International Journal of Scientific Research and Modern Education (IJSRME)
Impact Factor: 6.225, ISSN (Online): 2455 – 5630
(www.rdmodernresearch.com) Volume 2, Issue 2, 2017

THE ARCHITECTURE OF A DIGITAL LIBRARY SYSTEM – A STUDY

Dr. P. Giri Naidu
Assistant University Librarian, Rashtriya Sanskrit Vidyapeetha, Tirupati, Andhra Pradesh

Cite This Article: Dr. P. Giri Naidu, "The Architecture of a Digital Library System – A Study", International Journal of Scientific Research and Modern Education, Volume 2, Issue 2, Page Number 95-98, 2017.

Abstract:

Digital Library System design allows individual organizations to include their own material in the Digital Library System or to take advantage of network based information and services offered by others. It includes data that may be internal to a given organization and that which crosses organizational boundaries. This document presents a plan to develop such a system on an experimental basis with the cooperation of the research community. Digital Library System are easily as large as those derived from internal combustion engines and electric motors in the early part of this century. The principal function of the Indexing, Cataloging and Referencing Server(s) in the Digital Library System is to provide global cataloging and indexing services for the retrieval of Library content. The system is organized to support multiple, cooperating servers. It is also planned to accommodate alternative, specialized Indexing, Cataloging and Referencing Servers within this architecture to take advantage of new ideas and implementations without requiring the removal or replacement of existing services.

Key Words: Digital – Library, Document, Network – Cataloging, Database, Server & Import - Export

The Digital Library System design allows individual organizations to include their own material in the Digital Library System or to take advantage of network based information and services offered by others. It includes data that may be internal to a given organization and that which crosses organizational boundaries. This document presents a plan to develop such a system on an experimental basis with the cooperation of the research community. Finally, it addresses the application of a Digital Library System to meet a wide variety of user needs.

The productivity gains from having access to a Digital Library System are easily as large as those derived from internal combustion engines and electric motors in the early part of this century. Just as a car on an interstate highway is vastly more effective than one on a rutted dirt road, computer-based information "vehicles" can be made dramatically more effective given the proper operating environment. Computer and communications technology has made it possible for old fashioned, slow retrieval methods to be replaced by virtually instantaneous electronic retrieval. Each user of this technology can anticipate enormous potential benefit, but we lack the natural infrastructure to support this capability on a widespread basis today. This absence of infrastructure represents both a barrier and an opportunity of dramatic proportions.

Overview of Major Library System Components:

In the sections that follow, we will examine each of the major Library System components in turn, describing their functionality and relationships with other components. The rationale for the order in which these components are described is based on following a document [or, more generally, an object as it makes its way into the Digital Library and is then accessed and used.

The principal components of the system are:

- ✓ Import/Export Server
- ✓ Registration Server
- ✓ Indexing, Cataloging and Reference Servers
- ✓ Database Servers

In addition to these four basic components, there are two fundamental concepts which are intrinsic to the interaction of these various subsystems. These concepts are Knowbots and Shared Icon Geography which are discussed in more detail in Section 3. The initial information in the Digital Library system is assumed to be material which was originally intended to be printed (including multi-font text, graphics, bitmapped imagery) or otherwise displayed in static form. In addition to books, reports and periodicals, the system can include other material such as electronic mail, VLSI designs and organization charts. However, the underlying concepts will be easily extendable to allow more ambitious kinds of information such as holographs and digital films.

Import/Export Servers:

An Import/Export Server acts as a primary interface between the Digital Library System and the outside world. Contributions to and acquisitions for the Digital Library are presented through an Import Server. The method of interaction with an Import Server forms one of the most important interfaces in the system. An Import Server will be capable of accepting contributions to the Library in many forms. Contributions and submissions might arrive as part of an electronic mail message, as a CD-ROM, as a magnetic tape, as a PC

diskette or even as a facsimile scan. The common denominator is that the information has been converted to some definable digital form. One of the most important steps in the Digital Library design will be the determination of how many and which arrival formats will be acceptable. Conversion from analog to digital form, while an important consideration, is outside the scope of the library project.

The arriving objects (e.g., documents) must come with additional information if they are to be successfully entered into the DLS. Among other things, the Digital Library needs to know the origin of the object (bona fides); the owner of it (especially if any intellectual property rights are to be accounted for); terms and conditions for use, reproduction and access (including access control lists on an individual or organizational basis, for instance); descriptive information which might aid in retrieval; relation to existing information in the Library (e.g., part of a periodical series, book series, revision, etc.); and format definition.

Information which is not in a form which can be directly accepted at an Import Server will have to be prepared by services outside the Digital Library (an opportunity for any number of public agencies or private businesses). Similarly, Import Servers for particular classes of information might be implemented and operated or sold competitively.

An Import Server extracts the information relevant to registration from the arriving submission, packages it for processing by a Registration Server, and then forms and launches a Knowbot to deliver it there. At this point, the simple model is to send all of the information, including the actual submission, along with the registration Knowbot. This could prove impractical for significant contributions such as books. An alternative is for the registration Knowbot to carry only the information needed by a Registration Server and to carry references to the storage facilities at an Import Server for use when the information is to be transferred and incorporated into a database or catalogued by an Indexing, Cataloging and Reference Server.

