Content Management in Digital Libraries

Mohd. Nazi   Faizul Nisha

Abstract

Discusses briefly the concept and characteristics of digital library. Digital library is simply an online system providing access to a variety of contents such as various kinds of electronic media(text, image, video ed),licensed databases of journals, articles and abstracts and description of physical collections. The functions of content management such as selection and acquisition,indexing,storage,retrieval,maintenance and intellectual rights are discussed. Issues regarding the research in the development and management of digital contents are highlighted

Keywords : Content Management, Digital Libraries

0.   Introduction

Libraries have existed for centuries and since then they had been managed as warehouse of documents by acquiring, cataloguing and classifying books , journals and other materials and circulating them to their clients. But recent developments in Information Technology (IT), the Internet, World Wide Web (WWW), coupled with increase funding for research on creation, access and management of electronic information resources, have led to the development the new era of electronic and digital libraries. These technological innovations have improved the new breed of information professionals to select, organize, retrieve and transfer digital contents effectively and efficiently to their target audience (1).

1.   Digital Libraries : Meaning and Nature

The term electronic library, digital library and virtual library have bee used interchangeably and now widely accepted as description of the use of digital technology by libraries to acquire, store, conserve and make available their content to remote users.

In a broad sense, digital library may be defined as an organized and managed collection of high quality information contents in a variety of media (text, still image, moving image, sound or combination thereof), but all in digital forms accessible over different electronic networks. Such a digital library includes a number of search or navigation aids that operate library and allow access to other collection of information connected by network world wide.

The term digital library is best defined by Christine Borgman (2) as a set of electronic resources and associated technological capabilities for creating, searching and using information … they are an extension and enhancement of information storage and retrieval system that manipulate data in any medium.

The concept of digital library is rooted in age-old dream of creating a virtual library. But digital library is different from virtual library because of its physical identification. O'Donnel (3) differentiates digital library from virtual library as it can still maintain a physical presence, whereas virtual library is a vast, ideally universal collection of information and with instantaneous access to that information wherever it physically resides.

2.   **Digital Library Contents**

The most important components of digital library, however, is digital collection. Viability and extent of usefulness of a digital library would depend upon the critical mass of its digital contents. The information contents of a digital library include virtually any kind of electronic media(Text, image, graphics, video, etc.), licensed databases of journals, articles and abstracts and description of physical collection.

Theoretically any object from a text fragment to an animal in zoo may be rendered digitally and thus, there is no limit to the types of contents that may be held by a library. But in practice, digital contents are of three types:

- ✎ Contents created and existing primarily in machine readable format.

- ✎ Contents converted from the traditional format into digital(e.g., print text, pamphlets, manuscripts, motion pictures and recorded sound)

- ✎ Access to external contents, not held in-house, by providing pointers to web sites, publisher's services, password to consortium or other collaboration from commercial organizations

3.   **Management of Digital Library Contents**

Contents in digital library are organised and managed for the purpose of immediate access to the target audience. How contents are developed and managed is a critical issue to the long-term success of digital library services, especially when technical resources are limited. Content management includes the following key functions:

3.1   Selection and Acquisition

Libraries select contents according to a well defined collection development policy. Such policy manifests the mission of a library and determines how budgets on materials are expended. There are two key challenges in content selection i.e. cost and quality. As soon as decision about selection is made, content must be acquired. For objects, which are already in digital form, the file transfer through networks or mass storage is straightforward as long as file formats are well specified. In case of traditional objects, digitisation must be done. Scanners for text and images range in quality on several dimensions (i.e. output resolution, value and conditions of physical objects and speed) are required. In addition to these technical challenges, policy decision must be made. For example, which resolution and formats to adopt, how much text to OCR error is acceptable, how to link different representation for multiple media from single collections.

3.2   Indexing

Once content has been selected and acquired, it will be added to the collection in such a way that users may retrieve it easily. And thus, indexing is required for digital content to search and access in a selective way like OPAC for printed content. Decisions are to be taken regarding what to be indexed (author. keywords, phrase, etc), how the content and index files are linked, what sort of access points are provided, etc. Indexing strategy comprises not only the types of fields are to be indexed, but how they are to be treated (exhaustive or sparse).

Automatic indexing techniques are used to index the content of digital library. Several www-based services use a hybrid approach by manually creating classification system and then using automatic techniques to assign objects. Most retrieval systems for images, video, audio recordings and other non-textual objects have depended on items such as title, creator name or manually assigned subject headings for retrieval.

It seams certain that digital library research and development activity of 1990s will ensure that considerable progress is made in automatic indexing for textual and non-textual objects. New indexing challenges will emerge as more dynamic objects (e.g., virtual conference proceeding, active networks) are added to digital libraries. The temporal nature of such objects will require ongoing indexing techniques.

3.3   Storage

The next thing is how to store the content of digital library. Decisions regarding procuring suitable hardware, software, networking, etc. are to be made at this stage. Storage is mainly a technical requirement, although new media may complicate storage decision and costing. When data are to deliver continuously (e.g., streaming video or audio) rather than as discrete files, alternative technologies are required (4). Large digital repositories are required like multiple level of mass storage media (e.g., disk, tape, etc) and mechanical robots to locate and mounts the media. Various supercomputer centers are using tape robots that store and access to many terabytes of data. Digital libraries will surely apply such technology just as libraries of today apply movable shelving and complex conveyer systems to move physical materials.

