Design and Implement of Digital Library: An Overview

Kiran P Savanur
kiransavanur@rediffmail.com
Raman Research Institute
Bangalore-560080

Nagaraj M N
nagaraj@rri.res.in
Raman Research Institute
Bangalore-560080

1. INTRODUCTION

Several terms have been coined at different times to represent the concept of library without books, Libraries having influence on computer readable format or having access to influence on digitized or digital format. The terms which have been in vague at different times include: Paperless library, electronic library, virtual library, libraries without boundaries and now it has been moved to digital libraries. The term digital library on one hand, is used to refer to a system or application whose functions are chiefly to extend electronic access to material available in a conventional library to a remote user. On the other hand, it is used to describe both commercial and academic systems designed to enable electronic access to a large collection of electronic documents to authorized users.

The term's digital library and electronic library are used interchangeably and synonymously. The term "virtual library" or "library without walls" usually refers to the meta resources, subject portals that extend virtual accessibility of digital collections from several diverse sources without the users even knowing where the resource actually resides. A virtual library could potentially be enormous, linking huge collections from all around the globe, or it could be very small, consisting of few hundred links to digital resources maintained by an individual.

The hybrid library is in the continuum between the conventional and digital libraries where electronic and paper-based information sources are used alongside each other. The challenges associated with the management of the hybrid library is to encourage end-user to provide information in variety of formats and from a number of local and remote sources in a seamlessly integrated way. The hybrid library aims to bring a range of technologies from different sources together in the context of a working library. In effect a hybrid library maintains all or its major parts of its collection in a digitize form as an alternative to supplement the print material currently found in Libraries. It has a web-enabled computerized catalogue (WEBPAC) accessible through the Internet and most of the in-house services like acquisition, books processing circulation are computerized. A hybrid library has a strong presence in the internet with a home page for the library providing an integrated access interface, not only to digital collections available locally, but also to

the other commercial and non-commercial web-based digitized collections across the world.

2. CHARACTERISTICS OF DIGITAL LIBRARY

The term digital library may mean different things to different people, but it is not merely a collection of electronic information. It is an organized system of digital information that can serve as a rich resource for its user community.

A digital library emphases the equitable and timely access to a vast amount of diverse resources in a shared mode in a given specialty, lifting traditional barriers of time and space.

Digital libraries may have the following characteristics associated with them

1. Digital Libraries are the digital counterparts of traditional libraries and include both electronic
    (digital) as well as print and other (e.g. audio, video, graphics, animation etc) materials.

2. A digital library owns and controls the information.   It provides access to information, not just
    pointers to it

3.   A digital library has a unified organizational structure with consistent points for accessing the
    data.

4.   A digital library is not single entity, it may also provide access to digital material and resources
    from outside the actual confines of any one digital library.

5. Digital libraries support quick and efficient access to a large number of distributed but inter linked
    Information sources that are seamlessly integrated.

6. Digital libraries have collections that (I) are large and persist over time; (ii) are well organized
    and managed; (iii) contain many formats (iv) contain objects and not just their representation; (v) contain objects that may be otherwise unobtainable.

7. Digital libraries include all the processes and services offered by traditional libraries, though
   these processes will have to be revised to accommodate differences between digital and paper media

3. WHY DIGITAL LIBRARIES?

With the advent of the below technologies have forced traditional libraries to go for Digital Library and also these forms the basic requirement of Digital Library.

1. Emergence of Internet and web technologies as a media of information delivery and access. The Internet, particularly the World Wide Web (www), allows rapid access to a wide variety of networked information resources extending a uniform interface to a vast number of multimedia resources. The web, being a hypermedia-based system, allows linking amongst electronic resources.

2. Availability of highly evolved, extraordinarily simple and intuitive user interface, e.g.. Internet
   Explorer and Netscape Navigator for all prevalent platforms

4. Advances in online storage technologies enabling storage of large amounts of contents at increasingly affordable cost

The digital libraries offer significant and unparalleled improvement and value addition to library services. While providing workable solutions to problems traditionally associated with the management of print based collections in traditional libraries, improved information retrieval and enhanced document delivery capability is widely accepted. The cost of creating, storing manipulating and transmitting digital information has decreased considerably providing necessary urge to the digital library initiates worldwide. Rise in acquisition and subscription fees have forced the libraries to find other means to make information available to their users and content aggregators and electronic publishers are providing the means to do so.

