Thursday, 26 February 2004

Conference Bulletin

International Conference on Digital Libraries 2004

Knowledge creation, preservation, access, and management

Organized by TERI in partnership with Department of Culture, Ministry of Tourism and Culture, Government of India

New Delhi • 24–27 February 2004

http://www.teriin.org/events/icdl

In brief . . .

KEYNOTE ADDRESS 1

Chair:
Dr R Natarajan, Chairman, AICTE
Speaker
Mr Kiran Karnik, President, NASSCOM

Digital library in a broader context is nothing but a database. The objective of the digitization process should be 'empowerment of people'. The Indian IT (information technology) industry has done remarkably well globally in recent years. But, its orientation towards export has given the industry a negative image within the country.

There are technologies available, but use of the appropriate ones, not necessarily high-end ones, is the key. While conceptualizing any digitization project, it is important to have some measurement criteria for outputs as well as outcomes. Outputs should be in terms of number of documents digitized; outcomes should be in terms of ease of accessibility, the spread or the reach, and the ability to access information and knowledge. Both qualitative and quantitative goals should be incorporated into the project.

Another important dimension of the digitization process is the coordination between information technology and library science professionals. These professionals will have to work as a team to achieve the desired objectives.

A few successful examples from India were cited. In Madhya Pradesh, in an attempt to empower the rural populace, several kiosks were set up to provide information. These kiosks are totally entrepreneur-driven and a small fee is charged for their use. In Karnataka, about 20 million land records were computerized, bringing in more transparency and authenticity of ownership. In Andhra Pradesh, through the e-*Seva* project in Hyderabad, electricity bills, telephone bills, property taxes, etc., can now be paid through the Internet. In Rajasthan, many government policy documents have been digitized and made accessible to the public.

SESSION I

Session IA
Digital Library Concept

Development in information and communication technology has helped not only in managing classical libraries better, but also in creating, distributing, and accessing information resources in digital formats through computer networks. Libraries follow different standards and guidelines for digitization. Developing a set of adaptable yet comprehensive standards to meet the needs of all libraries is not a viable proposition. However, greater efforts are needed to ensure more standardization rather than more varieties of standards.

In the US (United States), one-third of academic libraries and one-fourth of public libraries have been digitized. Although preservation of historical resources is a primary goal in a number of digitization projects in the US, the common thread shared by many digital projects is to provide greater access than ever before to these collections.

For digitization to succeed in the developing countries, the socio-economic conditions have to be studied while choosing the correct approach or a road map to a digital library. Other challenges include technological obsolescence, media fragility, hardware and software compatibility, copyright, and problems related to formats and security.

Session IB
DL Services

As the increasing number of DLs (digital libraries) expands the overall size of the Web, with many of them not included in search engines, it becomes difficult to connect potential audience with available resources. Understanding linking and meta environments, and keeping up-to-date with search engines will make the available resources visible to users. This is necessary as the search engine industry continues to change its patterns of indexing, ranking, and providing access to information.

The digitized/computerized information services provided by the Indian Institute of Technology, Roorkee, India, was a case in point in the discussion, which highlighted the benefits in shifting

to a digital environment for information dissemination.

Panellists discussed the 'cache architecture' for LFDL (light-weight federated digital library) to address networking with other libraries and improving performance and usability. Caching allows processing the metadata, improves search services, and enables the exposure of metadata via 'open archive initiative protocol for metadata harvesting'. This way, it is feasible to build a federated search service that is compatible with non-cooperating DLs. It was stressed that there is need to investigate various approaches to caching and also the management of cache including replacement policies.

Session IC
Technology: Planning, Development, and Management - I

In the current digital era, when most of the information is easily accessible despite barriers of space and time, libraries are also going digital. Monash University in Australia, for example, has taken a lead and digitized all its collection and is now able to provide e-books, e-journals, etc., to its students, faculty, and other professionals. According to them, such features as search and better definition of metadata have helped their users. The GSDL (Green/stone Digital Library) software is an open-source software. With features like data management and publishing for the Web and CDs, it has become one of the most popular and most used softwares around the world. Integration of CDS/ISIS database with GSDL data has become easy with different kinds of approaches different people have developed. Library professionals are now required to widen the range of their skills if they are to effectively operate the software available for digital libraries. Information technology and library professionals have to go hand in hand to make the digital libraries a reality and a success.

