DESIDOC Bulletin of Information Technology, Vol. 22, No. 4 & 5, July & September 2002, pp. 9-15

Design and Development of Web-Enabled Databases in Libraries With Special Reference to RDBMS: Selection of Tools and Technologies

Ram Kumar Matoria & PK Upadhayay

Abstract

There are a number of ready-made library automation software packages in the market. However, many of these packages lack web interface and, thus, do not provide library database access (OPAC, etc.) through web. Moreover, these commercial packages provide less flexibility to librarians for make desired changes to the existing databases and to publish the same on the web. Therefore, designing in-house databases in libraries by using common relational database tools (RDBMS) like MS Access, SQL, DB2, etc., as back-end solutions is an area of deep concern.

This paper discusses the merits and demerits of the currently available RDBMS software for library applications as back-end solutions. The paper also discusses the various tools/technologies required/available for publishing the library databases over web. It has also been emphasized to use the technology that is more user friendly, easy to design, requires less programming skill thus, suitable for librarians to setup web-enabled solutions in the libraries.

1. INTRODUCTION

Application of computers certainly has improved the ways in which we acquire, process, store and disseminate information in libraries. As a result, we have automated house-keeping operations, catalogue and, user services in libraries. Moreover, with the advancement in the network technology during last decade, we have started providing user services through various online modes viz. e-mail, FTP, Telnet, etc. These static/interactive modes of data/information sharing over network have improved the library services and, their popularity among the users.

During 1994, libraries began to show their presence on the internet by setting up their web sites. In the beginning, libraries built their web pages by providing mainly static information. However, in the late 1997, libraries began re-engineering their web pages with the intention of providing information from databases in a dynamic manner[1]. Today the advancements in web and database technology have enabled the libraries to share/publish the data/information directly from the databases over the internet. Now web has become a dynamic platform for information delivery over the internet and very useful for libraries to provide various user services.

This paper intends to provide an overview of the database tools available for designing in-house databases in the libraries as well as the web technology for publishing such databases over the web. Besides, the paper discusses the need of developing in-house databases in libraries, advantages of web-enabled databases, and selection of database tools from RDBMS (relational

database management system) commonly available.

2. IN-HOUSE DATABASES IN LIBRARIES

There are a number of ready-made library automation software packages in the market which can be used to automate the library functions and user services. However, many of these packages lack web interface and, thus, do not provide library database access (OPAC, etc.) through web. Moreover, these commercial packages provide less flexibility to librarians for making desired changes to the existing databases and to publish the same on the web.

An in-house designed database system provides full flexibility to the developer/administrator/librarian and is simpler to install, maintain and, use without having to rely on external support, and also is highly customizable and can be designed to meet local needs[2]. Librarians can also participate in design/development process as now a days easy and user-friendly database tools are available with in-built web interface.

As per the survey conducted in order to get a clear understanding of the current role of library web master, 100 percent of the respondents felt that librarians should have some role in web site design and maintenance[3].

3. BENEFITS OF WEB-ENABLED DATABASES

Web-enabled databases have certain advantages over other means of sharing/publishing information over internet. These advantages are common to bibliographic databases also where libraries can put their OPACs and other databases on web. According to L.J.Haravu[4]; Kristin Aantelman[5] and; Chong Ho Yu[6], the benefits of web-enabled databases are as follows:

- Global/wide access to data with a single user interface
- Format independent access of data (text, graphics, audio, video, etc.)
- Up-to-minute updated access of data in real time
- Dynamic updation of data from many locations
- Cost-effective way of data sharing
- Instant feedback from users
- Seamless links from resource to resource
- Unlimited download/printing
- Environmental friendly.

Besides, Ortiz-Repiso[7] and Norm Medeiros[8] have written excellent papers on differences between traditional v/s innovative means (web OPACs) of library catalogs, while B. Ramesh Babu, *et al* [9] have presented an overview of web OPACS with their advantages in libraries.

4. DATABASE TOOLS

In the early days of database development, it was the flat-file system to store data in a single file by using field separators and record separators. Later, such flat-file systems were facilitated by an index in which a single index file stored keywords and pointers to the records that were stored in main file. This made retrieval more efficient, still flat-file systems were seriously inefficient. It showed data-redundancy, poor data control, and did not supported complex data types (audio, video, graphics, etc.)[10].