An Import/Export Server also provides a basic mechanism for the equivalent of interlibrary exchange services. It should be possible for several, otherwise distinct, Digital Library Systems to exchange information, queries, responses and library contents. Analogous to conventional inter-library loans, this capability is essential if the Digital Library System technology is to be independently proliferated to support a variety of products and services. Every effort must be made to assure that the architecture is free of the assumption that a single system is unique in the information universe. This does not rule out the need to tightly integrate some Digital Library components into a particular coherent system, but emphasizes the need to tolerate and accommodate diversity.

It is not yet clear whether the inter-library exchange facility can be implemented merely as an electronic message exchange or whether the interaction should also permit more immediate and direct forms of Knowbot exchange. The latter may require too much context sharing or accounting/billing and authentication mechanism to be implemented for essentially distinct Digital Library Systems. Additional research will be required on these matters. For the present, it is assumed that an electronic message exchange convention will be the basis for interactions among distinct Digital Library Systems.

All such systems, if they are to interact at all, must share a common name and address space to support message exchange. This could be provided by relying on international electronic messaging conventions which include provision of such a common name and address space for electronic mailboxes.

In addition to its import functions, an Import/Export server has the responsibility for exporting information (objects) from the local library environment to other environments. The latter may be other libraries or other presentation media (paper, CD-ROM, facsimile, etc.). An object may be exported either as the result of an action taken by a user (or a Knowbot acting on behalf of a user) or as a consequence of a request for service imported from another library system. Although the inter-library exchange mechanism is assumed to be based on electronic mail, other less general but perhaps more efficient choices are possible. Other media conversions (e.g., to print) may have to be handled in idiosyncratic ways.

Registration Servers:

Registration Server(s) are responsible for 1) receiving messages from or hosting arriving Knowbots carrying new information (or references to new information) to be added to the Digital Library, and 2) registering new users, sources of information (databases) or other components newly added to the system.

One of the most important tasks of a Registration Server is to associate a unique identifier with any new object in the system. Ideally, it should never be necessary to re-use any identifier; thus the identifiers need to be allowed to increase in length. If identifiers are to be assigned by more than one Registration Server, methods must be invoked to assure uniqueness (e.g., by prefixing the object identifiers with Registration Server identifiers).

A Registration Server reports the existence of a new object to the relevant Digital Library component. If the object is a new user, this is reported to the Accounting System and to the Indexing, Cataloging and Reference Server(s) so that queries regarding that particular user can be properly answered. New information to be added to the Library is likewise reported to the Accounting system in the event that charges are to be associated with its access and use. A Registration Server may also supply a description of the charging algorithm to be used for this information. This might be as simple as a reference to a standard algorithm or as complex as a program for computing use charges for the particular item.

If it is readily apparent which database server(s) should house the arriving object, a Registration Server will so inform the Indexing, Cataloging and Referencing Server(s) and direct the Registration Knowbot to ferry the data to the appropriate Database Server. Alternatively, if the information did not come along with the registration Knowbot, a Registration Server can form a new Knowbot to pick up the information from the Import Server and deliver it to the appropriate Database Servers.

Registration Servers interact directly with Indexing, Cataloging and Referencing Servers by providing them with an instance of the object being registered. An Indexing, Cataloging and Referencing Server determines which database can house the object (there may be more than one) and reports this information to the Registration Server. Other items, in addition to documents, which require registration in the reference database include, inter alia, all intra-library servers, users and other known Digital Library Systems.

Indexing, Cataloging and Referencing Servers:

The principal function of the Indexing, Cataloging and Referencing Server(s) in the Digital Library System is to provide global cataloging and indexing services for the retrieval of Library content. The system is organized to support multiple, cooperating servers. It is also planned to accommodate alternative, specialized Indexing, Cataloging and Referencing Servers within this architecture to take advantage of new ideas and implementations without requiring the removal or replacement of existing services.

An important design issue will be the control of potentially open-ended interactions between Registration Servers and multiple Indexing, Cataloging and Referencing Servers to avoid network congestion and deal with the resulting multiple copy database update problem. Criteria for selecting among alternative Indexing, Cataloging and Referencing Servers must be worked out, if several deal with the same or inter-related information. It is easier to deal with the case that knowledge about the content of the Digital Library is partitioned non-redundantly among several multiple servers. For instance, one server might specialize in cataloging and indexing electronic mail messages, another in books and a third in journals or other periodicals. Alternatively, if redundancy is to be supported, it might be based on multiple, complete, copies of identical indexing and cataloging information, rather than overlapping or partitioned components. Maintaining a consistent set of registration database copies is an interesting challenge in its own right.

