METAMORPHISM OF LIBRARIES

By

Aditya Tripathi*

ABSTRACT

Libraries have gone through several changes since inception. The present social, economical and technological stress and strains have put libraries in the process of metamorphism. This metamorphism has brought about a complete new form of libraries known as Digital Libraries. A model has been suggested for developing a Digital Library for small and specialized libraries.

* Junior Research Fellow, Documentation Research & Training Centre, Indian Statistical Institute, Bangalore - 59 E-mail : aditya@isibang.ac.in

0.   Introduction

During the movement of library system, developmental changes such as , closed access to open access, the role of library from a store-house to active information service organization and reading place to center of recreation have been seen. Now digital media is being used to store, process and retrieve information that causes Metamorphosis in the very form and function of library, though the philosophy of librarianship is still the same.

1.   Metaphorism of Library to Digital Library

Metamorphism of library means that , the library has passed through certain amount of pressure and environmental change to be addressed, in today's electronic age, as digital library. Basically the term Metamorphism is borrowed from geology and according to Webster's Dictionary it is "a pronounced change effected by pressure, heat, and water that results in a more compact and more highly crystalline condition."[1]
Factors that make the change are [2]:

1.1   Technology

Technology has influenced libraries in many aspects:

1.1.1 Media:
        Electronic storage forms have changed through magnetic tapes, floppies, CD-ROM and now DVDs. The very storage capacity of CD-ROM made 20 volumes of The Oxford English Dictionary available only on one CD [3].

The newly developed DVD technology stores 4.7 GB-17GB data, depending upon the specifications. These powerful storage devices have compressed the size of Library in Digitized form [4].

1.1.2 Cost:

Cheaper and reliable technologies have reduced the cost of products resulting in increased storage capacity of libraries in less cost. The cost of CDs has come down very fast in last two years. But new technology like that of DVD is still costly, but may come down after wider use.

Compared to traditional books or other materials, the digitized products are cheaper from the library point of view. The 20 volumes of normal OED cost £ 1800 (Rs. 1 lakh and 19 thousand) and in leather jacket cost £ 3750 (Rs. 2 lakhs and odd), while in CD-ROM cost £ 175 (Rs. 11 thousand and odd) and updation is £ 51.06 [3].

Duplication through computer is cheap. Practically one can copy any amount of data within short time. Besides, documents lose fidelity in traditional environment and offset printing is too costly. Digitized duplication does not lose single percent fidelity. But copyright issue is not to be overlooked.

1.1.3 Access:

In online digital environment information can be accessed at any time. Such facility can be extended to user's home or office. Besides, there is no danger of non-availability of document due to circulation.

1.2 Publication:

Technology has no doubt changed the trend of publication. Lots of products are available on Internet and in electronic form. Though most of the products are of low quality, content or presentation wise, but the advantage with these products is, ease of accessibility. So, such products force standard products to go online in order to keep them-self in market. For example, Encyclopedia Britannica couldn't cope the challenge posed by inexpensive, CD-ROM encyclopedia and thus it had to be published electronically.

In case of e-publishing, there is scope for improvement in performance, because such publications can be well reviewed by many persons. Better version of the work can come very fast compared to paper form. This particular characteristic is making digital or electronic publishing common and more feasible than the traditional one.

1.3 Barrier for use:

Studies show that as the distance to library increases, users become more unwilling to go to library. A study by Andrew Odlyzko shows that in Penn State, USA, the idea of off-site repository was disliked by users for all journals published before 1973, which where less used, though it takes only one day to retrieve an article from the off-site collection.

Tthe frequency of use of off-site collection was only 4 items per person per year among the total users in institute numbering 1050.

1.4   Usage of electronic information:

It is hard to measure online activities but studies show that usage of electronic information is increasing almost at the rate of 100%. For example, take the data of Library of Congress [2]:

Month	GB	Request (million)
Feb 1995	14.0	1.1
Feb 1996	31.2	3.9
Feb 1997	109.4	15.1
Feb 1998	282.0	36.0
Feb 1999	535.0	48.6
Feb 2000	741.1	61.3

Fig.1 Library of Congress Statistics

Usage of theses and dissertations is not restricted to institute. Now there are Internet based digital libraries, which provide full text access to theses and dissertations.