Several large-scale digitization projects are aimed at conserving and preserving old, fragile and deteriorating documents of high scholarly value and further it provides increased access and search possibilities. Digital libraries enable greater access to digital contents can be managed from remote locations and provide a way to enrich teaching and learning environment. Since information in the digital library is electronically stored and accessed it is not bound to space and time. Digital library systems can be accessed simultaneously by multiple users guaranteeing continuous availability of documents. Digital library implementation can dramatically reduce

floor space requirements as compare to conventional shelf – type storage of books and journals.

4. COMPONENTS OF DIGITAL LIBRARY

Digital Library consists of
1. User Interfaces or user
2. Storage Media or (Repository)
3. Identifiers

User Interfaces or user

A standard internet browser is used for the actual interactions with the user . This can be Netscape Navigator or Microsoft internet explore. Through the browser user connects to client server
(Repository) URL with TCP technology.

Storage   media   (Repository)   and   Identifiers

Stores and manage digital objects and other information. The interface for this is called Repository access protocol (RAP). These Repository will have Identifiers with the help of identifiers users can identify internet resources such as digital objects. In other words identifiers shows the identifier of the repository where the objects are stored.

Repository consists
1. Data
2. Metadata
3. Meta –object

Data

The things which are digitized from the data. The materials selected may include Image text, Audio and video. This material need to be acquired and process is called data acquisition this can be done by scanner and digital cameras. The place where all the digitized objects are stored is called repository.

Matadata

Matadata is the data about data that describe the content and attributes of any particular item in a digital library. Metadata describes the items that are digitized. Metadata is important for digital libraries because it is the key to resources. An

example of matadata is traditional card record of a book. Each digitized item will have metadata . For the creation of metadata MARC Or Dublin core Format is used. Some of the Dublin core elements are Title, subject, description , source, language, relation, coverage, publisher and creator. Normally metadata are created using HTML or XML

Meta –object

The repository may have plenty of digital objects in the form of data each of this data will be explained used metadata. The meta-object provides references to the set of
Digital objects.

5. GREENSTONE DIGITAL LIBRARY SOFTWARE (GSDL)

The Concept of Digital library can be better understood with the help of Digital Library Architecture. Some of the Digital Library Architecture are:

Architecture of California Digital Library (CDL)
Architecture of Green stone digital Library (GSDL)
Harvest
Architecture of NCSTRL (Networked computer science report Library)

University of Waikato in collaboration with UNESCO and the Human Libraries Project to provide information to the developed world developed GSDL.

Through HTML browser like Netscape Navigator or Microsoft Internet explore user Can send HTTP requests and receives HTTP response. The GSDL retrieves requested information and generates HTML page that contains the information required by user. With the help of Green Stone we can customize the page according to the need of the user. It has a Graphic editor, which can configure files.

Steps involved in GSDL

1. Building collection using GSDL
2. Assigning Metadata to the digital documents of a collection using Dublin core in XML
3. The collection configuration
4. Creating user defined indexes or search index
5. Finding information.

6. DEVELOPMENT OF DIGITAL LIBRARIES

Some of the important points to be considered in developing a digital library are

1. Digital collection or material selection
2. Conversion of existing Print, Audio and video into digital format.
3. Cataloguing or Metadata creation
4. Storing
5. Creating portals or gateway to the electronic collection available on the web
6. Integrated access interface.

Digital Collection

One of the important issues in the creation of a digital library is the building up of a digital collection. A digital library can have a wide range of resources. It can obtain both conventional documents and /or digital or computer processible form. The conversion of digital form is just to ensure better access and to reduce dependence on physical libraries.

The new digital resources are either deliberately created as digital or created in parallel to print. Publishers are increasingly moving to XML or SGML format. Future digital libraries resources are electronic journals, electronic books through databases and datasets in many formats.

The acquisitions of documents, which are already available in digital formats, like CD-ROM database is a part of the transition . Nowadays a large number of information products are available on CD-ROM, like MEDLINE, COMPENDEX, METADEX, LISA etc. Libraries can subscribe to any of these database for providing bibliographic or full text information forms an important input to the digital collection.

