Managing E-Database in E- Library Implementation, Problems and Prospect: A case study

Kamalednu Majumdar
Assistant Librarian
Central library
Indian Institute of Technology
Kharagpur – 721 302
West Bengal , India
Email: kamal@library.iitkgp.ernet.in

Prof. U N Singh
Head
Department of Lib. & Info. Sc.
Guru Ghasidas University
Bilaspur (C G)
India

Abstract

The automation of the Central Library Indian Institute of Technology, Kharagpur) had got off the ground with the acquisition of the Meteor III computer with 22 terminals, LIBSYS software, a few 486 PC's with CD-ROM drives, barcode scanners, laser printer etc. The mammoth task of database creation of entire holdings of 300,000 records, bar-code generation of books in active circulation etc have achieved. But in the mean time there was a steady growth of documents in electronic form and access will necessarily have to be provided to them. Databases are increasingly available in (and sometimes only in) machine-readable form. Publishers commonly have or could have the text of their books / Journals in machine-readable form. To cope up with this situation Library starts to subscribe the E-Database along with the print form of the same journal and provided the stand-alone service to the user who take the difficulty to come over the library. We find this type of user is less in number or our learned user don't find time to come over the library. So Information on their desktop is demand of time. By acquiring a high end server with Xeon processor , 1GB Network interface , 1GB RAM, 432GB SCSI HDD, the library tried to solve the problems. The available database and the Book CD copied to the HDD and through IIS server the same may be access from the remote client from various laboratory of Institute. . In this paper the author intended to discuss the bit and pieces of the problems and prospect of the networking E-database.

Keyword : Electronic Database, Electronic Library, High-end Server

01. Introduction

The term "E-Library" to describe the situation in which documents are stored in electronic form, rather than on paper. Paper copies of electronic documents can also be excerpts from them, can generally be produced for the user convenience. But the essence of the Electronic Library is that documents are stored and can be used in electronic form Digitization is used to mean text in which individual letters, numbers, and other characters are each separately coded and the term "digitized" is used to denote an image composed of lots of bits. One could well have a digitized image of a page of text, but in that case none of the individual characters would be coded as characters. In some cases where the characters are regularly formed, it may be feasible to use pattern recognition techniques to identify the individual characters within the

digitized image and, thereby, to derive a digital version of the text from a digitized image of the text. The other way is scanning of the document and with the help of OCR software the scanned image may convert to text. Here the error rate with best OCR also 7% to 9%. A digital record is more economical and, ordinarily, more useful.

The adoption of computers for libraries' technical operation can be viewed as an evolutionary development. The transition from the Traditional Library to the Electronic Library, much of the change represented the mechanization of previously manual procedures of the Traditional Library. Electronic Library, in which materials are stored in electronic form, may seem more revolutionary than evolutionary because of the implications for the provision and use of library services. But is it really so radical a change? Where are the impacts on the provision of library service? How are we to achieve a graceful and efficient continuity of service as electronic documents come into use? Frederick Kilgour [1] has characterized the evolution of electronic publishing as being in four stages:

i. **Preparation of text in machine-readable form for the production of paper copies with subsequent discard of the machine-readable version.**

ii. **Publication of paper copies and also in machine-readable form,**

iii. **Publication in machine-readable form only, and**

iv. **Publication of information in machine-readable form of CDROM database. (Abstract / Full text)**

02 Property of an E-Database

- E-Databases are not localized
- Given telecommunications connections, an electronic document can be used from anywhere, without one even knowing where it is stored graphically.
- **N numbers of people can use the same database or electronic records at the same time.**
- Electronic documents are easily copied or download.
- Documents stored electronically are very flexible.

021 Current Trend of E-Database

There is a steady growth of documents in electronic form and access will necessarily have to be provided to them. Databases are increasingly available in machine-readable form. Publishers commonly have or could have the text of their Books / Journals in machine-readable form even if they may not yet choose to publish them in that form. Presently the following E-Database are available with Central Library, IIT, Kharagpur.