After 1960s, two database models were developed in quick succession to solve the limitations of flat-file systems. These database models were Hierarchical Database Model and Network Database Model. The Hierarchical Database composes of a root segment, parent segments and child segments. This model depicts a set of one-to-many relationships between a parent and its children. However, this model has number of limitations. It does not include ad hoc querying capability and, a multi-parent relationship is difficult to implement, as in a many-to-many relationship. The Network Database Model attempted to deal with many of the hierarchical model's limitations. The network database structure easily accommodates the multiple child, although, its structural complexities often limit its effectiveness and efficiency. Both of these

database models are obsolete now mainly because of their incompatibility with web environment and have been replaced by RDBMS.

RDBMS is the current database implementation standard being used for publishing information on the web. It includes the following features which make it suitable for web environment:

(a) Provide extremely useful tools for database administration;
(b) Adhere to powerful query language SQL (structured query language) developed by Microsoft Corporation;
(c) Compliant with ODBC/JDBC and other database interface tools;
(d) Work in client/server mode;
(e) Offer distributed database and distributed processing options; and
(f) Have referential integrity controls to ensure data consistency and, thus, suitable for web environment.

It is easier to design applications in RDBMS due to the following features:

- Requires less programming skill
- Availability of third party tools, and
- Inbuilt web interface.

These all features of RDBMS have made them ideal for library applications to setup bibliographic database solutions and publishing them over the web.

In the early days of library systems development, vendors have designed library applications software by using pre-relational database technology. Still some library systems suppliers are continuing to maintain and develop such products. These proprietary library applications software have their own advantages and disadvantages. However, in the library management systems marketplace, there is a growing trend towards the use of RDBMS as shown in the figure 1. and figure 2. (Source: *Biblio Tech Review: Database Updates,* 12 April 2001, http://www.biblio-tech.com/).

As per the Gartner[2] list of database management systems, there are 140 products from 67 vendors ranging from pre-relational,

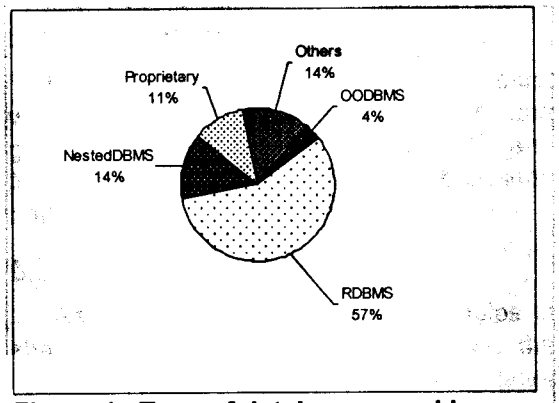

Figure 1. Type of databases used in library

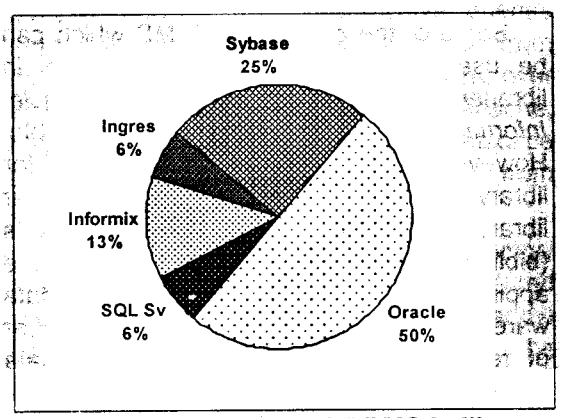

Figure 2. Use of various RDBMS in library systems

relational and, object-oriented databases (OODBMS). The RDBMS accounted three-quarters of the total database market. Table 1 shows that in the relational segment, there are 'Big Five' DBMS vendors which include Oracle, IBM, Microsoft, Informix, and Sybase (figure 3. Source: *Dataquest,* May 2001).

Table 1. Market percentage of vendors in DBMS sector

Company	2000 market (%)	1999 market (%)
Oracle	33.8	31.4
IBM	30.1	29.9
Microsoft	14.9	13.1
Sybase	3.2	3.3
Informix	3.0	5.0
Others	15.0	17.3

