

# CONTENTS

<b>Units</b>		<b>Page No.</b>
1.	Introduction to Information Technology and Applications	1-56
2.	Library Automation and Information Services	57-85
3.	Online Public Access Catalogue (OPAC)	86-113
4.	Database and Networks	114-158
5.	Digital Libraries	159-187

# SYLLABUS

## INFORMATION COMMUNICATION TECHNOLOGIES : APPLICATIONS

Mlib-06

### UNIT – I

**INTRODUCTION TO INFORMATION TECHNOLOGY:** IT–Origin, Development and Applications–Computer Systems: Hardware, Software, Data Communication–Data Processing and Programming–Concepts.

### UNIT – II

**LIBRARY AUTOMATION AND INFORMATION SERVICES:** Library Application software in General–Library Application Software: Acquisition & Cataloguing.

### UNIT – III

**ONLINE PUBLIC ACCESS CATALOGUE (OPAC):** Library Application Software, Serials Control & Circulation–Use of computers in Information Services.

### UNIT – IV

**DATABASES AND NETWORKS:** Introduction to Databases–Bibliographic Databases and Search Techniques–computer Networks–Library and Information Networks.

### UNIT– V

**DIGITAL LIBRARIES:** Introduction to Digital Libraries–Digitization of Library–Materials–Metadata–management of Digital Libraries.

### References

1. **A Manual of Library Automation and Networking**, N.R. Satyanarayana, New Royal Book Co, 2003
2. **Software Packages for Library Automation**, Vasantha N. and Mahesh V. Mudhol, Ess Ess, 2000 .
3. **New Media Technology and Communication**, B.K. Chaturvedi and S.K. Mittal, Global Vision, 2010
4. **Library Networking**, Edited by S.P. Singh, Omega Pub, 2008

---

# UNIT I INTRODUCTION TO INFORMATION TECHNOLOGY AND APPLICATIONS

---

## ★ STRUCTURE ★

- 1.1 Introduction
- 1.2 Origin and Development of Information Technology
- 1.3 IT:General Applications
- 1.4 Concepts and Terminology of Computer
- 1.5 Organization (Structure) of a Computer System
- 1.6 Software Basics
- 1.7 Data Communication and Networking
- 1.8 Data Communication
- 1.9 Programming and Data Processing
- 1.10 Summary
- 1.11 Review Questions
- 1.12 Further Readings

---

## LEARNING OBJECTIVES

---

After going through this unit, you should be able to:

- define technology and purpose
- discuss the history of computers, characteristics of computers
- explain the functioning of CPU
- define various types of printers
- discuss various types of transmission media
- define various switching techniques
- define different types of programming languages.

---

## 1.1 INTRODUCTION

---

Technology has been defined as "systematic knowledge and action, usually of industrial processes but applicable to any recurrent activity". In providing tools and techniques for action, technology at once adds to and draws from a knowledge base in which theory and practice interact and compact. At its most general level technology may be regarded as definable specifiable way of doing anything. In other words, we may say a technology is a codified, communicable procedure for solving problems.

Technology, Manfred Kochen observed, impacts in three stages. First, it enables us to do what we are now doing, but better, faster and cheaper; second, it enables us to do what we cannot do now; and third, it changes our life styles.

## NOTES

Information technology is a recent and comprehensive term, which describes the whole range of processes for generation, storage, transmission, retrieval and processing of information.

---

## 1.2 ORIGIN AND DEVELOPMENT OF INFORMATION TECHNOLOGY

---

### Early History of Computers

#### *Computers: Then and Now*

Let us discuss the history of computers because this will give the basic information about the technological development trends in computer in the past and its projections in the future. If we want to know about computers completely then we must start from the history of computers and look into the details of various technological and intellectual breakthrough. These are essential to give us the feel of how much work and effort has been done to get the computer in this shape.

The ancestors of modern age computer were the mechanical and electromechanical devices. This ancestry can be traced as back and 17th Century, when the first machine capable of performing four mathematical operations, viz. addition, subtraction, division and multiplication, appeared.

### Mechanical Computers

The very first attempt towards this automatic computing was made by Blaise Pascal. He invented a device which consisted of lots of gears and chains and used to perform repeated addition and subtractions. This device was called pascaline. Later many attempts were made in this direction, we will not go in the details of these mechanical calculating devices. But we must discuss some details about the innovation by Charles Babbage, the grandfather of modern computer. He designed two computers:

**The difference engine:** It was based on the mathematical principle of finite differences and was used to solve calculations on large numbers using a formula. It was also used for solving the polynomial and trigonometric functions.

**The analytical Engine by Babbage:** It was general purpose computing device which could be used for performing any mathematical operation automatically. It consisted of the following components:

**The Store:** A mechanical memory unit consisting of sets of counter wheels.

**The Mill:** An arithmetic unit which is capable of performing the four basic arithmetic operations.

**Cards:** There are basically two types of cards:

- (a) **Operation Cards:** Selects one of four arithmetic operating by activating the mill to perform the selected function.
- (b) **Variable Cards:** Selects the memory locations to be used by the mill for a particular operation (i.e., the source of the operands and the destination of the results).

**Output:** Could be directed to a printer or a card punch device.

NOTES

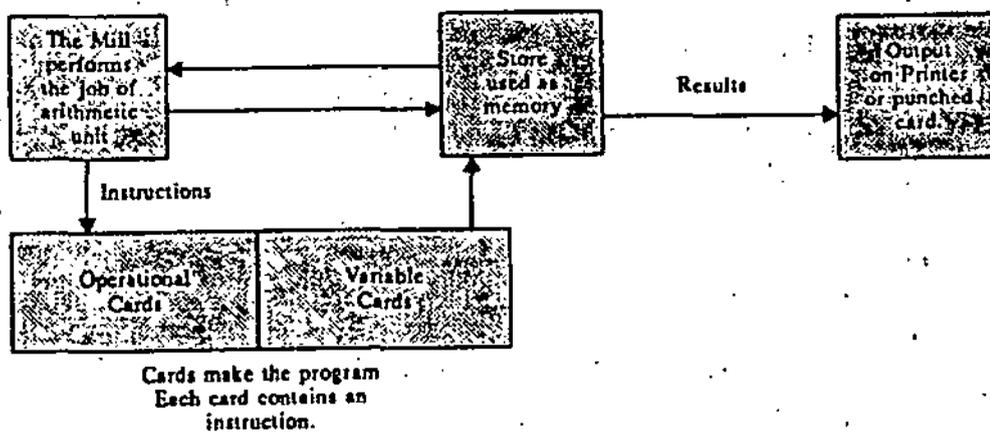


Figure 1.1 : Logical Structure of Babbage's Analytical Engine

The basic features of this analytical engine were:

- It was a general purpose programmable machine.
- It had the provision of automatic sequence control, thus enabling programs to alter its sequence of operations.
- The provision of sign checking of result existed.
- Mechanism for advancing or reversing of control card was permitted thus enabling execution of any desired instruction. In other words, Babbage had devised a conditional and branching instructions. The Babbage machine is fundamentally the same as modern computer. Unfortunately Babbage work could not be completed. But as a tribute to Charles Babbage his Analytical Engine was completed in the last decade and is now on display at the Science Museum at London.

Next notable attempts towards computer were electromechanical Zuse used electromechanical relays that could be either opened or closed automatically. Thus, the use of binary digits, rather than decimal numbers started.

### Harvard Mark I and the Bug

The next significant effort towards devising an electromechanical computer

## NOTES

was made at the Harvard University, jointly sponsored by IBM and the Department of UN Navy, Howard Aiken of Harvard University developed a system called Mark I in 1944. Mark I was a decimal machine.

Some of you must have heard a term called "bug". It is mainly used to indicate errors in computer programs. This term was coined, when one day, a program in Mark I did not run properly due to a moth short-circuiting the computer. Since then, the moth or the 'bug' has been linked with errors or problems in computer programming. The process of eliminating error in a program is thus, known as 'debugging'. The basic drawback of these mechanical and electromechanical computers were:

- Friction/inertia of moving components had limited the speed.
- The data movement using gears and liner was quite difficult and unreliable.

The change was to have switching and storing mechanism with no moving parts and then the electronic switching technique "triode" vacuum tubes were used and hence born the first electronic computer.

## Four Generations of Computer Development

### *First Generation Computers*

It is indeed ironic that scientific inventions of great significance, have often been linked with supporting a very sad and undesirable aspect of civilisation *i.e.*, fighting wars, Nuclear energy would not have been developed as fast, if colossal efforts were not spent towards devising nuclear bombs. Similarly, the origin of the first truly general purpose computer was also designed to meet the requirement of World War II. The ENIAC (the Electronic Numerical Integrator and Calculator) was designed in 1945 at the university of Pennsylvania to calculate figures for thousands of gunnery tables required by the US army for accuracy in artillery fire. The ENIAC ushered in the era of what is known as first generation computers. It could perform 5000 additions or 500 multiplications per minute. It was, however, a giant machine, occupying a number of rooms, and needed a great amount of electricity and emitted excessive heat. The main features of ENIAC can be summarised as:

- ENIAC was a general purpose computing machine in which vacuum tube technology was use.
- ENIAC was based on decimal arithmetic rather than binary arithmetic.
- ENIAC needed to be programmed manually by setting switches and plugging or unplugging. Thus, to pass a set of instruction to computer was cumbersome and time-consuming. This was considered to be the major deficiency of ENIAC.

The trends which were encountered during the era of first generation computer were:

**NOTES**

- The first generation computer control was centralised in a single CPU, and all operations required a direct intervention of the CPU.
- Use of ferrite-core main memory was started during this time.
- Concepts such as use of virtual memory and index register (you will know more about these terms in advanced courses).
- Punched cards were used as input device.
- Magnetic tapes and magnetic drums were used as secondary memory.
- Binary code or machine language was used for programming.
- Towards the end due to difficulties encountered in use of machine language as programming language, the use of symbolic language which is now called assembly language started.
- Assembler, a program which translates assembly language programs to machine language was made.
- Computer was accessible to only one programmer at a time (single user environment).
- Advent of Von-Neumann architecture.

### ***Second Generation Computers***

The second generation computers started with the advent of transistors. A transistor is a two state device made from silicon. It is cheaper, smaller and dissipates less heat than vacuum tube but can be utilized in a similar way as that of vacuum tubes. Unlike vacuum tubes, a transistor do not require wires, metal glass capsule and vacuum, therefore, is called a solid state device. The transistors were invented in 1947 and had launched the electronic revolution in 1950.

The generation of computers are basically differentiated by a fundamental hardware technology. Each new generation of computer is characterised by greater speed, large memory capacity and smaller size than the previous generation. Thus, second generation computers were more advanced in terms of arithmetic and logic unit and control unit than their counterparts of first generation. Another feature of second generations was that by this time high level language were beginning to be used and the provisions for system software were starting.

One of the main computer series during this time was the IBM 700 series. Each successful number of this series showed increased performance and capacity and reduced cost. In these series two main concepts I/O channels, an independent processor for Input/Output and Multiplexor, a useful routing device were used.

### ***Third Generation Computers***

A single self-contained transistor is called discrete component. In 1960s,

NOTES

the electronic equipments were made from the discrete components such as transistors, capacitors, resistors and so on. These components were manufactured separately and used to be soldered on circuit boards which then can be used for making computers of the electronic components. Since computer can contain around 10,000 of these transistors, therefore, the entire mechanism was cumbersome.

Then started the era of micro-electronics (small electronic) with the invention of Integrated Circuits (ICs). Use of ICs in computer defines the third generation of computers.

In an integrated circuit the components such as transistors, resistors and conductors are fabricated on semiconductor material such, as silicon. Thus, a desired circuit can be fabricated in a tiny piece of silicon rather than assembling several discrete components into the same circuit. Hundreds or even thousands of transistors could be fabricated on a single wafer of silicon. In addition, these fabricated transistors can be connected with a process of metalisation to form logic circuits on the same chip they have been produced.

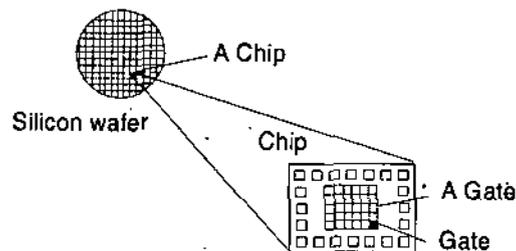


Figure 1.2 : Wafer, Chip and Gate

The major developments which took place in the third generation can be summarised as:

- IC circuits were starting to find their application in the computer hardware replacing the discrete transistor component circuits. This resulted in reduction in the cost and the physical size of the computer.
- Semiconductor (Integrated Circuit) memories were starting to augment fertile core memory as main memory.
- The CPU design was made simple and CPU were made more flexible using a technique called microprogramming, a scheme where programs controls the generation of central signals.
- Certain new techniques were introduced to increase the effective speed of program execution. These techniques were pipelining and multiprocessing.

The operating system of computers were incorporated with the efficient methods of sharing the facilities or resources such as processor and memory space, automatically.

## Later Generations

As discussed earlier with the growth in micro-electronics the IC technology evolved rapidly. One of the major milestone in this technology was the very large scale integration (VLSI) where thousands of transistors can be integrated on a single chip. The main impact of VLSI was that, it was possible to produce a complete CPU or main memory or other similar devices on a single IC chip.

This implied that mass production of CPU, memory etc., can be done at a very low cost. The VLSI-based computer architecture is sometimes referred to as fourth generation computers. Let us discuss some of the important breakthroughs of VLSI technologies.

### Semiconductor Memories

Initially the IC technology was used for constructing processor, but soon it was realised that same technology can be used for construction of memory. The first memory chip was constructed in 1970 and could hold 256 bits. Although the cost of this chip was high, but gradually the cost of semiconductor memory is going down. The memory capacity per chip has increased as: 1K, 4K, 16K, 64K, 256K and 1M bits.

### Microprocessors

Keeping pace with electronics as more and more component were fabricated on single chip, fewer chips were needed to construct a single processor. Intel in 1971 achieved the breakthrough of putting all the components on a single chip. The single chip processor is known as a microprocessor. The Intel 4004 was the first microprocessor. It was a primitive microprocessor designed for a specific application.

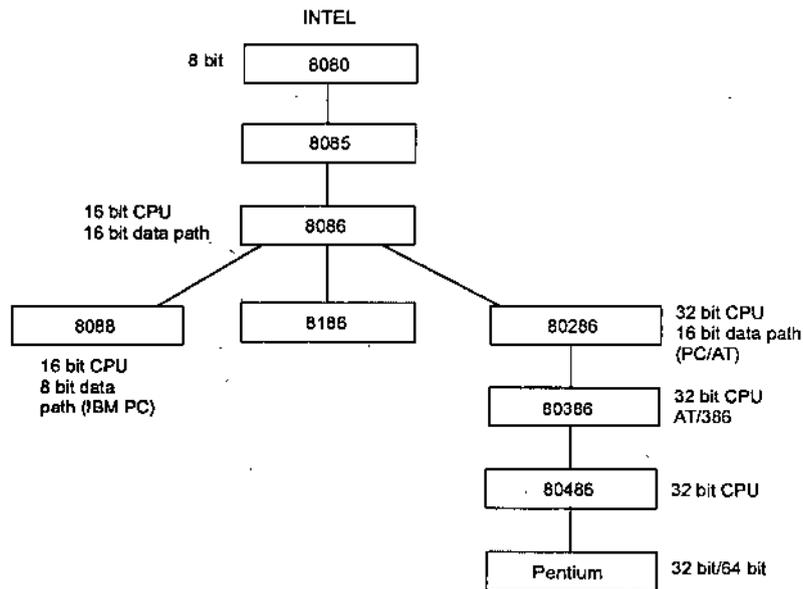


Figure. 1.3: Families of Intel microprocessor

Intel 8080 which came in 1974 was the first general purpose microprocessor. It was an 8 bit microprocessor. Motorola is another manufacturer in

this area. At present 32 and 64 bit general purpose microprocessors are already in the market. For example, Intel 486 is a 32 bit processor, similarly Motorola's 68000 is a 32 bit microprocessor. P6 which is announced by Intel 1995 can process 64 bit data at a time.

## NOTES

UNESCO defines Information Technology as "scientific, technological and engineering disciplines and the management techniques used in information handling and processing information, their applications; computers and their interaction with man and machine and associated social, economic and cultural matters". (Stokes)

This definition, while emphasising the significant role of computers, appears not to take into its purview the communication systems. It may, however, be stated that communication systems are as essential to information technology as computers. As a consequence, we have a convergence of **three strands** of technologies: **computers, microelectronics and communications**. In other words, a mosaic of technologies, products and techniques have combined to provide new electronic dimensions to information management. This mosaic is known by the name new information technology. It is important to bear in mind that information technology is **not just concerned with new pieces of equipment but with much broader spectrum of information activities**. Information technology encompasses such different things as book, print; reprography, the telephone network, broadcasting and computers.

In the following sections let us briefly consider the major components of information technology namely: computer technology, communications technology and reprographic and micrographic technologies.

### **Computer Technology**

Computer technology may conveniently be grouped into: processor technology, storage technology and software aspects.

#### ***Processor Technology***

Computers consist of electronic components assembled in a design or "architecture" that will perform necessary functions of input, output, computation and control (control of both the computer itself and of attached peripheral devices that perform input and output functions and store the files).

In the past, electronic components were expensive, so a minimum number were used in a single processor that alternately performs input, control, processing and output. Besides, the first generation of computers, operated by means of vacuum tubes or valves, were relatively bulky and energy consuming. The first major innovation, leading to microelectronics was the discovery of transistor, a product of solid state physics, which used semiconductor materials. The most important development of these today is based on the non-metallic element 'silicon'. Being much smaller than

NOTES

the vacuum tube, the transistor quickly replaced it in all electronic equipment. However, transistors and other equipment had to be wired together and a single piece of equipment might have thousands of such components. The wiring and assembly of such elements were a delicate and costly process. This naturally paved the way for research towards the concept of **integrated circuit (IC)**. At first IC's were simple but, as the technology developed, they rapidly became smaller and more complex. This led to the miniaturisation and refinement.

The central feature of microelectronics is the development of **micro-processor**, a special form of IC with functions of arithmetic, logic and control - similar to those of Central Processing Unit (CPU) of a computer and **contained in a single chip**. In addition, the microprocessor includes units to interpret instructions from the stored programme to supply the control memory the information necessary to retrieve instructions and send out data as required. The microprocessor is the **building block** from which modern computer systems are assembled. The microprocessor uses very little energy and has few environmental requirements of older machinery. Air conditioning, for example, might not be necessary for a general purpose computer using microprocessor technology. The practical significance of this is that it is now possible to bring the computer to the problem instead of bringing the problem to the computer. The aspects referred to above form a major hardware component of a computer.

### ***Storage Technology***

In the previous section, the recent innovations relating to the processing aspects of computer technology were discussed briefly. In considering some of the advances in devices for digital information storage, it may be stated that most of the primary storage in computers is now supplied by semiconductor circuits. There have been significant developments in memory technology affecting three areas of performance spectrum; the high speed, high performance; the midrange and the low speed bulk memory systems.

It is now possible that even a small computer system might have **cache memory**, a small associative memory retaining most recently referenced information and in a readily available place. In some cases, cache memory may be at the **top of a hierarchy of memories** having a wide variety of characteristics. Memory management, dynamic memory allocation, and virtual memory schemes, generally found in large computer systems, are now appearing on computers which are small and less costly.

The development of **Charge Coupled Devices (CCDs)** and **bubble memories** has filled the gap which previously existed in the continuum of memory devices such as fixed-head magnetic disks and these are slower than other semiconductor memories. These memories have advantage over magnetic disks in that they contain no mechanical parts and

could be used to store significant amount of information and can be treated as a structured file system.

There has been a continuous improvement in recording densities of magnetic media.

#### NOTES

Floppy disks and microfloppies provide a convenient media to store data. The development of video disk has added a new dimension to the information storage technology. Video disks could, be used to store large volumes of information in digital form. This kind of mass storage devices are believed to be very useful in the development of information storage and retrieval systems. It May be stated that all these innovations in storage technology provide us a variety of alternatives depending on the requirements of speed of operation. These developments add more capabilities to the storage aspects and may be considered advances in the storage technology.

#### **Software Aspects**

Software is a generic term covering the concepts, procedures and instructions which enable computer systems to do useful things. Usually, software is conceived in terms of computer programs, discrete units of software which make the computer to carry out specific tasks, and or systems or packages.