- IEL(IEEE/IEE Electronic Library (I E L) 1988- to date
- INSPEC : Physics 1989 to date
- Ei Tech Index 1987-1999
- Ei COMPENDEX 1995 to date
- Derwent Biotechnology Abstracts 1982-1998
- Diolog on Disc Biotechnology and Bioengineering 1984-1999
- Chemical Abstracts on CD (1999 to date)
- CAB Abstracts(Commonwealth Agricultural Bureau Abstracts) 1990-2001
- LISA(1999 to 2001)
- ITU - T Standards (June 2000)
- ITU - R Standard (March 2000)
- ASTM Standards(March 2000)
- BIS (March 2000)
- ABI/Inform (1989 to 1993)

CURRENT CONTENTS(CC)

- Agriculture, Biological & Environmental Sciences(1998 to date)
- Physical, Chemical & Earth Sciences(1997 to date)
- Engineering, Computing & Technology (1996 to date)
- Social & Behavioral Sciences(1996 to date)
- Life Sciences(only 1989 to 1999)
-

And more then 700 Book CD are also available.

03 Methodology

As we already have some experience what to expect in the Electronic Library from our experience with the transition from the Traditional Open access Library to the Automated Library and from our experience with on-line bibliographies. We consider the following point

- Remote access to Database.
- Multiple access of Same Document.
- The linking and combining of files,
- Access to numerous different files from the same terminal,
- Increased cooperative use of shared files. (Application Software)
- Data Security
- Optimized the downtime of the server.

031 System Installed to serve the E-Database Through LAN

We copied all the database in the same format which they are in the SCSI Hard disk (HDD) in a **Proliant ML530** System**.** The system configuration is as follows

- two Intel Pentium III Xeon 1.0 GHz processor with 256KB of level-2 writeback cache; one processor standard
- The Highly Parallel System Architecture providing greater throughput for improved overall system scalability and performance
- 133MHz front side bus and 133MHz registered ECC SDRAM
- Integrated ATI Rage IIC Video Controller with 4MB Video Memory
- Dual Channel Integrated Ultra2 SCSI Adapter
- Compaq NC3123 Fast Ethernet NIC PCI 10/100 WOL Controller in a slot
- 36GB*12 Altra SCSI HDD

The Smart Array 5302/32 is a dual channel controller with 32MB of battery-backed cache. The Smart Array 5302/32 supports Ultra3 LVD SCSI technology for data transfer rates of 160MB/s per channel, and is backward-compatible with devices using older SCSI technologies. A cache option kit and a two-to-four channel Wide Ultra3 option kit allow the Smart Array 5302 to be upgraded to 128 MB cache, four Ultra3 SCSI channels, or both. The cache Array Accelerator boards are battery-backed to maintain cache data for up to four days, with two replaceable battery packs.

0312 Internet Information Server (IIS)

IIS is software for serving up content to Web clients, In addition to supporting delivery of Web content via HTTP. IIS also supports the File Transfer Protocol (FTP) Internet Information Services 5.0 (IIS) is the NT based Web service that makes it easy to publish information on the intranet for the Internet.

Features of IIS

Digest Authentication: Digest authentication allows secure and robust authentication of users across proxy servers and firewalls. In addition, Anonymous, HTTP Basic, and integrated Windows authentication.

Secure Communications: Secure Sockets Layer (SSL) 3.0 and Transport Layer Security (TLS) provide a secure way to exchange information between clients and servers. In addition, SSL 3.0 and TLS provide a way for the server to verify who the client is before the user logs on to the server

IP and Internet Domain Restrictions: Grant or deny Web access to individual computers, groups of computers, or entire domains may possible

Restarting IIS: The IIS service can Restart without having to rebooting computer .

Process Accounting: Provides information about how individual Web sites use CPU resources on the server. This information is useful in determining which sites are using disproportionally high CPU resources or which might have malfunctioning scripts or CGI processes and many more.