The Brazilian SciELO (Scientific electronic Library Online) project started out in early 1958. The number of pages transmitted from Brazilian SciELO (Scientific electronic Library Online) project grew from 4943 in January 1999 to 63,695 an year after [5].

Similarly, data from Firstmonday (http://firstmonday.org) shows that in early 1999, full paper download growth ranged from 50,000 to 60,000 per month, which went to 110,000 - 120,000 per month in early 2000 [6]. That shows how fast electronic form is picking up.

2.   What is it?

In digital libraries information is stored in digital form i.e. on computer or CD-ROM or magnetic tapes and is retrieved through digital information access tools.

3.   Modelling the digital library

Development of digital library involves several issues. It may extend from administrative level to accessibility of information. The development of digital library can be taken as a project in the traditional library. The major issue to be taken care in the development of digital library are[7]:

3.1   Digitisation

This is the heart of the project. Following questions must be answered before digitization:
- What element of the original to be captureed?
- How to capture?
- Copyright permission

The exposition of recorded knowledge takes place in four forms:
- Image
- Text
- Audio
- Video

3.2   Cataloguing

Library catalogue, as we know, gives the information about the holdings of library. In Digital library, there are two ways of doing the work:

3.2.1 Creation of catalogue :
Catalogue pages can easily be created in HTML format, having the information about title, author, publisher, place and so on. Or even library automation packages, having OPAC module, can be used for the same, provided they have facility of web interface.

3.2.2 Metadata :
The term metadata stands for data about the data. There are many popular metadata approaches and the most popular, particularly for the Internet resources, is Dublin Core. It contains the following 15 elements [8]:

Content	Intellectual Property	Instantiation
Title	Creator (Author)	Date
Subject	Publisher	Type
Description	Contributor (Other agent)	Format (Form)
Source	Rights	Identifier
Relation		
Language		
Coverage		

There are two different approaches of using metadata.
a.   Approach 1:

    Metatags can be embedded in the HTML documents [9]. For example,
<meta name = "DC.Author" content = "S.R. Ranganathan">
<meta name = "DC.Title" content = "Prolegomena to Library Classification">
<meta name = "DC.OtherAgent" role = "Assisted by" content = "M.A. Gopinath">
This tag resides in the <head> and before <title>.
Disadvantages:

    i.   <meta> tags is being used for many things like reloading a page or directing a page to some other URL.

    ii.   It doesn't give flexibility. For example, there are well defined set of tags, which gives robustness to document structure.

b.   <u>Approach 2:</u>

Warwick framework: Resource Description Framework

Warwick framework supports the existing metadata. Basically it uses the "Container Package model" [10]. That means, there is a Container, inside which there are packages of different nature. Packages can be of different types:

    i.   <u>Primitive package</u>: Contains one or more elements of metadata. Each of them has a type. Ex: MARC package, Dublin Core package, etc.

    ii.   <u>Indirect package</u>: Basically these are links, giving direction to the other information resource.

    iii.   <u>Containers package</u>: A package itself can be a container.

Resource Description Framework (RDF):

This was instantiated by W3C. A project called Platform for Internet Content Selection (PICS), to rate the webpages was started by W3C.. Later on, the scope of this project increased and one can set its own parameters to rate the webpages. The second approach leads to description of resource and thus Resource Description Framework (RDF) was born.

RDF Model:

Resource is an object that is uniquely identified by a Universal Resource Identifier (URI). Resource has Properties. Properties have Property types. And these Property types have Values.