The importance of software is obvious, since it is the software which applies the power of the computer to solve the users' problems. Many of the users need a clear understanding of the capabilities of software more than hardware aspects. It is known for some years now that the "rapid increase in the capabilities of computer systems has not been matched by corresponding increases in the development and quality of software. This situation has caused much disenchantment with computer systems.

The methods by which computer software is produced have changed considerably in recent years with the emergence of "software engineering", which enabled improvements in programming practice, such as structured, or modular programming.

As one of the solutions to the software problem increased production and availability of packaged software is encouraged. Another solution to the problem is the use of fourth generation languages and flexible integrated software to produce prototypes of programmes to meet the user needs. It is hoped that these solutions would be able to meet fairly standard requirements. For the average user they mean that there will be an increasing number of packages to meet most of his needs.

---

### **1.3 IT : GENERAL APPLICATIONS**

---

During the last four decades, computers have revolutionised almost all disciplines of our life. Computers have made possible many scientific, industrial and commercial advances that would have been impossible otherwise. Computers are being used in many areas of applications viz.

business, industry, scientific research, defense, space, communications, medicine, education etc. In this chapter, we are discussing the utilisation of computers in different fields.

## **Role of Computers in Business**

Computers are widely used in business for processing volumes of data of an organisation. They are used in various areas of business functions such as Accounting, Inventory, Sales and Marketing, Manufacturing and Human Resource Development. They are widely used for automation of general office tasks. Let us discuss the role of computers in these areas.

### ***Accounting***

Accounting is the most important service activity in business. An organisation uses computers to maintain its accounting records in order to generate various financials statements and reports. A computerised accounting information system satisfies the information needs of management and other people. Managers use computers to access information about the organisation's assets, liabilities, revenues and expenses. Today, all financial accounting functions in most large organisations are computerised. Various financial statements viz. Trial balance, Trading account, Profit and loss account, Balance sheet, etc., and MIS reports viz. Cost analysis, Forecasting etc., are generated through computers. Computers are also used to calculate salaries/wages of employees and to generate payslips and payroll.

### ***Inventory***

Inventory is concerned with the stock of raw materials and finished goods available in the firm. The improper stock levels (low or high) cause many problems to the company. Therefore, maintaining of optimum level of inventories becomes critical for an organisation. Computers are used to maintain optimum level inventories in the firm. A computerised inventory control system generates purchase order, purchase book, inventory status reports, materials return report, materials transfer report and purchase analysis reports.

### ***Sales and Marketing***

Sales and Marketing departments use databases to store the names, addresses, telephone numbers, buying habits and other details of potential customers. Marketing Managers and executives use computers for the following purposes:

- Generating invoices and cash-memos
- Checking and executing customer orders
- Designing advertisements and promotional materials
- Mailing promotional materials, bills, reminders etc., to customers
- Maintaining customers credit histories

## **NOTES**

- Devising pricing and discount strategies of the products
- Planning and analysing the results of market surveys.

### ***Manufacturing***

#### NOTES

Computers are widely used in the manufacturing industries for:

- Designing products ranging from leather shoes to airplanes by using CAD (Computer Aided Designing) and CAM (Computer Aided Manufacturing) software.
- Generating production, planning and control reports
- Creating simulated versions of finished products such as cars, planes, medical equipment etc., using Virtual reality.
- Assembling and testing of products using Robots.

### ***Human Resource Development (HRD)***

- In the area of human resources, computers are used for:
- Selection and recruitment of candidates for jobs
- Training of the employees
- Salary and wages structures design
- Performance analysis of employees
- Data processing of routine personnel activities.

### ***Office Automation***

Office automation is the application of computer and related technologies like communication and networking to integrate the general office tasks so that the efficiency of office work is improved. Although all the work of a small or big office can be performed manually, but it is very difficult or even impossible today for an organisation to compete in the market without office automation. There are many essential requirements of today's office environment, which are listed below:

- To reduce cost of administrative overheads.
- To increase the efficiency of office tasks.
- To provide better service to the customers.
- To provide accurate information to the management.
- To provide the best and fastest way of communication.

The above requirements cannot be achieved without using latest technologies and therefore, office automation is needed for an organisation.

### ***Office Automation Systems***

Many types of functions are performed in an office. The basic functions, which are needed to be automated in any office are -

- (a) *Document Generation.* In all offices, many documents are needed to be prepared, typed and printed. Typewriters, computers and printers are widely used in automating this routine task of offices.

- (b) *Document Processing.* Documents are also needed to be processed in order to extract useful information required for MIS and other official purposes. Many office automation tools like word processing, desktop publishing, etc., are used to perform this task.
- (c) *Document Distribution.* All offices require an electronic distribution system for transferring documents and data within and outside the organisation. The main office automation tools for distribution of documents are Photocopiers, Teletax and Fax machines.
- (d) *Archival Storage.* The office documents are also needed to be stored for a long period, so that they can be retrieved when required. This task is achieved by the use of different storage devices like tapes, disks, etc.

For achieving the basic functions of an office, different types of office automation systems are used. These systems can be broadly classified into following four types:

- (a) *Document Management Systems.* These systems include computerised tools for generation, storage, processing and distribution of documents. Document management is the first important office task, which is needed to be automated. The commonly used office automation tools for document generation are typewriters, computers, printers and scanners. Photocopiers are widely used for preparing multiples copies of documents.

Computers have revolutionised the system of generation, processing and storage of documents in offices by a technology, called *word processing*. *Desktop Publishing (DTP)* is the another popular office technology, used widely for generation of documents using computers and laser printers. This office automation technology is used to produce a high-quality document for commercial printing. Using DTP technology, text and graphics can be combined into a single document that is printed generally on a laser printer.

Besides printers, there are some computer systems which are used to electronically capture, store, process and retrieve images of documents. These systems are called as *image-processing systems*. Optical scanners are often used as image-processing systems.

Archival Storage is generally very expensive and inefficient to store large volume of archival data on paper or on-line computers. Therefore, offices need efficient storage devices to store such data. The commonly used archival storage devices are *magnetic tapes, optical disks, floppy disks, hard disks and computer output microfilms*.

- (b) *Support Systems.* Certain support systems for managing the activities of work groups are also used in some offices. An automated office also needs certain systems that can help to manage the activities of work groups. These systems, which are actually

## NOTES

NOTES

software packages, are known as *office support systems* or *office automation tools*. Word processors, Spreadsheets and Database Management packages are also available as integrated packages generally called *Office Automation Packages/Software Tools* or *Office Suites*. MS Office and Lotus Smartsuite are two most popular examples of office automation packages (For details on Office Automation Packages refer Chapter 8–Software).

- (c) *Communication Systems*. These systems are used for sending messages, documents and data within and outside the organisation. We will discuss about these systems in the later part of the chapter.
- (d) *Teleconferencing Systems*. An electronic means of communication for conducting seminars and training programmes in an organisation is achieved through various teleconferencing systems. We will discuss about these systems in later part of the chapter.

### Role of Computers in Science

Computer are extensively used in various scientific research programmes and medicine for:

- Storing and analyzing scientific data collected from experiments and field work.
- Doing complex scientific calculations
- Representing data graphically
- Creating models and simulations
- Controlling various scientific devices
- Predicting earthquakes
- Forecasting weather
- Designing compounds for treatment of various diseases
- Analysing human DNA (a genetic molecule) for treatment of genetic diseases
- Maintaining database of patients' history
- Scanning body organs by CAT (Computerized Axial Tomography) scan and MRI (Magnetic Resonance Imaging) machines
- Doing various clinical tests of blood, urine, stool, etc., in a laboratory
- Providing help to physically disabled persons
- Controlling devices (e.g., pacemakers) implanted in human body
- Conducting nuclear tests without physical explosions
- To simulate potential military scenarios and determine defensive strategies
- To guide equipment from satellites to nuclear submarines
- To analyse satellite photographs for searching locations of missile

sites ... and so on, There is no end of this list as the use of computers in scientific research, medicine and defense is enormous.

## **Role of Computers in Education**

Computers have brought dramatic changes in the field of education. Today, almost all universities, colleges, institutes and school systems use computers for the following purposes:

NOTES

- To perform accounting functions like salary and fee calculations
- To prepare time-tables, date-sheets for examinations and question papers
- To prepare syllabus and course materials
- To generate documents like letters, circulars, memos and mailing lists
- To provide Computer Assisted Instruction (CAI) or Computer Based Teaching (CBT)
- To provide facilities to students for submitting applications for courses and examinations on-line through Internet
- To provide distance learning programmes through CDs and Internet
- To provide E-mail and Internet facilities to students
- To computerise Library information services
- To provide training through educational software and electronic textbooks.

## **Role of Computers in Entertainment**

Besides commercial, scientific, industrial or any other professional purpose, computers also provide entertainment to the user. Now a day, many special application software are available which can entertain you. The popular entertainment software available for PCs are:

- Computer games programs
- Graphics software
- Multimedia and animation software
- Internet web pages and chatting.

Using graphics, multimedia and animation software, you can draw pictures, make movies and games along with the audio. Computer games are not only popular among children but also among professionals. Some of the popular games are Prince, Chess, Cat, Kingkong and Bricks. Actually, there is a long list of computer games available in Windows and Internet. You can entertain yourself by browsing interesting web pages on Internet. You can also watch movies and some TV channels on Internet. People spend hours on Internet Cafe for chatting and making friends.

## Role of Computers in Data Communications

### NOTES

Computers provide data communication facilities to geographically separated offices through networking. They are used to transmit all forms of data and information, including digital data, voice, sound and video from one location to another over some form of transmission media. The major application areas of computers in the field of data communications and telecommunications are:

- Networking and client/server computing
- Internet
- Communication systems
- Teleconferencing systems

We shall discuss the role of computers in data communications in Chapter 12–Data Communications and Networking and Chapter 13–Overview of Internet in detail. However, we are providing below a brief overview of Communication and Teleconferencing systems.

### **Communication Systems**

Modern offices need computer-based message or communication systems to transfer messages and data rapidly from one location to another. Telex, Teletex, Videotext, FAX, EPABX, Workstations, E-mail, Internet/Intranet are commonly used communication devices/technologies. Though Telex is the oldest means, E-mail and Internet/Intranet are the most modern means of telecommunications. We will now discuss mainly computer-based communication systems.

**Videotext.** Videotext is the modern means of computer-based telecommunication. In videotext, the data is superimposed upon TV signals and the message is displayed on television sets. Teletext and Viewdata are two common types of Videotext, which differ from each other in their way of communication. In teletext, the message or information is sent out as pages to the receivers' teletext in single way and cannot be received back, while in viewdata, the users can send as well as receive messages.

**Workstations.** Any Personal Computer (PC) or minicomputer in a single or multi-user environment with a wide range of features is known as workstation. Workstations or computers are the essential part of an automated office.

**Electronic Mail.** Electronic mail (E-mail) is most widely used technology for sending messages or documents from one location to another by using electronic workstations or computers. E-mail services are either available within intra-office network (e.g., local area network) or through an outside vendor. Within an organisation, the employees use a workstation for sending their messages from one electronic mailbox to another. By using E-mail services from a vendor, the E-mail subscriber enters the message into the computer and addresses to the recipient's computer by quoting the E-mail code. The message is then transmitted through the modem to

the recipient's mailbox where the recipient can download it. (For more details refer Chapter 12—Data Communications and Networking and Chapter 13—Overview of Internet).

**Internet and Intranet.** Internet and Intranet are the most recent telecommunication technologies, which have brought a technological revolution not only in all offices but also in homes. Internet is the world's largest network of millions of computers all over the world connected through telephone lines. Intranet, on the other hand, is a network of computers within the closed perimeters of the office. (For more details refer Chapter 13—Overview of Internet).

### **Teleconferencing Systems**

Teleconferencing systems are the latest office automation technologies for conducting meetings of *widely separated people through a communication channel*. These systems enable people to communicate audio, video or image information in a conversation taking place between two or more locations. The teleconferencing systems are of three types—Audio Teleconferencing, Video Conferencing and Computer Conferencing.

**Audio Teleconferencing.** Audio teleconferencing is simply a conference phone calls system. With such type of conferencing, participants can only hear the voice and cannot see the participants. Audio teleconferencing is used in most of the companies, as it is the least expensive medium for conducting meetings among the participants sitting at far away places.

**Video Teleconferencing.** Video teleconferencing has completely changed the atmosphere of a modern office. With videoconferencing systems, the participants not only hear the voice but also see each other. The communication takes place either in one-way or multi-way modes. In one-way mode, which is also known as *point-to-point videoconferencing*, one of the locations sends the information and others receive it. In multi-way mode, which is also known as *multi-point videoconferencing*, two or more locations can send or receive the information simultaneously.

**Computer Conferencing.** Some types of teleconferencing systems use computers for conducting meetings, which are known as Computer Conferencing Systems. In these systems, the participants use either E-mail or Electronic Bulletin Boards for sending and receiving information. The messages are entered into the computers using E-mail facility and an electronic conversation takes place among participants. The messages can also be posted on a computer system, called *Electronic Bulletin Board*, that maintains the list of messages.

NOTES

---

## 1.4 CONCEPTS AND TERMINOLOGY OF COMPUTER

---

### **What is a Computer?**

People call a lot of things, computers. Some of them are cheap, but can

## NOTES

only process numerical information. Such devices would better be described as 'calculators' rather than computers. Historically speaking the word computer has been derived from the Latin word 'computere', which means to calculate. Therefore, the term computer can logically be applied to any calculating device. However, in technical parlance, the term has come to refer specifically an electronic computer. In other words, it may be stated that the 'computer' uses electronic devices such as transistors, resistors, diodes based on two measurable states (*i.e.*, 'on' and 'off').

As generally understood computer has some defined characteristics. These are:

- (i) **It is an electronic device.** That is to say that electronic devices are used as component parts in making a computer.
- (ii) **It has internal storage.** Computer has an internal storage (memory) for storing both the program as well as the data processed by the computer.
- (iii) **It has a stored program.** Every computer has a program of instructions, which specifies the sequence of operations to be followed, stored in its internal memory. It is this stored program which makes the computer automatic because the entire set of steps to be taken is determined in advance and no human intervention is required during its execution.
- (iv) **Program modification.** A distinguishing feature of the computer is its ability to change the stored program of instructions during the execution of the program steps. The modification is based on the form, quantity or value of the data being processed. In other words, a computer is essentially an electronic device that can receive and store data and a set of instructions and then act upon these instructions in a pre-determined and predictable fashion to process the data in a desired manner. A computer contains a large number of electronic circuits arranged in such a way as to enable it to:
  - (a) perform arithmetic operations;
  - (b) evaluate logical conditions;
  - (c) store values and retrieve values from its memory;
  - (d) process input and output data.

Although there are many shapes and sizes, makes and prices, computers have some features which are common.

Firstly, computers are able to store internally not only the information to process, but also the set of instructions which tells them how to process it. Both instructions and information are represented in the computer as patterns of electrical states in electronic circuits.

Secondly, computers are extremely fast. Since the computer processes information electronically without the use of mechanical moving parts,

it can, in principle, process several millions of instructions in a second. Thirdly, computers process information in coded form. Information is often prepared for computer processing by using a keyboard, and is automatically converted into coded form.

## **Categorisation of Computers**

Depending on computing power and other capabilities, computers may be grouped into different categories: super computers, mainframes, minicomputers, microcomputers and personal computers. Let us try to understand some of the basic characteristics associated with these machines.

### ***Super Computers***

The specialised demands and requirements of science, industry and military have led to the creation of powerful super computers. For numerically intensive computing (geological exploration, wind tunnel simulations, molecular physics, and space and defence systems), computers capable of performing multiple giga flops (billion floating point operations per second) are being produced. These computers, of which Cray is a well-known example, use vector and parallel processing to achieve the necessary number-crunching power. *Vector and parallel processing require unusual computer architecture and special programming.* For vector processing, the computer architecture works with vector arrays rather than discrete scalar elements, and the numerically intensive portions of a program must be written to use this facility. To achieve full speed and power in vector arrays processing, the super computer must not be interrupted in the middle of its program, which limits the interactive use of super computers. Scientists and engineers have found that they can experiment with more design and research solutions in a short period of time on interactive engineering and graphics workstations.

Super computers built with a parallel processing architecture are designed to overcome the processing bottleneck caused by using a Von Neumann-type of single instruction stream. The parallel computer architecture must coordinate communication and timing across an array to allow simultaneous computation of subroutines. Although, massively parallel super computers, such as those made by Sequent and Floating Point Systems, are in principle faster and less expensive than the serial super computers, the unconventional programming required to utilise the parallel architectures has been an obstacle to wide use of these machines.

It may be stated here that while research continues in the United States and Japan to create faster super computers, some of the techniques employed in super computer architecture are being transferred to microprocessor design. For example, Intel has produced a RISC (Reduced Instruction Set Computing) microprocessor, the 80860, modeled on the

## **NOTES**

NOTES

Cray super computer. The design uses one million transistors to provide a 64 bit processor, data and instruction caches, integer and floating point math units and a graphics processor. The math units can operate in parallel using a pipe lining system and the data cache. Integrating these functions on a single chip makes it possible to sustain high processor speed overtime. The Intel 80860 and similar RISC chips are produced by Motorola and other companies.

### **Mainframe Computers**

Mainframe computers are large and powerful machines. They usually support many users with varied applications. They can process large quantum's of data at greater speeds (i.e., 10 million instructions per second) and support many input, output and auxiliary storage devices. The mainframes are multi-user facilities and therefore support large network of terminals and remote job entry stations. Generally, they play the role of central machine of a large organisation.

### **Minicomputers**

The minicomputers are intermediate in power, and may function as small mainframe computers. They are often dedicated to a particular purpose such as data base access and support several users at a time. Cost wise, they are less expensive compared to mainframes.

### **Microcomputers**

Microcomputer systems have received a great deal of attention in recent years. They are physically small, and relatively cheap. Originally they were of low power and only single user.

Recent developments in chip and processing technology has resulted in the emergence of new microcomputers capable of out performing mini computers in certain applications, usually heavy computational or simulation applications.

Descriptions of microcomputers usually refer to the type of chip that is at the heart of the CPU and there are two main 'families': The Motorola series and the Intel series.

The Motorola 68000 series are particularly adept at handling graphics and are used in personal computers such as the Commodore, Atari, and Apple Mackintosh ranges.

The Intel series was adopted by IBM and IBM compatibles or clones and is the most widely used in business microcomputers.

For example : 8088/8036 was used in the First IBM PC and then in IBM PC XT and compatibles

80286 : Formed the basis for the IBM PC AT and compatibles. Software for PC and XT can run-on AT machines.

80386 : At this juncture the microcomputer manufacturers were

divided into two camps: IBM Micro Channel Architecture (MCA) ps/2 range and Extended Industry Standard Architecture (EISA). Special software had to be written for the ps/2 range to make full use of its capabilities. EISA machines can make use of existing software.

#### NOTES

- 80486 : Introduced in 1989 and capable of out performing mini computers if suitably configured. Available in both MCA and EISA machines.
- 80586 : Under development since 1990.
- 80860 : Microchip using RISC (Reduced Instruction Set Computing) technology. This is hailed as bringing the 'power of a Cray to your desktop'. Suitable only for 'number-crunching' because of the special software that is required, but being incorporated into microcomputer as a maths co-processor rather than as the main. CPU. The Motorola equivalent is 88000 series.

*It is important to remember that the overall performance of a microcomputer is not determined solely by the processor.* The memory architecture, disk system, graphic system, and input/output channel also play important roles. The safe rule is to see the machine working with the software of interest to you.

### ***Personal Computers***

These are basically microcomputers used for day-to-day personal applications of individuals. They have become popular due to their low price. In fact, many predictions put forward before 1975 about the ways computers would be used have had to be adjusted to take land account several noteworthy changes brought about by the personal computer:

- (i) the PC, by allowing individuals to have their own machines, removed many of the obstacles that separated users from mainframes;
- (ii) through ease of access and portability, the PC user was encouraged to try using the machine in new ways, which stimulated the development of new applications;
- (iii) the development of a user friendly operating environment allowed refinement of techniques for using the computer as a personal tool in word processing, data management, and a variety of other specialised areas. The wide spread use of personal computers only goes to prove that the computer has transformed itself from a tool of only large institutions and corporations to a tool that can be used generally in society. Users, once passive recipients of a computer service, have been actively involved with the computer and its software. These changes have really altered the relationship between society and computers and given a net meaning to the computer revolution.