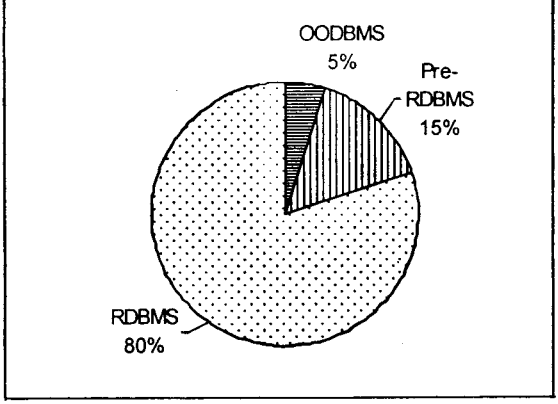

Figure 3. Worldwide DBMS new license revenue

Some of the common RDBMS which can be used to design in-house databases in libraries are *MS Access, SQL Server, Oracle, Informix, Sybase, File Maker Pro, SAS,* etc. However, the selection of suitable RDBMS for library applications depends on the particular library requirements, kind of applications (bibliographic/full-text/numeric), the application level (database level/data warehouse level), magnitude of data (number of records/length of fields), nature of data analysis (basic statistics, i.e., sum, average, mean/advance statistics) and, comprehensiveness and inter-operability. Other points to be considered while selecting database tools for library applications are easiness, user-friendliness, flexibility and low cost.

In most of the cases in libraries where bibliographic databases are dealt with, commonly available consumer database products such as *MS Access*, and *File Maker Pro* are sufficient. These products can support up to 1GB of data and provide cheaper solutions. However, these database tools are suitable for small databases and not useful for large databases, hence provide limitations in web environment. There are specific database solutions for large databases available from *MS SQL Server, DB2, Oracle, Informix, Sybase,* etc. These database tools support up to several terra byte of data; permit thousands of simultaneous users; work in client/server mode; show natural affinity towards web environment and, thus, suitable for large bibliographic, web-enabled databases in libraries. However, to design full-text databases in libraries where large files are to be stored and retrieved, these tools do not offer good solutions as most of these tools support 200-255-400 characters field length. Although, many of these tools allow unlimited entries in memo field, the data inside a memo field are not searchable and thus, offer limited utility. In this respect, *MS SQL Server* has extended the searchable field length up to 8063 characters and thus, provide solutions for moderate full-text databases.

As a common trend , a number of libraries have been using open source software available either free or on nominal prices. These database management systems include the most common *mySQL, mSQL* and, recently updated *PostgreSQL*. These database tools are easy to work and run on various platforms such as UNIX, Lynux, NT, etc.

Besides, there are some library specific text retrieval database tools for management of full-text large files, generally known as 'Free Form Databases', provide powerful search engine. Although, in some ways they are at the opposite end of the RDBMS, little database structure imposed on the designer and, thus, provide an easy and readymade database structure. However, they provide some limitations in achieving integration in the modules other than searching such as cataloguing, classification, management module, etc. These software are *BRS/Search, Basis Plus, Status,* etc.

5. WEB TECHNOLOGY

World Wide Web or WWW or web, sometimes wrongly mentioned as synonymous to internet, is a sub-set of it and is a most promising mode of publishing/sharing of data/information. Now, the web has emerged as the dominant protocol over internet due to its special features like easy interfacing of diverse platforms/technologies; provision to incorporate text, graphics, audio and video, etc.; permitting to access data/information in a non-linear way and; jumping from one

document/resource to another through hyperlinks. The superiority of www over other modes of online sharing viz. e-mail, FTP, Telnet, Gopher, etc., has been well explained by Paul Scully[11]. In fact, web has provided a single user interface to facilitate the user interaction with the databases or systems resources.

The web uses unique URL (universal resource locator) address which facilitates the access of information seamlessly across the diverse platforms in a client-server mode. The actual data represented in a structured way reside in a powerful computers called server, while the latter is equipped with server tools/software to populate the information over web. The client/user equipped with web browser such as Microsoft Internet Explorer (MSIE) and Netscape Navigator, etc., which interact with the server through common communications protocol TCP/IP over high speed network/internet. In fact, the client/user sends its query to the server while the latter processes the request and gets the response/results from the database and sends it back to the client/user over the web.

6. COMPONENTS OF WEB-BASED SYSTEMS

Following components are required for web-based systems (figure 4):

6.1 Web Browser

It is a software that helps users to interact with the server for sending queries and getting results over the web. MSIE and Netscape are the most popular web browsers, both of these are available free.

6.2 Web Server

Web server is the software that helps in sharing the system resources/files over the internet through web browsers. The latter communicate with a web server via the standard Hyper-Text Transfer Protocol (HTTP). In fact, web servers receive queries from the client, process it, get data from databases and send it back to the client over web. The popular web servers are Apache, Microsoft IIS, iPlanet from Sun-Netscape alliance, and newly developed *Zeus* from Zeus Technology Ltd.