If we generate an RDF model, resources are at nodes, property types are links and values are in quotes [11] [12] (Fig.2).
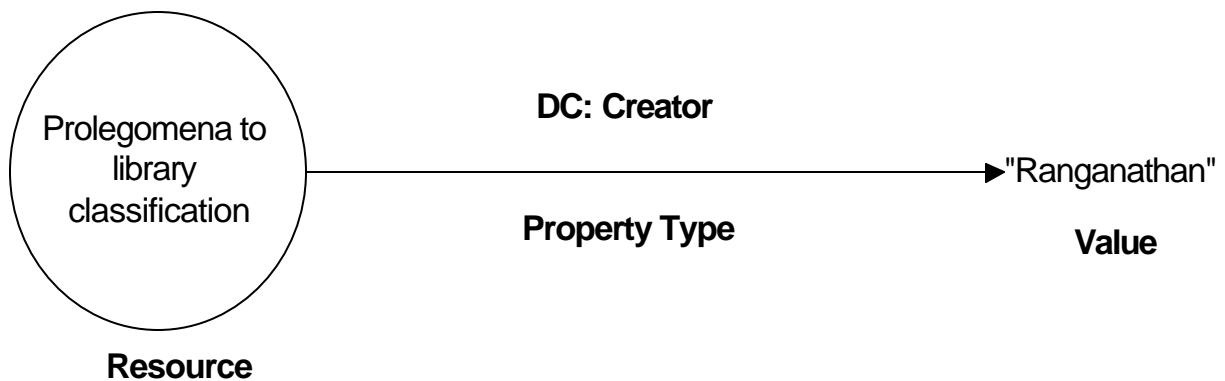
DC: Creator

Prolegomena to library classification

Property Type

"Ranganathan"

Value

Resource

Fig. 2 RDF Model Representation

Basically RDF uses XML (Extensible Markup Language), where one can define its own tags unlike to HTML.

3.2.3 Databases :

The current approach is to store information in databases. These databases are made accessible through web-browsers or any other interface depending on the library. Any user can search these information using search techniques and a search engine. In such databases search can be done on a specific field and it removes the need of catalogue.

Disadvantage:

i. The major disadvantage of this system is that the reader has to browse the whole document to find its relevancy.
ii. It makes mandatory to use bibliographic database system, because commercially available popular database systems do not give sub-field level description of document.

4.   An experiment

A catalogue-based approach has been adopted for developing Digital Library model. The library gives access to all sorts of documents including text, audio, video and image in different formats. Browser is used as front end for search and delivery of documents. For searching the document, search engine, provided within MS-Frontpage, is used.

The formulated query uses catalogue for search and displays the hit records in a list format (Fig. 3). Each item in list is hyper-linked to the catalogue entry page.

Each catalogue page in-turn is hyper-linked to the original document.

Establishing a digital library is definitely a Herculean task. It involves so many decisions and problem solving. There are certain issues, which hinder the development of digital library.

Advantages:

i. This type of cataloguing is very much helpful for the audio, image or video collections where it is difficult to index the collection. So such catalogue can be used as document surrogate and can be searched easily using any search engine.
ii. An abstract gives a clear view to reader about the usefulness of document for his/her study.
iii. It is a simple approach and can be built using common softwares like, Frontpage.
iv. It does not use any database.

Disadvantages:

i. This approach is only feasible in the small libraries and specialized libraries, which have collection that can't be indexed.

ii.   For large size library, powerful search engine is required, because search engine indexes the content of whole catalogue pages and searches through them, which takes time.

iii.   The results are only based on pattern matching so there are no semantics associated to search.

5.   Challenges for Digital Library

5.1   Intellectual Property Rights (IPR):

No doubt the primary aim of life is to earn. Due to this approach of life the concept of IPR came in picture, which suggests the benefit of creator (monetarily or other means) if his/her work is used for the monetary gain or so. This monetary gain became the chief concept in publishing industry.

This particular concept really hurts the library philosophy, because it stops copying of documents or its any part for use, without the prior permission of copyright holder.

In digital library, actually the media of document is changed which is even more serious problem than the earlier. At this particular juncture many of the digital library projects die.

In such an environment, even if library somehow gets the permission to change the document form or accepts the publication in digital form, it is difficult to save the document from unauthorized distribution and copying. Watermarking is a sort of a solution of it. There is wide variety of softwares, which can be used for watermarking like Digimarc, SysCoP etc. But side-by-side there are softwares, which can change watermarks, for example, StirMark [13]. IPR and security are the two major issues, which dampen digital library projects.