NOTES

## Recent Trends in Computer

**Pocket PC/Personal Digital Assistant (PDA):** These are very small computers with limited functionality. The general function of such computers is organizing. Many people find them incredibly handy and run them in different circumstances and locations. Popular examples would be Palm Pilot, the Psion Organizer, Compaq Aero, and so on. Some actually run a compatible version of Windows so that you can swap files with your normal size computer. These devices sometimes are named "handhelds" and have some basic features:

- Date book to input the time of meetings, appointments, etc.
- Memo Pad to place notes and write whatever.
- Calculator to solve problems via calculations.
- To-Do List to help accomplish some tasks.
- Address Book to keeps track of all your important contacts.
- On Screen Key/board to let typing with the touch screen and your stylus.
- Graffiti (Palm)/Character Recognizer (Pocket PC) to let handwriting on the screen instead of typing on keyboard.
- Email option to let composing and sending email.

Note: Graffiti is the trade name for the specialized set of characters that the Palm recognizes.

There is also a lot of entertainment value in the PDA. You can find many games to please any taste, and use software to carry your pictures, listen to music, watch a movie, etc., major manufacturers are Sony, Handspring, Palm, Hewlett-Packard, Casio, Compaq, Toshiba, Handera, ViewSonic.

**Ultra personal computer (UPC):** The OQO issued the model 01 of uPC as a fully-functional Windows XP computer. OQO model 2+ has Intel Atom Performance, 2GB of RAM, OLED display. The OQO Docking station allows connection to a full size display, keyboard, mouse, and so on. This type of computer is powerful enough to most complex applications at work, sending email at home, listening to music on a train, or watching a movie on an airplane.

**A new generation of minicomputers** differs from minicomputers, which was developed in the last Century by Digital Equipment Corporation (DEC). Hewlett-Packard, Honeywell, and others, In size and power modern minicomputers lie between microcomputers and normal PC of workstation (desktop or laptop).

**Netbooks.** Dell issued Mini 10 and Mini 10v Netbooks with a keyboard 92% the size of a traditional laptop keyboard, wireless options, build-in GPS, et cetera.

Mini Book PC, Book PC., Cube PC (Shuttle) and others, developed by The Book PC. com, Inc. are space saving minicomputers with performances like some traditional larger desktop computers.

Little PC com Puters, developed by the Little PC, Inc (a division of Stealth Computer Inc. (www.stealthcomputer.com)) available on the Market in the different configuration.

NOTES

## 1.5 ORGANIZATION (STRUCTURE) OF A COMPUTER SYSTEM

In the present section, an attempt is made to acquaint you with some hardware basics regarding computer.

The schematic diagram provided earlier illustrates the integration of basic elements of a computer system. They are:

- (i) input devices,
- (ii) the central processing unit (CPU),
- (iii) the auxiliary or backing storage, and
- (iv) the output devices. Let us try to understand their functioning.

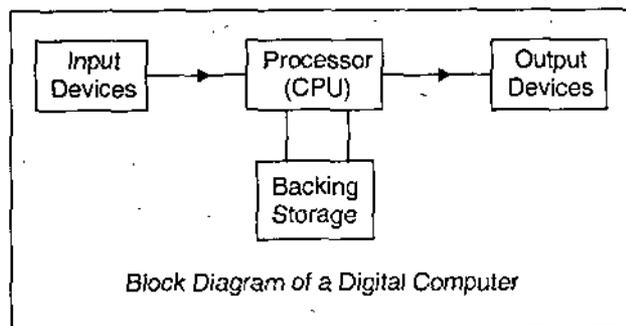


Figure 1.4

- (a) **Input Devices.** Each input device of a computer system reads data from a specific form and converts the data into electrical pulses. It then transmits these pulses to an input area in the CPU. There are different types of input devices such as: Keyboard, Pointing devices, Voice input, Barcodes, OCR, and Image Processing.

### **The Central Processing Unit (CPU)**

The CPU is the physical device that controls computer operations. It is considered as the 'brain' of the computer system.

The CPU comprises three components: The control unit, the arithmetic/logic unit, and the main memory (main storage) All these are composed of microelectronic devices, virtually always based on silicon semiconductor technology.

The main memory is the largest component of CPU, and is divided

## NOTES

into individual computer words, each of which may be accessed individually by its address. All present day computers are based upon semiconductor technology, the ability to place many thousands of electronic components within a limited area of silicon, which has been doped with carefully controlled amounts of other elements. The main memory holds program instructions and data required as the programs operate.

The **arithmetic/logic** unit contains a set of registers or accumulators, individual words of very high speed storage into which words from the main storage are copied, prior to instructions being applied to them. Attached to the registers is the circuitry which carries out the few hundreds of basic operations available to the computer. In essence, when an individual instruction needs to be carried out on some word in the main storage, that word is copied into a register, the appropriate circuitry activated, and the register's contents copied back to the main location.

The control unit extracts program instructions one at a time in the required sequence from their locations in the main storage, decodes them, and issues instructions to the main storage to access the required word, and to the arithmetic/logic unit to activate the required instruction circuitry.

### ***Computer Storage***

Computer systems include two types of digital information storage: internal storage, within the CPU, and the backing (back up) storage on external devices such as disks or tapes. Different types of storage media differ according to a number of criteria such as speed of operation, capacity, cost, reliability, the degree to which information is immediately accessible, etc.

### ***Internal Storage***

Internal storage is also known as main, primary, or (for historical reasons) core storage, or, memory. It is used to hold those instructions and data required at any moment while a program is running, which must be available instantaneously. Silicon semiconductor chips are now invariably used for internal storage; these are categorised as either RAM (random access memory) or ROM (read only memory) chips.

ROMs are used for data which is never altered for example, a computer's operating system instructions, while RAM is used for data which is liable to change often, and instructions used in the execution of a program.

Internal storage capacity is measured in bits or bytes and multiples thereof. The size of individual chips comprising the internal memory has rapidly increased from 4 KB (Kilo Byte chip) to 16 MB (Mega Byte) chips.

### ***Backup Storage***

Backup (backing) storage, also termed **external** storage, is used to hold

**NOTES**

programs and data which are read into internal storage when required. The most common form now, as for the last thirty years, is magnetic storage media, either tapes or disks. With both of these, data is recorded onto a plastic surface, coated with a varnish containing an oxide which can be magnetised in one or two directions; each magnetisation representing one bit of data. Data is written or read with the help of read/write heads, small electromagnets, close to the surface.

**Magnetic Tapes** are very commonly used for storing large quantum of data for which rapid access is not necessary; specially for archival data backups etc. Tapes are cheap means of storing data but access is generally slow. Small cassette tapes, of relatively low capacity are used with microcomputers.

**Magnetic Disks** are most widely used form of backing storage, suitable for holding information required rapidly for example, data for a running program. In the type of disk drive in larger computer systems, data is recorded on the flat surfaces of a stack of circular disks revolving on a common spindle, with one read/write head for each disk. Access times, typically 0.01 seconds, although much faster than tape, are rather slow, by comparison with other computer operations, and disk access times are one of the main limiting factors in the speed of operation of retrieval systems with information stored on disk.

Disk capacities have increased greatly, from original values of about 10 MB (Mega Bytes to present units with 1 GB (Giga Bytes which is one thousand million bytes). Smaller computer systems use **Winchester** disks, with a single hard disk in a sealed unit, floppy disks, which are compact, cheap and convenient for transfer of programs and data. They are limited in storage capacity, with relatively slow access times, and less convenient than a hard disk in day-to-day use. Floppy disks now are available in two sizes, 5.25 and 3.5 inches.

**Optical Disk** storage is relatively new alternative to magnetic storage. Digital data is burnt into the surface of a plastic disk coated with a low wetting point metal by a laser beam. The presence or absence of a hole corresponds to one bit and can be read by another laser. This technology offers a high capacity storage at low cost, and in a form allowing rapid access (analogous to magnetic disks).

**WORM** (Write Once Read Many Times) disks, are ideal for archival purposes allowing organisation to store large volumes of documentation on one disk. For instance a 12 inch disk can hold about 2.5 GB at a cost of few hundred rupees. Storage capacities of 5.0 GB is now available. But, the major problem with WORM technology is the absence of any industry standards. Now 10 inch disks (compact version CD-WORM) is being developed.

The CD-ROM (Compact Disk Read Only Memory) is a 4.75" disk capable

of holding 650 MB. The CD-ROM has become a means of distributing databases and documentation (including journals and patents) in electronic form.

### ***Input and Output***

#### **NOTES**

After having understood the characteristics relating to CPU and storage media, let us try to learn about I/O aspects of a computer system.

Input/output (I/O) devices now account for more than 50% of the cost associated with microcomputer system. They also have an impact on the ease of use of computer systems, and their popularity with users. Therefore, it is necessary to think carefully about the sort of I/O hardware required. To enable appropriate choice of the equipment based on requirements of applications software which might be used, a range of available devices are discussed. Many packages, for example, require a colour screen to be fully effective, while to take advantage of the fonts and type faces available in some word processing packages requires a laser printer. However, a brief account of some of the I/O devices is provided in the following paragraphs.

### ***Input Devices***

- (i) Keyboard is the most common form of input devices: It was originally designed in the last century. Since then, only minor improvements have taken place in key board design.
- (ii) Pointing devices: These are used to indicate a point on a screen for example, to make a selection from a menu. They include: The light-pen (where the tip of the pen is touched on the screen); The touch-screen (touched directly by finger, pen etc.), and the mouse (moved around the desk or other flat surface to move a cursor on the screen, with the required position marked by pressing a button on mouse). These devices will obviously work with appropriate interactive computer software.
- (iii) Voice input is now in a stage of development where commercial systems are becoming available. It may take some more time for these systems to be trained to recognise each user's voice and are presently limited to vocabularies of a few thousand words. In other words, voice input is only in experimental stage of development. It may take some years before it is made operational.
- (iv) Barcodes. These incorporate information in a pattern of stripes, produced and read by special equipment. These are widely used in library issue systems and are slowly gaining wider use for general data transfer, particularly as a means of 'publishing' computer programs. To the extent they are used in library automation, they can be considered as data input devices.
- (v) Optical Character Recognition. This method has now developed to a stage, when it can be used as a practical means for routine entry of large quantum of print on paper information into computer systems.

Unless special steps are taken, error rates are very high in this technique. Hence, standardisation of input is highly necessary.

### ***Image Processing***

This technology is quite advanced and devices are now available for routine scanning and storage of printed pages, graphics, etc., which can then be retrieved and displayed when required.

In the foregoing paragraphs of this section attempt has been made to explain to you some of the developments taking place as far as input devices are concerned. These developments are to be studied along with other input devices such as card reader, terminal, magnetic tape drive and magnetic disk drive., etc., which existed since long time.

Future developments in this direction might include more alternatives to the key board, particularly voice input for commands which is presently available, and wide use of OCR and image processing for handling large volumes of information.

### ***Output Devices***

Each output unit of computer system transmits information from the CPU and converts the electrical pulses to an appropriate output form. A printer, for example, is an output device that transmits output from the CPU and converts it to a printed form. There are many other output devices that can perform the same electrical transmission and conversions to produce other output forms.

The Visual Display Unit (VDU) is the most common form of output device. Mostly, it is based on the Cathode Ray Tube (CRT). But these have markedly increased in quality and decreased in price. Standards over, colour monitors developed over the years include the Colour Graphics Adaptor (CGA), Enhanced Graphics Adaptor (EGA) and Video Graphics Array (VGA). These have become widespread. Alternatives to CRT including light emitting diodes, liquid crystals, and gas plasma discharge panels, are coming into use, particularly for portable computers.

**Printers:** Printers exist in a variety of forms:

#### ***Line Printers***

These have been widely used for many years in large computer installations. They are designed for the rapid printing of a large amount of information. Till recently, they have had the disadvantage of being noisy, inflexible, and limited to low quality output. However, because they are extremely reliable, they have remained popular with computer departments and now manufacturers are producing quieter and more flexible models of line-printers.

#### ***Character Printers***

These are known as character printers because they print one character at a time.

NOTES

NOTES

These are of two kinds: impact printers and non-impact printers.

- (i) **Impact Printers.** These are printers in which the print-head strikes a ribbon, and include the daisywheel and thimble printer (good quality, but relatively inflexible and slow), and the dot-matrix printers. Dot-matrix printers have traditionally generated comparatively poor quality output, but the 24 pin and 48 pin versions are capable of producing high quality material. Colour options and graphics are available in some of the recent models.
- (ii) **Non-impact Printers.** These are silent in operation and are improving in reliability and quality. These include ink-jet and ink-dot printers (particularly suitable for colour graphics); electrostatic printers (used on facsimile machines, etc.), and thermal transfer printers (suitable for smaller computer systems because of low cost, versatility, and portability). Colour printers are available in this type of printers, which are used in publishing and advertising.
- (iii) **Laser Printers.** These printers can print a whole page at one command, offer highest quality and can combine text, graphics, etc. They are generally restricted to A4 paper and black and white, though transparencies can also be printed. Colour laser printers are now available but at very high prices. Laser printers are mostly used in electronic publishing.

Future prospects for output devices include: high quality display screens of different sizes, more wide spread use of laser, and similar page printers, greater availability of cheap and versatile thermal printers and a greater role for voice output, etc.

---

## 1.6 SOFTWARE BASICS

---

Software is a generic term covering the concepts, procedures and instructions which cause the computer systems to accomplish the required job. Generally, software is thought of in terms of programs, discrete units of software which enable the computer to carry out a particular task, and or systems or package and integrated collections of programs.

The importance of software lies in the fact that it is the software which applies power of the computer to solve the problems faced by the users. Most users require a detailed understanding of the capabilities of software than of hardware.

It has been clear for some years now that the rapid advances in the capabilities of computer systems has not been matched by corresponding advances in the availability and quality of software.

Software may conveniently be divided into two categories: systems software (*i.e.*, programs designed to control the execution of other programs and to utilise hardware effectively), and applications software (*i.e.*, programs

## NOTES

which enable to solve users' problems). The methods by which software is produced have changed considerably in recent years with the emergence of the new discipline known as 'software engineering'. The traditional pattern of software development used to be a systems analysis to undertake a detailed study of user needs, leading to an elaborate specification for the software, which would then be converted into a programming language. This process is considered inadequate, and leading to backlogs. It is also felt that this procedure does not meet user needs effectively and is poorly maintained.

Increased availability of packaged software and use of fourth generation languages, and flexible integrated software to produce rapidly modified 'prototypes' of programs to meet user needs, are suggested as remedies to the software backlog problem. However, for the average user they mean that there will be an increasing array of packages to meet his requirements.

### ***Systems Software (Pre Written Software)***

Systems software is generally supplied by the hardware manufacturers. It includes operating systems, assemblers, compilers, and interpreters (to convert programming languages into machine code) programs for controlling input and output devices, copying data between storage media, etc., and utilities for sorting, merging, and editing files, controlling program libraries, etc. In other words, it is all "supporting" software, which enables the power of the computer to be applied to user's problems.

### **Operating Systems**

The operating system is the software that mediates between the applications programs and a level of instructions nearer to the machine's operations. In other words, it is "software that controls the execution of a computer program, and that may provide scheduling, debugging, input/output control, accounting, compilation, storage assignment, data management, and related services".

Operating system software is written specifically for the type of function the computer is going to be used for. Generally, the more complex the function is, the more complicated the operating system will be. There are many types of operating systems available: some are made to work on a variety of different types of hardware, and some are designed for only one specific type of machine. Mainframe and minicomputers generally have their own proprietary operating systems. In the early days of personal computing, all operating systems were single user. Apple, Amiga, Commodore and IBM are all computer hardware manufacturers with machines that run different single-user operating systems. The most commonly used operating system in the world today runs on the IBM PC. This is the Microsoft disk operating system (MS-DOS). Since 1980s, DOS has been improved and refined many times:

NOTES

A second type of operating system, described as multitasking, is still only capable of having one user but more than one task, operating at a time. The main task that requires the user's attention (for example a Word processing program) remains on the computer monitor, while the other task is working away in the background. The background process could be a complex statistical analysis program that takes a long time to complete, and the user could spend the waiting time for some other work. A multitasking successor to DOS is called Operating System 2 (OS 2).

Microsoft has gone on to develop its Windows environment, an operating system with multitasking and other advanced capabilities. Multitasking operating system must divide the CPU resources between the different tasks. In addition, multitasking itself is a more sophisticated operating system than the single user type, and can take more CPU resources just to run.

The third type of operating system is multi-user. By definition, multi-user systems are also multiprocessing, since each user who is working on the machine accounts for a separate process. Depending on the power of the CPU and the size of the computer, the number of users that can be working at one time might go into hundreds. The most common examples of multi-user systems are in the large mainframe environments. Systems like VMS from DEC and MVS from IBM are examples of operating systems with the potential for large numbers of users.

UNIX is a multi-user operating system that has gained high popularity. Developed originally at AT&T Bell Labs, UNIX was designed to be used by engineers and scientists. Because of its efficient design and built-in security features, and its reasonable price, application software companies have considered UNIX as an attractive development platform. Probably the most attractive feature of UNIX is that it is an open system and multiplatform. This means that the operating system will work on a range of hardware configurations from microcomputers to mainframes. A new competitor to UNIX on personal computers, workstation and networks is the Windows NT operating system, a new operating system from Microsoft Corporation. Multiplatform open operating systems are very attractive because applications that are developed for one platform can be run on another with minimum conversion problems. A UNIX word processing software package on a Sun Workstation could also be used on a UNIX P.C. The application is said to be 'ported' (portable) from one machine to another. From user's point of view, much of this may not be important. Factors which must be considered are the ease of use of the operating system, and any restrictions it places upon **choice of software**. The latter point is particularly important, since any application program or package is restricted to specific operating systems. For example, something designed for the Macintosh will not run under an IBM operating system and vice versa.

## **Applications Software (Customised Software)**

In this section, we will try to learn some general aspects on applications software.

Applications software is the program that allow the user to get the required work done. In other words, applications software comprises the procedures and instructions which enable computer systems to do what the user requires. The software design essentially involves three abstract concepts: algorithms (*i.e.*, procedures, or recipes, for computation —may be numerical or non-numerical, *e.g.*; sorting, text searching): Data structures which indicate the way in which information is organised in the computer's memory. In other words, the structure indicates whether it is organised in array or tree pattern. File structures represent the way in which large amounts of information is stored on the storage media for example, whether it is organised sequentially, or in direct access, inverted file.

Even though users do not need any in-depth knowledge relating to the above mentioned three concepts, it is desirable to have some understanding of the terminology, since it is commonly used in software suppliers' literature.

Once a software procedure is expressed in such terms, it can be coded into a form which the computer can use, by means of a programming language. There are a number of programming languages in use. It must be emphasised that a computer can only operate on instructions in machine code (binary), and early computers were indeed programmed this way. Assembly language, a more convenient mnemonic form of machine code, is still us id today for specialised applications.

**High level languages** (converted to machine code by compiler or equivalent) are much easier to program, as well as to understand. Several high level languages are in use today. Best known are the FORTRAN (Scientific), COBOL (business) and BASIC (interactive) languages. Most recently introduced general purpose languages include PASCAL, ADA and C. Most recent programming languages are called the 'fourth generation 'or' very high level' languages, generally as part of integrated program systems, such as DBMS, which facilitate users to program their own applications very easily:

It is not necessary for users of a program or package to have any knowledge of the language in which the package is written, or indeed to know what that language is.

Let us, learn about some commonly used software packages.

### **Software Packages**

- (i) **Word Processing (WP).** Word Processing is one of the most widespread application software types in use today. Developed as a successor to primitive text editors that were popular on mainframe computers, a word processing program allows interactive

NOTES

NOTES

editing of documents, enabling easy redrafting and merging of chunks of existing documents, without the need for extensive retyping. Most of the popular programs have features such as 'spell' checkers, outlining, choice of fonts, line drawing, as well as page layout settings to produce any type of document desired. One feature that most of the popular packages offer is a way to just see how the document will look on the printed page; this is a handy feature when laying out complicated tables or columns of text. Some of the examples of word processing software Word Star, Word Perfect, MS Word etc.

