

**Digital Library of IITM-K – Experiences of Next Generation resource centre
K.Srinivasan, B.V. Laxmi, Manjul Sahay, Abhay Singhal, K.R.Srivathsan***

**Indian Institute of Information Technology and Management, Kerala
Park Centre, Technopark, Trivandrum 695584.**

Abstract

Digital Library has many features that make it distinct from traditional libraries. These features make it an excellent medium to address information and interaction services related to quality of education, and management of learning environments. While the paper libraries have become very expensive and are the prerogative of only a few well endowed institutions, Digital Libraries make these services and more at affordable costs. This will remove the digital divide between developed and developing nations in long run. Transversal E-Networks [TEN] a company under incubation at IITM-K has developed an academic aggregation server concept in which several academic functions such as Course Management, Authoring, and collaborative group-work are built around their unique metadata standards compliant Digital Library implementation. This server called “ACADO” is being field-tested as central information server for Indian Institute of Information Technology – Kerala. The server has more than proven its effectiveness in increasing the productivity and quality of academic collaboration, management of learning environment and research in the institution. The same server can network itself with similar servers in other institutions and form as Information, Knowledge or Educational Grids across the different Digital Library spaces.

Introduction

Broad perspectives of Library are storage, acquisition, preservation, organization, retrieval and dissemination of pertinent information to the clients. The traditional Library functioned as storehouses of knowledge. With the revolution in computer, information and telecommunication technologies the world of library has witnessed various developments. We may classify this into four phases, such as Traditional Libraries (First Generation), Electronic Libraries (Second Generation), Digital Libraries (Third Generation) and Fourth Generation Library systems-out come of new technology using WAP and virtual space¹.

Manjul Sahay and Abhay Singhal are Graduates of IIT Kanpur and two of the founders of Transversal E-Networks [TEN], a incubated company of IITM-K. Others are part of IITM-K.

Acknowledgements: Authors acknowledge gratefully the assistance and valuable comments provided by Mr. Sainul Abideen P, Librarian, IITMK.

Email: ksrini@iitmk.ac.in, manjuls@mail.com

Digital Library

A great deal of work has already been done in defining “Digital library” the most commonly accepted one is “*Digital Libraries are a set of electronic resources and associated technical capabilities for creating, searching, and using information...they are an extension and enhancements of information storage and retrieval systems that manipulate digital data in any medium (text, images, sounds, static or dynamic images) and exist in distributed networks* ((Borgman, 1996))”. A digital library is not merely a collection of electronic information; it is an organized and digitized system of data that can serve as a rich resource for its user community (Shalini Urs, 2001). As a Concept Digital library has different connotations for different professional groups. There are many stakeholders in this type of system with multiple roles played by users of the library.

Digital Library is an emerging field in the area of information and library services. With rapid developments in information technology, particularly, the web technology, the world of digital information resources has changed rapidly and exponentially. This new technology approach has increased or forced interest in the libraries around world. Digital information resources include not only rapidly growing collections of electronic full text resources, but also images, video, sound, and even objects of virtual reality.

There are many advantages in going for digital library system, materials can be delivered directly to the users’ computers, searching information is easy, any number for documents can be given to any users, multiple copy distribution at anytime to anywhere, easy maintenance and much more. The first step in the search for information has to be digital, as it will allow exploring more sources with in limited time. Internet is a huge repository of electronic information used widely. In the present day context the physical sources of information are being digitized, for its inherent advantages such as multiple usage, easy access.

Feature and characteristics of Digital Library

There is increasing demand for improving information and knowledge management solutions in enterprises, which can integrate access to disparate information resources and provide staff personalized information. Lets look at some of the features and design framework for the development and management of enterprise-wide digital information system. The basic functionality in such systems usually have, content organisation, content storage and processing, resource discovery, manage content publishing, usage and monitoring, access management,

preservation and maintenance, interoperability and networking. The broad categories are given below.

Content Organization:- There should be a system for content organization which is metadata driven approach. This metadata will carry different information about the resources like:- descriptive (subject, category etc.), administrative, rights management, security and authentications, bibliographic information. A system which supports for auto generation of indexing, classification, metadata for search, submitting information, mentoring of information submitted, Knowledge sharing, etc.

Digital Library at IITMK

Digital Library paradigm of IITM-K is not just collection of Books, Audio Visual Aids and back volumes but facilitating access of resources at all levels. This includes online education, discussion forums, references and reviews, authentication of resources, web transactions and so on. After months long discussion initiated by us along with various educationalists at IIT Kanpur, IIM Kozhikode and Kerala University we were able to build the paradigm. IITM-K library is powered by Transversal E Networks (TEN)- a company incubated by the institute.

IITM-K Library is focusing not only on procuring, classifying, and managing resources, but also facilitating continuing education and easy access of resources to different geographical locations, and resource submissions. This helps one develop case studies/ histories and scenarios. Why the case studies are important in the Knowledge Management era? Actually it reduces efforts by not reinventing the wheel. This further minimizes the efforts by learning from similar experiences.