Access to external digital collection

Digital libraries can acquire permission to digital collections provided by external sources like other institutions, commercial publishers, resources of other libraries, and electronic journals through on-line access. Many of the commercial publishers like Elsevier, Academic press, ACM, SIAM are making their journals available on-line through web sites. Many of the journals are available in print and as well as in electronic form.

Conversion of existing Print, Audio and Video into digital format

Nowadays, a part of conventional collection of a library is being converted into digital form. The process of conversion of paper documents into digital format is mainly with the help of scanners. Printed text, pictures and figures are transformed into computer accessible forms using a digital scanner or a digital camera.

Scanners are mostly using for converting print resources into digital format. Most scanning software generate by default TIFF (Tagged Image File Format). The scanned textual images (TIF) are not searchable nor can be manipulated like text file document (ASCII). The scanned TIF format are converted to text by the process of Optical Character Recognition (OCR). The OCR software allows the option of maintaining text and graphics in their original layout as well as plain ASCII and word processing formats. Through OCR software we can save the file into html, doc and other formats. The images can be browsed through a table of contents composed in HTML providing linked to scanned images.

The important step involved in the process of digitization is scanning, as explained above with the help of scanners. Some of the important scanners being used for capturing digital images are

(a) Minolta PS 7000
(b) HP scanjet 6,300 C,
(c) Bell and Howell 1000 FB,
(d) Kodak 500S
(e) Digital camera - Zentschel omini scan 3000 Minolta PS 3000,
(f) Slide scanner - Kodak PCD Scanner 4045
(g) Microfilm scanner – Mekel M 500 XL sunrise SRI – 150

Some of the image scanning software are

Quick Scan	Altris software	Power office
OPTIX	Documentum	Caere's Omni page
File net	Java system	ABBYY
Finereader		

Storing

Digital resources can be stored in CD, DVD, Tape and Hard disk. Usually the things to be digitize will be in Image, Text, Audio and Video. The Text, Image and Photographs after scanning is stored in the in **JPEG** (Joint Photographic Experts Groups) and **GIF**(Graphics Interchange Format). These two formats are widely used for storage Images because they are small, fast and capable of displaying any type of picture. Audio files can be saved in .wav, .mp3, midi etc format. Recently developed MP3 format is very compact and takes less space while quality of audio is also better compared to other formats. The digitized Video files are saved in .mov or .avi, divx, mpeg file formats.

Access to digital information availability on web

The web provides the hyper media based systems that allow rapid access to a wide variety of networked information resources. One can browse the different web sites which are scattered geographically and have access to the major resources from which one can download the information. Some of the major portal sites or gateways that provide access to electronic resources are

> http:// www.edoc.com/
> http://mel.library.mi.us/
> http://bubl.ac.uk/
> http://sunsite.berkeley.duke/internalindex

So digital libraries can develop their collection through the integration of a number of resources and media types. Digital libraries can also provide access to electronic resources through library home pages

7. DIGITAL LIBRARIES AND ITS USES

The important functions and uses of the digital library in the context of users are that it

1. Provides access to a very large information collection in a digital form
2. Supports multimedia content
3. Is network accessible
4. Provides a friendly interface
5. Offers links to local/external objects
6. Supports advanced search and retrieval
7. Supports the traditional library mission of collection, development, organization, access to presentation
8. Supports publishing, annotation of new information

9. Brings together people with formal, informal and professional learning missions
10. Provides faster access to information resources
11. Provides an easy mechanism for resource sharing with other libraries sharing of digital files in much easier.

8. MAJOR ISSUES / CHALLENGES

Creating effective digital libraries poses serious challenges. Some of the serious issues facing the development of digital libraries are

8.1 Technical architecture

Libraries need to enhance and upgrade current technical architecture such as

- High speed local network and fast connection to internet
- Relational database that supports a variety of digital formats
- Full text search engines to index and provide access to resources
- A variety of servers such as web servers and FTP servers
- Electronic management system

8.2 Building digital collections

One of the most important issues in creating a digital library is building of digital collections.
One of the major issues is the degree to which libraries will digitize existing material and acquire original digital works. This is the old access versus ownership issue. How is the specific material to be digitized to be acquired or identified by a given library. Who collects and /or digitizes which material could be based on factors such as collection strength, unique collections, the priorities of user groups, manageable portions of collection, technological resources and skills of the staff.