Although web servers are increasingly found as components within full-featured server operating systems, there are some basic selection points that should be highlighted. These points are scalability and performance; high availability and reliability; interoperability and open standards; system administration and; security[12]. In fact, there are not much choice in the web server segments as many of them come along with database tools, show natural affinity with the database tools from same vendor, and some

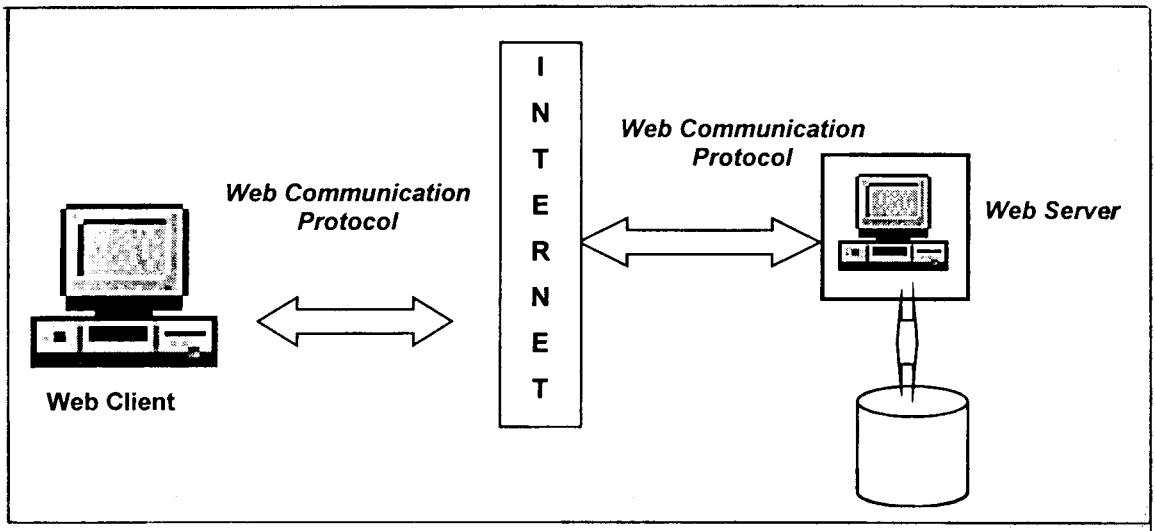

Figure 4. World Wide Web Architecture

other work better with certain platform/operating systems. According to the widely followed Netcraft[13] (www.netcraft.com/survey) web survey, approximately 60 percent of web servers run Apache, with Microsoft's IIS and the Netscape/iPlanet web server as its major competitors. However, Gartner[2] studies have given Netscape/iPlanet a larger share of public web sites, followed by Apache and Microsoft/IIS. Apache is being used for the public sites of companies such as Merck, American Express, Texaco, Chrysler, and others. However, it is easier to find both Microsoft/IIS and Netscape/iPlanet as the servers for large corporate public web sites when using the lookup facility offered by Netcraft (www.netcraft.com). Also, for intranets, most counts credit Microsoft/IIS and Netscape/iPlanet as the leaders, followed by Apache, IBM's Lotus Domino Go server, and others. Among hot e-commerce sites, E*Trade runs the Netscape/iPlanet server, eBay runs Windows/IIS, and Amazon runs Apache (the Apache-based Stronghold Secure Server), according to the Netcraft web survey[14].

6.3 Web Authoring Tools

Basic web server software does nothing except it sends HTML data in response to a browser request. However, certain other web-authoring and scripting tools are required to extend the functionality of web server to extract the data from databases. These web-authoring tools (*HTML* and its derivatives; and other Markup languages, etc.) and scripting tools (*Perl, C++, Java Script, JScript, VBScript,* etc.) are required to write web pages and scripts to interact with back-end databases. In early days of web-based applications, *CGI* (common gateway interface) programs were used extensively to extend the server functionality by allowing web users to interact with the web server to request specific data from the databases. Although, writing *CGI* program is difficult and requires programming skill, however, it works with high reliability. Now a days, a number visual editors are available to generate the web pages and write scripts automatically. The common editors are *Microsoft Front Page, Visual Interdev, Netscape Composer, Web Builder,* etc. Besides, other technologies have been introduced to communicate with the back-end databases. These technologies are Microsoft *Active Server Pages (ASP)* and *Java Server Pages (JSP)* from Sun Micro Systems. Both the technologies have combined the HTML codes with the scripts in a single page and thus, making it easier to interact with back-end databases.

