27 February 2004 (Concluding Day)

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 3

Chair
Prof. Ian H Witten, Department of
Computer Science, University of
Waikato, New Zealand
Speaker
Prof. N Balakrishnan, Professor and
Chairman, Division of Information
Sciences, IISc., Bangalore

If technology push, crashing costs, and global connectivity are the signposts for the coming decade, digital libraries are the next step in the convergence revolution. The overall goal is to have all public knowledge online, available for free to all, everywhere. At present, there are only 50 000 000 books in the world and a few billion dollars could bring these online. However, there are several obstacles to reaching the goal of universal access. Use of copyrighted material, privacy, reliability of information, and, most important, the cost-effectiveness of the entire operation are still issues to be discussed and resolved.

In the 20th century, governments created BBC (British Broadcasting Corporation), PBS (Public Broadcasting Service), AIR (All India Radio), and also the public library system in the interest of public good. These provided compensation for artists and writers while providing free access to public. The total global expenditure in public broadcasting and public libraries exceeds a hundred billion

dollars. Our ancient kings supported poets and scholars so that people at large can benefit. It is time to find the 21st century equivalent of BBC, AIR, and PBS.

Perhaps the solution lies in the four Cs (Consortium for Compensation of Creative Contents), which suggests setting aside 25% of the current national expenditure on public broadcasting and public libraries. Authors are encouraged to put the work on the Web after a few years of commercial exploitation and in return may be exempted from tax.

Knowledge multiplies whenever bits are circulated on the web. The Internet and the four Cs are necessary to push each other towards the exponential increase in connectivity and knowledge. Technology follows the law of accelerating returns, not the law of diminishing returns. The 'expanding universe theory' means that the time span of a technology will be shortened.

SESSION IX

Session IXA
Manuscripts: Archiving and Digitization

Medieval manuscripts hold a high potential to be preserved in digital format for knowledge organization and accessibility. However, it needs planning and execution at the government

level, so that it is done in a systematic way. At present, Indian manuscripts have a global audience and their invaluable nature calls for preservation, conservation, and distribution. There are two ways of doing this: through microfilming and through digitization.

Digitization is cost-effective; it also allows bibliographic control, unhindered access to contents, and effective retrieval. Many initiatives have been taken to digitize manuscripts in the past. Oxford University, UNESCO Memory of World Heritage, and DELNET are a few players who have gone ahead and digitized their manuscripts collection.

Digitization of manuscripts also calls for image enhancement. Pre-processing includes enhancing and refining the raw image, which involves identifying and extracting blurred, stained, faded, bled-through, or transferred characters. The new PRISM method specifically identifies and analyses linear structures (line strokes). This technique works in both 3D (CT, MRI) and 2D (images) domains.

But at the ground level, especially in India, all these applications are possible only if there is a strong restructuring of all universities and public libraries. So far, a librarian has been a guardian of books, but now her role may change. Going by the speed of digitization, she may become a broker between

information and the user.

Library collections of Indian universities are in transition; a new era is beginning. The Information and Library Network – INFLIBNET – has so far helped 142 university libraries that are on the way to computerization. ERNET India and UGC (University Grants Commission) are setting up the UGC Infonet. Under this programme, it is proposed to use ICTs (information and communication technologies) and the Internet to transform the learning environment from a mono-dimensional to a multi-dimensional one.

University libraries in India have come a long way. Their collection has grown in quantity as well as quality with the changing times, which will also include digital documents in the near future.

As scholarly communication changes and as users' needs change, librarians and libraries must also change to meet those needs. They will have to be flexible enough to continue changing and adapting to change as they have done over the past decades, but more rapidly and more creatively.

Session IXB
Storage and Retrieval System

The quantity of information available is increasing rapidly every year. Hence, managing and archiving huge data and answering user queries are the main problems one comes across while extracting information across multiple documents. The main aim of information retrieval technologies is to provide a higher level of information to the user. IEMD (Information Extractor from Multiple Documents) is a step in this direction. IEMD synthesizes information common to a set of related documents, either qualitatively or quantita-

tively, and helps users find relevant information without having to go through the entire document. It is an improvement over the conventional information extraction methods. The potential applications of IEMD are identifying research trends, expanding cited references, etc. However, issues that still need to be addressed are information extraction, document standardization, and handling more abstract-level queries.

IR or information retrieval is concerned with automatic storage and retrieval of documents. IR systems and the Library of Congress Classification are currently being used. There are different mining techniques available (for example, sequence, web, or text mining). Clustering is one of the techniques used in data mining. Clustering groups similar types of data to facilitate easy access. Hence, based on the composition of keywords, simple and investigative methods to analyse and compare documents are possible.

Session IXC
User Interaction

One of the principles declared at the WSIS (World Summit on Information Society; Geneva, 10–12 December 2003) is related to access to information. With this background, the session began by raising such questions as: Do users have access to information? If so, is information easily intractable?

When a survey (55 postgraduate students were asked to carry out a literature survey on hybrid libraries) was conducted to understand the human information-seeking behaviour, it was found that users need some basic skills in ICTs, different interfaces for e-books and for e-journals, good connectivity, and the ability to formulate a proper query.

It was suggested that the user should be trained in using digital libraries; there should be a one-stop window for searching various collections, a single gateway to search the Internet, intranets, digital libraries, and KM (knowledge management) repositories. There should also be a provision for a user-driven information access system and also a task-based information access system.

A UCD (user-centred design) should involve capturing user requirements through user research. It would help in developing user profiles. Before releasing the end-product, there should be a prototype, which would be thoroughly tested by the users. In other words, users should be involved from the beginning until the end and at all stages of designing a digital library system.

