneous Series, no. 60. (Grand Forks, N.D.: North Dakota Geological Survey, 1981).
6. L. Greenwood and others, *Bibliography of the Geology of North Dakota 1980–1993* North Dakota Geological Survey Miscellaneous Series, no. 83. (Bismarck, N.D.: North Dakota Geological Survey, 1996).

Related URLs

Linux Homepage: www.linux.org/
MySQL Homepage: www.mysql.com/
Perl Homepage: www.perl.com/
Apache Homepage: www.apache.org/
NDSU Forum Index: www.lib.ndsu.nodak.edu/Forum/
Collective Bibliography of North Dakota Geology: www.lib.ndsu.nodak.edu/ndgs/