- (ii) **Desk Top Publishing Program.** Desktop publishing programs contain many of the features of word processing software, but go even further in the ability to layout the format of a printed page. These packages are being used to control complex components of the printing process right from the PC. Desktop publishing does what used to be called type setting, and has truly changed the printing business. Sophisticated packages like Ventura Publisher and Page Maker allow, in-house departments to produce page layouts to include graphic images that were digitised with scanners, and print the output to laser printers or high quality devices that produce camera-ready output. Though this procedure was more difficult to use than a word processor, many of the features of desktop publishing software have been streamlined and are now incorporated into word processing packages. While variety of type font and size is not to be *confused* with sound typographical design, in general, application of DTP techniques has improved the look of documents and manuals in many of the organisations that use these techniques. A new development in DTP is the ability to exchange various personal computer (PC) platforms final documents in published formats. It is stated that this may pave the way for electronic distribution of widearray of materials, such as newsletters, which are now distributed in print farm.
- (iii) **Database Software** Another major type of application comes under the heading database software; which allows for the collection of, searching for and manipulation of information about unique entities. The topic (subject) might be the characteristics of competitor products, the bibliographic and abstract information relating, to relevant research publications and activities dealing with an area of R&D, etc. All topics on which files of information may be collected are susceptible to be controlled by database management software. One of the early arrivals to the PC data base management software market was a product from Ashton-Tate called database. Still in, use today, database also provided a fairly sophisticated third generation programming language. Everyday users can use the software in its standard form but the programming features

allow more advanced users to develop customised applications that are tailored specifically to the needs of an organisation.

Each of these approaches has its own advantages and disadvantages. Standard packages are by far the most common way of acquiring software. They offer simple and relatively cheap option, with the advantage of a large group of users of common package. Turnkey systems, with all components of hardware, software and communications provided by the vendor (supplier) are simpler for their users as every aspect is looked after by the supplier. This is a common method pursued by some larger libraries. It is desirable that the package acquired is an integrated one-that is to say it can carry out a number of functions within a common framework. Examples could be library automation packages, incorporating acquisitions, cataloguing, circulation etc. There is one possible drawback associated with fully integrated packages, apart from not being flexible; some elements of integrated packages might be inferior to a program dedicated for the same purpose. For example, if database applications are particularly important, it could be a mistake to rely on the database component of a standard package. The 'customised' approaches provide flexibility and a product geared precisely to your specific needs. Though popular, software specially written for information services is too expensive and difficult to maintain and modify for it to be a realistic option in many cases. Some ability to modify software is desirable and standard packages offer the chance to modify such aspects as output formats, and help message. Also, some larger systems are available in modular form, with distinct pieces of software carrying out specific functions. Modules for particular purposes can be acquired and then combined as required, giving a fair degree of flexibility in the look of a software system. The obvious example is the library house keeping package with separate modules for acquisitions, cataloguing, circulation, etc.

## NOTES

---

## 1.7 DATA COMMUNICATION AND NETWORKING

---

Large organisations are generally spread over several office locations. In order to computerise such geographically separated offices, various data communication facilities are required. The transfer of data and information from one computer to another over a communication link is known as *data communication*. Data communication has become essential for an organisation due to several reasons. The major reason for needing data communication is that it is the fastest and most cost-saving medium for sending data and messages to various geographically separated offices. There are several types of communication media through which data can be transmitted from one place to another. Some of the common data transmission media are Wire Pairs, Coaxial Cables, Microwave Transmission and Communication Satellite. *Wire pairs* are used in local telephone

## NOTES

communication for transmitting digital data for short distances. *Coaxial cables* are groups of specially wrapped insulated wires, which transmit data at very high rates. *Microwave transmission* is the popular medium, which is used to transmit data through the electromagnetic waves. A *communication satellite* is a microwave relay station placed precisely at 36,000 KM above the earth. It is the most widely used medium for sending data across a wide area on the earth.

In this chapter, we are discussing below some of the widely used data communications and network technologies.

### Overview of Data Communication

Data communication refers to the electronic transfer of data from a sending device to a receiving device. For instance, transmission of voice from one telephone to another; transmission of radio programmers from a radio station to your radio; and sending of data from one computer to another, all are considered as data communications. Data communication is commonly called *data transmission*, so both terms are synonymous. Data are transmitted in the form of *signals*, which can be defined as the physical forms of transmitted data. There are two types of signals— analog and digital. Before discussing the methods of data communication, let us first understand the difference between analog and digital signals.

### Analog and Digital Signals

An analog signal is one that is continuous with respect to time, and may have any value within a given range of values. In an analogue mode of data transmission, the variations in originating audio and video correspond to variation in the electrical signals. So, the transmission of data is in a continuous wave form as voltage varies with the time as shown in figure 1.5. The transmission of data (sound, picture, etc.) over telephone cables is an excellent example of analogue mode of transmission.

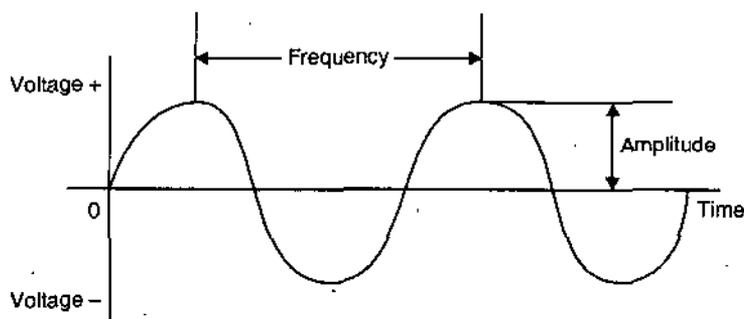


Figure 1.5 An analog signal

Waves are measured using one or more of the three characteristics viz. Frequency, Amplitude and Phase. *Frequency* of a wave is the time taken by it to complete one cycle. It is measured in Hertz (HZ) or cycles per second. *Amplitude* of a wave is the distance from the baseline (when voltage is zero)

to the peak (when voltage is maximum) of the waveform. It is measured in meters and is computed by dividing speed of transmission with frequency. *Phase* of a wave is a relative measurement that describes the temporal relationship between two signals having same frequency. It is measured in degrees, with one full oscillation cycle having 360 degrees.

A *digital signals* may have only two values *i.e.*, zero and one only. The ON pulse of electricity represents one while OFF pulse represents zero as shown in figure 1.6. The transmission of data between the computers is always in digital signals as computers understand and process data only in binary form. However, when data is sent from one computer to another through a network of telephone cables (*e.g.*, in Internet), it is converted into analog form through a device called *modem* and the process of converting analog signals to digital signals is called *digitalization*. We will discuss about modem in the later part of the unit.

## NOTES

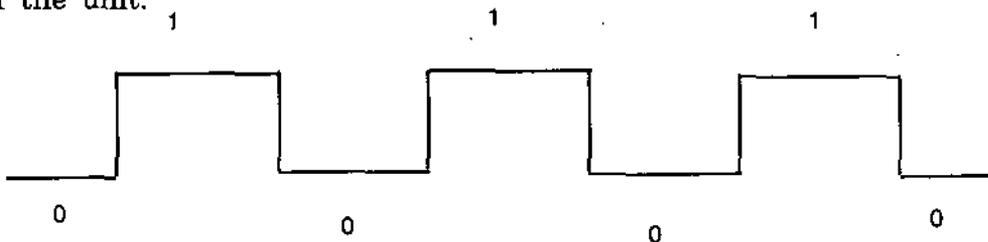


Figure 1.6 A digital signal

### **Modes of Data Communications**

There are two modes or methods of transmitting data—Asynchronous and Synchronous Communications.

(i) **Asynchronous Communication.** In asynchronous mode of data transmission, each character (byte) is transmitted one at a time using two types of special bits, called *start* and *stop bits*. The start and stop bits co-ordinate the flow of data so the time intervals between individual characters do not need to be equal. A start bit precedes each individual character. The purpose of start bit is to inform the receiving device that a character transmission has begun. The transmitted character ends with one or more stop bits. The stop bit(s) suggests that a character has been transmitted. The stop bit(s) may be preceded by a *parity bit*, which checks the accuracy of data received as shown in figure 1.7. There are generally two types of parity bits—Odd and Even. *Odd Parity* indicates that the sum of all 1 bits in the byte sent plus the parity bit must be odd. If the total is already odd, the parity bit is set to 0 otherwise it is set to 1. Similarly, in *Even Parity*, if the sum of all the 1 bits is even, the parity bit must be set to 0 otherwise it is set to 1. There are several kinds of bit representations of a character such as ASCII (American Standard Code for Information Interchange) and EBCDIC (Extended Binary Coded Decimal Information Code). In widely used ASCII representation, 7 bits uniquely represent a character and the 8th bit is called parity bit.

NOTES

In asynchronous mode of communication, there is no indication of the speed at which the character is being transmitted. Therefore, there is no perceptible method for the receiving device to sample the incoming signal precisely to recover the transmitted data. For this lack of synchronisation of transmission, this mode of data transmission is called asynchronous transmission or asynchronous communication. This anomaly of non synchronisation is overcome by providing internal clocks at each end, transmitting as well as receiving. Both internal clocks work at the same rate such that on the receipt of start bit the receiver device starts its internal clock to instruct receiver's interface to sample the incoming signal.

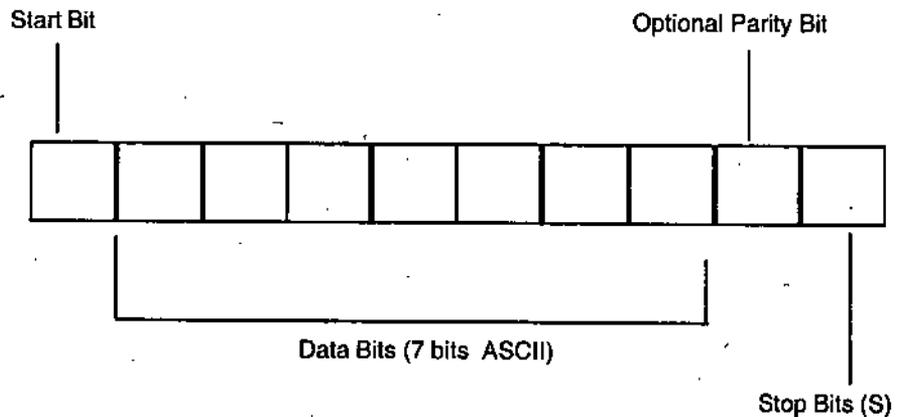


Figure 1.7 Asynchronous communication

(ii) **Synchronous Communication.** In synchronous mode of data transmission, a block of characters is transmitted one at a time using a few synchronisation (SYN) characters, which precede each block of characters as shown in figure 1.8. As these SYN characters contain the timing information that allows automatic synchronisation, this type of data communication is known as synchronous communication. In synchronous communication, the size of blocks of characters may vary, so the signals contain the ASCII control characters, called *STX* (start of text) and *ETX* (end of text), which are used as delimiters.

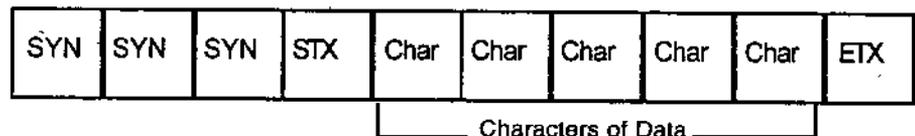


Figure 1.8 Synchronous communication

The synchronous mode of data transmission is useful when large volume of data needs to be transmitted at a higher data transfer rate. The efficiency of synchronous transmission varies according to the block size as described below:

- If the block size is too small the overheads of protocol is more due to additional SYN characters and delimiters but there are less chances of error in transmission.

- If the block size is too large the general efficiency for large data transfer may improve in terms of data transfer rate but there are more chances of error in transmission.
- If the block size is neither too small nor too large (512 bytes in most of protocols) a reliable error free data transfer takes place over a long distance.

## NOTES

*Error Detection in Synchronous Data Transmission.* In terminology of data communications, a block of data is called a *packet*. A sequence of blocks that is transmitted from a sending device need to be reassembled at the receiving device. Each packet contains a packet or block identity number and routing address for destination for accurate reassembly as the original data. In case of identification of error on receipt of a packet, a request is sent to the sending device for retransmission of that packet again. In synchronous transmission, the common methods of error detection in data transmission are:

1. *CRC (Cyclic Redundancy Check).* CRC error checking uses a complex calculation to generate a number based on the data transmitted. In this method, the sending device performs the calculation before transmission and includes it in the packet and then it is sent to the receiving device. The receiving device repeats the same calculation after receiving the packet. If both devices obtain the same result, it is assumed that the transmission was error-free. This procedure is known as a *redundancy check* because each transmission includes not only data but extra (redundant) error-checking values.
2. *Checksum.* Checksum is a calculated value that is used to ensure error-free data in transmission. It is created by adding the binary value of each alphanumeric character in a packet and sending it with the data. At the receiving end, a new checksum is computed and matched against the transmitted checksum. A non-match indicates an error. Checksums detect single bit errors and some multiple bit errors, but they are not as effective as the CRC method.
3. *FEC (Forward Error Correction).* It is a means of controlling errors by inserting extra (redundant) bits into a stream of data transmitted to another device. The receiving device uses the redundant bits in detecting and correcting error in the data. The redundant bits are also called *Hamming Codes* (named after R. W. Hamming of Bell Labs). FEC Technique is used mainly in a satellite link connection.

## 1.8 DATA COMMUNICATION

### NOTES

A wire pair connecting two telephones is the simplest circuit, allowing communication in both directions. Other channels include coaxial cables, optical fibres, satellites, a light beam, microwaves, and others. Whatever the physical medium over or through which message travels, transmission may take place in a number of different modes. They are : Simplex, Half-duplex and Full-duplex.

**Simplex :** In this type of channel, the message travels in only one direction, that is from point A to point B but not back again.

**Half-Duplex :** Transmission can take place in either direction, but not at the same time. In other words, message can be transmitted either from A to B or by B to A one at a time.

**Full-Duplex :** In this type of arrangement messages can be transmitted and received over the same channel at the same time.

Many librarians and information scientists are familiar with the latter two types because database search terminals must be set to transmit full or half duplex in order to communicate with database vendors' computers.

### Synchronous and Asynchronous Transmission

Another method of setting of terminals denotes synchronous or asynchronous transmission. Many terminals can only communicate in one or the other mode. Asynchronous transmission is also known as "start-stop" transmission. This means that every transmitted character (or group of characters) is defined by a special character at the beginning and end denoting 'beginning' and 'end' of that character.

Thus, the receiving terminal does not have to operate "in phase" with each other and transmission is interpreted or decoded at the receiving terminal exactly as it was encoded by the sender in other words, in synchronous communication data is sent from one machine to another in a continuous stream, the sending and receiving equipment must be in step with each other throughout the transmission. This type of transmission is used for high speed data transmission, generally between two mainframes.

Asynchronous form of communication is used by most information systems attached to the public telephone network.

### Modems

As it was explained earlier, communications channels can operate in either the Analog or digital mode. However, a given circuit can operate only in one mode not both of them at the same time. Thus, there is need for translating from one to the other of the two modes. All voice communication occurs in the analog mode; yet, when we use voice-grade channels to

transmit data we need to communicate with the receiving or sending computer in the digital mode.

Modems are devices which function as "translators", changing the digital signals coming out of the terminal to analog so that it can travel along a *voicegrade channel*. At the receiving end, there is another modem, retranslating the analog signal into the digital signal the computer can accept. Modem stands, for modulation-demodulation. Some modems are built into the terminal, while others are separate devices interposed between the terminal and the telephone line.

## NOTES

### **Transmission Media**

When we speak of transmission media, we usually mean a mix of physical lines ranging from wire pairs to cable, and over the air transmission media, such as microwave and satellites.

The transmission media used in telecommunications networks vary both physically and in their carrying capacity. The "commonly used performance measurement of a telecommunications link is its **bandwidth**. This gives an indication of the range of frequencies which can be transmitted by these channels. The greater the bandwidth, the greater the rate at which data can be transmitted.

The channels or media can be grouped into two main divisions, *i.e.*, line and free space. In the line group, we have two metallic media, twisted pairs and coaxial cables and one optical medium—glass or optical fibres. In the free space category, we have radio propagation which includes broadcast (*e.g.*, TV) and point to point (*e.g.*, microwave signals). Let us discuss these in a greater detail.

### **Twisted Pairs**

Twisted pairs are familiar to all of us as the copper wire telephone lines. These are of low frequency, and support a limited bandwidth (one voice channel) but can also be used for data communication at the lower speeds of 300bps. For higher speed or rates of data transfer, 4-wire leased lines may be used. The problems associated with this kind of transmission are: noise on the line, parity errors and so on. Twisted pairs are used for conventional voice telephone and telex services.

### **Coaxial Cables**

Another metallic transmission medium which offers a large bandwidth, is the coaxial cable. Such cables consist of an outer protective shell made of copper tube, less than one inch in diameter, which actually carries messages. There are two types of coaxial cables—the base band and the broad band. Base band cables use digital transmission and are suitable for fairly short distances. These have a data rate of about 50 Mb/s in half-duplex mode. Broad band coaxial cable on the other hand, is more familiar as CATV and is suitable for long distance (up to 15 KM)

NOTES

transmission. It can support data rates of over 100 Mb/s. Coaxial cables, with their better electrical properties, can transmit at much higher frequencies and support wider bandwidth than the twisted pairs. They are suitable, therefore, for high quality bulk transmission of data at high speeds as well as text, picture and voice. For example, one such cable can support the equivalent of up to 10800 telephone (voice) channels.

### ***Optical Fibres***

Coaxial cables have limitations such as broadband transmission medium which can be overcome by the use of optical fibres. Optical fibres carry light waves (representing electrical impulses) by the principle of internal reflection. These fibres carry signals in the form of light, and hence can carry large volumes of data at high speed, and are resistant to interference. Thus, they are 'clearer' lines with less noise. Though they were expensive when first introduced, but nowadays costs have fallen to such a level that they are being used by some organisations for their local area networks.

### ***Microwave Transmission***

Using space as transmission medium, microwave emanates from an origination point on earth, such as telephone exchange, where many individual messages have been concentrated. Because the microwave beam travels in a straight line, it is advantageous to place microwave towers on hills or mountains to minimise interference from land mass. Usually, towers are placed between 25 to 30 miles apart to remain in sight of each other ('line-of-sight'). Microwave transmission can carry 600 to 1800 voice channels. Advantages of microwave transmission include lower cost than coaxial cable, where right-of-way must be obtained and higher costs are incurred when the physical cable is laid over long distances. As with other systems that transmit through free space the signals are subject to fading caused by absorption and scattering from moisture and precipitation in the air. In other words, atmospheric interference is greater and rain can cause severe transmission problems.

The great advantage is that the microwave transmission is a broad band facility able to carry several thousand channels. It is suitable for bulk transmission data over long distances.

### ***Satellite Transmission***

Satellites use microwave frequencies and techniques. Satellites used for message transmission (as opposed to those, which travel around the earth on strategic and meteorological missions) remain in constant position in relation to a fixed location on earth; they are said to be in a geostationary orbit. This position allows the satellite to send data over a long distance. Used extensively for general telecommunications, as well as for television, satellites were originally contemplated only for long distance transworld links but are now increasingly employed for high speed high volume

transfer for example document delivery, electronic publishing. In some countries like India, satellites are used instead of underground cables because of the high cost of installing the latter.

### **Choice of Medium**

All the transmission channels discussed in the foregoing paragraphs can be used as point to point connections or as broadcast systems. Transmitting information electronically or photonically is considered in two aspects - telecommunications and broadcasting. The former is a direct interchange of information within limited groups of talkers and listeners (which may be terminals of computers as well as people). Thus, there is a certain amount of interaction. Broadcasting is the mass distribution of information to large markets with one 'speaker' and many "listeners" and no real interaction. Because of the increasing linking up of the two forms with computers, the traditional distinction is being blurred and now you get systems like electronic mail, which can either be broadcast of telecommunication and cable TV broadcasting systems which carry telecommunications. And any given transmission circuit could be made up of a mixture of terrestrial microwaves, cable and satellite links. Thus, a given service could be transmitted over more than one of channel. For example, a television program can be broadcast let us say normally—as you get it at home via satellite or via broadband coaxial cable using radio frequency. The decision on which medium to use is determined by factors such as the distance involved, the area to be covered and the type of information to be sent. For example, terrestrial microwave radio transmission requires line-of-sight communication and for long distances repeaters must be added. If the medium has to accommodate video or high speed data then a broad bandwidth such as is provided by satellite and coaxial, fibre optical cables will be required.

### **Switching Mechanisms**

Switching mechanisms are techniques devised to send messages in many directions at once and to ensure that these messages are received with a minimum of delay. In other words, if a bottleneck or queue occurs in message traffic, bypasses or alternate paths to the destination are provided by taking recourse to switching mechanisms.