During last few years, object-oriented and component based solutions were added to further make it easier to interact with the back-end databases. These solutions include *Microsoft COM/DCOM, ActiveX Data Objects (ADO); OMA* and *CORBA (Common Objects Request Broker Architecture)* from Object Management Group; and Sun's *EJB (Enterprise Java Beans), Java Applets (for client) and Java Servlets (for server),* etc. Besides, some database connectivity tools such as ODBC and JDBC have also emerged as in-built features of operating systems, etc.

6.4 Free Wizards and Third Party Tools

Writing HTML codes and scripts for database interaction is the most difficult part, requires programming skill, systems logic and manipulation planning, especially for non-computer professionals. However, a number of free wizards and third party tools are available on internet which can generate hundred of lines of codes automatically with a single click, thus, making it easier for non-skilled users to set up web-enabled database solutions in libraries.

Some of the free wizards are Rule Zero (http://www.Rulezero.com), Power ASP (http://www.Powerasp.com), Active Server Pages Tools & Components (http://www.web-savant.com/users/kathi/asp/tools.html), etc. Besides, *Microsoft FrontPage* also provides database wizard to generate *ASP* pages/*VBScripts* automatically to interact with back-end databases, while *Visual InterDev* provides a visual interface to develop a web-enabled web site with minimum efforts.

7. CONCLUSION

Now a days, libraries are using web technology extensively for user services to provide dynamic services to the remote users. Web technology currently available is helping the libraries to achieve their purposes. Although, web technology available, especially from Microsoft, is easier to setup the applications. It requires less programming skill and, is suitable for librarians. However, there is a need for librarians to get extensive exposure with the computer and communications technology in general and web technology specifically. The advancement in database technology especially RDBMS has made them ideal for library applications. These RDBMS provide simple architecture, easy to design and are equipped with flexible data exchange interface and, thus, present ideal solutions for web-enabled databases in libraries.

REFERENCES

1. Jordan, William. My gateway at the University of Washington libraries. *Information Technology and Libraries,* 2000, **19**(4), 180-85.
2. Gardner, Mike & Pinfield, Stephen. Database-backed library web sites: A case study of the use of PHP and mySQL at the University of Nottingham. *Program,* 2001, **35**(1), 33-42.
3. Taylor, Mary K. Library webmasters: Satisfactions, dissatisfactions, and expectations. *Information Technology and Libraries,* 2000, **19**(3), 116-24.
4. Haravu, L.J. Web-enabled databases: An overview of technology and problems. *In* Vision of Future Library and Information Systems, *edited by* T. Ashok Balu; L.S. Ramaiah; S.C. Sexena & D.S. Bedi. Viva Books, New Delhi, 2000, 61-66.
5. Antelman, Kristin. Getting out of the HTML business: The database driven web site solution. *Information Technology and Libraries,* 1999, **18**(4), 176-81.
6. Yu, Chong Ho; Jannasch-Pennell, Angel & DiGangi, Samuel A . Opportunities and options for web-enabled databases: Comparing in choosing the right software for virtual courses and communities.
7. Ortiz-Repiso, Virginia; & Moscoso, P. Web-based OPACs: Between tradition and innovation. *Information Technology and Libraries,* 1999, **18**(2), 68-77.
8. Medeiros, Norm. Delivering the goods: Web OPACs and the expanding role of cataloguer. *Issues in Science and Technology Librarianship,* Spring 1998, 1-6.
9. Babu, B. Ramesh. Web OPACs in the West: An overview. *In* Vision of Future Library and Information Systems, *edited by* T. Ashok Balu; L.S. Ramaiah; S.C. Sexena; & D.S. Bedi. Viva Books, New Delhi, 2000, 128-36.
10. Bhartiya, Chandra Shekhar. What databases are all about. *Information Technology,* Nov. 2001, 20-27.
11. Scully, Paul. Weaving your web: Practicalities of web site planning and design. *LASIE*, Sept. 1998, 33-39.
12. Lubrano, Cynthia R. Web server technology—Unix vs. Windows: Perspective. *Datapro Reports,* 15 Nov. 2000, 1-8.
13. Market share for top servers across all domains, August 1995-October 2001, Oct. 2001. http://www.netcraft.com/survey/.
14. Osmundsen, Sheila. Apache web server. *Datapro Reports,* 1 June 2000, 1-9.

Contributors: **Shri Ram Kumar Matoria** is Scientist-C at Library & Information Services Division, National Informatics Centre (NIC), A-Block, CGO Complex, Lodhi Road, New Delhi-110 003. e-mail: niclib@hub.nic.in; http://lib.nic.in; http://news.delhi.nic.in:90/niclibrary

Shri P. K. Upadhyay is Scientist-C at Library & Information Services Division, NIC, New Delhi.