Session II

Session IIA
Digital Divide

Digital divide refers to the gap between individuals, households, businesses, and geographic areas at different socio-economic levels with regard to the opportunities available to them to access information and communication technology. Five types of digital divides have been identified: (1) between rich and poor nations, (2) between haves and have-nots within a country, (3) between those who can speak English and those who cannot, (4) between commercial and individual rights, and (5) between digital elite and the urban or rural poor in terms of digital development itself.

The factors that produce this division are (1) technology, (2) economics, (3) information, and (4) tools. Metasearching and link resolution are two new tools that make it easier for end-users to search for information and remove much of the drudgery in searching for information and using it. Also, we need technologies to break language barriers. Participation of non-governmental organizations and public–private–academic interaction are also important in the digitization process. Specialized training programmes to upgrade skills of the professionals are essential.

The case study of Gulbarga in Karnataka highlighted the nature and extent of digital divide in an urban setting in India. For bridging the digital divide, the developing nations could use both online and off-line digital libraries. Off-line digital libraries are less expensive compared to online libraries.

Session IIB
e-Publishing and Multimedia

Context-driven personalization of DL reduces the gap between the content provider and the user. The session highlighted various personalization methods that enable context-driven library access and illustrated three advanced approaches: (1) personal reference libraries, (2) collaborative content annotation, and (3) modelling and exploitation of personal web context.

Rather than seeing a 'document' in the conventional sense of being a paper-based entity, it was urged that we consider a document as 'an entity consisting of any media type appropriate to store or exchange information in a given context'.

In the context of the changing process of scholarly publishing, it was observed that there were tremendous expectations from electronic publishing. Achieving higher quality and quantity of retrieval can meet these expectations. This could be realized by creating a 'new culture': new rules, standards, and technology. The discussion also covered the development of DLs by utilizing freeware like Apache as web server, PERL packages and modules, and the Greenstone Digital Library software.

In the developing countries, where general literacy level is low, electronic literacy is a dream. But considerable effort is being made to connect to most regions of the world. Application of information technology to the collection, selection, distribution, and preservation of archival contributions offers the possibility of fostering international cooperation in this regard.

Session IIC
Technology: Planning, Development, and Management - II

Digitization of the library has become a buzzword. Several

softwares are available off the shelf that can help libraries in managing and publishing data. DSpace is an open-source institutional repository system developed by Hewlett-Packard Labs and Massachusetts Institute of Technology libraries in response to the need to cope with increasing amounts of born-digital information artefacts in institutional settings. Since its release, thousands of organizations have downloaded the DSpace source code, and many DSpace sites have sprung up around the globe. The extensive functionality of DSpace, which includes data model, metadata, user management, and authorization, in addition to a robust architecture, has helped many organizations. The 'COLLATE Project', on the other hand, is set out to design and implement a Web-based collaboratory facility for archivists, researchers, and end-users working with digitized historical or cultural material. The COLLATE system offers a comfortable working environment and user interfaces for domain experts and end-users for indexing, annotating, and searching multi-format, multimedia archive material. On top of the distributed digital repository, it functions as a 'collaboratory in use', supporting distributed user groups through such dedicated knowledge management facilities as content-based access, comparison, and in-depth indexing and annotation of digitized sources.

Session III

Session IIIA
Case Studies - I

Metadata are the key component of digital libraries. To build digital library repositories, caution and cooperation are needed. Risks can be minimized by using open-source software like Fedora, which is available for free on the Internet. While digitizing is easy, scalability and quality control can become a problem if these issues are not addressed earlier. Test cases on digital library developed by C-DOT include digitization of newspaper clippings, sharing technical knowledge, and technical review and updating. Open-source software can be modified to suit organizational needs, and reusing designs is cost-effective. The Tata Memorial Hospital's case study showed that doctors relied on the library for immediate and critical information. Doctors had a tendency to search less and browse more. Journals were accessed either in portable document format or as HTML, so that figures could easily be adapted. The Tamil Nadu case study highlighted the importance of digitizing ancient manuscripts and showed how two initiatives, namely Project Madurai and TADILNET, have been successful. It was observed that there is a need to develop an OCR (optical character recognition) package for Indian languages; there is also a need for standardization.

