Lib Sc 22; 1985; Paper X

INFORMATION TECHNOLOGY AND ITS IMPACT ON INFORMATION RETRIEVAL SYSTEMS

M A GOPINATH, *Documentation Research and Training Centre, Indian Statistical Institute, 31 Church Street, Bangalore 560 001.*

Developments in Information technology are delineated. Impact of these technologies in various aspects of information storage and retrieval systems are analysed. The total picture of information technology and conducive use in information dissemination are also presented.

1 Introduction

Libraries have been repositories of a variety of documents and process them for use for a clientele such as scholars, scientists, engineers and other professionals, and in larger context to any citizen. The documents have been a carrier of information or knowledge packages to be mediated to users. Thus, documents have been a kind of physical forms which can facilitate recording, storing, and transmitting of knowledge. The annual production of articles has been estimated to be about 500,000 (Chemistry), 350,000 (Medicine) 350,000 (Biology) 85,000 (Physics) 85,000 (Engineering) and 25,000 (Psychology) (4). It is obviously not possible for any specialist to keep track of everything published in his field. This fact has made way for production of abstracting and indexing periodicals such as the Chemical Abstracts. Even to handle the production of these secondary periodicals, it is no longer feasible to have manual methods. We have to have machine-readable records. At present, printed version lives side by side with electronic version.

According to F W Lancaster (5), "We are at present in an interim phase in the automation of science communication. This interim phase is one in which machine readable data bases exist side by side with printed data bases. The computer is used to produce a conventi-

onal printed publication which is also distributed conventionally. But this must change. At some date in the near future, there will begin a natural cross over from electronic production of print to electronic publication and dissemination (i.e. to the paperless mode of operation). The same evolutionary process will apply to the production of primary publications, but the evolution in primary publication will log some years behind the development in secondary publication. By the year 2000, it seem entirely reasonable to suppose that formal communication in science and technology will be almost exclusively electronic, and that a substantial move to machine-aided informal communication will also have occurred. Such changes seem to be simply a matter of economic necessity" (2). Thus a scenario, which appears to be slowly gaining momentum as the current developments indicate. The growth rate as calculated by Hall and Brown (3) is as follows: It was 250,000 entries in 1968, 3,00,000 in 1972, 24 million in 1976, 70 million references with an annual addition of 10 million references. According to Hall and Brown again, Applied Sciences have 34 million, Pure Sciences about 22 million, Agriculture 5 million, Philosophy and Psychology 635,000 references, Religion and Theology 50,000 references, Linguistics and Language 300,000 references, Arts/Recretion/Music 220,000 references, Literature 300,000 references, and History, Geography, Biography 380,000 references. However, such a vast input into on-line information storage and retrieval does not cover entire spectrum of knowledge. There are still many fields in which the electronic memory is yet to pick up. But as Hall and Brown say "The reality of the present database industry is that the very existence of the present enormous on-line knowledge pool is a staggering intellectual and technical achievement, accomplished in less than 20 years. It has required the combined efforts of thousands of people in fields ranging from computing and programming to surface and satellite communications technology. To this, must be added truly enormous amounts of effort poured into the production of primary databases, whether on-line or paper-based. The existence of the on-line knowledge pool is a remarkable tribute to the small number of original on-line pioneers, and to the even smaller number of far-seeing financial sponsors... It is arguable that of all the information retrieval methods available only one on-line retrieval can closely approach

the desirable ideal of putting the enquirer instantly in touch with a substantial part of man kind's collective memory". These fascinating developments in information retrieval has been due to the impact of Information Storage Technology and the logic of search and retrieval of information.

2 Contents of Information Technology

'Information Technology' connotes an ensemble of technologies. They particularly cover the computer's capability to store and process information, known as information-processing and the telecommunication technology which are capable of transmitting information to distances. The breakdown of these technologies is presented by James Williams in his paper "Information Technology—A state of the art (7). He presents the data in terms of the following six major technologies that appear to be relevant to libraries.

(1) Processors, memory, and input/ouput channels;

(2) Micro, mini and large-scale computers;

(3) Mass storage technologies;

(4) Data communication, networking, and distributed processing;

(5) Data entry, display and responds technology;

(6) Software.

The detailed development of these technologies are provided in the text of the Williams state-of-the-art mentioned earlier. We will discuss the impact of these technologies in information retrieval systems.