If traffic between two exchanges is too heavy, an alternate path is chosen. These exchanges perform the switching function for the telephone system. Switching is carried out by computers in most cases. Computers also serve to concentrate (collect) messages for transmission over higher capacity lines. The lines used in the telephone network are wires, coaxial cables, microwave and satellite links.

The different techniques relating to switching mechanisms are discussed below.

### **NOTES**

NOTES

### ***Circuit Switching***

Many switching techniques are in use. One of them is line or circuit switching in which a communications path is actually established before the message is accepted for transmission. For voice communication circuit switching is the norm where by a path through connected circuits is established. While this method is suited to telephone traffic, it is not so, economical or efficient for data which is transmitted at different bit rates and in varying amounts.

### ***Message Switching***

This is a different technique associated with switching. According to this method, the system accepts all messages, stores them in a computer, and transmits them as and when channels become available.

### ***Packet Switching***

This is one of recent techniques of switching. Packet switching was originally developed for use by ARPA network by the Defence Department of US. According to this method, each message is broken into packets of approximately 1000 bits each. Each packet carries information about its origin and destination. Packets are stored and forwarded by means of computers in a steady stream, but not in any order. Specially designed computers sort out the packets at each node in the transmission network and retransmit them along the best path to their destination, where they are put in the right sequence for decoding. Because of the standard size of packets, transmission is highly efficient and less prone to error. As a result, economies are effected which benefit the end user. Packet switching is the technology used in value-added networks like Tymnet and Telenet, which provide libraries with connections to many database search services. The advantages of packet switching include transmission economies because of shared utilization, high quality, error reduced service, and the facilities for speed changing and procedures conversion thus allowing communication between different types of terminals and the interconnection of networks.

### **Networking and Telecommunications**

Computers can now communicate with each other and with a range of peripheral devices, over distances with increasing speed and reliability. Technological advances in worldwide communications have ushered in a new era not only of computing power but also of access to information services at least in those countries which have been able to develop extensive and reliable telephone networks. The transmission over telephone network, not only of voice, but more recently also of text and illustration through telefacsimile, laid the foundations for this age in which communication is taken for granted in developed countries. Predictions made earlier,

that rapid growth in computer networking would be the emphasis of the 1990s seem to be borne out in the developments of networks at local, regional, national and global levels. Some of the developments are discussed in the following paragraphs.

## **Networks**

There are different interpretations for the term 'network'. The Oxford English Dictionary states that 'a network is an interconnected chain or system of immaterial things'. On the other hand, the National Commission on Libraries and Information Science (NCLIS) USA defines network as: "Two or more libraries and/or other organisations engaged in a common pattern of information exchange, through communications, for some functional purpose. A network usually consists of a formal arrangement whereby materials, information and services provided by a variety of types of libraries and/or other organisations are made available to all potential users. Libraries may be in different jurisdictions but agree to serve one another on the same basis as each serves its own constituents. Computers and telecommunications may be among the tools used for facilitating communications among them".

UNISIST II working document defines Information Network as: "A set of interrelated information systems associated with communications facilities, which are operating through more or less formal agreements and institutional arrangement in order to jointly implement information handling operation with a view to pooling their resources and to offer better services to user. They generally follow identical or compatible procedures." (UNISIST II).

---

## **1.9 PROGRAMMING AND DATA PROCESSING**

---

### **Data Processing**

Data are a collection of facts-unorganised but able to be organised into useful information. Processing is a series of actions or operations that convert inputs into outputs. Hence data processing can be defined as any process that converts data into information. The processing is usually assumed to be automated and running on an electronic computer. According to Encyclopaedia Britannica data processing is manipulation of data by a computer. It includes the conversion of raw data to machine-readable form, flow of data through the CPU and memory to output devices, and formatting or transformation of output. Any use of computers to perform defined operations on data can be included under data processing. In the commercial world, data processing refers to the processing of data required to run organisations and businesses.

We have established that organisations need information; the production of this information is known as data processing. Data can be defined

### **NOTES**

## NOTES

as a set of symbols that are used to represent objects, events or activities within the real world as we know it. We can refer to people by their name, *e.g.*, barry or we can give them a number such as a student number. In computer based systems, where precision in identifying objects is required, we usually find that numbers are used to identify them partly because numbers can be easily represented within the computer and partly because they can be readily manipulated.

Just as objects can be coded, so can events and activities. For example, when you make a purchase in a department store, that activity can be coded by its date, the amount involved, the code number of the item purchased, and the type of transaction (*i.e.*, whether you paid by cash or by credit card).

Normally, data can be seen as a collection of meaningless symbols until they are processed. The purchase in the department store mentioned above might be recorded as

19047900126913876547

which is just a string of digits until use is made of it. The processing of data involves the execution of various operations on the data (*i.e.*, manipulation of the data) into a form which is meaningful to a human being or another machine. When it is meaningful, it is usually described as information.

Processing can include arithmetic operations (add, subtract, multiply, divide), logic operations (checking, testing, comparing) or simple movement operations (editing, transmitting, displaying, recording). Thus data processing can be considered to be the manipulation of symbols to produce information that is useful to the recipient.

### ***Data Processing System***

When we speak of a data processing system, we are concerned with the set of elements which come together in a systematic way to process data into information; these elements will consist of rules, procedures, files, computers, pens, paper, and above all, human beings. Humans are involved at various stages of the processing, and often machines are involved also. People provide input to computers, operate them, and use their output. Materials, such as boxes of paper and printer ribbons, are consumed in great quantity. Facilities are required to house the computer equipment, people, and materials. A typical data processing system in an organisation is the payroll system; in simple terms it collects data about hours worked, relates it to rates of pay and calculates wages. In addition to producing pay slips, this data processing system can also generate information for management decision-making about job costs, manpower scheduling, absenteeism, recruitment needs, and so on.

Usually a data processing system is based around a set of files, containing all data relevant to a particular task. A manufacturing organisation will have files of data about the structure of its products; a local authority

**NOTES**

will have a file of data about its electors; a bank will have a file of data about customer accounts; and so on. The data is filed because it is anticipated that it will be required for some future purpose. For example, the product structure file will be used to plan production; the register of electors will be used to send out polling cards; the customer accounts file will be used to provide the current balances of accounts. Most organisations will have lots of files and frequently it is necessary to access more than one file in order to acquire the necessary information. For example, in order to plan production, not only is the product structure file needed but also the orders file, the raw materials file, the machine availability file, and so on.

In this context, data is defined as a collection of numbers or characters that represent a measurement from the real world. Information is defined as a meaningful answer to a query. Data processing or information processing, operations (*e.g.*, handling, merging, sorting, and computing) are performed upon data in accordance with strictly defined procedures, such as recording and summarising the financial transactions of a business. In automatic or electronic data processing the operations are performed by a computer. In distributed data processing some or all of the operations are performed in different locations at computer facilities connected by telecommunications links.

Data processing also consists of a number of stages. Data processing begins with data entry. By entering data from the checkbook ledger into an accounting book or computer program in accordance with the chart of accounts it begins to take on structure. However, raw data provides little information, even when organised, without being further processed. The financial data that was collected in the checkbook ledger and entered into an account book has little meaning until all the figures are processed. Account entries must be summed for each account category; expenses subtracted from income, profits added to equity, ratios calculated and so on.

### ***Methods of Processing Data***

Computers are only one of several methods by which to process data. Data can be processed manually, mechanically, electromechanically, and electronically.

#### **Manually**

Originally, all data was processed manually (by hand). Even prehistoric cave dwellers kept records by drawing pictures and symbols on cave walls to record such information as the number of animals they had killed for food. Throughout history, people have processed data manually. When small amounts of data are involved, this is often the most efficient method. You are processing data manually when you take notes in class and when you write a check to pay a bill. By 3500 B.C., the

## NOTES

Babylonians were using clay tablets to record information and to improve communications. About the same time, the Egyptians developed papyrus (a crude type of paper) and the calamus (a sharp pen) for use in recording information. One of the first calculating devices was the abacus. Prior to this century, most data processing was done manually and main tools used to process data were pens, pencils, rulers, journals, and ledgers.

### **Mechanically**

Use of a manual typewriter is an example of processing data mechanically (by machine). The typewriter, first introduced in the early 1900s, is a mechanical recording device which increased writing speeds and improved legibility.

Blaise Pascal, a Frenchman, led the way in mechanical processing with his development in 1642 of the first mechanical adding machine, called Pascal's Calculator.

In the early 1800s Charles Babbage, an English inventor and mathematician, decided to build a machine that could solve mathematical equations. He designed a model of what he called a Difference Engine on paper. Then he went on to design the Analytical Engine in 1856, another mechanical machine capable of performing even more complex mathematical calculations. The device contained the major elements found in modern digital computers: an input device, a processing unit, a control unit, a storage unit, and an output device. Due to these important contributions, Charles Babbage is now recognised as the "Father of the Computer". Following World War I, customised accounting machines appeared. Most were designed for specific purposes, such as recording retail sales or calculating a payroll, and they even allowed the user to combine processing steps.

### **Electromechanically**

Both mechanical and electronic devices are used to process data electromechanically. Unlike a mechanical device, a truly electronic device does not have internal moving parts.

Professor Howard Aiken of Harvard University built an electromechanical machine in 1944. This machine, called the Mark I, contained both mechanical and electronic parts. Data was stored electronically inside the machine with the use of electromagnetic relays. Mathematical calculations, however, were performed by the machine's internal mechanical counters. This machine was the largest electromechanical calculator ever built. It used over 3000 electrically actuated switches to control its operations. Although its operations were not controlled electronically, Aiken's machine is often classified as a computer because its instructions, which were entered by means of a punched paper tape, could be altered. Many machines used today are electromechanical. For example most computer printers contain both mechanical and electronic parts.

## **Electronically**

A computer is an electronic device that contains no internal moving parts. It consists electronic circuitry carefully designed to allow data to be processed by directing the flow of electricity through the circuitry using tiny "on" and "off" switches.

Even as the Mark I computer was being constructed, work was underway to introduce electronics into the design of computers. Dr. John Atanasoff, a professor of physics at Iowa State College, developed an electronic machine to solve certain mathematical equations. This machine was called the Atanasoff-Berry Computer, or ABC, after Atanasoff and his assistant, Clifford Berry. It used 45 vacuum tubes for internal logic and capacitors for storage.

In 1946, J. Presper Eckert and John W. Mauchly built the first all-electronic computer, called the ENIAC (Electronic Numerical Integrator And Calculator). It was developed as a result of a military need. The ENIAC took up the wall space in a 20 × 40 foot room and used 18,000 vacuum tubes. Though, it was fully electronic, the ENIAC had two major shortcomings: it could store and manipulate only a very limited amount of information, and *its programs were wired on boards*. These limitations made it difficult to detect errors and to change the programs. In recent years, several technological improvements have resulted in the development of modern digital computers with tremendous capabilities.

## ***Basic Data Processing Operations***

Five basic operations which are characteristic of all data processing systems are inputting, storing, processing, outputting, and controlling. These are defined as follows:

1. ***Inputting***. The process of entering data, which are collected facts, into a data processing system.
2. ***Storing***. Saving data or information so that they are available for initial or for additional processing.
3. ***Processing***. Performing arithmetic or logical operations on data in order to convert them into useful information.
4. ***Outputting***. The process of producing useful information, such as a printed report or visual display.
5. ***Controlling***. Directing the manner and sequence in which all of the above operations are performed.

## ***Data Processing Activities***

Thus Data processing consists of three basic activities: capturing the input data, manipulating the data, and managing the output results.

1. ***Capturing the input data***: Data must be recorded or captured in some form before they can be processed. They may first be recorded on source documents — paper forms designed to accept

## NOTES

NOTES

input data — or they may be captured by keyboards, automated teller machines, or other direct input devices that accept input data in a paperless, machine-readable form.

2. **Manipulating the data:** One or more of the following operations may then have to be performed on the gathered data.

(i) *Classifying.* Organising similar items into groups or classes is called classifying. Classifying is usually accomplished by assigning predetermined abbreviations or codes to the items being arranged. The types of codes used are numeric (Postal Zip codes used for geographic classification), alphabetic and alphanumeric.

(ii) *Calculating:* Arithmetic manipulation of the data is called calculating, and is a common processing task.

(iii) *Sorting:* Usually, it's easier to work with data if they are arranged in a logical sequence. Examples include first to last, biggest to smallest, oldest to newest. Arranging classified data in such a sequence is called sorting.

(iv) *Summarising:* Reducing masses of data to a more concise and usable form that is easier to interpret and analyse is called summarising. For example representation of data in the form of charts is one of the many types of computer-generated graphics that are used for this purpose.

3. **Managing the output results:** Once data have been captured and manipulated, one or more of the following operations may be needed:

(i) *Storing and Retrieving:* Retaining data for future reference is storing. Recovering stored data and/or information is the retrieving activity. Data may be stored on magnetic disks as well as paper, microfilm, tapes, optical disks or other storage media. A computer can easily retrieve data stored on these media.

(ii) *Communicating and Reproducing:* Transferring data from one location or operation to another, for use or for further processing, is data communication. This is a process that continues until information in a usable form reaches the final user. Data may be communicated from one location to other locations electronically. And sometimes when it is necessary to copy or duplicate data, the reproduction activity can be done by machines such as printer, scanner, etc.

### ***Advantages of Data Processing***

Although data processing systems that involve computers perform the same functions as those that do not, the differences between the two are extremely significant. Computer-oriented data processing systems, often called computer data processing systems, should not be designed to imitate manual systems. They should blend the capabilities of both humans and

computers. Computer data processing systems can be designed to take advantage of four capabilities of computers.

NOTES

1. **Accuracy:** Once data have been entered correctly into the computer component of a data processing system, the need for further manipulation by humans is eliminated, and the possibility of error is reduced. Also, computers, when properly programmed, are unlikely to make computational errors. These are important reasons why we see so many terminals placed at locations where sales occur, such as fast-foods counters and department store sales stations. These terminals, called point-of-sale, or POS, stations, capture data as the transactions that create the data take place. Of course, computer systems remain vulnerable to the entry by humans of invalid data.
2. **Ease of Communications:** Data, once captured at a POS stations or entered by any other means, can be transmitted wherever needed by communications networks. These may be either earth-or satellite-based systems. A travel reservations system is an example of a data communications network. Reservation clerks throughout the world may make an inquiry about transportation or lodgings and receive an almost instant response. Another example is an office communications system that provides executives with access to a reservoir of data, called a corporate database, from their personal microcomputer work stations.
3. **Capacity for Storage:** Computers are able to store vast amounts of data, to organise it, and to retrieve it in ways that are far beyond the capabilities of humans. The amount of data that can be sorted on devices such as magnetic disks is constantly increasing. All the while, the cost per character of data stored is decreasing. Some examples of very large files maintained by computers are those of banks. Also, databases and data base management systems will be widely used by corporations in the decades ahead.
4. **Speed:** The speed at which computer data processing systems can respond adds to their value. For example, the travel reservations systems mentioned above would not be useful if clients had to wait more than a few seconds for a response. Similarly, the response required of a complex computer-controlled manufacturing process might be a fraction of a second.

Thus, an important objective in the design of computer data processing systems is to allow computers to do what they do best and to free humans from routine, error-prone tasks. The most cost-effective computer data processing system is the one that does the job effectively and at the least cost. It is not necessarily the least expensive system. Nor is it necessarily the system with the greatest data processing capability, if that capability is not needed and adds extra cost. By using computer

in a cost-effective manner, we will be better able to respond to the challenges and opportunities of our postindustrial, information-dependent society.

### Programming—Languages

#### NOTES

A computer can neither think nor make decisions on its own. To solve a problem using computers, a set of instructions should be given to the computer in a form which it can understand. The process of writing these instructions is called programming.

Understanding what is programming can be explained by taking a simple and well-known example. Consider searching the name of an author from the catalogue in the library.

Manual solution to this problem is to flip through the catalogue for the initial letter of the name. Once found, search for the subsequent letters of the name and on finding the name, note the details of the book by the author.

Let us assume the catalogue is available in a computer. We need to write a program to solve this problem of searching a word in the dictionary. The algorithm to solve the problem is given below:

1. Take the name to be searched as an input
2. Match the names available in the catalogue one by one until the exact match is found. (It is simplest logic but we can apply more efficient logic)
3. Display the details of the book as output for the input name.

This solution is in natural language which computers cannot understand. Therefore, we need to write it in a language, as a set of instructions, which computers can understand. Such a set of instructions that are arranged in a sequence that guides the computer to solve a problem is called a program.

The process of writing a program is called Programming. It involves *solving a problem and writing it in a form which a computer system can understand and execute*. While considering how to solve a problem, there are three main specifications which should be taken into consideration:

- What information flows into the system?
- What information flows out of the system?
- What the system does with the information?

Programming is a multi-step process and includes the following five steps:

1. *Identify the programming needs:* Specify the program objectives and program users, and state the input, output and processing requirements.
2. *Design the program:* Design the details using pseudocode or flowcharts.
3. *Code the program:* Select the appropriate high level programming language and code the program in that language following its syntax.
4. *Test the program:* The fourth step is to test the program and "debug" it of errors so that it works properly.

5. *Document and maintain the program:* Written descriptions, procedures about a program and how to use it need to be developed.

Computers cannot understand natural languages such as English or Hindi for solving problems. To communicate instructions and commands we need programming languages. And learning programming languages requires learning the symbols, words and rules of the language. There are many languages available to do programming. These are developed for specific applications.

Based on these applications, programs are of two types-*system programs* and *application programs*.

*System program:* These are the programs which makes computer easier to use.

Operating Systems such as DOS and Windows which consist of many other programs for controlling input/output devices, memory, processor, etc., is system program. The *C language* is widely used to develop system software.

*Application program:* Application programs are designed for specific application such as *library automation, payroll information retrieval, processing, inventory control, etc.* These applications should be able to get input, produce output, do calculations and store and retrieve data. There are many languages which contains instructions to support all these operations.

### ***Types of Programming Languages***

Programming languages are designed to meet specific needs. Consequently, one language may be better suited than others for writing certain kinds of functions. There are programming languages that are used for solving engineering problems, writing out complicated sales reports, interact with databases, manipulate graphics or even design web pages. So, with proliferation of programming tasks, proliferation of programming languages have also taken place. Let us see the various families of programming languages and the types of programming languages that belong to them.

#### **Procedural Languages**

Languages in which programs are written, as sequence of statements that manipulate data item and change the contents of memory cells, are called procedural languages (also called imperative languages). The fundamental operation of these languages are storing and retrieving data values. For example,

$$a = 1$$

This statement stores value 1 in location a

$$a, c = a + b$$

This statement retrieves a and b, add and stores the result in c. Some of the procedural languages are discussed below.

#### **NOTES**

NOTES

### Fortran

The name FORTRAN derives from FORMula TRANslation. The name itself indicates that the applications of language with "formula" of engineering-type application. The first commercial version of FORTRAN released in 1957. This makes FORTRAN the first high-level programming language. FORTRAN has some features ideally suitable for problems that are heavily mathematical or computational oriented.

### Cobol

The name COBOL derives from *Common Business Oriented Language*. COBOL was designed to serve business needs such as managing inventories and payrolls. In such applications summary reports are important output. Much of the business world concerns updating master files with change in transaction files. For example, master file contain names, manufacturers, and quantities available for various items inventory; a transaction file would contain names, quantities and items sold out of inventory over a period of time. The master file would be updated from the transaction file on weekly or daily basis to reflect new quantities available and print summary report.

### Pascal

The programming language Pascal was named after Blaise Pascal, the inventor of Pascaline calculator.

The design of Pascal was to easily learn and enforce good programming techniques. Pascal looks very similar to pseudocode. So it is easy to read and the syntax is easy to learn. Even though Pascal is not much useful as commercial language but, programming environment like Delphi uses Pascal language with a facility to develop windows based applications with modern graphical user interface.

### C

C was developed in early 1970s by Dennis Ritchie at AT&T Bell Laboratories. It was originally developed for system programming, particularly for writing operating system UNIX. It is a popular general purpose language for two reasons. One is the relationship with UNIX. Second reason is its efficiency that is, speed with which its operations can be executed. The efficiency derives from the fact that C programs can make use of low-level information such as knowledge, where data is stored in memory, yet has powerful statements and portability to machines that high-level languages offer.

