
BUILDING A DIGITAL LIBRARY

O. Sessaiah

Abstract

The idea of converting library materials into digital format for creating digital collections has advanced rapidly in the last few years, thus leading to the concept of a digital library. A digital library is a library that contains materials in digitized form or access to internal or external digital resources. Digital information resources include not only rapidly growing collection of electronic full text resources, but also images, video, sound, and even objects of virtual reality. The most important component of a digital library is digital collection it holds or has access to. A digital library can have a wide range of resources. A digital library is not a single entity although it may have digital contents created in-house or acquired in digital formats stored locally on servers. A digital library may also act a portal site providing access to digital collections held elsewhere. The digital imaging technology offers a number of choices that can be adopted to translate from print to digital. Digital conversion process includes Document, Data Capture, Data Processing, Storage, Indexing and retrieval. Dublin Core standard is adopted internationally to create Meta data of digital resources. The advent of the Internet acted as catalyst for Digital Library initiatives. Digital library environment is the context of rapidly changing computer and communication technology. This paper explains digitization methods, process of digitization and tools required for digitization.

Keywords: Digitization, Digital Library, Electronic Library, Dublin Core

1. INTRODUCTION

The idea of converting library materials into digital format for creating digital collections has advanced rapidly in the last few years, thus leading to the concept of a virtual library or a library without walls. It is very difficult for a single library to acquire all these resources and provide the same to their users. Yet at the same, it is the duty of the librarian to provide the required latest information, wherever available and in whatever form and whenever required.

The availability of software, hardware and networking technology, the advent of WWW, its ever increasing usage and highly evolved browsers have paved the way for creation of digital libraries.

Digital information resources include not only rapidly growing collection of electronic full text resources, but also images, video, sound, and even objects of virtual reality.

1. DIGITAL LIBRARY

A digital library is a library that contains materials in digitized form or access to internal or external digital resources- that can be manipulated and delivered in many ways that a conventional version of the materials cannot be given.¹

2. WHY DIGITIZE ?

The growing of information and Telecommunication technologies, web technologies and database technologies has compelled library and information centers to use these technologies effectively to render services. With the growing number of e-sources, it has become imperative for information professionals to redefine their role in disseminating information to the users.²

3. BUILDING-UP DIGITAL COLLECTION

The most important component of a digital library is the digital collection it holds or has access to. A digital library can have a wide range of resources. It may contain both paper-based conventional documents and information contained in computer-processible form.

Information a content of a digital library, depending on the media type is contains may includes a combination of structured/unstructured text, numerical data, scanned images, graphics, audio and video recordings.

4. METHODS OF DIGITIZATION

A digital library is not a single entity although it may have digital contents created in-house or acquired in digital formats stored locally on servers. A digital library may also act a portal site providing access to digital collections held elsewhere.

A. Acquisition of digital media

Thousand of CD-ROM databases are currently available from multitude of CD-ROM producers including silver platter which alone producers more than 250 information products on CD_ROM. CD-ROM networking technology is now available for providing web-based simultaneous access to access to CD-ROM databases on the LAN as well as WAN. The libraries have on option to subscribe to these full text databases as a part of their DL.

B. Buying access

Majority of digital collections would be provided by external sources like commercial publishers, collection mounted by scholarly societies, resources at other libraries, electronic journals sites etc. Internet has long been a favorite media for experimenting with electronic publishing and delivery.

C. Content Creation

Institution may have dissertations, in-house journals, annual reports technical reports etc. that may be included in digital collection. The documents composed on word processing packages or desktop publishing packages can be converted into HTML, Postscript and PDF using tools like Acrobat or Acrobat exchange. HTML, as a defacto language of the web and PDF as a defacto standard for online distribution of electronic information, can be employed to facilitate transition from computer processible files to a format accessible on the web.

D. Conversion of Existing print media

The existing sources of information may; be in print media. These sources can be converted into digital media by scanning the materials. This is the most popular method of converting the existing sources of information. Scanned documents are stored like pictures or graphics and this should be converted as text documents. Much software is available for this purpose.

E. Portal Site

Portal means a gateway or entrance the portal sites or gateways redirect a user to the holders of the original digital materials. It may provide its own indexing and search services or may combine original resources from a number of different providers.

F. Integrated access interface

Digital Libraries typically integrate multitude of resources and media types. Constituents at a digital library may have:

1. Collection acquired in digital form
2. Collection digitized in –house
3. Buying access to electronic resources
4. subject gateways and the Library OPACs.³

5. DIGITIZATION PROCESS

The digital imaging technology offers a number choices that can be adopted depending on the objective of scanning, end users, availability of finance, etc. there are four basic approaches that can be adapted to translate from print to digital.

1. Scanned as image
2. OCR
3. Retaining page layout using acrobat capture
4. Re-keying the data

Tools required for digitization is as follows:

Hardware:

1. Computer Systems
2. Scanner
3. Digital Camera

Software:

1. Image editing software
2. File compression software
3. OCR Software

Digital conversion process includes

- a. Document
- b. Data Capture
- c. Data Processing
- d. Storage
- e. Indexing and processing
- f. Retrieval and display

a. Document

It is including text, bibliographic or full text, photographs, diagrams, charts, maps, colors, images, etc. They exist either in print or non-print form or also as single unit or collection.

b. Data Capture

Includes manual data entry OCR or imaging using scanners

c. Data processing

The text in the convertible document may require conversion of diacritics or special characters: images may need enhancements, amplification or compression.

d. Storage

The digitized information needs to be recorded in proper digital storage medium, which may be hard disk, magnetic tape, optical CD-ROM, or networks with workstation to access.

e. Indexing and processing

Digitized documents need to be processed using standards; protocols and indexing systems metadata application should be a major component of the digital information processing. Dublin Core standard is adopted internationally to create Meta data of digital resources. Dublin Core recommends a minimal of 15 data elements to describe any documents. There are: Title, Subject, Description, Type, Sources, Relation, Coverage, Creator, Publisher, Contributor, Rights, Data, Format, Identifier and Language.

f. Retrieval/Display

It is the process that which an array of technologies for browsing, displaying and applying packages that ultimately helps in access.⁴

6. DIGITIZATION INITIATIVES IN INDIA

Digital libraries in the developed countries during the 1970s, but in India, it began in the mid 1990's with advent of it on a large scale and the support extended by Central Govt.

The advent of the Internet acted as catalyst for Digital Library initiatives. Digital Library projects and developments in the country are so many, and a large number of them are at an aggressively enthusiastic preliminary stage.⁵

7. CONCLUSION

Digital library environment in the context of rapidly changing computer and communication technology. It is proposed that future collection development should be in the digital media. Digitization is one of the technological advancements. It brings the resources nearer to the benefit of users.

8. REFERENCE

1. Hariharan, A. Lecture on Bulding a Digital Library Collection using Greenstone Digital Library and dtSearch Software. AICTE-ISTE-STTP on “ Management of Information Systems in Technical Libraries” . Coimbatore. July 05-16, 2004.
2. Ibid.
3. Gopalakrishna, S. Digitization Process and methods. Workshop on Recent Trends in Digitization.Tirupati. June 25-26,2005
4. Karisidappa, C.R. Digitization: The New Revolution in the Information Transfer Chain. Recent Trends in Digitization. Tirupati. June 25-26,2005

About Author

O. Seshaiiah is currently working as a Librarian in PBRVITS, Kavali, Andhra Pradesh

Email : osheshy@yahoo.com