5.2 Economic issues :

Development of digital library involves lot of resources and in terms of money it is not a small amount, which has to be dealt with.

5.2.1 Operating cost :
        It is true that to set up a digital library, big money is required. Once it is set up the operating cost for maintenance has to be considered. No doubt duplication and printing cost is negligible, but it really doesn't cut some of the costs, which are there in traditional library. Readers are still very comfortable in reading through printed form and on demand it is library's duty to give in print. So paper stationeries and printing may still be required. Side by side money is required to have backups, buying electronic stationeries, system maintenance, electric charges etc. Besides, library has to consider payment to ISP and telephone line and any other communication mode [13].

5.2.2   Revenue :

Copying can be done in digital environment with 100% fidelity without cost. But to distribute a non-digital product in a digital form, library must have to pay to copyright holder, and if all agree that library should generate its own funds, the question of HOW comes. Charging to user is big problem, because as it is known that publishing industry works on nonlinear pricing, which influences libraries to have nonlinear prices. For example, for two different products if library is paying different amount then it is required to charge differently to users. Thus publishing industry plays havoc for digital libraries [13].

5.2.3 Monitoring:

If digital library, working on Internet, gives links to other Internet resources, due to volatile nature of web documents, it is a must to keep track of presence of web documents. There are few softwares, which do this work like MoMSpider and Web:lookout, which can trace any changes on the linked webpages. But the real issue is that if the link is entirely changed how to trace that which is quite common on Internet [14].

5.2.4 Communication:

If library delivers its service through the Internet or Intranet it has to have a user interface. Now it should be taken care that how fast the interface get loaded to reader's machine and how fast the results of searches are displayed. For that, librarians should have sound knowledge of technology and changing web environment.

6.   Conclusion

In library development movement, digital library is a milestone. Though developing a digital library seems to be easy from the point of view of technology but the real challenge lies in the culture of society. To propagate the knowledge it is required to have a sharing of knowledge and in which the publication industry poses biggest challenge. That is why, many of the digital library projects went up to putting the home publication on digitized form not the collection of library. It seems that such problems can be solved if learned bodies or libraries start publishing themselves. But this revolutionary concept may rouse many moral and ethical issues in the librarian's lobby. It is a matter of debate or deep discussion which librarians have to go through and finally again we have a Yaksha Prashana- Are we really ready for this change?

7.   References

1. Merriam Webster collegiate dictionary. http://www.m-w.com/cgi-bin/dictionary
2. ODLYZKO (Andrew). The rapid evolution of scholarly communication.
    http://www.research.att.com/~amo
3. Oxford English dictionary. http://www.oed.com/
4. DVD FAQ. http://dvddemystified.com/dvdfaq.html
5. SciELO library collection. http://www.scielo.br/scielo/bibsur.htm

6. Firstmonday. http://firstmonday.org
7. The digital library toolkit. 2nd ed. Sun Microsystem. http://www.sun.com/edu/libraries/digital_library_toolkit.pdf
8. Dublin core metadata initiative. http://www.purl.oclc.org/dc/
9. HOLZNER (Steven). HTML black book. Dreamtech : New Delhi, 2000. pp.26-29.
10. LAGOZE (Carl). The Warwick Framework: a container architecture for diverse set of metadata.
11. LASSILA (Ora). Introduction to RDF and metadata. http://w3c1.inria.fr/TR/NOTE-rdf-simple-intro
12. Resource description framework (RDF) models and syntax specification. http://www.w3.org/TR/REC-rdf-syntax/
13. Intellectual property and economic issues for digital library: a framework for future research. http://www.iei.pi.cnr.it/DELOS/NSF/ipe.htm
14. ACKERMAN (Mark S.) and FIELDING (Roy T.). Collection maintenance in the digital library. http://www.ics.uci.edu/~ackerman/pub/95b15/dl95.html