REVIEW OF RESEARCH

IMPACT FACTOR : 5.7631(UIF) UGC APPROVED JOURNAL NO. 48514 ISSN: 2249-894X

VOLUME - 8 | ISSUE - 7 | APRIL - 2019

A FRAMEWORK OF DIGITAL LIBRARY

Dr. B.V. Chalukya
Librarian, SCS College, Omerga, (MS) India.

ABSTRACT:

Digital libraries are built around the Internet and web technology; hence, they need to follow the Internet standards and protocols so as to determine interoperability, portability, modularity, scalability and seamless accessibility. A typical digital library implementation follows client-server framework as does the Internet and web technology. The objectives of this section are to discuss and impart knowledge on the following aspects of the technical infrastructure of a digital library:Client-server framework as useful to the digital library;Important features like scalability and sustainability, seamless access, interoperability, federation, capacity to handle several files & formats and location- independent identifiers, that should be measured while designing a digital library;Digital library design models and framework with examples of digital libraries built using these models;Important Internet protocols and standards as beneficial to digital libraries;Computers and Network Infrastructure: Along with Server-side Hardware Components, Server-side Software Components, also Client-side Hardware & Software Components; andInteroperability in Digital Library

A typical digital library in a distributed client-server environment comprises of hardware and software elements at server side as well as at the client's side. It briefly characterize interoperability in digital library. The paper describes all components with examples of software products that are available in the market place.

KEYWORDS: *Digital Library, API, Client Server Framework, NCSTRL, CRADDL, FDBS, National Science Digital Library, NDLTD, Software Agents Framework, OAIS,etc.*

1. INTRODUCTION:

The Internet and web technology are precept system deployed in a digital library to search, navigate and deliver electronic resources across the globe. The digital contents in a digital library can be reachable on a single location or distributed across the network. End users or clients receive direct and seamless access to the information requested from a collection regardless of where the data is physically stored. A typical digital library implementation follows client-server framework as does the Internet and web technology. The online information search services like DIALOG, BRS Search and STN worked on "Host-terminal Technology", wherein hosts were huge mainframe computers that controlled all aspects of search and communication sessions and dumb terminals were connected to the host computer. A centralized server managed communications, user query interaction, database management and data presentation. The data from different sources had to be converted into a single homogeneous structure and organization. In contrast, enabling technology behind digital libraries provides for seamless access to heterogeneous digital objects created on different platforms and hosted in a diverse environment distributed at different locations on the Internet.

2. CLIENT-SERVER FRAMEWORK AND MIDDLEWARE:

The development of client-server framework is the major enabling technology behind distributed computing and databases. Client and server connect to equally computer programs as well as to the computers. The client program commonly resides on the user's personal computer, when the server program resides on a server that hosts information contents. The server program and client program impart over a telecommunication network applying a well-defined protocol. The client program (web browser) is reliable for making a request to the server and for starring the information it retrieves from the server. The server is responsible for receiving request from the client, controlling access to the information, performing the computation needed to recover the information, sending desired information to the client after authentication, if required and recording usage statistics. Equally client and server have their tasks to execute and, therefore, workload is balanced between the two. Today's PC-based clients can handle multiple tasks, such as maintaining simultaneous connections to assortment of sources. This results in clear approach to information resources distributed beyond the Internet regardless of its location. The server controls the database management tasks and processing of requests after client.

Middleware are a computer program that affixes software elements or applications on clients and servers. Middleware resides equally on client and server. Middleware certify that the client and server can impart with each other irrespective of different hardware and software involved. This is made achievable by use of standardized arrangement of messages called protocols. The Application Programming Interface (API) is the middleware basic that facilitates the moving of messages between clients and server based on protocol. The APIprotocol explains a set of messages that equally the client and server understand. The client's APItransfers the message within a form that is platform independent and transmits it to the server over the network. The server's API receives the message and translates into a form that the server understands. The server receives the message and responds to the client through its API. Middleware are used most often to support complex, distributed applications. It involves web servers, application servers, content management systems, and comparable tools that support application development and delivery.