0313 How it works

The Client/Server Model has two entities, ***Web client*** requests files, data and services and the ***Web server***, which fulfils the client's requests for content using the protocol. The Web browser and the Web server send information back and forth using the Transmission Control Protocol/ Internet Protocol (TCP/IP). This is the same network protocol used by many other Internet services, such as e-mail and the file Transfer protocol (FTP) . It's a common language at all computers connected to the Internet/Intranet speck. A digital software program or socket connects computers using the TCP/IP protocol.

A socket is a line of "plumbing "that is established by either the client or the server software that is used to move packets of data back and forth between them. It is through this "channel" that requests and responses are sent inside packets of data. A packet is a fragment of data that includes information about its origination, its destination, and so forth.

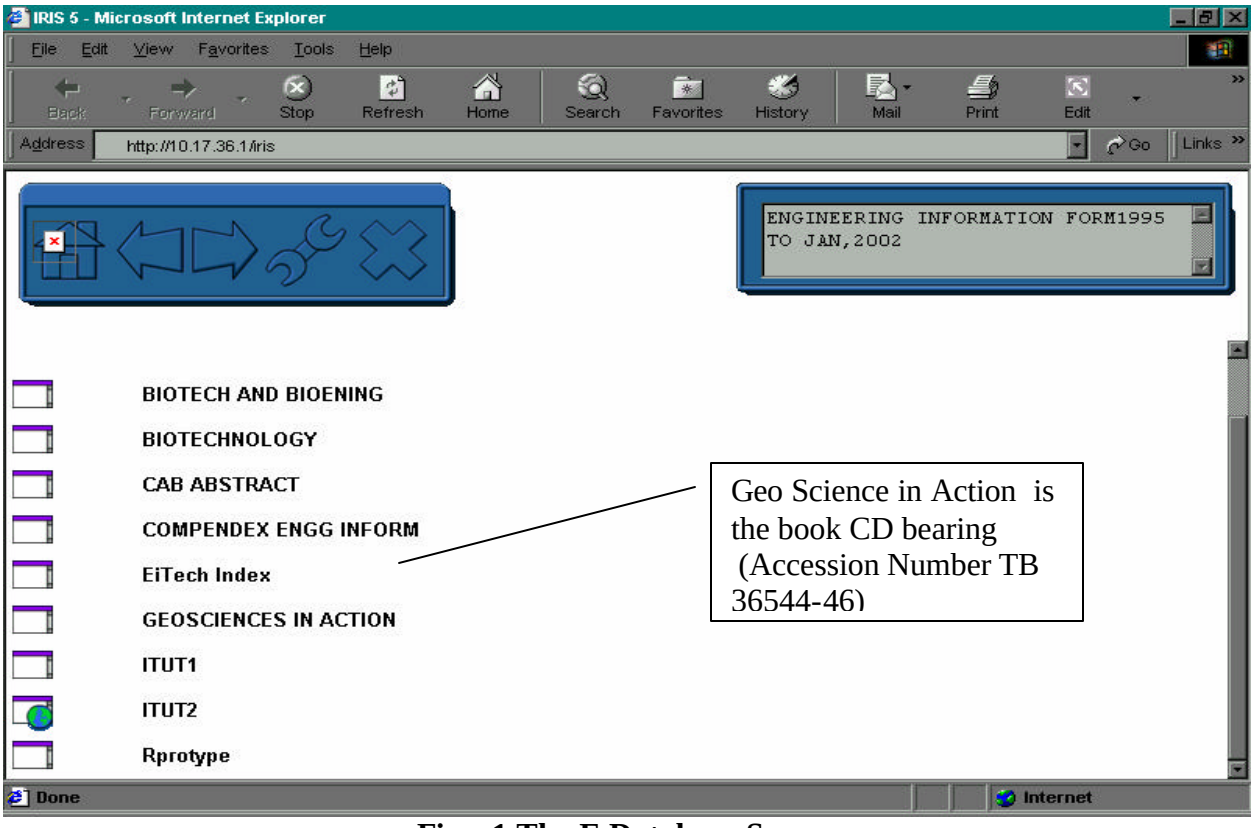

Fig – 1 The E-Database Server

A socket is a line of “plumbing “ that is established by either the client or the server software that is used to move packets of data back and forth between them. It is through this “channel” that requests and responses are sent inside packets of data. A packet is a fragment of data that includes information about its origination, its destination, and so forth. The TCP/IP protocols handle the encapsulation of user data into these packets for transmission. Packets transmitted along sockets arrive at their destination in exactly the order in which they were transmitted. Web application s communicates over sockets using a higher-level protocol – the Hyper Text Transfer Protocol. (HTTP).