The strong feature of C is that it provides a data type called pointer. Variables of pointer type contains memory addresses instead of integers, real numbers and characters. For example the statement,

```
int intPointer;
```

declares *intPointer* as a pointer variable that will contain the address of a memory location containing integer data.

```
int A = 3;
```

```
intPointer = & A.
```

The first statement declares a variable called A and assigns the value 3 to that. The second statement makes the *intPointer* pointing to A. We can also assign the integer value 10 to variable A by the statement

```
intPointer = 10;
```

The powerful use of pointers are in the areas of writing system programming, operating systems, assemblers, and programs that allows the computer to interact with input/output devices and so on.

For example, consider a problem of writing a device driver (software to interact with a hardware device) for mouse on a PC. The port in which the mouse is connected reads the changes in mouse position by voltage levels. It stores voltage levels in a fixed location of memory. C provides facilities to access this memory and manipulate it.

Even though C is powerful for system programming it is widely used for general purpose programming also.

### **ADA**

ADA development was for a common high-level language needed for various branches of United States armed services for defense contracts. ADA is a large language like C++ and has been accepted not only in the defense industry, but for other technical applications and is a general purpose language as well.

ADA is known for its multi-processing capability—the ability to allow multiple tasks to execute independently and then synchronize the communication when directed.

### ***Object Oriented Programming Languages***

Object Oriented Programming started by Alan Kay's work at Xerox Palo Alto Research Centre in early 1970s. The resulted language is smalltalk.

The concept of Object-oriented programming is dividing the major task into sub tasks. The program can be thought of a giant statement executor designed to carry out major tasks, even though the main program may simply call the various modules that do the sub task work.

To understand objects, let us take an example of library system. Circulation, Acquisition, Catalogue, etc., are major objects. Each object is an example of a task performed in library system.

### **Special Purpose Languages**

The procedural languages that we saw so far are more or less general purpose languages. There are several special purpose languages designed

### **NOTES**

for specialized tasks. Here we take three representative special purpose languages, which are very popularly used.

### **Structured Query Language (SQL)**

#### **NOTES**

SQL is designed to be used with databases, which are collections of related facts and information. A database stores data; the user of the database must be able to add new data and to retrieve data already stored. For example, a database of library catalogue may contain titles of book, its author's name, publisher name, year of publication, etc. The database of catalogue user should be able to add information on a book and to retrieve information on a book already in the database. In addition, databases can also be queried with the user posing questions to the database. For example, the library catalogue could be queried to reveal books by a particular author or by a publisher. Such queries may be framed in SQL. SQL is the language used to frame database queries. Sample SQL statement is as follows:

```
SELECT BOOK_TITLE, PUBLISHER, YEAR  
FROM CATALOGUE  
WHERE AUTHOR NAME = "GUHA";
```

This SQL statement will retrieve all the book titles with publisher and year details of the author Guha.

Examples of prominent database management systems that use SQL are Visual FoxPro, Oracle, MS-Access and SQL Server. Visual FoxPro is a commonly used database management system in libraries. For example, the library catalogue can be created in Visual FoxPro and the SQL can be used to query the database by name of the author, title of the book, keyword, etc.

### **Practical Extraction and Report Language (PERL)**

PERL is designed to scan arbitrary text files, extract various kinds of information that is contained within the text, and print reports based on the extracted information. The language syntax is somewhat based on C. Perl uses sophisticated pattern matching techniques to speed up the process of scanning large amounts of data for a particular text string.

### **Hyper Text Markup Language (HTML)**

This is the language used to create HTML documents that, when viewed with Web browser software, become Web pages. An HTML document consists of the text to be displayed on the Web page, together with a number of special characters called tags that achieve formatting, special effects, and references to other HTML documents. Tags are enclosed in angle brackets (<>) and often come in pairs. The end tag, the second tag in the pair, looks like the begin tag, the first tag in the pair, but with an additional / in front.

The format for an HTML document is

```
<html>
<head>
<title> text that is appear as title </title>
</head>
<body> text that is to appear as body on the page
</body>
</html>
```

NOTES

---

## 1.10 SUMMARY

---

- This Unit introduces you some basic concepts necessary to understand a modern computer. It explains to you what a computer is and how over the years, it has developed into its present form. Categories and classes of computers and their specific characteristics have been discussed to provide you the backdrop required for a proper understanding.
- In this Unit, an attempt has been made to explain to you the basic terminology and concepts in telecommunications. The main emphasis, however, has been particularly on data transmission, communication links, switching mechanisms, terminals and networks.
- The focus has been on basic concepts and simple explanations as far as possible, with examples easily understood by persons not acquainted with telecommunications. Information professionals must remain abreast of new technological advances as they become practical and affordable to ensure that their telecommunications needs are met by the best possible combinations of telecommunications processes and devices at the lowest cost possible.

---

## 1.11 REVIEW QUESTIONS

---

1. Explain in a few sentences the main characteristics of a computer.
2. Explain briefly what you understand by the expression 'computer generations'.
3. How are computers categorized or grouped?
4. Describe the basic components of a modern computer system and explain their role in its working.
5. Briefly explain the significance of different types of computer storage media and indicate their use.
6. What do you understand by the term 'Software'? Briefly explain the difference between 'Systems software' and 'Application software'.

NOTES

7. What is an operating system? Explain in simple terms the difference between single user operating system and multi-user operating system.
8. Explain the difference between simplex, Half-Duplex and Full-Duplex modes of transmission.
9. What is the difference between synchronous and asynchronous transmission?
10. What is meant by the acronym modem? Explain its use in transmission process.
11. Explain the meaning of the term 'Switching Mechanism' and mention the advantages of 'Packet Switching'.
12. List the four methods of data processing.
13. Briefly discuss the primary historical contributions of Pascal, Babbage, Eckert and Mauchly, and Alan Turing and Berry to data processing.
14. What is data communication? Explain the components that are typically found in a simple data communication system.
15. Illustrate asynchronous and synchronous data transmission modes. Which of these is used transmission modes, which of these is used in high speed data transmission and why?
16. What is LAN? What is the difference between a star LAN and a Bus LAN?
17. Explain the difference among a LAN, a MAN and a WAN.

---

### 1.12 FURTHER READINGS

---

1. Arora, Jagdish (2001). Building Digital Libraries : An Overview. *Desidoc Bull of information Technology*. 21;3-24.
2. Harter, Stephen P (1997). Scholarly Communication and Digital Library: Problems and issues. *Journal of Digital information*. 1:4-14.
3. Army, W Y (1995). Key Concepts in the Architecture of the Digital Library. *D-library Magazine*. July. URL: [http://www.dlib.org/organization/July 95/07 arms.html](http://www.dlib.org/organization/July%2095/07%20arms.html).
4. Chapman, S and Kenny AR (1996). Digital Conversion of Research Library Materials. A Case for full informational Capture. *D-Library magazine*. Oct 1996. URL: [http://www.dlib.org/organization/dilb/october96/10 chapman/html](http://www.dlib.org/organization/dilb/october96/10%20chapman/html).

---

# UNIT II LIBRARY AUTOMATION AND INFORMATION SERVICES

---

*Library Automation  
and  
Information Services*

NOTES

## ★ STRUCTURE ★

- 2.1 Introduction
- 2.2 Library Application Software
- 2.3 Computerised Acquisition Subsystem
- 2.4 Computerised Cataloguing Subsystem
- 2.5 Summary
- 2.6 Review Questions
- 2.7 Further Reading

---

## LEARNING OBJECTIVES

---

After going through this unit, you should be able to:

- Explain different modes of adopting a library application software.
- Describe the history of library Automation and its present Service.
- Describe the Popular library Automation software.

---

## 2.1 INTRODUCTION

---

A computer system has two basic units—physical unit *i.e.*, the hardware and logical unit *i.e.*, the software. The whole array of software can be grouped into two fundamental categories—system software and application software. System software are responsible for the overall management of computer resources whereas application software are designed to perform certain tasks and thereby make computers able to perform different predefined jobs. Library automation software, as application software, performs day-to-day library activities through human interventions. Library automation packages are developed in view of the two most essential activities of any library—housekeeping and information retrieval. An automated library is one where a computer system is used to manage one or several of the library's key functions such as acquisitions, serials control, cataloguing, circulation and the public access catalogue. Automated library system centers on library automation package. Such automation packages are also called 'Library

Management Software (LMS)'. LMSs are now established as an essential tool in the support of effective customer service, stock management and management of services offered by libraries. These are based on knowledge and experience of library professionals over the centuries.

NOTES

---

## 2.2 LIBRARY APPLICATION SOFTWARE

---

Library automation process can either adopt a piecemeal approach, converting individual operations one at a time, or it can go directly for a fully integrated system. Sometimes the process can work towards the integrated system progressively, using a 'planned installation' approach.

### Library Automation Packages: An Overview

Software may be viewed as digital version of human knowledge. LMSs are based on knowledge and experiences acquired by library professionals over centuries. These are used as intelligent tools for performing housekeeping operations, information retrieval and MIS (Management Information System) activities. Current LMSs are integrated systems based on relational database architecture. In such systems files are interlinked so that deletion, addition and other changes in one file can automatically activate appropriate changes in related files. The market place for LMSs is now a matured one in India. Almost all special libraries and large academic libraries in India have either adopted a computer-based system or planning actively to go for library automation.

### Development of Library Automation Software

Software upgradation is a continuous process. LMSs are no exceptions. A critical study of development of LMSs over the years suggests that LMSs may be divided into four generations on the basis of sophistication of their facilities for integration and interconnectivity. The LMSs developed in all parts of the world from mid 1970s till date may be fitted into one of the four compartments described below:

- The first generation's LMSs were module-based systems with no or very little integration between modules. Circulation, cataloguing modules were the priority issue for these systems and were developed to run on specific hardware platform and proprietary operating systems;
- The second generation LMSs became portable between various platforms with the introduction of UNIX and DOS based systems. The LMSs of this generation offered links between systems for specific function and were command driven or menu driven systems;
- The third generation LMSs were fully integrated systems based on relational database structures and client-server architecture. They embodied a range of standards, which were a significant step towards open system interconnection. Colour and GUI features, such as windows, icons, menus and direct manipulation became standards and norms in this generation; and

- The fourth generations LMSs were based on web-centric architecture and facilitated access to other servers over the Internet. These systems are UNICODE complaint and allow accessing multiple sources from one multimedia graphical user interface.

Thus we see that the progress of LMSs through the generations was towards an effective and straightforward user interface, which supports access to multiple sources and services from one multimedia interface. Moreover, the latest LMSs allow customised report generation, data manipulation, and investigation of various scenarios. Therefore, they have all the potentials to be a decision support tool.

## NOTES

### **Features of Library Automation Softwares Available in India**

The automation of library activities in India started in full swing with the introduction of CDS/ISIS. CDS/ISIS is a menu-driven generalised information storage and retrieval system designed by a team of experts under UNESCO/ PGI programme. It is specifically meant for the structured non-numerical databases. In India, erstwhile NISSAT (ceased existence since 2004) with the help of other professional bodies organised a number of training courses on application of CDS/ISIS (DOS and Windows version) in information organisation activities. As a result, a large pool of trained manpower developed all over the country. Some organisations from the experience of use of CDS/ISIS, MINISIS etc., developed their own LMSs e.g., DESIDOC developed DLMS (Defence Library Management System), INSDOC (now NISCAIR) came with CATMAN (Catalogue Management) and SANJAY was developed by DESIDOC under NISSAT project by augmenting CDS/ISIS (Version 2.3) for library management activities.

### **Library Automation History**

The effects of ICT on libraries and information centers are characterised by:

- Mechanisation—doing what we are already doing more efficiently;
- Innovation—experimenting with new capabilities, that the ICT makes possible; and
- Transformation—fundamentally altering the nature of the library operations and library operations and services through the capabilities extended by ICT.

It could be said that library automation began in 1930's when punched card equipment was implemented for use in library circulation and acquisition. The computer technology first arrived in libraries in the late 1960s in the form of locally developed software or commercial packages intended to automate library processes. The whole phase of library automation development i.e., 1960 to date may be grouped into four distinct eras:

## NOTES

**First Era:** This era is characterised by computerisation of library operations by utilizing either commercial automation package or software developed in house. The development of shared copy-cataloguing system is another significant achievement of this phase that utilised computer and communication technologies for collaboration and cooperation within the library community.

**Second Era:** This period of library automation is characterised by the rise of public access i.e., the arrival of OPAC as a replacement for the traditional card catalogue. This period also saw major developments in online access to abstracting and indexing databases, union catalogues, resource sharing networks and library consortia.

**Third Era:** This era is characterised by the full text access to electronic documents over high-speed communication channels. The advent of Internet as global publishing platform and largest repository of information bearing objects revolutionised the ways and means of delivering library services.

**Fourth Era:** It is known as 'networked information revolution' era. This era supports a vast variety of digital contents and services that are accessible through the network at any time, from any place, can be used and reused, navigated, integrated and tailored to the needs and objectives of each user. Digital libraries, multimedia databases and virtual libraries are major achievements in the present era.

### **Library Automation Packages of Foreign Origin**

This group includes library automation packages developed by foreign or multinational vendors and distributed in India either through approved agents or value-added resellers. The group covers four most popular packages of foreign origin, which are discussed below.

#### **Alice for Windows**

This LMS developed by Softlink International, Australia, is a global software package and is marketed worldwide through a number of agencies based in America, Australia, Britain, Iceland, India, Malaysia, New Zealand and Singapore. This software is marketed under the name of Embla in Iceland, Alice elsewhere in Europe, OASIS in South East Asia and Australia, Annie in America and other parts of the world. Recently Softlink International decided to call the software Alice for Windows all over the world to maintain consistency in nomenclature. The main features of Alice are as follows:

- It has four distinct versions—Public library version, Special library version, Academic library version and School library version.
- The package is modular and modules are grouped into one of the three sets as mentioned below:
  - *Standard Set:* Includes Management; Reports and Utilities; Circulation; OPAC.
  - *Advanced Set:* In addition to standard set it includes Acquisition;

Periodicals; Journal Indexing; Multimedia; Web Inquiry.

- *Special Set:* In addition to Standard and Advanced set, it includes Reservation; Interlibrary loan; Patron self checking; Rapid retrospective conversion; Multilingual features; Self circulation; Union catalogue.

## NOTES

- The LMS is backed by a number of support services which include on site training programmes, continued R & D, feedback system through user groups, free newsletters etc.
- Besides traditional library materials, it can be used to manage slides, audio and video cassettes, paper clippings, maps, charts, electronic documents and websites. Location of documents in library can also be seen with the help of the library map. It is possible to maintain consistency in recording items through the use of authority files. Alice has a capacity of holding 99 lakhs records.
- It supports a total of eleven search criteria to search the database from any machine (UNIX/MAC/Apple etc.) through Internet or Intranet. It helps to generate customised reports in addition to 800 preformatted reports available with standard set.
- It supports barcode technology and has inbuilt communication function. As special features, the LMS provides data protection function, rapid retro conversion facility and online tutorial and help system.

## **BASISplus and TECHLIBplus**

BASISplus and TECHLIBplus are products of Information Dimensions Inc. (IDI), USA. National Informatics Centre (NIC), New Delhi is the value-added reseller of these packages in India. BASISplus is a client-server relational database system for text and mixed object documents that adhere to fundamental principles of open systems including interoperability, portability and scalability.

The database engine provides user authentication, document access control, concurrency control, deadlock protection and recovery. The features of the LMS are as follows:

- Relational DBMS.
- Client-server architecture.
- Active data dictionary.
- Enhanced security feature.
- Complete backup and restore capabilities.
- Power search facility.
- Full text retrieval.
- Mixed object management.

## NOTES

- Thesaurus and controlled vocabulary.
- Screen customisation.
- Document converters.
- Immediate and Deferred updating (online and batch).
- Content based retrieval.
- Component-level retrieval and image management.
- Networking (LAN and WAN).
- Seamless Internet support.
- Open Application Programming Interface (API) to support client access to server databases.
- GUI based easy user interface for retrieval, display and data entry.
- Intelligent search assistance and thesaurus manager.
- Converter technology which allows user to import and export over 30 different word processor file formats.

TECHLIBplus is a comprehensive library automation package developed over BASISplus and customised to perform all the operation and activities of a fully electronic library. TECHLIBplus supports OPAC, Catalogue maintenance, Circulation, Serials management, Acquisition, Processing and MARC cataloguing. The LMS provides direct access to information in Current Contents.

### **Library Automation Packages of Indian Origin**

This group includes packages designed and developed by Indian vendors and software agents. The features of nine automation packages are discussed here. Packages are selected for discussion on the basis of their customer base and popularity.

#### **AUTOLIB**

AUTOLIB is fully integrated multi-user software on Windows Environment, designed to automate various activities of University Libraries, College Libraries, R & D Libraries, Public Libraries and Special Libraries. The software is developed by Auto Lib Software Systems, Chennai and the product range includes:

- MS-Access with Visual Basic Version
- MS-Access with Visual Basic & WEB Edition
- MS-SQL server with Visual Basic Version
- MS-SQL server with Visual Basic & WEB Edition
- MS-SQL server with Visual Basic

The LMS is module based system, designed and developed by a team of library and information specialists, system analysts, software professionals, network specialists and database designers. The features of AutoLib may be listed as:

NOTES

*General features:* Module based, User Friendly, GUI Environment, Based on client server architecture, Uses Visual Basic 6.0 as front end and MS-SQL V.7.0 RDBMS as back end, Uses TSQL Query Language, Module level Security, Z39.50 Protocol Support, Export/Import of data in ISO 2709 format, Cataloguing of digital resources, Implementation of AACR, CCF, Dublin Core, TCP/IP and Dial-up Network support, Web based reports, Menu based operations, Incorporated Mandatory Fields of CCF, Powerful Search Facility/Query Builder, Printing various reports in several formats, Simple data entry, User ID and password protection, Online help/documentation, Continuous product up gradation, Customer support and maintenance.

*Minimum hardware and software requirements:* Server configuration- Pentium II/III 64 MB RAM/4.2 GB HD, 32x CD-Drive/1.44 FDD, Network accessories, SVGA Monitor, Windows NT Server 4.0 and SQL Server 7.0. Client Configuration - Pentium II with 32 MB RAM/4.2 GB HD Windows NT workstation/Windows 9X.

*Modules:* Cataloguing, OPAC/Advanced OPAC, transaction (circulation) management, book ordering/acquisition control, journal/serials control, system administration, report management, article indexing, digital library, Web OPAC.

The functional features of this LMS are as follows:

- Database Management—data entry/updating of database for user, author, publisher, supplier, member, book, journal issues and back volumes, article report, thesis, standard, non-book materials, budget, subject, department etc.
- OPAC—powerful and versatile search facility, Simple search for beginners, Query builders for advanced users, query windows for complicated search, boolean search, field level search (single field/multiple fields, author/title/keyword/subject, accession no /classification, journal name/article name etc).
- Circulation—transaction, issue, return, renewal of books, journals, back volumes, recall, reservation, cancellation, reminders, reports.
- Serials Control—subscription of new journals, renewal of journals, receipts of new issues, reminders for missing issues, invoice processing, payments, browsing issues, reports generation.
- Acquisition Control—duplicate checking, indent processing for new books, book ordering, reminders, receipts of books, invoice processing, payment, budget management.
- System Administration Module—user ID and encrypted password protection, module level security, budget management, stock verification, global updating, fixing due dates, overdue charges etc., holiday maintenance, reports, new additions, catalogue (main/author/title), accession register/bibliography, list of books by author/title/publisher/year, subject/call number (by any order), books

## NOTES

by unique titles, frequently issued books, frequently accessed books, books issued/returned/reserved, receipt for fine amount/ deposit/ loss of book, etc., list of users/publishers/suppliers/departments, no-due certificates, stock verification report, budget details, orders, journal list, journal subscription/order report/missing issues.

- Article Indexing—allows to create journal article database, allows to create author index and keyword index, allows to search and retrieve journal articles, allows to create index and abstracts, allows to publish CAS bulletin, allows to generate contents pages;
- Digital Library Module—allows to catalogue multimedia digital resources such as text, images, audio file, video clippings, etc. Allows to catalogue based on Dublin Core standard, Allows to handle various file formats such as .bmp, .jpeg, .pdf, .doc, .avi, etc.