8.3. Digitization

Another aspect is what portion of collection to digitize. Digitization is conversion of any fixed or analogue media – such as books, journal articles, photos, paintings, microfilms into electronic form either through scanning or networking

The approaches are

- Retrospective conversion of collections
- Digitization of a particular special collection or a portion of it
- Highlight a diverse collection

- Highly used materials
- An ad hoc approach (one digitizes and stores material as they are requested)

There is the problem of naming identifiers and persistence. Naming is required to identify digital objects. Any system of naming should be permanent, lasting. The names cannot be bound with specific location. A global scheme of unique identifier is required. Three schemes proposed to get over the problems of persistent naming are PURLs, URNs, and digital object identifiers.

PURLS – These are persistent URLs, a scheme developed by OCLE to separate document name from its location

URNs – Uniform Resource Name have been developed by Internet Engineering Task Force (IETF)

Digital Object Identifier (DOI) - Developed by association of American Publishers and corporation for National Research Initiatives to provide a method by which a digital object can be reliably identified and accessed

8.4. Copyright : Copyright is one of the most important barriers to digital library development. For
current works , copyright breaks down in the digital environment because the control of
copies is lost. Digital objects are less fixed, easily copied and remotely accessible to multiple users
simultaneously. The problem of libraries are that they are for the most part simply caretakers of
information . They do not own the copyright of the material they hold . So libraries will never be
able to freely digitize and provide access to the copyrighted material in their collection. They have
to develop a mechanism for managing copyright

8.5. Preservation : Another important issue is preservation. In the preservation of digital material, the real issue is technical obsolescence. Preservation of the storage medium tapes, hard drives, floppy
discs have a short life span when considered in terms of obsolescence. This form of preservation involves preserving access to the content of the document regardless of the format. While files can be moved from one storage medium to another.

9. CONCLUSION

Due to rise in acquisition and subscription fees, libraries have forced to find other means to make information available to their users. Since cost of creating storing, manipulating and transmitting digital information has decreased considerably providing necessary urge for the creation of digital library initiates worldwide. Digitization is the first step in building digital libraries. Besides digitization of documents achieves the purpose of preservation for the future generations and also supports the traditional library mission of collection development, organization, and access to presentation. Digital documents facilitate search and retrieval and can easily accessible world wide once they are made available on the Internet. It should be noted that digitization task time consuming and involves high quality acquisition of hardware, software and also manpower.

REFERENCES

1. Association of Research Libraries ARL proceedings 126 : Annual meeting 17-19 1995 (http://arl.cni.org/arl/proceedings/126/2- defn.html)

2. Cole Timothy W creating a framework of guidence for building good digital collections. FirstMonday, 7(5) 2002 (http://www.firstmonday.org/issue7-5/cole/index.html)

3. Guthrie, Kevin M Jstor , From project to independent organization Dlib july/ Aug , 1997 (http://www.dlib.org/dlib/july97/07guthrie.html)

4. Hulser, Richard P digital library Content presentation in a digital world. DESIDOC bulletin of information Technology 17(6) 7-14 1997

5. Pandey Richa, Digital Library Architecture 9-25p, In DRTC Workshop on Digital Libraries, March 2003

6. Das Jaba, Greenstone Digital Library Software (GSDL) : A Tutorial 235-269p . In DRTC Workshop on Digital Libraries, March 2003

7. Rusbridge, chris. Towards the hybrid library. D.Lib Magazine, July/Aug, 1998 (http://dlib.org/dlib/july98/rusbridge.html)

8. Arora, Jagadish (2001) Building digital Libraries : An overview DESIDOC bulletin of Information technology 21; 3-24

9. Army w y (1995) key concepts in the architecture of the Digital Library. D Library Magazine July (http://www.dlib.organization.july95/07arms.html)

10. chapman , S and keny A R (1996) digital conversion of Research library materials a case for full information capture. D- Library magazine OCT 1996 (http://www.dlib.organization/dlib/october06/10chapman/html)

11 lest, M (1996) Going digital scientific american March 1996, 58-60

12 Schatz, Bruee R (1997) Information retrieval in digital libraries, bringing search to the net. Science . 275 ; 327-33

13 Sahoo, B B, Digitization of Print materials, Audio and Video, 154-164p . In DRTC Workshop on Digital Libraries, March 2003