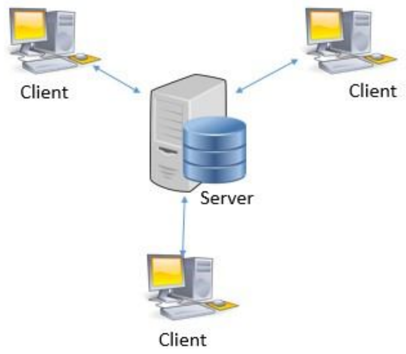

Figure 1. Client Server Framework

The very commonly consumed clients for Internet browsing, called browsers, are Microsoft's Internet Explorer and Netscape Navigator. The clients (or browsers) recognize different communication protocols so as to relate to different type of servers. Browsers which assist access to resources anchored on the web servers are growing in popularity due to their relative ease-of-use, user-friendly interfaces, free availability and increasing capabilities.

3. KEY PRINCIPLES: DIGITAL LIBRARY FRAMEWORK:

Major problems of digital library design are caused by separations in the computer systems, file structure, formats, information organization and various information retrieval requirements of collections (such e-journals, e-books, reference sources, online courseware, GIS, etc.) accessibleover the digital library. While the web has emerged as the preferred media of information delivery and access, the use of standards and protocols compose it achievable to make digital collections interoperable and accessible seamlessly. Some of the important features that should be considered while designing a digital library framework are as follows:

Open Framework:

Open framework concerns to computer framework or software framework that gives adding, upgrading and swapping components or software modules. As opposed to open framework, software and hardware with closed framework have pre-defined modules or components that are not generally upgradable.

Open framework gives potential users to see internally all or components of the framework without any proprietary constraints. Typically, an open framework publishes all or parts of its framework that the developer or integrator wants to share. Digital library project should treat open framework and a set of well-defined standards and protocols so as to facilitate scalability and interoperability.

Scalability, Extensibility and Sustainability:

The scalability, extensibility and sustainability are three most significant design features of a digital library that addresses the issue of the ability of a digital library to handle the increased volume of digital objects and its ability to sustain it for a long period of time. The digital library design should ensure that software should be able to handle large quantity of data, and hardware and network should be scalable to handle large quantity of digital objects and its transmission over the network. Moreover, digital library design and planning should provide for human and financial resources required for sustaining the digital library on long-term basis.

Seamless Access:

The digital libraries should provide transparent, seamless and platform-independent access to distributed array of information resources to users.

Interoperability:

Interoperability addresses the issue of ability of digital libraries and its components to work together effectively in order to exchange information in a useful and meaningful manner. Use of open framework and a set of well-defined standards and protocols ensure interoperability amongst heterogeneous digital libraries in a distributed environment.

Federation:

Federation refers to distribution of responsibilities for content creation, management and administration of various functions and service of digital libraries. It needs to be ensured that the participants follow the agreed standards, technologies and tools.

Digital Preservation:

The framework of the digital library needdetermine persistent and long-term access to its collection.

Modularity:

It is a design approach that attach to four fundamental principles of cohesiveness, encapsulation, self-containment and high binding to design a system component as an independently operable unit.

Platform Independence:

The digital library framework should occur platform- independent equally at hardware and software level.

Multiplicity of Files & Formats:

Digital library should be capable to operate multiple files and formats such as unstructured / structured text, audio, video, images, graphics, animation, etc.

Location-independent Identifiers:

Digital intents in the digital library should uphold location-independent identifiers such as handles, PURLs, DOI and OpenURL.

4. Key Components: Digital Library Framework and Design:

Different digital libraries have their own underlying design and framework. Most digital library frameworks provide for the five key components.

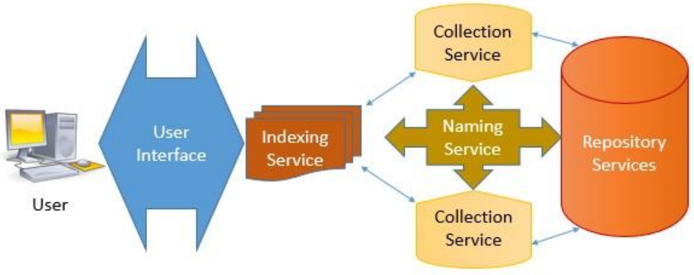

Figure 2. Digital Library Framework and Design

4.1 User Interfaces:

Digital libraries are required to provide interfaces for the user facilitating them to examine its collection, conduct searches, navigate through hierarchical menus of subjects, select and deselect searchable options, and sort search results in a fashion required by them.

4.2 Digital Repository:

Digital repositories accumulate and operate digital objects and metadata. The digital targets in a digital library may be "born digital" or digitized from the legacy document through the process of scanning. The metadata, that explains the digital objects to accelerate searching and discovery, may be extracted automatically or created manually. A bigdigital library may have various distributed repositories depending on collections it keeps. The digital library creators collaborate with digital repository using Repository Access Protocol (RAP). RAP appreciates rights and permissions to apply intellectual property rights, if required. E-commerce functionalities may also be performing, if required to handle accounting and billing.