The web client established a link through HTTP and will get the links file . Now when the user click on the desirable link the related application software file getting execute in the server it self and the client get the environment as the execution is happening in his own system

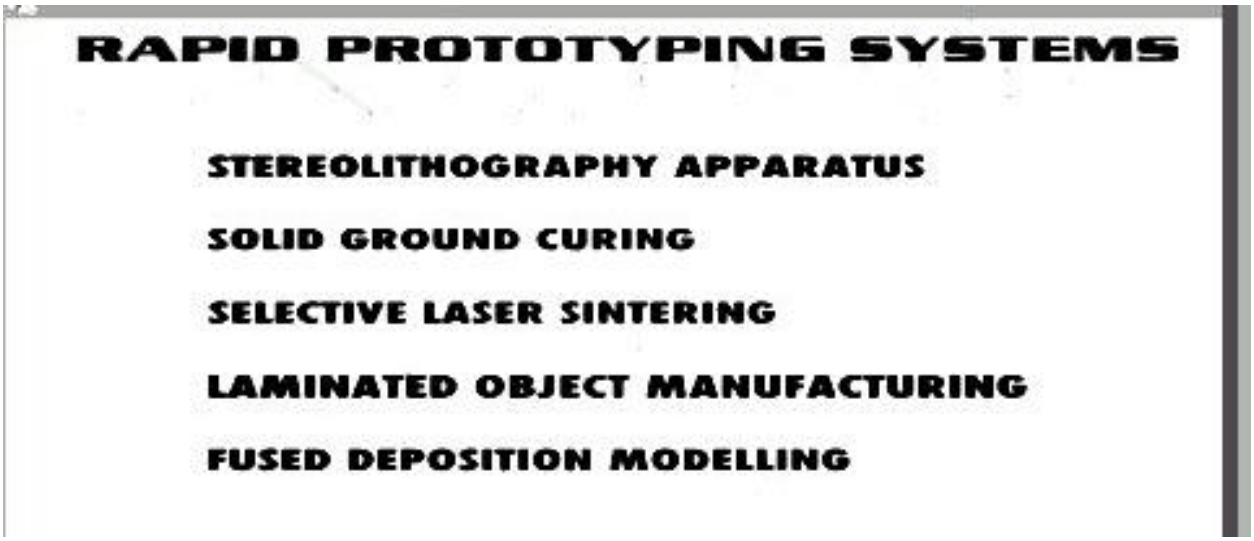

Fig – 2 Book CD getting access through the Browser with Flash5

04 Problems Encounter

The E-Databases available with us are in two modes

- Stand alone
- Networked

But when we tried to make the stand alone database accessible through INTRANET. We encounter the problems that the abstract part of the database can be made accessible from the client system but in case of “full text “ , for the source by default the application program use to search the local ‘D’ drive in the server itself instead of searching for the file in the Hard disc . This problems may be solved by a CGI (Common Getaway Interface) program and the development for this program is in progress.