3 Information Retrieval Systems

The primary purpose of information retrieval systems (IRS) is to provide information to the searchers of information. For this purpose, the IRS has to acquire information that have potential utility for the users of IRS, process this information into a storage system, methodology for searching and provide a display of the retrieved information and transmit the same long distances. The specific steps can be stated as follows:

1. Understanding Users Needs
2. Acquiring the needed information
3. Creation of Data Base

3.1 Data Structure for storage and retrieval

3.2 Coding and compression of Databases

4. Searching and Search Strategies

5. Display and communication

31 Understanding User Needs

It is hardly necessary to emphasize that this is a crucial step. Modern developments in survey and interview techniques has helped to ascertain the information seeking behaviour of the clientele of information retrieval systems. The development of computer-technology has helped us to store the profile of users of information in detail as well as uptodate.

32 Acquiring Information

Acquisition of information calls for the identification of the existence of information its procurement of the same, and incorporating it into the system. The identification and procurement is facilitated by the large on-line databases such as Books-in-print and several secondary services in the on-line. Library automation, particularly acquisition control and serial control techniques, have provided quicker pace and better efficiency for document acquisition.

33 Creation of the Database: Storage Media

The creation of database calls for the storage of information. Print-on-paper media storage is well known. We will discuss the other media for storage. The storage technologies include the following:

(a) Magnetic Storage
Tape/Disk storage
Floppy Disks
Winchester and removable hard disks

(b) Optical Storage
Optical Digital Disks
Digitally Encoded Optical Videodisc
Digitally Encoded Compact Disks
Erasable Optical Storage

331 Magnetic Disk Storage

Magnetic Disk Storage is still the predominant technology for computer information storage. The fundamental physical elements are the magnetic surface and the read/write head(s). Magnetic tape was one of the earliest form of storage. The tape provides only serial access wearout. Hence Disk/Drum (rotating) have random accessibility. Improvement in Disk/Drum is in three directions:

(a) increasing the number of tracks per inch through improvements in head technology;

(b) increasing the bit density along a track by the development of new and improved magnetic materials; and

(c) decreasing the head-to-surface gap.

Further developments in the operational aspects of magnetic disk technology are:

1. Rotational position sensing which frees the data channel while the heads are positioned correctly for a read/write operation;
2. Low-cost fixed head storage such as IBM 3340 and 3350 which provide several cylinders of rapidly accessible data;
3. Switched read heads which permit a number of physical disk cylinders to be read without head motion;
4. Improved servo control;
5. Improved surface coating using thin film technologies and thin film media.
6. Increased track and bit densities.

Floppy disks or flexible disks has been adopted in microcomputers. The floppy disk is either a 5¼ inch or 8 inch platter that can be recorded on one side or both sides. The read/write heads on a floppy disk actually contact the surface of the disk, so that wear has always been a problem. Floppy has today one megabyte and access time is about 100 milliseconds. Technological improvements have brought in:

1. Higher densities—a new 5 megabyte floppy is available with the posibility of a 10 megabyte drive by 1984;

2. Longer-life new materials such as Al_2O_3 abrasive, and thermally stable magnetic coatings. New lubricants and surface finishing also add to durability;

3. Faster access time-improvements in the mechanical motion and head positioning permit access times to be halved.

Microfloppies have less than 4 inches format. The Sony microfloppies range from 3 inches to 3.8 inches. It incorporates a hard plastic jackets. It is likely to have 1.6 MB equal to that of 8 inch floppies. The smaller microfloppies have improved bit densities and can rotate at a higher speed (upto 600 rpm) than their bigger brethern.

Winchester Disk drives consists of hermetically sealed entire disk peak and heads in a single, non-separable enclosure. In each such unit, head to surface gaps are down to tenths of microns, allowing ever greater storage densities. When not rotating, the lighter Winchester heads rest on lubricated disk surface. As the drive is turned on and comes up to speed, the heads rise on a cushion of air caused by the spinning disks. Such disk are cost beneficial. Mini Winchesters are of 5¼ inches, while large ones are 14 inches one. Mini Winchesters range from 5 to over 300 MB. They are less flexible for handling than microfloppies. However, in recent year Winchester Disk catridges have been developed which are removable. They are in the range of 3.9 inches (100 mm) disk drums with capacities from 6.38 to 38.2 MB including a removable disks.