4.3 Digital Objects Naming Service: Unique Identifiers:

Digital objects in a repository need location-independent unique identifiers. These identifiers must endure valid whenever documents are replaced from one location to another, or are migrated from one storage medium to another. A number of registry or resolver-based applications are being consumed currently for providing persistent URLs to digital objects. These unique identification schemes do not directly describe the location of the resource to be retrieved, but instead straight a user to an intermediate registry or resolver server that maps a static persistent identifier to the current location of the object. Yet, "mapping table" in the registry or resolver server must be renovated whenever the object is moved. Samples of the most-consumed registry or resolver-based applications are: PURL, handles, DOI and OpenURL.

4.4 Index Services:

The procedure of indexing digital objects comprises linking of database of digital objects to a text database consisting of keywords and subject descriptors. Digital objects are essential to be connected to the associated keywords and subject descriptors so as to facilitate their retrieval. A digital repository commonly stores a big amount of unstructured data in a two file system for storing and retrieving digital objects. The first file stores keywords or descriptors of digital objects along with a key to a second file. The second file contains the location of digital objects. The user selects a record from the first file using a search algorithm. Once the user selects a keyword or a descriptor from the first file, the location index in the second file finds the digital object and displays it. It is assumed that a digital repository has several indices and catalogues that can be searched to discover information for subsequent retrieval from a repository.

4.5 Search System and Content Delivery:

The design of the digital library system should support searching of its collection. The search engine should support features like Boolean searching, proximity searching, phrase searching, etc. that are supported by the traditional information retrieval system. Most digital library software integrates external search engines. Dspace, for example, uses the Apache Lucene search engine. The digital library should also support content delivery via file transfer or streaming media.

5. DIGITAL LIBRARY MODELS AND FRAMEWORKS:

While different digital libraries have their own underlying design and framework, most of them support key components mentioned above. Some of the important digital library frameworks are discussed below:

5.1 Kahn-Wilensky Framework:

Kahn and Wilensky (1995) defined a general-purpose framework for a distributed digital library comprise of anexceedingly works. Kahn and Wilensky specified the basic entities stored, accessed, disseminated and managed in appropriated digital repositories. Presentation of naming conventions for identifying and locating digital objects in digital repository was one of the most important contributions of this framework.

5.2 Dienst and NCSTRL:

The Dienst (server in German) emerged as one of the first digital library framework situated on three basic principles of a distributed digital library system, i.e. open framework, federation and distribution (Davis and Lagoze, 2000). Grown by the Digital Library Research Group at Cornell University, the Dienst model was implemented in the “Networked Computer Science Technical Research Library (NCSTRL; www.ncstrl.org)”. The NCSTRL has other than 150 participating institutions and 20,000 digital objects.

5.3 CRADDL:

Grown-up by the Digital Library Research Group at Cornell University, the Cornell Reference Framework for Distributed Digital Libraries (CRADDL) is component-or- service-based digital library framework. The CRADDL extends the following five primary services (Lagoze and Fielding, 1998):

- Repository Service that provides mechanisms for depositing, storing and access to digital objects.
- Naming Service to recognize digital objects by Unique Resource Numbers (URNs) and registered then with the naming service.
- Indexing Service that allows a mechanism for discovery of digital objects throughout query.

- Collection Service that supply mechanisms for the aggregation of access to sets of digital objects and services into meaningful collections.
- User Interface Services Gateways, which supply a human-focused interface to the functionality of the digital library.

5.4 FDBS Framework and the NSDL:

The NSDL and NDLTD are examples of digital libraries based on loosely-coupled Federated Database System (FDBS) that are cooperating but are autonomous database system. Individual participants in the FDBS proceed their local operations as specify by their own Database Management System (DBMS). The FDBMS is the middleware software that commands and coordinates whereby the component databases cooperate.