### **E-GRANTHALAYA**

This LMS is developed by National Informatics Centre (NIC), Bangalore centre to suit the requirements of small and medium sized libraries. It is an easy-to-use software package and supports all the routine library operations. The General and Special features of the LMS are as follows:

- The package includes functional modules for administration, serials control, acquisition, circulation, OPAC, reports and index;
- Generates customised reports and statistics of library usage;
- OPAC allows simple and advance search options, supports web-enabled searching.
- The package has bilingual capabilities and can be customised to suit all Indian languages supported by ISM 2000 developed by C-DAC, provides options to control access through login id and password and supports quit-in privileges for users and staff.
- Supports both stand alone and networked operation mode. Recommended server configuration is Pentium III processor, 128 MB RAM and 4.3 GB Hard disc.
- Requires Windows NT/Windows 2000 for server and Windows 98/XP/2000 for client machines and uses MS SQL server as backend database.
- Requires ISM2000/Leap office 2000 as bilingual tool.

### **GRANTHALAYA**

This CUI based (DOS & UNIX) modular LMS is developed on FoxPro by INSDOC (now NISCAIR) for medium range libraries. It includes all the modules required for day-to-day library operations. The package is made of seven modules-library administration; query; circulation; acquisition; serials control; technical processing and data administration. The salient features of the LMS are:

- Based on object oriented design.
- Supports CCF and ISO 2709 for import and export of data.
- Supports Boolean operators and range searching.
- Provides online help through screen messages.
- Generates a dictionary for various data elements for easy searching.

## NOTES

### **LIBSUITE**

This GUI or CUI LMS, developed by SOFT-AID Computer Ltd., Pune, is based on web-centric architecture and designed to work with different media. LIBSUITE is based on three-tier web centric architecture in which server machine uses Windows NT/2000 and Internet Information Server (IIS)—as web server. The database server relies on Oracle 8i and clients use web interface for accessing server through Internet or Intranet. LIBSUITE extensively uses latest technologies like Active Server Pages (ASP) and Component Object Modeling (COM). The web-centric architecture ensures that any machine with a web browser can be a client. It also ensures working independent of operating systems in client machines. The following are the significant features of LIBSUITE:

- It provides all the standard modules and supports customised report generation and standard protocols Z39.50;
- The package bundles follow fully featured modules-acquisition, cataloguing, circulation, queries, serials control, set up and maintenance.
- Cataloguing module, apart from supporting regular activities manages multi-format materials, generates entire status *i.e.*, total number of books, number of books issued and number of books available on stack.
- Circulation module supports all the required operations including ILL and generation of photograph of the member in circulation panel.
- System administration module supports controls over the access, creation of authority entries and setting of parameters for cataloguing, circulation, etc.
- Supports stock verification and global addition and deletion.
- Acquisition module supports all media and production of accession register.
- Web-centric architecture ensures use of any machine as client as it does not require the installation of client-side software.
- Provides easy user interface and ensures seamless navigation through Intranet, and login and password based access as security measure.
- Includes various utilities like calculator and calendar.

## LIBSYS

### NOTES

LIBSYS is a fully integrated multi-user library management system based on client-server model and supports open system architecture, web-based access and GUI. This indigenous LMS is designed and developed by LibSys Corporation, New Delhi. LIBSYS has seven basic modules - Acquisition; Cataloguing; Circulation; Serials; OPAC; Web-OPAC and Article indexing. The leading features of the different LIBSYS products such as LIBSYS 4.0, LS-Premia, LS-Digital, LSmart and LSEase are as follows:

- Based on client-server model and TCP/IP for communication and networking.
- Provides ANSI Z39.50 compliant web access for making the server accessible through Internet/Intranet.
- Supports web OPAC for accessing bibliographic databases through Internet/Intranet.
- Supports standard bibliographic formats like MARC 21, UNIMARC, CCF etc.
- Includes images and multimedia interfaces with LIBSYS search engine.
- Supports barcode technology for membership card production and circulation.
- Offers SDI, CAS, fine calculation, e-mail reminders etc.
- Provides flexibility in choosing operating platforms (UNIX, Windows, NT, Novell NetWare) and backend RDBMS (SQL sever, Oracle).
- Supports Web-OPAC through PERL/CGI access mechanism.
- Offers a range of products suitable for different types of organisation e.g., LSEase is an affordable solution for small and medium sized organisation.
- Provides RFID technology (in cooperation with TAGSYS—the largest smart card solution provider) for inventory management and smart card technology integration for identification of individuals.
- The LSPremia provides full UNICODE support and management of multi-site libraries.
- Offers digital resource management system through LS-Digital suite. It supports resource structure definition, scanning, PDF conversion, multimedia database management and metadata based searching and retrieval.
- LIBSYS supports operating systems Unix, Windows NT and Novell NetWare in client-sever environment using TCP/IP protocol. It does not require any RDBMS as it uses proprietary database but user library may opt for SQL server or Oracle as backend RDBMS. The server hardware may be Pentium; Alpha, HP-9000 or Sun (SOLARIS) and client machine may be Windows 9x/NT, X-window or Web-client.

## NEWGENLIB

NEWGENLIB is the result of collaboration between a charitable trust called Kesavan Institute of Information and Knowledge Management (KIKM), Hyderabad and Verus Solutions Pvt. Ltd. The following are the features of NEWGENLIB.

### NOTES

- Completely web based, and adheres to International standards, supports web services and allows networking of unlimited number of libraries, database and operating system independent and uses open-source, *n*-tier, and Java based technologies for scalability, reliability and efficiency.
- It includes seven modules: cataloguing, circulation, acquisition, serials management, OPAC, network configuration, administration and setup, and available as single user (small library version), multi-user, single library LAN/Intranet version, multi-user single library web version, multi-user, multi-library consortium version.
- It adheres to international standards like MARC 21, ISO 2709 (Communication format), and AACR-2R. Cataloguing database design is based on well proven database design to adhere to MARC and also supports Unicode 3.0 and UTF-16 encoding format, by which it can support all the possible languages (if available in Unicode 3.0 code charts).
- It supports functionalities like—Import of MARC data from sources such as OCLC and freely available webbased resources, extensive use of setup parameters in configuring the software to suit specific needs, *e.g.*, in management of fines, multi-user and multiple security levels, automated email facility integrated into different functions of the software to ensure efficient communication between library and users, vendors, module-specific querying in all modules.
- Acquisition supports—Online requests by users, firm orders, on-approval purchases, standing orders, solicited gifts, unsolicited gifts, exchange-triggered acquisitions, web service interfaces to supply sources such as amazon.com, management information reporting to enable better decisions in acquisitions management.
- Cataloguing supports data-entry using MARC tags, fields, sub-fields, etc., or simple label and form based data-entry; import of MARC records from sources such as OCLC or from free MARC download sites on the web; access to authority files during data entry and catalogue database searching; catalogue record attachments enabling access to related data, *e.g.*, multimedia, web-based resources, scanned images, and full text digital documents; provision of a search engine to search full text documents, plug-ins for specialised thesauri; automatic validation etc.
- Network functionalities support sharing of hardware, server and application software between the host and one or more associate

## NOTES

libraries. It helps users of branch libraries: (i) to download metadata or full text of records, where records are available, into their desktops, (ii) in acquisition of new publications from the host library, (iii) to access their circulation records, (iv) to access electronic journals across all the libraries in the network, (v) to improve services to both the end user and the library staff.

- OPAC option supports—browser-based access to the library's catalogue database, extensive search, retrieval, display, print, download and formatting options for patrons (customised, text format (brief), text format (Full), MARC tagged, ISO 2709, MARC-XML, Dublin core). Patrons can request new additions, access their circulation data, make reservations and go to the web via the OPAC, patrons can trigger interlibrary loans, interact with library staff via instant messages/e-mail.
- Circulation, apart from traditional functions, supports—setting of a wide range of circulation options, fines, user privileges, etc., needed in different library environments, rapid charging, discharging, renewal and reservation operations, built-in traps for delinquent users, reservations, etc., on-the-fly circulation, interlibrary transactions, binding management, management information reporting for better management of collection and assistance in stock verification.
- Serials control module includes facilities like—integrated management of serials subscriptions, registration, cataloguing and binding, rapid registration of incoming serials using a kardex-like interface, batch and on-demand claiming for missing issues, support for Union catalogues, MIS reporting for better serials management.

## NEXLIB

This window based LMS is designed and developed by NexEvolue Logic Solutions Pvt. Ltd. It provides a simple point-and-click navigation interface backed by a powerful database engine capable of maintaining millions of records. Nexlib provides all the basic utilities required for the management of libraries namely acquisition, cataloguing, circulation, serials control and OPAC. The important features of NexLib are as follows:

- Acquisition module is fully integrated with the cataloguing module. It can manage a variety of library materials and fund accounting.
- Cataloguing module helps to define location of items by floor, shelf number etc.
- The entire circulation task can be carried out from one screen.
- Provides facility to create unlimited number of user types and member data can be transferred from any existing user information system.
- OPAC can be accessed through any standard web browser.

**NOTES**

- Generates over 50 pre-defined reports related to library MIS.
- Ensures easy-to-use library staff interfaces.
- Full-featured serials control with the support for variety of formats and advance search option for OPAC.
- Digital Media Archive (DMA) module for the management of full text articles, newspaper reports, images etc.
- DMA allows incorporation of standard metadata scheme and metadata-based retrieval.
- NexLib is a platform-specific LMS and profoundly depends on Microsoft products. The hardware and third party software requirements are: Server should be Pentium III or higher with minimum of 256 MB RAM and pre-installed with windows 2000 server. Client machine should be Pentium II or higher with 128 MB RAM and preinstalled with Windows 98/NT/2000/XP.
- Follows client/server architecture and security of library database access is based on Windows NT model.
- Supports MS Access, MS SQL or Oracle 8i as back end RDBMS and uses Visual Basic (VB) as scripting language.

**SLIM 21**

SLIM (System for Library Information Management) a software suite from Algorithms Consultants Pvt. Ltd., Pune is a module-based LMS that offers wide range of functionality for library management. SLIM 21 is the latest product of the series that succeeded SLIM, SLIM++, SLIMEX and SLIMLX. SLIM 21 supports multiple operating systems and backend databases. The relevant features of SLIM 21 are:

- SLIM21 is a module-based system. The basic modules are acquisition, cataloguing, circulation, serials control, OPAC and article indexing.
- Enterprise module of SLIM21 supports usage statistics, current awareness service (CAS Publish), web aware OPAC (WAOPAC), web proposals for new books, interlibrary loan (ILL) and selective dissemination of information (SDI).
- Supports export/import through MARC/CCF/ISO-2709 standards and downloading of bibliographic data from online databases through DB Bridge module and Z39.50.
- Generates customised reports on screen/printers/RTF or as text/PDF/HTML files with auto e-mailing facility.
- Unicode based LMS that supports multi-script sequencing for Indian scripts.
- Generates shelving order for documents as per colon classification, supports smart card/RFID based circulation and touch chip (biometric) interface for user authenticity.
- Creates library map for easy location of items.

## NOTES

- Provides user-friendly online help and reference manual.
- Supports digital library environment and transformation of bibliographic data into XML, XHTML and DCMES.
- Supports both standalone and network architecture. Minimum requirement for server machine is P IV processor and 256 MB RAM with Windows XP/2000 as OS and SQL server as backend RDBMS. Client machine may be any P II machine with Windows 9X/2000/ME/XP or NT workstation.

### SOUL

The story of SOUL (Software for University Libraries) started with the development of ILMS (Integrated Library Management Software) by INFLIBNET in collaboration with DESIDOC. Two versions of ILMS (DOS and UNIX) were developed for university libraries in India. But with the introduction of GUI based system and other revolutionary changes in the field of computer software, INFLIBNET decided to develop a state-of-the-art, user friendly, Window based system which will contain all the features/facilities available with other LMSs in the market. As a result INFLIBNET came out with a LMS called 'SOUL'. The package was first demonstrated in February 1999, during CALIBER-99 at Nagpur. SOUL uses RDBMS on Windows NT operating system as backend to store and retrieve data. The SOUL has six modules - Acquisition; Cataloguing; Circulation; Serial Control; OPAC; and Administration. The modules have further been divided into sub-modules to take care of various functions normally handled by the university libraries. The features of SOUL are:

- Window based user friendly system with extensive help messages at affordable cost.
- Client-server architecture based system allowing scalability to users.
- Uses RDBMS to organise data.
- Multi-user software with no limitation for simultaneous access.
- User friendly OPAC with web access facility.
- Supports bibliographic standards like CCF and AACR II and ISO-2709 for export import facility.
- Provides facility to create, view, print records in regional languages.
- Supports LAN and WAN environment.
- Available in two versions—university library version and college library version.

---

### 2.3 COMPUTERISED ACQUISITION SUBSYSTEM

---

Computerised library acquisition systems are designed to handle considerable amount of paper work involved in purchasing library resources. The typical functions of computerised acquisition system are:

NOTES

- Print purchase order.
- Maintain book fund accounts and print book fund reports.
- Provide information on orders outstanding, and sometimes on works in process—that is, book received but not yet catalogued.
- Payment works such as generation of vouchers or payment orders, and
- Accessioning of received documents.

Apart from these fundamental functions, modern LMSs also support automatic conversion of foreign currencies, checking for possible duplicates, discount calculation, and provision of information for cataloguing or circulation functions etc. The LMSs also differ from each other in various aspects such as:

- By type of materials handled (most of LMSs handled monographs but some extend supports for serials, electronic resources etc.)
- By type of orders handled (many LMSs do not support standing, blanket or approval order or gifts and exchanges); and
- By type of payment (LMSs differ in dealing advance payment, deposit account or membership).

The ordering and acquisition process involve some basic routine clerical operations, which are applicable to all categories of libraries. As a result, the procedures related to acquisition subsystem have benefited from computerisation. Generally, acquisition subsystem concentrates on monographs and other documents (available in many formats) excluding periodical publications. The basic activities of any computerised ordering or acquisition subsystem can be specified as follows:

- receive records of items to be acquired.
- check whether items requested are already in the library or on order.
- print orders or despatch order electronically to suppliers/publisher.
- check when orders are overdue.
- follow up overdue order.
- maintain a file of records of items on order.
- note the arrival of ordered items.
- process for payment.
- maintain book fund statistics and accounts.
- generate printed and electronic listing of various reports.
- control currency conversions.
- maintain vendor performance reports and statistics.
- The acquisition module of modern library management software should also

## NOTES

- accommodate a variety of materials, including but not limited to—monographs, monograph in series, annual and cumulative indexes, loose leaf materials, supplements, reports, musical scores.
- accommodate and identify items in a variety of formats, including but not limited to—print, microform, film, videotape, audio cassette, CD-ROM, magnetic tape etc.
- record, store and display bibliographic information, acquisition type (order, gift, approval etc.), status (reported, received etc.), library/branch/copy/fund information, invoice information, vendor information, accounting information, requestor information etc.
- provide facilities for unlimited number of funds/budget head, vendors, orders, claims and transactions, and
- accommodate different types of order—regular order, membership, approval, blanket order, deposit account etc.

All the above-mentioned functions of computerised acquisition subsystem may be arranged into three groups for systematic study of each individual task. These three groups are pre-acquisition work, acquisition work and generation of outputs.

### **Pre-acquisition Works**

Acquisition module of any library automation software facilitates a library to procure and process new books from different sources. It includes necessary functions to enhance the collection development of a library. This module checks to detect duplication of records and helps to keep track of library's annual book budget and allocation of budgets under different budget heads. Books can be procured either on the basis of request from a member or books sent by suppliers on approval. All these functions require some essential works that need to be done before proceeding with actual acquisition work. These are termed as pre-acquisition work and may be identified as:

**Creation of master file for supplier:** The acquisition module must incorporate a vendor/supplier file supporting an unlimited number of vendor records including at least the following information—vendor name, address, code, phone, fax, e-mail ID, contact person, vendor discount etc. The system must be capable of printing the entire contents of the vendor file.

**Currency conversion:** This facility is required to assist in procuring foreign documents priced in various currencies of the world (e.g., US Dollar, Euro, UK Pound etc.). The conversion of foreign currencies into Indian rupees is necessary for fund accounting and payment on the basis of the current exchange rates. In this step a small database with the fields like name of currency, conversion rate, country and date is created and maintained.

**Budget process control:** One of the major functions of library ordering

and acquisitions subsystem is to record and to control expenditure from the library's accounts. Funds are committed for spending when orders are placed, and actually spent when the items are received in the library. Fund accounting helps to keep track of library's annual book budget and its allocation.

## NOTES

The management of fund accounting includes four basic steps:

- **Creation of budget heads:** In this step various budget heads are created as per the prevailing practice in the library (e.g., book procurement fund, serial subscription fund, electronic resource procurement fund etc.). Each budget head is described in detail and accessed through a code for easy recall as and when required.
- **Main budget allocation:** This is related to allocate the amount to the main budget along with other necessary information such as financial period, budget head, opening balance and total amount allocated or sanctioned amount. This minimum dataset is to be entered before activation of the budget process in the acquisition module.
- **Budget allocation in different heads:** This step is for receiving the amount in different budget heads.
- **Budget division:** Sometimes it is necessary to divide a budget head into several sub-heads (e.g., a book procurement head may further be subdivided into reference books and text books). This step allows a user to divide the budget into sub-heads or even divide the budget sub-heads further.

### ***Creation of Letter Formats***

An automated acquisition subsystem should generate and print various letter formats such as approval letter, purchase order, cancellation of order, reminder letter, intimation letter, payment letter etc. In this step templates of respective letters are created and maintained by the user.

### ***Creation of Member Database***

This step is to create and maintain a member system. It is required to link and integrate suggestions given by the users (for procuring various materials) with the member database. Creation of member database is based on some master entries. These are—Category and associated privileges, Name of the affiliated institute, Departments/Branches/Divisions/Sections under the institute etc. New members can be added after these steps. Member codes are either generated automatically or may be entered manually as per the practice of the library.

### ***Acquisition Work***

The acquisition work includes the functions that facilitate a library to procure and process new books from different sources. Document can

NOTES

be procured either through request from a member or books sent by vendors/suppliers on approval. The bibliographic details (available from request form or books themselves) provide maximum information to place orders. The flow of acquisition work for document procurement in computerised libraries irrespective of type or size may be represented diagrammatically.

**Group I Tasks**

The work of this group starts with entering bibliographical information and information about requestors from the suggestion slips and books submitted by the suppliers on approval. Bibliographical data given by the requestors in suggestion slips should be verified by consulting book selection tools. The online databases of virtual bookstores on the World Wide Web (WWW) may also be utilised for checking bibliographical information of recently published documents. Bibliographical details of documents received by libraries in ex-gratis are also entered into the database. A library normally receives a large number of suggestions and documents for ordering. The library staff shortlists these requests depending on need, availability of fund etc., by clicking the appropriate option(s) available in the package. Finally a report is generated for all the short-listed suggestions and documents indicating number of copies required, budget code, budget head and unit price of the items requested. The library committee approves the list officially and on the basis of the final approval list, library staff either select or reject the short listed titles. Books on direct approval and gratis items do not have to go through approval process from library committee or any such authoritative body.

Group I	Group II	Group III	Group IV
Processing of data related to suggestions and books on approval	Preorder Searching & Order Processing	Receiving and Accessioning	Processing of Payments
Deals With	Deals With	Deals With	Deals With
- New suggestions	- Preorder	- Receiving of items	- Invoice processing
- Updating of suggestions	- Creation of order	- Accessioning	- Advance payment
- Books on approval	- Order	- Intimation	- Release of payment
- Direct approval	placement and	generation	- Process for
- Selection for approval	print order		payment records
- Check for duplicates	- Cancellation of order		- Budget commitment
- Approval	- Intimation of order status		

- Gratis items
- Intimation of request status
- Reports for approval
- Reminders
- Budget commitment
- Report generation

## NOTES

**Fig. 2.1.** Flow of acquisition work

### **Group II Tasks**

The first step of this group is to select listed vendors (available from master files) for placing orders of approved documents. Order letters are then printed as per the format created in the pre-acquisition stage indicating name of supplier with address, reference number, terms and conditions and expected date of delivery etc. This group also includes the tasks of reordering, reminder generation (for a particular order or to a particular supplier/publisher) and report generation (for ordered items, overdue orders, budget commitment etc.).

### **Group III Tasks**

This group the works of receiving and accessioning of ordered documents. In case of barcode based circulation system, barcode labels for accessioned items are also generated in this sub-module of the package. The requestor or department may be informed about the arrival of requested documents in the library by generating intimation letter from the package.