332 Optical Storage Devices

Larger scale mass storage of information for both micros as well as maxis is the optical disk technology. They are of three types Optical Digital Disk (ODD), the digital encoding of information on optical videodiscs (DE/OVD), and finally the storage of digital information on the digital audio compact disc (DE/CD). Apart from disk format, there are other categories of optical recordings such as microform, aperture cards, ultrafiche, ultrastrip and video disk.

Optical recording has several advantages over magnetic recording. Because of the use of laser beam, there is no need to maintain closeness between the reading and disk surface. There is no track width specification. In ODD, it is the order of wavelength of light which, in the visible range, is just under a micron, and decreases to less than one-tenth of a micron in the ultra-violet. Furthermore,

while the ability to read a magnetic signal is dependent upon the magnetic energy in the media, in ODD the media is passive and the energy is in the light beam.

Holography is another optical system. It is the process of recording an reconstructing a wave front of light patterns. A hologram is the record resulting from the holographic process. Holography is like photography in that an image is recorded on film but it offers many advantages including three-dimensionality, that traditional photography cannot.

34 Data Structure

Data Structure for storage of information for retrieval purpose, calls several models. such as:

1. Sort Tree Structure;
2. Dictionary Storage Using Character tree;
3. Table structure to allow truncation specification;
4. Invested File Structure;
5. Scatter storage;
6. Stack structures;
7. Representation of Queues;
8. List storage structure;
9. Dynamic storage.

These storage structure have varied values to different situations.

35 Vocabulary Characteristics

The vocabulary control in the form of controlled input, based on a system language such as a thesaurus is a useful device. The study of vocabulary characteristics such as vocabulary frequency, distribution of term lenghts, vocabulary growths etc are important factors in design of the database. The varied impact of these factors in Information Retrieval depends on the subject field, size of the references or texts etc.

36 Coding and Compression of Data Bases

One of the method of reducing the size of a dictionary and sequential file for storage purposes is to use coding as a method of compression of character length in database. The dictionary would

code and decode whenever the input/output process arises. The development of the codes must be based on a sound theory.

37 Searching and Search Strategies

Searching is an important function in an Information Retrieval Systems. It involves the following steps:

1. Clarifying and Negotiating the Information Need and Search Objective.

Interviewing the information seeker clarifies the narrative form of the request determines search objectives:

(a) Retrieve all relevant items

(b) Retrieve only relevant items

(c) Retrieve some relevant items.

2. Identifying relevant data bases

Determining which online data base to use first, which next, and so on.

3. Formulating Basic Search Logic:

Planning Search Strategies

Analysing the search topic into parts called facets or concept groups. Planning approaches to search strategy for combining concepts of topic.

4. Compiling Search Terms

Choosing index terms from the database's thesaurus or other printed word lists. Selecting terms for free text searching of the subject conveying fields (title, abstract etc). Deciding to use thesaurus and alphabetic word lists online.

5. Making Choices

Limiting and printing output (offline and online).

Selecting an approach to search strategy which best satisfies the search objectives expressed by the information seeker.

6. Conceptualising the Search as Input to the Retrieval System.

Arranging the search terms into concepts or facets for search strategies using features of the retrieval systems for example, truncation, word proximity.

Noting most important and less important concept groups and deciding on sequence input to access these concept groups efficiently.

Retstricting or limiting output based on search objectives.

7. Evaluating Preliminary Results

Reviewing search results, step by step.

Considering alternative search strategies to meet search objectives.

8. Evaluating Final Results

Determining information seekers satisfaction with search results.

38 Display and Communication

Communication plays an important role. The display of wanted information to users at the end of a search has enriched values to the users. The form and format of representation of information to the intended users is critical for an information system. The technological developments in this area began as line printer, has given way to alphanumeric and graphic displays using a CRT (cathode ray tube) devices. These are like the home TV. Newer display methods have even given up keyboard methods. The interactor has only to touch the display screen. The Optical Mark Code Readers (OMR), Bar Code Readers (BCR), Optical Character Readers (OCR) and voice recognition units (for audio-communication) have been developed as dynamic inputting and outputting devices. Today, even graphical data can be digitised as input data and displayed at various manipulations. Technological developments in the areas of graphics, digitizers, videodisk, and computer input from micromedia (CIM), along with digital coding, analog-to-digital conversion, and data compression techniques are beginning to impact on this area. The Liquid Crystal Display Technology, and Light Emitting Diodes can be used in limited display situations. Using the home TV as a display device, for an interactive computer-based information system is feasible and possible. It can be a teletext terminal or a display unit for micropro-

cessors. Development of on-line communication device, telematics, and satellite communication technologies can carry information to anywhere. Global Distance is no longer a factor for physical transfer of information.