IITM-K Portal

The IITM-K Digital Library built on top of a Web Server powered by the “Info-Space Operating Systems” concept developed by TEN and is internally central to the four portals it serves. The four portals are, IITMK, Telehealth Kerala, Transversal E-Networks, and E-Governance Promotion. There are four major services, besides the digital Library functions that are managed by our portal. The four are,

1. Information and collaboration across multiple groups of participants.
2. Continuing Education
3. Courses Management
4. Institute Information Services.

There are three major areas, which are addressed by the portal. They are, online Library, online Education, Management, and Security. The portal addresses three levels of security .

Open space- This is the source which are openly available and having no restrictions. These are mainly contributed by the collaborating organisations and contain non copyright resources.

Protected – The resources which are protected by the copyright and acknowledged.

Restricted- The subscribed materials which are part of IPR.

The links are the URLs which are providing the links, without violating IPR issues. The person whoever is using the portal will not be given access to the other features. In case of online education, only the registered candidates will be allowed to access the resources.

Web Space

There are common web spaces area along with the common groups message board. This allows one to store the files which can e downloaded whenever required for their requirements.

The Technology

The first and foremost issue for any solution is the architecture. Our digital library implementation is based on an *n-tiered web-based architecture*. Figure 1 shows the vertical dissection of this architecture.

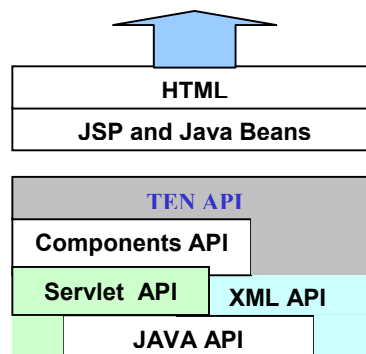


Fig 1.

The concept of n-tier architecture ([3tier](#)) is to separate the whole software into logical layers. The layers are kind of independent. The top layers talk only t layers just below them and above them. This makes layers independent of all layers other than their neighbors. So a change in any layer or the way it is programmed will only affect its two

adjoining layers and have no effect on the rest of the solution. For an example, let us consider that the metadata of resources of digital library are stored in files today and need to be shifted to databases like Oracle or MySQL at a later date. The n-tier (Fig.1) architecture will allow us to make changes only in lowermost layer of the shown architecture.

The above-mentioned architecture is a web-based architecture. All the data stored and processed are on the server. The users access the data from server using their browsers like Internet Explorer or Netscape Navigator. The server returns HTML to all client requests, which can be seen in the browser by users. All user functions including administrative and management actions are done using browser. There are two advantages in this architecture. First, it is browser enabled, and second, accessed from any computer on a network (or through internet).

The whole solution has been built with Java programming language and using Java Server Pages technology. This choice was made despite of a more difficult learning curve as the Java technology is mature and used for most enterprise-level applications because it can provide scalability and robustness. The XML is other cornerstone of the solution that has grown more with Java. XML or extensible markup language is a W3C standard and is a data representation language. We also made the choice of Java because solutions made with Java can be easily ported on both Windows and Linux platforms and even Solaris platform for that matter with minimal effort. Many libraries use Windows for its ease of use and administration while at the same time many libraries have opted for the open-source Linux platform. The ease of porting the solution on all platforms was a major deciding factor for Java. Java Server Pages (JSP) technology was used for the front-end presentation pages, which will finally output HTML for the client browsers. JSP pages again closely resemble HTML pages and logic or programs can be written in Java and embedded in them.

XML is a metalanguage, which allows tagging of data with custom meta tags. The primary reason for using it was to make our digital library solution future-proof. Today

the software industry is embracing XML and related standards as a means of making diverse applications interoperate. The objective is to make all applications and services talk to each other using XML and gradually ease out all proprietary and application-specific protocols. Thus to make our solution ready for the future all data is represented as XML. All metadata about the resource, user and comments ratings are represented as XML. There are specific programming methods available to import or export any resource as XML. The other reasons for choosing XML were it allows better data representation, better data retrieval and separation of data and presentation. XML represents data in hierarchical manner with parent-child relationship as against linear relationships in RDBMS (Relational Database Management System) like Oracle. This allows much easier design and data handling. XML by tagging data also allows better retrieval and search. For example, it will allow distinction between “Manjul Sahay” as the author of this paper and “Manjul Sahay” as CTO of the company TEN. Finally, data representation as XML allows usage of style sheet to present the data. Thus data and presentation are segregated as against HTML pages where data and style is interwoven and creates problems.

The other important technological aspect of our digital library solution is that it has been built on free and open-source software. The default configuration, which is installed at IITM-K, uses Apache web server, MySQL database and Tomcat servlet engine. All the above software are from the open-source domain. However, any of the above software can be replaced with any other proprietary software, if desired or required. For example for high-volume libraries Oracle can be used as database.

The boxes in the appendices give the glimpses into key features of the system and some issues addressed for conformance of Copyright/IPR concerns.