The National Science Digital Library (NSDL) is a programme supported by the US National Science Foundation (NSF), Division of Undergraduate Education. The objectives of NSDL are to establish a digital library for training during science education, mathematics, engineering and technology. The NSF is funding a number of projects below this action, every of these projects are making its own contribution to the Library. Many of mentioned projects are building collections when others are developing services supporting NSDL. The challenge of the NSDL is to ensure that each of the collections and services developed below the NSDL project is associated as a single consistent library, not directly a set of unrelated collections and activities. In order to perform interoperability, three sets of agreements are essentialbetween members:

a. Technical agreement among NSDL participants to choose on formats, protocols, security systems etc., to achieve interoperability between collections and services;
b. Content agreements to choose on the data and metadata, along with semantic agreements on the interpretation of the information; and
c. Organizational agreements to choose on the basisregulations for access, preservation of collection and services, payment, authentication, etc.

The NSDL scheme aims to accomplish interoperability at the following threesome levels:

i. **Federation:** Federation has toseparation of responsibility amongst participating members. The participant's accord to succeed sets of standards, protocols and technologies. The process assures interoperability, except participants are constrained to use an agreed set of standards, technologies and tools.
ii. **Harvesting:** The participants agree to make enable some basic shared services, without being required to adopt a complete set of agreements as in the federation.
iii. **Gathering:** Gathering uses web search engines approach. In this approach, even if participants do not co-operate in any formal manner, a base level of interoperability can be accomplish by gathering openly accessible information using a web crawler.

Metadata from all the collections is stored in the repository and made available to provide NSDL services.

5.5 The NDLTD: Federated Digital Library Design:

The NDLTD is another example of digital libraries based on loosely-coupled Federated Database System (FDBS). The Networked Digital Library of Theses and Dissertation (NDLTD), the digital library of theses and dissertations of masters and doctoral students from various universities in the USA and around the globe has adopted a federated design approach. To avoid the work and negotiation involved in adding protocol support to diverse search systems, the NDLTD team created an intermediate application that mediates search requests, and has access to descriptions of the search engines' user interfaces, the types of queries supported, and the operators that define and qualify those queries (Fox and Powell, 1998). The NDLTD has also explained the Searchable Database Markup Language (SearchDB-ML), an application of the stretchable Markup Language (XML), for describing a search site. Initially the model was tested on five sites using different software: two sites used Open Text, one consumed Dienst, another used HyperWave and the fifth used a Perl-based searchable script (search.pl). All could easily be described with SearchDB-ML Lite, and the Federated Searcher application was efficient to support cross-language retrieval, for sample to submit queries in English to the German site and request translations (Fox and Powell, 1998). The federated search system distributes a query to multiple sites and then gathers the results pages into a cache for browsing, results are show for the user without merging (Fox et al., 2001).

5.6 Common Object Request Broker Framework (CORBF):

Common Object Request Broker Framework (CORBF) represents one of the widely known models of distributed object-oriented computing. The CORBF standards have been incorporated in the middleware of several commercially available network system products. The CORBF relies heavily on object-oriented and client-server technologies. It uses an open systems approach where digital library designers can implement the CORBF specifications in a variety of ways depending on their requirements. Applications of CORBF are platform-independent both at hardware and software level. Components of a digital library may be distributed among different servers.

5.7 Software Agents Framework and UMDL:

The University of Michigan Digital Library Project (UMDL) uses a proprietary framework to support the federation of loosely-coupled digital library collections and services. The core of the framework is the concept of the software agents that is based on object technology. An agent is extremely encapsulated module of software describing an element of a collection or service with very specific capabilities. These software agents may powerfully team together to unite their capabilities to handle more sophisticated tasks such as the process of performing a complex search request. Software agents are intimated into the following three groups (Ferrer, 1999):

i. **User Interface Agent (UIA):** User Interface Agents mediate user approach to the system. They transform queries and other user interactions into a form that can be implied by other agents. UIAs produce and maintain user's outline that other agents can use to support searching. User outline are consulted by the agents to facilitate transmit of SDI services.

ii. **Collection Interface Agent (CIA):** Collection Interface Agent (CIA) settles access to collections. Collection may involve full-text documents, web sites and different multimedia objects. The major character of CIAs is to supply the registry with information regarding collections. They supplyanexact description of the content and structure of each collection. CIAs represent the indexing systems connected with each collection and how to search them, depending upon the syntax used. CIAs as well describe how to access the collection and what protocols to be consumed for accessing collection.

iii. **Mediation Agent (MA):** Mediation Agents (MA) manage all the necessary tasks that support the system, such as those tasks that ultimately direct a user to a collection based on specific query or user profile. Mediation Agents communicate entirely with other agents. Types of Mediation Agents contain registry agents (to manage registry) and remora agents that provide SDI services. Mediation Agents are also assigned with the task of maintaining statistics of various activities. Tasks Planner Agent in an MA is responsible for managing tasks and other agents.