Indexing, Cataloging and Referencing Servers are also used to locate services and users as well as information in the Digital Library. This function has an analog in the electronic mail domain in which name servers make it possible to find mailboxes associated with users. Search criteria for the name servers may be as simple as first and last personal names or complex conditional expressions, involving job title and/or function, company name, special interests (if known), locale and other identifying characteristics.

There are two distinct questions which can be answered by the Indexing, Cataloging and Referencing Server when it is dealing with Library content:

♦ "Here is the data, where should it be stored?" ♦ "Here is the kind of data I want, where is it?"

These two functions are, in fact, very similar and require the same, base level input information. Thus, any tools developed for one function can potentially carry over to the other.

Each Indexing, Cataloging and Referencing Server is capable of carrying out a repertoire of functions which can be invoked by Knowbots arriving at the Server. Knowbots arriving at a Indexing, Cataloging and Referencing Server will usually be performing one of several specific tasks:

♦ Cataloging/indexing of a new Library acquisition. ♦ Searching for a cataloged or indexed item.

♦ Collecting statistics about the content or usage of the Library.

When a new item is registered, a Knowbot is dispatched to an Indexing, Cataloging and Referencing Server for guidance in cataloging and indexing. The arriving Knowbot carries with it any key word or other cataloging and index terms that may have been assigned on publication (e.g., by the Library of Congress, the journal publisher, the author, etc.). It may also carry the actual item content so as to support cataloging and indexing algorithms which operate on the full "text" of the new item. Of course, the Knowbot also carries information such as the source (author), copyright owner (if any), International Standard Book Number (or other identification of this type), publisher, date, place (and time?) of publication. Both published and unpublished works could be included.

The indexing or cataloging information may vary depending on the nature of the new item. For example, arriving electronic mail would typically be indexed by origin, To: and CC: recipients, date and time of origin, unique message identifier, originating mail system, subject matter, and depending on the Indexing, Cataloging and Referencing Server, by key words or user-provided search terms.

Database Servers:

The design of the Digital Library System is intended to accommodate existing databases and database services and to provide a framework for new databases organized around the concept of Knowbotic information storage and retrieval. Database Servers bridge the gap between already existing, database services and the Digital Library System by providing support for resident and arriving Knowbots and exchange of inter-Knowbot messages. The principal tasks of the Database Servers are:

♦ To accept and store new information, and

♦ To house arriving Knowbots bearing queries

Some Database Servers may only provide the second of these functions as is likely to be the case if the actual database is managed and updated essentially outside the Digital Library System context. For database systems which are designed to operate within the Knowbotic paradigm of the Digital Library System, the functions of the Database Server may actually be combined with the database system itself. It is possible, of course, that these functions might still be supported by a separate Database Server for efficiency reasons.

Another motivation for including the Database Server in the architecture is to utilize new parallel processing technologies to speed the search and retrieval Process for both new and existing database systems. Full text databases could be searched in their entirety at very high speed. Coupled with the Knowbot concept, such special purpose servers could revolutionize the utility of existing databases. To achieve this goal, it would probably be necessary to collocate the Database Server and the database system it serves so as to provide an economical but very high speed interconnection between the two. For existing databases, such a specialized Database Server would absorb the entire database so as to permit ultra-high speed and novel searching algorithms to be applied independent of its pre-existing computational base.

Such an intimate link between the Database Server and the database will doubtless require both technical and business arrangements, particularly in cases where the database is considered to be proprietary. Where such an arrangement proves infeasible, the alternative is to configure the Database Server so that it looks to the database as an ordinary user but provides all the of required framework for interfacing to the Knowbots of the Digital Library System.

Conclusion:

The architectural framework includes the DLS functional components, the methodology by which the participating systems communicate with each other, and active, mobile software components, called Knowbots, which perform services for the users. Subsequent volumes will address detailed technical aspects of the architecture such as the design of Knowbots and the protocols required to bind the DLS components together. This research to specify the overall structure and function of a DLS and to provide a basis for subsequent creation of an experimental system to evaluate the concept with real users. The DLS provides each user with the capability to use other cooperating digital libraries and provides the necessary search, retrieval and accounting capabilities to support ready access to local and network-based information. The various digital libraries and the associated access to network based capabilities are integral parts of the Digital Library System. Convenient access to local and remote information, without regard for its location, is an essential goal of the system design.

References:

1. Cornell University Library/Research Departments. (2000), Moving theory into practice: digital Image for libraries and archives. Research Libraries Group. Available at http://www.library.cornell.edu/preservation/tutorial
2. Digital Library Federation. (2001), Registry of Digitized Books and Serial Publication, Available at http://www.digilib.org/collections
3. Ian, H. Witten & David, Brainbridge. (2003), How to Build a Digital Library, London: Morgan Kaufman Publishers
4. Sitts, Maxine K. (2000), Handbook for Digital Projects: A Management Tool for Preservation and Access. Northeast Document Conservation Center, Andover, Massachusetts. USA. http://www.nedcc.org/digital/dman.pdf
5. Smith, Abbey (2001), Strategies for Building Digitized Collection. Washington, D.C. Digital Library Federation, Council on Library and Information Resources. Available at http://www.clir.org