References

1. **Gopinath, Saji, Pathak, A.Kumar, M. Sreekumar M.G.**, “Fourth Generation Libraries” IIML-MANLIBNET Third Annual National Conventin: March 12-14 2001, IIM, Lucknow.
2. **T.B.Rajashekar** Digital Information Services in Enterprise: A design Framework and Strategies” ICADL 2001 , Bangalore.

3. **Urs, Shalini R.**Digital Libraries: An overview , United States Educational Foundation in India, DRTC/Indan Statistical Institute.
4. **K.R. Srivathsan**, “Establishment of Educational Grid Across Colleges and Institutions and IT Facilitation of Higher Education in Kerala” – a Project proposal for implmentation under the next Plan by the Government of Kerala, Dec. 2001. (available on request).

For useful information related to technology issues, the following links are provided.

5. <http://www.corba.ch/e/3tier.html>
6. www.w3c.org
7. <http://www.apache.org>
8. www.mysql.org
9. <http://java.apache.org>
10. Institute Website www.iiitmk.ac.in will give a feel of the system. Entry into the site may be given on specific requests.
11. For the features of ACADO Server, visit www.transversalnet.com

Appendices

Notes on Metadata Representation

A xml file showing metadata representation for a resource –

```
<?xml version="1.0" encoding="UTF-8"?>
<main>
  <resourceinfo>
    <title>Digital Library Paper</title>
    <institution>IIITMK</institution>
    <author>Srinivasan Laxmi Sahay Srivathsan</author>
    <publication />
    <keywords>digital library</keywords>
    <source/>
    <metadataauthor>manjuls</metadataauthor>
    <softwarerequirements />
    <accessrights>Free</accessrights>
    <usageremarks/>
    <courseAndresource />
    <category>Published papers</category>
    <expirydate />
  </resourceinfo>

  <no-of-user-comments>0</no-of-user-comments>
  <average-rating>0</average-rating>
</main>
```

iSOS - the TEN Core Engine

The digital library solution runs on iSOS (stands for Info-Space Operating System core engine developed by TEN (www.transversalnet.com)). All the basic functionalities like user management, session management, permission and access, database connectivity, task scheduling etc are provided by iSOS and used by the digital library application. iSOS is the core engine on which applications and services can be easily built and deployed. More details about iSOS is available at <http://www.transversalnet.com/techzone/about-tech.htm#arch> . An excerpt from the given link -

TEN core engine lies at the heart of the whole architecture. It provides a set of core components and functionalities on the top of which distributed web services can be built. It is a framework and not software or an application. It is extensible and easily scalable to meet diverse needs through the use of adapters to connect to other middleware applications. The core engine hides the data sources and systems beneath it from the top layers of

IPR Protection Implementation

In the current era of knowledge or information, the fundamental unit of almost all the products and services is information in one or another form. Most notions of the knowledge society like virtual products and virtual enterprise rest upon the cornerstone of 'information', in digital and non-digital form. In several cases such information is of proprietary in nature, based on the unique value derived from usage, research, development, design, and so on. Hence, the investment in that information product or knowledge product needs to be protected to encourage other similar initiatives. The initial investment is of critical importance given replicates such products, and could be created with relative ease and without incurring a large expense. This initial investment and effort is known as Intellectual Property and Intellectual Property Rights or IPR are the rights of the person or group, which made the initial investment in the creation of the knowledge product.

Following features of the digital library provide required IPR protection –

1. Complete access control – Librarians/administrators have complete control to provide access or deny access to any user or a group of users. Groups can be created for library or existing group distribution can be used. All resources are controlled by the access policies.
2. 40-bit encrypted login for administrators/librarians – 40-bit encryption based SSL login can easily discourage eavesdropper and hackers from getting access to the password of administrator/librarian.
3. Complete audit trail – The digital library software maintains a complete audit trail of the usage. Logs of usage and activity can be accessed based on user index and web-page index.
4. Encrypted URLs – The URLs of the resources are encrypted to preserve their identity.

Hardware and system configuration of IITM-K Digital Library Server

The digital library implementation at IITM-K is running on the following system configuration
(**Compaq Server** (Proliant ML 530 Intel Pentium III Xenon @933 MHz)

RAM: 512MB ECC SDRAM

Ethernet card:

HDD: 36GB+36GB+36GB+36GB =144GB

Tape Drive: 12/24 GB DDS-3 DAT Drive

CD ROM: 32 X CD ROM Drive

Graphics: Onboard graphics with 4MB

Power Supply: 450 watt

Additional Power Supply: Hot Plug RPS ML350

Additional Fan Kit: Hot Plug RED FAN Kit.

Windows 2000 Server, MySQL databse, Apache web server, Tomcat servlet engine.

Load testing was done on a system with following configuration - PIII 500 MHz processor, 192 MB RAM, 20 GB Hard Disk, Windows 2000, Apache web server, Tomcat servlet engine, MySQL database. For the above configuration, the application broke down at about 460 simultaneous user connections. The response time was within the limits of 5 ms to 600 ms. Conclusively; the above configuration is good enough for a scenario where at most there are 400 simultaneous user configurations. So any organization/institution with less than 4000 users can work easily with the above configuration.