Session X
Session XA
Copyright Issues and Digital Rights Management

Copyright is a legal device that provides the creator of a work of art or literature, or a work that conveys information or ideas, the right to control how the work is used. Intellectual property is a capital asset albeit an intangible one but it is protected by contracts, patents, and, sometimes, by encryption and watermarking (fingerprinting) technologies. However, certain uses of a work protected by copyright do not require permission of the copyright owner when the purpose is essentially non-commercial (a teacher taking photocopies for distribution in a class, for instance).

In the pre-digital-copying age, photocopying was private, labour-intensive, and un-remunerative until high-quality dry processes and copy shops came into being. By then, consumers

had acquired explicit legal rights over intellectual property owned by others. Today, many rights holders view the basic technology of the Internet as their enemy. Analogue sources provide no unambiguous clues about possible infringement. As long as any digital work can be printed or played in analogue form, it can be copied. The line between copying for private use and copying for worldwide distribution is dangerously blurred.

Copyright enforcement is a purely private matter within civil law. Some rights holders' organizations have taken on an active vigilant role. Universities are working to increase copyright awareness among students and faculty.

Widespread access to digital reformatting tools and the Internet has changed the balance between the publisher and the consumer. Whatever legislation is passed, and whatever technological protection is used, the management of intellectual property cannot eliminate reasonable exercise of consumer rights without risking widespread popular resistance and potential economic harm.

Computer specialists and library science specialists must be represented on all authorities, councils, committees and similar bodies concerned with formulating any public library act. The Press and Registration Act, the Delivery of Books Act, and the Copyright Act must be suitably amended. Depositing a digital version of every document should be made compulsory. In the beginning, all digital versions can be offered at a fee to cover costs.

Session XB
Storage and Retrieval Systems II

Improving access to information, so that it addresses the needs of people, is a major issue in the

storage and retrieval systems in digital libraries. Removing unnecessary data while keeping only relevant information, especially when there is a heavy information load (with the use of the World Wide Web and digital libraries), is also vital. In order to address these needs, a 'query preview system' and a 'collaborative filtering' method have been useful. Also, collaboration and communication are vital to successful information retrieval. Query previews give an overview of the data distribution in a data collection, preventing users from applying fruitless queries. Collaborative filtering keeps a tab on how users formulate queries. These methods help improve information access. But, in order to overcome the flaws of both the systems, a new CQP or collaborative query preview system is useful.

'Federated search' is also a path to better digitization. In this process, a user can search a library without logging on to a portal.

The use of multilingual search engines is also vital, especially with regard to countries like India where the rural masses are deprived of the knowledge available through the Web because most of it is in English. An integrated search system will also help make search engines more user-friendly.

Session XC
DL Consortium

A consortium of digital libraries will help librarians, users, vendors, and publishers tremendously, saving not only money but also effort and time. The INDEST (Indian National Digital Library in Engineering, Science, and Technology) consortium, which has 125 members, was set up in 2003 by the Ministry of Human Resource

Development on the recommendation of an expert group appointed by the ministry under the chairmanship of Prof. N Balakrishnan, Indian Institute of Science, Bangalore.

Apart from regular services such as offering access to subscribed e-resources and providing technical help and in-house training for optimal use, INDEST offers specific services such as shared technology (federated searches, an application service provider, a union catalogue), joint storage (journal achieves and back files), shared core collections, and much more.

The consortium also regularly monitors usage in several ways including analysis of statistics and asking users to fill questionnaires.

There is no universally acceptable pricing model for e-journals; subscriptions are available in different forms, such as electronic only, print and electronic, pay-per-view, membership-based access (to current and archived files), and extra charges for value-added services.

PANEL DISCUSSION
Digital Library Policy and Strategic Planning

The penultimate session of the conference touched upon the current status of digitization in India and went on to assess why and how India should build up its digital libraries. Panellists highlighted approaches and concerns that policy formulation should take into consideration.

It was felt that digital libraries should not be just about the abundance of knowledge but should also consider whose knowledge, what kind of knowledge, and at what cost it is embedded in digital formats. As these libraries are meant for different groups of people, their usefulness to society should be

ascertained and some mechanisms put in place so as to plough back the benefits to communities that are the real holders of knowledge.

Experts, while discussing the issues of technology development and compatibility, did acknowledge that the issues of scaling, speech input and output, resource allocation, and of scaling intellectual property access to all kind of users still need to be resolved and so do the research issues of OCR (optical character recognition) for Indian languages, machine translation, speech recognition, automatic summarization, and handling large distributed databases, to name a few.

Issues identified from the perspective of the user were: identifying what user wants, structure of digital libraries, ownership and accountability of data available in these libraries, and a system in place to digitize indigenous knowledge without jeopardizing the owners' intellectual property rights so that digital libraries become a vehicle of the transformation towards a knowledge society.

ICDL 2004 SUPPORTERS

Partner Department of Culture, Ministry of Tourism and Culture, Government of India

Associates Department of Information Technology, Ministry of Communications and Information technology Government of India • Commonwealth Educational Media Centre for Asia • Centre for Development of Advanced Computing

Co-Associates (Speaker Sponsors) Emerald • Goethe-Institut, Max Mueller Bhavan • Italian Institute of Culture • Public Affairs Section of the American Embassy • French Information Resource Centre, Embassy of France in India • The British Council, India • Embassy of Austria in India

Function Hosts Blackwell Publishing • THOMSON-DIALOG • Institute of Electrical and Electronics Engineers • Global Information Systems Technology Pvt. Ltd • Murugan Magazine Distributors Pvt. Ltd • Microsoft Corporation (India) Pvt. Ltd

Sponsors Nature Publishing Group • Tausco Book Distributors