3.4   Retrieval

Retrieval is another major issue, as far as digital library content and its access is concerned. Ultimately, users must be able to retrieve the content, which have been selected, indexed and stored by the librarians. During 1970s to 1980s, a large number of libraries invested heavily in computerizing cataloguing and circulation functions to give users better access and services. Online Public Access Catalogues (OPACs) have long provided author, title, and limited subject access to local holdings (and more recently to union holdings across to multiple libraries). The expectation for digital collection is that catalogue should seamlessly link to the digital content itself so that remotely located users can find and display not only bibliographic records, but also primary information objects.. In physical libraries, the card catalogues or OPAC is physically distinct from the items on shelves. These distinctions are difficult to make in electronic environments because everything is displayed on the same physical screen and thus, the boundaries between metadata and primary data are often blurred. Expectation to provide primary data with metadata yields several challenges to librarians. The challenges are first to extract and provide multiple levels of representation and second to provide users with control mechanisms to move from high -level surrogates to detailed objects (5). Today most retrieval is facilitated through words, titles, captions, manually created subscriptions, automatically extracted extracted keywords and so on. There is enormous attention focused on creating non textual surrogates such as colour and shape characterisation for images and speaker identification schemes for audio recordings, but there are more difficult metadata issues looming as more contents are not stored at all but created on the fly according to the specification of the users.

3.5   Maintenance

Maintaining buildings and systems and preserving content are important and costly activities in physical libraries. Digital libraries may avoid some of the cost of wear and tear on buildings and books but still have significant maintenance costs, including some unique to electronic environments. New equipments, improved or alternative network solutions (e.g. ISDN, ATM, Wireless), and software upgrades will require excellent technical personnel.

Just as the computational system changes, digital content may also change. A digital document may have numerous versions, especially given the ease with which electronic documents may have changed. Maintaining the most essential document requires that versions be well managed, which include updating and deleting the links to those objects (6).In addition to this version control problem, digital librarians must manage the multiplicity of indices and file formats. Requirements for link managements are more

problematic, as hypertext links are created among distinct documents. Although much research and development efforts in digital libraries have been devoted to maintaining the content. but further improvements are required to maintain security, updating versions, tools for automatically checking links, database tools for property rights, etc. for the smooth library functions and services.

3.6   Rights Management

Intellectual property right and information security and authority are two global interdependent issues, which influence research and development in digital libraries. Copyright exists to promote intellectual production by providing economic incentives. Security protects unauthorized access as well as ensures the veracity and authority of digital information objects. The misuse that can be put to digital content is far more serious and voluminous than for printed content. Efforts have been made to change copyright laws to protect the illegal use of digital objects and also to develop technical solutions that protect copyright either through copy protection or automatic billing mechanism. Research on encryption algorithms, digital watermarking and electronic commerce are leading to the development of trusted system that protect intellectual property rights by managing the necessary financial transactions while protecting consumers by providing authoritative information securely (7).These techniques ensure the veracity of an object and may help to prevent copying and distribution in an open market place.

4.   Conclusion

There is no doubt about the utility of digital libraries as they facilitate live and interactive access to wide variety of content online. But the problems of managing digital library content and its development are manifold. Management of digital library content requires two prolonged strategies(i) to digitize local content;(ii) to devise options for accessing external resources. Generally there is a feeling that publishers copyright most of the contents available in our library, and we are not in a position to provide online access to those contents. Though our libraries are facing a shortage of content, there is a wide spectrum of formal and informal sources available with them but could be converted into digital form by devising suitable action plan. Image format, compression schemes, network transmission, monitor and printer design, and image-processing capabilities are all likely to improve dramatically over the next decade. But technology alone will not determine the future; relationship, economic and pattern of behavior are equally important.

5.   References

1.   Shemeent (Jorge Reina). Encyclopedia of Communication & Information. Vol.2;2002.p.449-553.

2.   Borgman (Christine L.). What are Digital Libraries? Computer Vision. Information Processing. Vol.35;1996;p.227-43.

3.   O'Donnel (J.J.). The Virtual Library: An idea whose time passed. Philadelphia, University of Pennsylvania,1995,p.1.

4.   Arora,(Jagdish).Building digital libraries:An overview.DESIDOC Bulletin of information technology.Vol.21 (6);2001;p3-24

5.   Marchiomini (G).information Seeking in electronic environments.New York,Cambridge Press,1995.

6.   Richvalsky and Walkins.Designing and implementing a digital library.ACM Crossroad student magazine,Jan 1998.http:// www.acm.org/crossroads.rds5.2/digital.html.

7.   Wiederhold (G).Mediation in the architecture of future information systems.IEEE Computer,38-49 (March 1992).

About Authors

Mr. Mohd. Nazim is doing research in Department of Library & Information Science, Aligarh Muslim University, Aligarh, Uttar Pradesh. He is Junior Research Fellow of UGC.
Email : monis_naz@yahoo.co.in

Faizul Nisha is doing research in Department of Library & Information Science, Aligarh Muslim University, Aligarh, Uttar Pradesh.
Email : momi_sonu@yahoo.co.in