Software agents communicate with each other using a proprietary language developed by the UMDL called Conspectus Language.

5.8 Open Archival Information System (OAIS):

The OAIS Reference Model was grown by the Consultative Committee for Space Data Systems (CCSDS) targeted to digital preservation projects. It is a framework for knowledge and applying concepts desired for long-term digital information preservation. It is also a first point for anexample addressing non-digital information. The model establishes terminology and concepts applicable to digital archiving, identifies the key components and processes native to mass digital archiving activity, and proposes an information model for digital objects and their associated metadata. The source model does not mention an implementation, and is, therefore, neutral on digital object types or technological issues. The model can be useful at anextensive level to records digital image files, "born-digital" objects, or even physical objects (Sayer, 2001). OAIS has directly been taken up as an ISO standard (ISO 14721:2003).

The OAIS frameworks appreciate the prestige of a de facto standard in digital preservation. The OAIS reference model provides a high-level overview of the types of information needed to sustain digital preservation that can extensively be grouped below two major umbrella terms called i) Preservation Description Information (PDI); also ii) Representation and Descriptive Information.

i) Preservation Description Information

The preservation description information comprise of four leading types of metadata elements, especially reference information, provenance information, context information and fixity information.

ii) Representation and Descriptive Information

Representation information facilitates appropriate rendering, understanding, and interpretation of a digital intent's content. At the almost fundamental level, representation information transmits meaning to an object's bit-stream. In case, it may specify that a chain of bits represents text encoded as ASCII characters and furthermore, that the text is in French. The depth of the representation information required depends on the designated community for whom the content is intended. Descriptive Information metadata contains more ephemeral metadata, the information used to aid searching, ordering, and retrieval of the objects.

6. INTEROPERABILITY IN DIGITAL LIBRARY:

Interoperability is a critical problem in the network environment with an increase in the number of diverse computer systems, software applications, file formats, information resources and users. It is particularly more important in a digital library implementation given the fact that digital conversion activities are distributed amongst libraries that hold traditional print-based resources and the digitized information is to be made accessible universally. Collaboration amongst participants is, therefore, necessary in order to adopt a framework for achieving a suitable level of information sharing.

Interoperability is ability of digital library components and services to be functionally and logically interchangeable by virtue of their having been implemented in accordance with a set of well-defined publicly known interfaces. In this model various services and components can communicate with each other via open interfaces, and clients can interact with them in an equivalent manner. The ultimate goal of interoperability is to create and develop components of digital library independently yet be able to call on one another efficiently and conveniently (Paepcke, A., 1998)

Interoperability in digital library achievement addresses the challenges of creating a common framework for information access and integration across many domains. Digital library created using the principles of interoperability result in repositories of digital contents which may have different credits but can be address in the corresponding manner due to their shared interface definition. There are many approaches to perform interoperability in digital library implementation. Paepcke (1998) recognizes the following approaches to accomplish interoperability:

➢ Standardization:

Standardization is a proven approach to achieve interoperability. MARC and it different versions and Dublin Core are the recognized standards for bibliographic description of records. Z39.50 is a well-known standard for information retrieval. Standards and protocols suitable for a digital library are explained in other module.

➢ Families of Standards:

Families of standard approach extend the choice of implementing one or many of several standards. The International Standardization Organization (ISO) standard supporting Open Systems

Interconnection (OSI) produced an interoperability framework based on the family of standard approach. OSI in its seven layers structure supply a family of standards regard with a given set of interoperability issues in the area of interconnection. TCP / IP is an OSI protocol.

➢ Specification-based Interaction:

Interoperability can also be executed by representing the semantics and structure of all data and operations. The specification-based communication circumvents the necessity of mediation systems. Some of the well-developed authorizing technologies to accomplish this goal involve Agent Communication Language (ACL).

➢ Mediation

Interoperability can also be accomplished by deploying mediation machinery andcollaborates for translation of data formats and interaction modes between components. In the area of interconnection of different networks, network gateways play the role of mediators. However, translations in the sense of simple mapping is not always sufficient to achieve complete interoperability. For example, two sets of digital libraries may sometime completely lack certain data types or operations and, therefore, cannot interoperate without further work. However, mediation interfaces can be designed to augment functionalities and services that may search two digital libraries and present the results with its own value-additions. Such mediation facilities are called "wrappers" or "proxies". Mediation technology thrives on standardization. For example a single mediation system can cover all Z39.50 compliance sources at once.