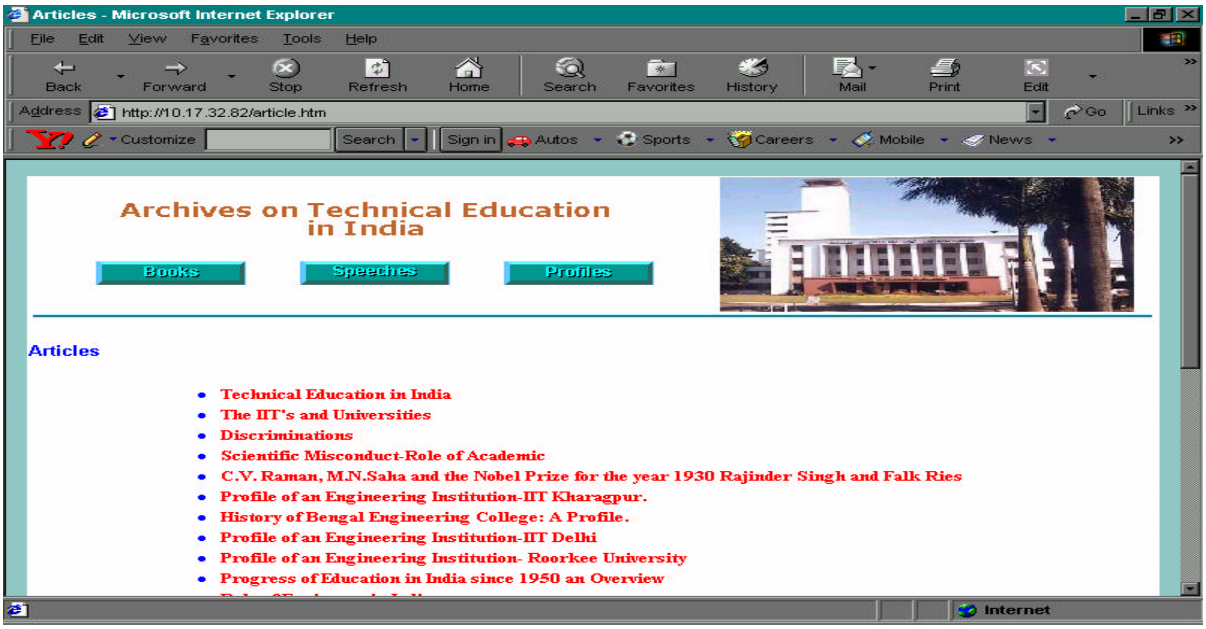

Fig 3 : List of Articles Database

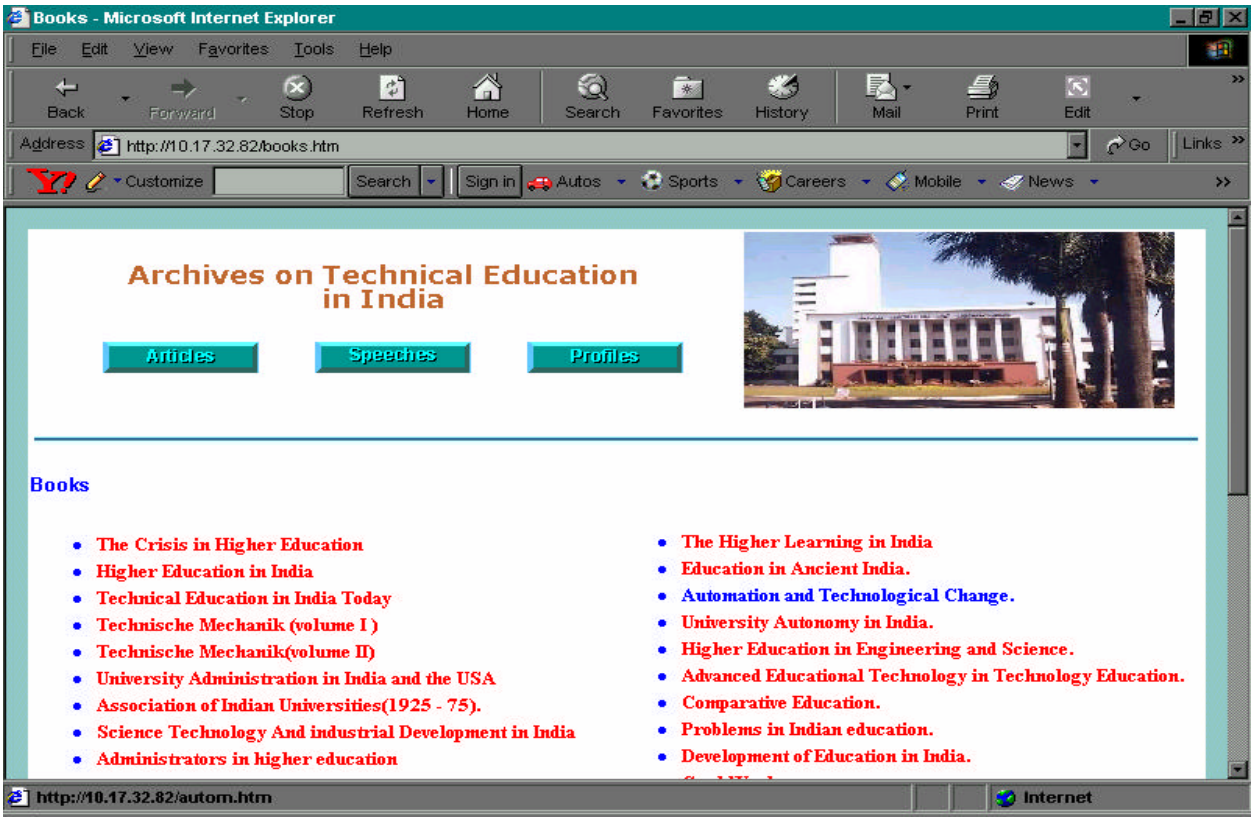

Fig 4 : List of Books Database

05 Conclusions

This presentation is a reflection of our practical experiences in exploring of Microsoft NT technology towards handling to-day's popular demand from the learned user community related

to E-Database. The significant change faceable with an electronic document is that if the user has to call the server (IP address of the server) to gain access to it remotely, view it, download it, print it and, in brief, "use" it. It may be take in consideration how much simpler and quicker it would be if librarians and, even better, library users could obtain their own interlibrary "loans" (now, technically, copies or excerpts) on a self-service basis, requiring the tolerance but not the time or energy of the staff of the library from which it is obtained. This change would be rather like the change from having closed library stacks, in which library employees had to fetch each book for users, to open stacks in which library users could obtain and examine books by themselves. Similarly, in the Electronic Library, library staff would be mainly concerned with creating and sustaining the system so that users could serve themselves. Self-service, however, is a mixed blessing. It also assumes standardized, intelligible procedures, presupposes some expertise on the users' part, and may make it less easy for the service providers to know what is going well and what is not going well. Yet it may be the only affordable way to support large-scale library use.