4 Microcomputers and their Use in Libraries

Many of the properties of an IRS can be effectively minaturised in a Micro-computer. Microcomputers are also called as home computers or personal computers. It is used for a small volume of data processing, small data base management systems, software development, low-cost computing, personal computing etc. The typical microcomputer system costs about $ 4,000 to $ 6,000. They are desk top computers. Microcomputers include software devices which is sufficient to permit the layman to gain productive use of the system simply by plugging it into the wall outlet. J. H. Katayama says that "Special librarians are often faced with micro-budgets, insufficient staff, and mini-collections but with macro demands for services, the microcomputer can play a significant role in abbreviating many, but not all, of the problems" (1).

The library functions in relation to microcomuter application can be divided into four types: (a) Communication; (b) Word Processing; (c) Administrative; (d) Data Base Management.

As a communication device, microcomputer can act as an input and output terminals to external data base. It can download the data relevant and retrieved from the databases and edited to reformat the information to the target audience. However, we need modems and communication software for this kind of interaction.

As a word processor, microprocessors can effectively replace a typewriter. It can store, process, reformat, and hold a variety of configuration of data-structure. With this, it is easy to produce and transmit a variety of services such as data on periodical holdings, policy and procedure manuals, and updating of several rules and procedures in the library.

Microcomputer can of course relieve the library staff from clerical routines. These have been discussed by I. K. Ravichandra Rao, in his Seminar on Library Automation (6)

As database management system a computer can store a variety of information. It helps to sort, reorganise, reformat a test to the required needs without much problems. One can effectively produce reports from collection of abstracts of a variety of media such as chapters of books, periodical articles, video tape recording, computer program receipts, or any other media and software to produce good state-of-art for a subject or a mission, or any other kind of information analysis and consolidation products.

Thus, microcomputers, act as a mini-library or information retrieval systems. They can be used independently or they can also networked for wider communication.

5 Computer Networks, Databases, and Document Delivery

From the document delivery point of view, the on-line system to a database management system would be a welcome. A user can order document from his terminal and the document can be sent to him electronically. On-line access to libraries book holdings will become common place. The automatic circulation data will provide this quite regularly, Some studies on the effect of on-line networks significantly increased inter-library loan activity. In on-line circulation systems the status of each items availability is current at all time. The system also eliminates paper work. The computerised text storage can also provide translated version of foreign languages through machine translations. This will reduce the time lag in overcoming language barrier. Thus, translated documents can be provided in a short span of time. However, the machine aided translations of language may need editorial touch ups, it is already continously improved for automated on line editing.

6 Developments in Communication Technology

Telecommunications has achieved impressive advances in recent years, channel capacities, reliability and error rates have improved dramatically concurrent with major cost reductions in both switching and communication means such as coaxial cables, optical fibres, microwave laser beams and satellites. The satellites can transmit information for high speed and bulk transmission of data over a long distance. More complex information such as data images and documents can be transmitted and received via satellite. With the possi-

bility of a satellite carrying nnearly 6,00,000 telephone channels, an universal satellite communication system is expected by the end of the century.

The developments in the optical fibre communication, glass fibres are used to convey pulses of light energy packets as signals. The media offers very wide band widths and thus high transmission rates and low losses. A single hair-thin optical fibre measuring 0.125 mm in diameter can transmit 10,000 telephone circuits or several hundreds of TV channels. The optical fibres will become a commercial reality before the end of the century. The laser beams having a high frequency can transmit one lakh times more information than a microwave.

Several numerical and non-numerical data bases have emerged in electronic data storage. The entire text of wellknown reference books, whose future editions are going to be computerised with on line transmission facilities would help fast transmission of information. Even the encyclopaedias are likely to be put into computer data storage, for updating, editing and transmitting information for on-line interactive transmission to likely users. The directory type information has to be transmitted fast to the users. Several on-line directories are in use in India itself.