➢ Mobile functionality

Mobile functionalities consist of software agents that travel over the network to sites where they access the service that they need. These software agents reach back to their original sites with the results of their works. Java applets and servlets facilitate such mobile functionalities that deliver new capabilities to client components at run time. Instead of depending upon standardization or third-party mediation, mobile functionality accomplishes interoperability by exchanging codes that facilitates communication amongst components.

7. CONCLUSION:

Digital libraries are built around the Internet and web technology; as a result, they demand to follow the Internet standards and protocols so as to ensure interoperability, portability, modularity and scalability. A typical digital library implementation follows client-server framework as does the Internet and web technology. Client-server framework as useful to the digital library is discussed.

Major problems of digital library design are caused by differences in the computer systems, file structure, formats, information organization and different information retrieval requirements of collections accessible through the digital library. While the web has emerged as the preferred media of information delivery and access, the use of standards and protocols makes it possible to make digital collections interoperable and accessible seamlessly. Important features that should be considered while designing a digital library include: scalability and sustainability, seamless access, interoperability, federation, capacity to handle multiple files & formats and location-independent identifiers. These features are discussed briefly. It describes various digital library design models such as Kahn-Wilensky Framework, Dienst, Cornell Reference Framework for Distributed Digital Libraries (CRADDL), Federated Database System (FDBS) Framework and National Science Digital Library (NSDL) and NDLTD, Common Object Request Broker Framework (CORBF), Software Agents Framework and UMDL, Metadata harvesting Framework and OAIS Reference Model as well as examples of digital libraries built using these models.

A typical digital library in a distributed client-server environment comprises of hardware and software elements at server side as well as at the client's side. It briefly describes interoperability in digital library. The paper describes all components with examples of software products that are available in the market place.

REFERENCE:

1. Davis, J.R. and Lagoze, C. NCSTRL: Design and development of a globally distributed digital library. Journal of the American Society for Information Science, 51(3), 273-280, 2000.
2. Ferrer, Robert. University of Illinois: the federation of digital libraries: Amongst heterogeneous information systems. Science and Technology Libraries, 17(3&4), 81-119, 1999.
3. Fox, E.A. and Powell, J. Multilingual federated searching across heterogeneous collections. D-Lib Magazine, September,1998.(http://www.dlib.org/dlib/september98/powell/09powell.html)
4. Fox, E.A. et al. Networked Digital Library of Theses and Dissertations: bridging the gaps for global access. Part. 1: Mission and progress. D-Lib Magazine, 7(9), 2001. (http://www.dlib.org/dlib/september01/suleman/09suleman-pt1.html)
5. Fox, E.A. et al. Networked Digital Library of Theses and Dissertations: bridging the gaps for global access. Part. 2: Services and research, D-Lib Magazine, 7(9), 2001. (http://www.dlib.org/dlib/september01/suleman/09suleman-pt2.html)
6. Kahn, Robert and Wilensky, Robert. A framework for distributed digital object services. cnri.dlib/tn95-01, May 13, 1995. (http://www.cnri.reston.va.us/k-w.html).
7. Kardorf, B. SGML and PDF: Why we need both. Journal of Electronic Publishing, 3(4), 1998. 14p. (http://www.press.umich.edu/jep/03-04/kardorf.html)
8. Lagoze, C. and Fielding, D. Defining collections in distributed digital libraries. D-Lib Magazine, November, 1998 (http://www.dlib.org/dlib/november98/lagoze/07lagoze.html)
9. Paepcke, A., Chang, C-C.K., Garcia-Molina, H. and Winograd, T. Interoperability for digital libraries worldwide. Communications of the ACM, 41(4), 33-43, 1998.
10. Payette, S., Blanchi, C., Lagoze, C. and Overly, E.A. Interoperability for digital objects and repositories. D-Lib Magazine, 5(3), May1999. (*http://www.dlib.org/dlib/May99/payette/05payette.html*)
11. Sayer, Donald, et al (2001). The Open Archival Information System (OAIS) Reference Model and its usage.http://public.ccsds.org/publications/documents/SO2002/SPACEOPS02_P_T5_39.PDF (last visited on 4th Oct., 2006)
12. Sheth, A.P. and Larson, J.A. federated database systems for managing distributed, heterogeneous and autonomous databases. ACM Computing Surveys, 22, 183-236, 1990.

Dr. B.V. Chalukya
Librarian, SCS College, Omerga, (MS) India.