But here we want to put our view regarding a burning question, what of the older materials on paper that occupy so many miles of libraries' shelves? Libraries have undertaken a major, systematic effort at the retrospective conversion of older catalog records from cards to electronic records. What of the retrospective conversion of the texts of older paper documents themselves? Present technology permit us to convert the same in to digital format easily. Therefore, it is technically feasible that very substantial amounts of older as well as future library material may be converted to E-Database form when the Storage costs for electronic documents are decreasing steadily, while the building costs for storing paper documents are not.

05 References

1. FREDERICK (G. KILGOUR) . Beyond Bibliography. London: British Library, 1985.

2. TANENBAUM (A.S). Computer Networks. New Delhi ;Prentice Hall of India ;1986, p.596

3. DELISLE (N. M). and SCHWARTZ (M. D). Contexts--A Partitioning Concept for Hypertext, ACM Transactions on Office Information Systems 5 (1987): 168-86.

4. MICHAEL (BUCKLAND K). Library Materials: Paper, Microform, Database, "College and Research Libraries 49:(1988): 117-22.

5. BRAIN (SHACKEL).Information Exchange within the Research Community. Information Technology and the Research Process: Proceedings of a conference, Cranfield, 1989, 147-62.

6. MAJUMDAR (KAMALENDU). Social Care Information System (SOCIS) for Information Age- FID ’98 Article No – 0202119 pp-II-1-5