7 Computer Communication in Document Creation and use

The impact of this ever-renewing information technology on communication media and information can be summarised in the following manner:

Work	Technology	
	Old	New
1. Communication	Personal travel Postal communication	Teleconferencing Teletex Satellite transmission
2. Inputting text (primary format)	Oral presentations Writing and Typing	Word processing Optical scanning through computers

Work	Technology	
	Old	New
3. Multiplication	Printing and Near printing process	Computere visuals Display Terminals Videodiscs and Cassettes, Computerised photoprinting
4. Storage	Book shelves Pamphlet cabinet Microform storage devices	Computer-based Digital and Analogue storage. Discs, tapes and drum storages
5. Information Retrieval	Browsing through the surrogates in a catalogue and through the shelves	Browsing through on-line terminals compterised-Data base softwares

Note: The data presented in this table must be interpreted to project a transition phase. In the total human civilisation we always see old coexist with the new technologies.

8 Development In India

Information technology has begun its advent in India. Satellite transmission technology and teletransmission technology have become almost part of home life. There are about 1000 computer installations in various research centres, industrial establishments, government departments. Many of the technical information centres in India utilise on a job basis the computer-communicatin facilities. Exploratory work on utilisation of computer for preparation of directory type of information and bibligrophic information retrieval services have been done by INSDOC. INSDOC has at present several different kinds of directory. Of particular interest to us is the Directory of Periodical Holdings in special libraries and university libraries in India. Experiments have also been made on the use of online transmission of information through bibliographic database from Europe and USA. In 1976, an interesting project linking to world sources of information was conducted at the TIFR Bombay and it was sponsored by several organisations under the initiative of INS-

DOC. For this purpose. INTELSAT communication satellite was used. In 1977, a similar attempt was made at IIT, Madras to connect the data bases of Systems Development Corporation and the DIALOG of Lockheed. This was followed in 1982, when the Computer Society of India organised demonstration of on-line interaction with ESRIN Database of Frescati, Rome in Delhi and the National Aeronautical Laboratory, organised a similar interaction at Bangalore. These experiments generated lot of awareness among specialist users of information in India.

The INSDOC also pioneered the computerised SDI facilities. It began with CAN/SDI software and the databases like chemical abstracts, INSPEC tapes and Compendex tapes were used. This project provided an experimental base for the study of SDI services in India. The current Science Information Centres, Programme under UGC located at the Indian Institute of Science, Bangalore, and Roorkee University, Roorkee have in addition to the SDI services, a document delivery system for users. Almost all special libraries and university libraries are today provided with reprographic and micrographic machines. These developments although are not well-integrated and continuous ones, they can be energised by good financial backing by the governments, institutions and the professional bodies in India.

The National Informatics Centre has developed Information Transfer Systems and are planning to develop a National Data Centre. These projects should also be strengthened to provide good document delivery system in India. In fact, India should have a good network for document transmission system for a national hook up of Telefacsimile Transmission of Text of Documents. Such a service would help not only libraries but more the media centres such as the daily newspapers. Investment on such a network would certainly improve the pace of development in our country.

9 Conclusion

The potential of information technology for document delivery service is enormous. Means and methods for overcoming legal and organisational hurdles have been suggested and pursued. The Gutenberg revolution brought in the book as a predominant media for knowledge communication. Today, the electronic telecommuni-

cation media is pushing fast knowledge through variety of its media. It has become a space and time binder. Thus, utilisation of knowledge through its instant transmission has added value and variety to our life. Here, I quote Richard De Gennare:

> "What is needed, and what is being developed and implemented, is a new library technology based on electronics as well as fundamental restructuring of traditional library goals, relationships, and dependencies. This restructuring will force libraries to undergo a major transformation in the coming decade".

10 Bibliographical References:

1. CHEN (Ching Chin), Ed. Microcomputers in Libraries. 1982. (Neal Sehuman. P127).
2. De GENNARO (R), Research Libraries enter the information age. (Library Journal. 104; 1979; November).
3. HALL and BROWN, On-Line bibliographic data base: International Bibliography, Ed. 1981. ASLIB).
4. HATRICS, Notes on Information retrieval.1979.
5. LANCASTER (F W), Towards paperless information systems. 1980. New York, Academic press.
6. RAVICHANDRAN RAO (I K), Library automation, (DRTC Refresher seminar. 1983).
7 WILLIAMS (JAMES), Information technology — a state of art report. (Kent (A) and Galrin (T J). (informatin technology 1982 P342-458).