Session IIIB
Metadata Issues

As organizations go digital, thanks to the all-pervasive technology, the challenges to provide meaningful information to users in the shortest possible time become immense. That is why metadata, which is defined simply as 'data about data', are becoming important in the context of the World Wide Web.

Metadata are well-organized, descriptive information about web resources, like catalogue cards in the case of printed documents. Constantly changing content and information resources on the Internet and the lack of a single format make cataloguing, search, and retrieval difficult. A universally accepted standard format for identification and organization of data elements of Internet resources is, therefore, crucial.

XML, a mark-up language for structured information, is the basis for many new and emerging technologies. Whether it is for bibliographic databases – the print equivalents of indexes and abstracts – or RDBMA (relational database management) for storing and managing different forms of digital content, XML becomes the preferred syntax for library standards. It is a very efficient format for publishing and exchanging data and a powerful tool to access the contents of any type of relational database.

Session IIIC
DL: Architecture, Network and Technology - I

Networked digital libraries must have a vision, namely 'access any document from anywhere without language and format barriers at any point of time at a low cost'. It was categorically stated that the lack of high bandwidth connectivity and proper training within the country would hamper the spread of digital libraries to remote and rural parts of the country. Creation and exploration of 'musical information spaces' for retrieval of music and searching of content effectively and efficiently are new dimensions in digital libraries. High-speed networks are essential for migration to high-tech libraries. However, for effective digitization of libraries, a 'traditional-level' (10 mbps) network has to be set up to begin with, which can later be converted into a 'middle-level' (100 mbps) network before moving into a 'high-level' (1000 mbps) network. The session, however, concluded by accepting the need for high-end networking devices and powerful IT infrastructure for developing intangible (digital) libraries from the tangible (printed books) ones.

Session IV

Session IVA
Case Studies - II: Special Libraries

The process of creating digital libraries is similar to that of the print media. Interoperability is the main challenge in the development of digital libraries. Different concepts of digital libraries exist and a clear-cut definition needs to be evolved. There is a need for well-developed national policies along with policies on preservation. Copyright is another issue that needs to be addressed especially from the users perspective. However, copyright of academic work should be addressed differently as against entertainment work. The challenge of digital libraries is to arrive at a consensus as regards technical and workflow formats. Vidyanidhi was one such project developed in Kannada, which overcame the problem of fonts wherein the user did not have to worry about downloading fonts.

The digital library maintained at IIT, Kharagpur, provides access to online journals, e-books, digital documents, portals, web sites, and so on. The Dublin Core has been strongly influenced by librarians and is referred as the library standard. The library has initiated to preserve existing digital resource for long-term use by applying modern preservation techniques.

The TERI library initiative is an automated, networked library with bar coding done in house. The library also maintains various specialized online networks such as ENVIS, Mycorrhiza Network, Regulateri, and many others.

Session IVB
DL Communities

The advent of IT has revolutionized the way a library functions, with digital libraries emerging as a faster way to search, read, and disseminate information. A step beyond are digital library communities, which bridge the gaps in global information access and dissemination. While the developed countries have taken great strides in developing and strengthening information access to communities, the developing countries need to undertake more initiatives to reap the benefits.

To give a fillip to the development of a digital library community at the national level, there is urgent need for countries like India to put in place a national apex body to coordinate and support strategic, technical, and infrastructural plans.

The key to success would lie in the selection of contents, collaboration with other partners in a sustained and organized way, putting in place proper infrastructure, and a composite approach to the digitization process.

Apart from academics, designers too are using digitization to bring on one platform text, audio, and visual resources for multiple reference, quality presentation, and archival storage.

Session IVC
DL: Architecture, Network and Technology - II

Digital environment encourages users to discover, support, create, and interact with resources from any place at any time. The basic requirements for creating and operating a digital library are collection, infrastructure, deployment, evaluation, and monitoring. Development of interactive learning objects was projected as the unique feature of the model described in the given session. In BITS (Birla Institute of Technology and Science), Pilani, metadata have to be very urgently digitized to move towards a digital library. Two types of trends, namely techno-driven (driven mainly by computer scientists) and user-driven (driven mainly by library professionals), were elaborated. However, the techno-driven changes had an edge over user-driven changes because of greater power and efficiency. The discussion covered the growth and evolution of medical digital libraries in our country. It is necessary to integrate the existing conventional libraries with digital libraries. The session concluded by agreement on the positive aspects of digital libraries and their networking.