### **Group IV Tasks**

The work of this group starts with the processing of invoices received from the suppliers along with the documents by entering necessary elements into the database. Release of payment is the next step in which letters/reports containing all the necessary administrative and financial details are generated against supplier or order number or invoice number for requesting appropriate authority (generally Finance Section) to release payment to the supplier. After release of payment, the financial details of payment are entered and stored into the database.

### **Output of Computerised Acquisition Subsystem**

This step includes the tasks of generating outputs from acquisition module for management operations and user services. Every computer-based system includes only three basic operations—input, processing and output. Computerised acquisition subsystems are no exceptions. Data entering and processing tasks in various pre-acquisition and acquisition works are primarily meant for the generation of various outputs in the form of list, reports, letters and statistics. In summary the outputs from a computerised acquisition subsystem typically are:

## NOTES

- List/Report of item(s) requested.
- List/Report of item(s) from supplier/publisher.
- Item(s) selected for approval.
- Item(s) approved by the authority/library committee.
- Item(s) rejected in the approval process.
- List of gratis item(s) received by library.
- Report on request status.
- Printout or soft copy of letters for approval.
- Printout or soft copy of order letters and query letters.
- Printout or soft copy of reminder letters.
- Printout or soft copy of order cancellation letters.
- Printout or soft copy of reordering.
- Letters for adjustment of advance payments.
- Letters to bank for foreign exchange rate.
- Report on order status.
- List/Report of item(s) selected for order.
- List/Report of overdue item.
- List/Report of item(s) actually ordered.
- Reports of budget commitment.
- List/Report of item ordered against advance payment.
- List/Report of item(s) received against orders.
- Letters of intimation (on arrival of documents).
- Printout of accession register.
- Printout of barcode labels.
- List of supplier/publishers.
- List of currency and exchange rates.
- Budget with commitments
- Report of detailed annual budget of library.
- Report of amount received in different budget heads.
- Report/statistics of vendor performance.
- List of recent additions.
- Generation of book cards (in case of integrated ordering and cataloguing system).

### **Advantages of Computerised Acquisition Subsystem**

The advantages of computerised acquisition subsystems in an integrated automated environment are manifold. Such a system can :

- generate financial and statistical reports in the desired format automatically to help planning and management of libraries;

**NOTES**

- ensure quicker and cheaper data processing;
- contribute in the development of integrated library system;
- reduce the workload of processing section;
- minimise routine clerical operations and related paper works;
- lead towards better management and more productive use of library staff;
- help to introduce new user services;
- interact with other library systems/networks to download bibliographical data of items on order;
- produce number of reports, letters, statistics and list to support MIS activities of libraries; and
- communicate different outputs of acquisition works electronically to members, suppliers, publishers etc.

---

## **2.4 COMPUTERISED CATALOGUING SUBSYSTEM**

---

Development of computerised cataloguing subsystem is one of the most important and intense facets of library automation. In an automated cataloguing system the intellectual work of describing an item or document and analysing its subject content has been done by the library staff, whereas machines have been used to generate a wide variety of products derived from such description and analysis. Automated cataloguing systems are useful to control the clerical and technical processes involved, and to promote the exchange and use of cataloguing data locally, regionally, nationally and globally.

Over the last 20 years, databases and online public access cataloguing (OPAC) have gradually replaced the conventional catalogue such as card, sheaf and microform catalogues. The catalogue records have become the central bibliographic record for the library management system. These records are used in the cataloguing subsystem, circulation control and acquisition control. A computerised cataloguing system requires to be standardised for the interchange of cataloguing records. Standardisation and structuring of cataloguing records are based on content designators such as MARC 21 (Machine Readable Cataloguing), CCF (Common Communication Format), UNIMARC (Universal MARC) or other similar standards.

### ***Basic Requirements for Catalogue Module***

In view of the recent developments in ICT sector, a modern LMS should provide appropriate facilities on its catalogue module. The basic necessities of module for machine-readable cataloguing may be formulated with the following requirements:

NOTES

**Bibliographic requirements should:**

- support different types of classification schemes and vocabulary control devices (*e.g.*, subject heading lists or thesaurus and electronic resources);
- support standard bibliographic and authority record formats (*e.g.*, MARC, CCF, etc.);
- enable records to be exported or imported to and from tape, disk or other devices;
- enable a bibliographic record to be retrieved and modified at any time, after entering the data;
- support items like monograph, serial, government document or any other type of materials;
- enable catalogue data to be added, validated, updated and removed online via the workstation; and
- support withdrawal of items, export and import of records.

**Authority control requirements should:**

- be capable to support and incorporate changes in the MARC authority format or other national/international standard formats;
- generate various kinds of references from authority records;
- accommodate:
  - personal, corporate and topical name heading in a name authority file;
  - title, uniform title, and series entries in a title authority file;
  - subject headings in a subject authority file.

**OPAC requirements should:**

- allow both simple and expert searching;
- support users to enter multiple words or phrases to be searched in one, more than one or all fields;
- support Boolean operators within and across all fields such as:
  - OR (either one or both terms must be in the record)
  - XOR (either term, but not both, must be in the record)
  - AND (both terms must be in the same record)
  - NOT (following term must not appear in any record)
- support positional operators
  - SAME (terms must be in the same field)
  - WITH (terms must be in the same sentence within a field)
  - NEAR (terms must be adjacent to one another, but in either order)
  - ADJ (terms must be immediately adjacent to one another)
- support relational operators (less than, greater than, equal to etc.);

NOTES

- include user self services, features including but not limited to reservation of items, self-renewal, cancellation of reserved items, change of address, phone number, creation of personal profile for CAS and SDI etc.;
- enable users to limit searches by: publication year, language, item type, item category, location and access;
- support multilingual features;
- enable searchers to specify which fields are to be displayed and to indicate brief or full display of hits;
- provide facilities to search global resources in the form of web documents, subject gateways, web-OPAC of other libraries, network resources, bulletin boards, discussion forums, news items, book reviews, tables of contents, best seller lists etc.

**Downloading requirements should:**

- be Z39.50 compliant cataloguing system [ANSI/NISO Z39.50 (1995) or ISO 239.50 (1998)];
- enable to capture bibliographic and authority records from any Z39.50 server through Z39.50 client;
- allow local manipulation of captured data.

**Reports and backup requirements should:**

- produce a count of all records added, edited by a specific operator or over a specified time period;
- generate lists, statistics and counts of items added or tabulated by call number, item categories, item location etc.;
- produce a list of all citations with authority file violations;
- support backup of all cataloguing records in suitable media (magnetic, optical etc.), and easy recovery of records at the time of need.

**Basic Concepts Related to Computerised Cataloguing**

Before discussing the works related to computerised cataloguing subsystem, it is essential to introduce you to some fundamental concepts related to database, standardisation of data entry format, authority file, retrospective conversion, shared cataloguing and other relevant ideas. Let us look at them one by one.

**Data:** A collection of raw facts that become information after proper organisation or processing.

**Data element:** A meaningful collection of related characters also called a field or data item. It is the smallest unit of the record that is explicitly identified.

**Data structure:** A logically related set of data that can be decomposed into lower level data elements. It is a group of data elements handled as a unit.

## Computerised Cataloguing Works

The objective of any computerised cataloguing procedure is to create appropriate catalogues. To this end records may be drawn from any of the following sources:

### NOTES

**Title in process.** This provides the cataloguer with a list of titles, which has been already procured and accessioned by the library staff through acquisition module of the package. These records are then upgraded to the cataloguing standard through necessary addition of new data elements and modification of existing data elements.

**Retrospective conversion.** It is the machine-readable cataloguing of old and existing library stock. Here the cataloguing data is first entered on a worksheet or datasheet designed by the library on any standard content designator scheme and then data transferred to the database through manual keying.

Existing library catalogue in machine readable form Development of library OPAC started long back in many libraries of the world and at that time integrated library automation packages were not available. As a result cataloguing data in digital format is available in many libraries which requires to be merged with the catalogue database of newly installed LMS.

Union catalogue Union files of the stock of several libraries, or another shared database may be imported, converted into local standard format and finally merged with the catalogue database.

Commercially available files of MARC records In this process records from external databases may be added from tape, or by downloading directly from the files through network. A further option is to acquire records on CD-ROM or DVD-ROM and to download records from them.

The key features of a cataloguing module are:

- Authority control,
- Data entry:
  - for newly acquired document
  - for existing old stock.
- Downloading.

These basic works of a cataloguing module should also be supported by regular backup of the catalogue database in suitable media. In the family of magnetic storage devices, Digital Audio Tape (DAT) is the most popular in libraries all over the world. DVD-ROM—an optical storage device with a typical storage capacity of 17 GB is coming a big way to replace all other backup media suitable for the storage of library records.

### Authority File Creation and Maintenance

Authority file is essential to control form of index terms or headings, such as author headings, or subject index terms for better retrieval efficiency. Records in this file may be created locally or drawn from externally

NOTES

available files such as the name and subject authority files of the Library of Congress or other agencies. Library automation packages provide facility to create and maintain authority file in the catalogue module. This file is acting as a master database, where entry is to be made once. This gets reflected in various modules of the package. The master file containing authority entries can be consulted during cataloguing, possibly by display in a separate window. The new headings are added immediately to the authority file with an opportunity to review the headings, and the same is authorised for the users to access locally or remotely.

### **Data Entry for Cataloguing**

This facility of the catalogue module of automation packages is utilised for updating and standardisation of bibliographical data elements of newly procured documents and entering bibliographical data of existing stock of the library. Easy and structured data entry form design on the basis of standard content designator scheme is important for local creation of records. An integrated automation package use the same record for cataloguing function as it is used in the acquisition module. In the catalogue module the record is standardised entering additional data elements and rendering of access points with the help of authority file.

The transformation of bibliographical data elements of existing stock of any library into machine-readable form is called Retrospective Conversion or simply RECON. The work of RECON starts with recording of bibliographical data elements on a worksheet. The worksheet is designed as per the internal data format of the automation package. These internal bibliographic data formats are based on internationally adopted standard content designator schemes such as MARC, UNIMARC or CCF. Finally bibliographical data of each document as recorded on the worksheet is entered into the catalogue database.

### **Copy Cataloguing and Downloading**

Computerised cataloguing provides a unique advantage of loading and merging of bibliographic and authority records from external databases. This feature of an automated system leads to a reduction in cataloguing effort and a consequent saving in the unit cost of cataloguing. This mode of shared cataloguing is popularly termed as copy cataloguing and implemented in libraries thorough:

- merging of existing machine readable form of cataloguing data with the catalogue database;
- relevant and appropriate records from external databases (union catalogues or commercially available catalogue datasets) in magnetic or optical media may be added to the catalogue database, after modification to match local requirements (e.g., indication of added entries, references) and the addition of local data (e.g., call number, accession number, location); and

## NOTES

- downloading of bibliographical data of relevant records from any Z39.50 server (version 2 or version 3) including but not limited to OCLC, RLG, and Library of Congress etc. This facility is achieved through the incorporation of Z39.50 copy cataloguing client in the catalogue module of the package.

### **Output of Computerised Cataloguing Subsystem**

The development of computerised cataloguing subsystem has followed the progression from offline to online systems. Present systems use online input and searching techniques, and end products are normally produced offline in a batch mode—that is, all at one time rather than the data are entered. However, the whole array of products or output may be grouped as:

---

## **2.5 SUMMARY**

---

- Library automation encompasses mechanisation of housekeeping operations, sophisticated information retrieval, integrated access interface and MIS activities. The process of library automation centres on automation package or library management software. Modern packages are integrated in nature and offer all the facilities essential for day-to-day library management. Automation packages available in India are basically of three types - packages of foreign origin, packages developed over foreign LMSs and packages developed in India. These packages are modular in structure and most of them supports all the housekeeping operations, OPAC, Web-OPAC, digital media archiving, Z39.50 copy cataloguing and modern data capture devices like RFID, smart card, etc. The process of selection of LMS should be based on some well- defined criteria because implementation of LMS is a big investment in terms of money, time and manpower.
- In an integrated automation package, acquisition and cataloguing work in harmony. The bibliographical data of newly acquired items are transferred from acquisition module to the catalogue module. In the cataloguing module, the bibliographical data elements of documents are standardised through necessary addition and modification. The bibliographical data format is based on any internationally adopted content designator scheme to allow exchange of cataloguing data. Computerised cataloguing includes three groups of works namely authority control, data entry and downloading. The catalogue records act as the central bibliographic database in a library system.

---

## **2.6 REVIEW QUESTIONS**

---

1. Compare the features of any two library automation packages of Indian origin.
2. List main works of a computerised acquisition subsystem.
3. What is the role of authority file in computerised cataloguing?

**NOTES**

---

## **2.7 FURTHER READINGS**

---

1. Arms, William Y. (2000), *Digital Libraries*, The MIT Press: Cambridge, MA.
2. Arms, W.Y. (1995). Key Concepts in the Architecture of the Digital Library. *D-lib Magazine*.
3. Haigh, S.(1996). Optical Character Recognition (OCR) as a Digitization Technology. *Network Notes. no.37*.
4. IMLS: A Framework of Guidance for Building Good Digital Collections. (<http://www.ims.gov/pubs/forumframework.htm>)
5. Jantz, Ronald. (2001). Technological Discontinuities in the Library: Digital Projects that Illustrate New Opportunities for the Librarian and the Library. *IFLA Journal 27,74-77*.
6. Kenney, Anne R. and Stephen Chapman. (1996). *Digital Imaging for Libraries and Archives*. Ithaca: Dept. of Preservation and Conservation, Cornell University Library.
7. Kessler, Jack. (1996). *Internet Digital Libraries: The International Dimension*. Boston: Artech House Publishers.

## UNIT III ONLINE PUBLIC ACCESS CATALOGUE (OPAC)

### NOTES

#### ★ STRUCTURE ★

- 3.1 OPAC and other Catalogue forms OPAC
- 3.2 Serial Control
- 3.3 Circulation and Binding
- 3.4 Use of Computers in Information Services
- 3.5 Summary
- 3.6 Review Questions
- 3.7 Further Reading

#### LEARNING OBJECTIVES

After going through this unit, you should be able to:

- Explain different modes of adoptly a library application software.
- Describe the history of library Automation and its present Service.
- Describe the Popular library Automation software.

### 3.1 OPAC AND OTHER CATALOGUE FORMS OPAC

OPAC (Online Public Access Catalogue) has been described as the shop window of a library. It is the interface between the catalogue database of all library materials and users. OPAC meets all the objectives of the library catalogue and allows rapid information retrieval. Because of the importance of OPAC as an user interface, it is not only designed as an information retrieval system but also as a module of an integrated library management system. OPAC allows browsing and searching of the catalogue databases of all the library materials. Once records have been identified, there are a number of ways in which they may be displayed. Usually the full record display includes holding information relating to individual copies, as well as the basic bibliographic data in desired format. The searching facilities of OPAC include keyword searching (post-coordinated searching), phrase searching (pre-coordinated searching), combined searching (using Boolean, relational and positional operators) and truncated searching

(left truncation or right truncation). OPACs of the modern LMSs also allow seamless access to Internet resources, online databases, web-OPACs of other libraries, e-books and digital book reviews Acquisition and Cataloguing etc.

### **(a) Other Catalogue Forms**

In addition to online access to the catalogue databases, most packages support the generation of hard copy catalogues. Hard copy can typically be produced in offline mode in the form of:

- card catalogue (main entry and added entries);
- printed book catalogue;
- microform; and
- computer output on microform.

Off-line production of catalogue records also utilises magnetic (tape, disk etc.) and optical (CD-ROM, DVD-ROM etc.) media as mass storage devices to store and retrieve catalogue database.

### **(b) Reports, Lists and Labels**

The catalogue module of automation packages supports the generation of various reports, lists and labels that are required for the management of catalogue section.

These are:

- Reports with a count of all records added, modified or edited by a specific operator or over a specific period of time;
- Reports that produce statistical account of items added and tabulated by call number, item categories, item location etc.;
- Lists of items catalogued by class number, subject heading, collection type, language etc.;
- Spine labels, shelf list, book cards etc.; and
- List of authority file entries, and all citations with authority file violations.

### **(c) Information Products for User Services**

The catalogue module also generates information products that form on the basis of a number of user services like bibliographic service, current awareness service etc. These are:

- List of books received in the library:
  - during a particular period
  - on a particular subject
  - by a particular author
  - by a particular author on a particular subject in a particular period.

NOTES

- Bibliographies of documents received by the library in standard format or as per the format specified by users.

### **Advantages of Computerised Cataloguing Subsystem**

#### **NOTES**

The application of advance level ICT in the management of library processes leads to a significant change in the nature and role of catalogue records. The impact of these changes has contributed towards standardisation of entry format, resource sharing and efficient access to documents and their contents. The advantages of computerised cataloguing may be illustrated as follows:

- In an integrated set up, circulation module and acquisition control programs utilise cataloguing records. Similarly catalogue module uses bibliographical data elements of records created in acquisition procedure and also utilises transaction records from circulation control to notify users about the availability of selected documents;
- Computerised cataloguing ensures greater standardisation in catalogue records;
- It reduces routine clerical operations required for maintenance of catalogue;
- It supports interchange of catalogue records and thereby ensures reduction in unit cost of cataloguing;
- It supports seamless access to not only library resources but also web resources, OPACs of other libraries, online databases and a variety of information services including subject gateways;
- It provides opportunities to take output in variety of forms and formats;
- It enables users to retrieve relevant records through the application of variety of search techniques and search operators and display the retrieved records in desired formats; and
- It helps library staff to generate variety of information services.

---

## **3.2 SERIAL CONTROL**

---

### **Introduction**

As per the definition of International Serials Data System (ISDS), a serial is a publication issued in successive parts and intended to be continued indefinitely. Serials include periodicals, newspapers, annuals, proceedings, transactions etc., and are differentiated from monographs by their ongoing or continuing nature. Monographs are received, paid for, catalogued, shelved and there the matter ends. Serials on the other hand, continue to be received; they have to be ordered and paid for periodically; cataloguing must include additional information, such as

the frequency of publication and often the cataloguing information must be changed; information on the library holdings must be constantly updated; and even binding must take place repeatedly and precisely. As a result, separate serials control systems have been devised solely for managing these types of materials. Serials management subsystem of an automated library system has to deal with the features unique in serials control.

These features may be summarised as below:

- Periodicals are procured through various subscription modes and by gift or exchange;
- Successive issues are received at regular or irregular intervals and it is necessary to ensure that successive issues arrive when they have been published;
- Subscriptions to periodicals must be renewed recurrently;
- Catalogue data that describe serials must be extensive;
- Serials change their titles, are published under variant titles and may change their frequency of publication. References must be inserted to link associated periodical titles;
- Precise control over the binding of successive issues is very important; and
- Indexes, special issues and supplements must be controlled.

The automated serials control subsystem which attempts to provide mechanical means for checking in serials issues, issuing claims, handling binding and other such functions has to be designed very carefully because of the complex nature of serials management.

### **Basic Facilities Available in Computerised Serials Control Subsystem**

- Management of serials requires frequent and repetitive record addition or amendment. Computerisation is an attractive proposition for serials control because of this reason. The serials control module of any Library Management Software (LMS) should provide following basic facilities to cope with the complexities of the management of periodicals and other serials publications. The typical facilities available in general for handling serials control activities are:
  - New subscription
  - Renewal of subscription
  - Cancellation of subscription
  - Budget control—Department/unit-wise budget;
  - Invoice processing—Invoice for individual issues, or for annual (or other period) subscription;

### NOTES

## NOTES

- Recording the receipt of journal issues—Formula for generating expected issues;
- Managing (sending claims for) missing issues—Sending reminders;
- Needs to be able to cope with "special editions", supplements, and indexes;
- Should also be able to cope intelligently with name changes (of publication, publisher) and merges or splits (*i.e.*, one journal becomes two, or two join together);
- Binding control;
- Accessioning bound volumes—Barcoding of accession numbers;
- Report generation;
- Listing the periodical for browsing—Hyperlinking the e-journals;
- Editing and updating of records;
- Searching:
  - By title
  - By publisher
  - By distributor
- Sorting by date or volume/issue number;
- Printing;
- Options for Web-OPAC;
- Table of contents and other personalised information services;
- Article indexing—The serials control module should support indexing of journal articles by author, title, and subject keywords. There should be option for bulk import in some format; and
- Union list and union catalogue—The module should support creation of union list and union catalogues. The union list is the list of journals currently subscribed in different affiliated libraries. In union catalogue the complete holding information is given along with all its missing issues, discontinuation in subscription, changes in title etc.

### Computerised Serials Control Operations

The fundamental work of any serials control subsystem, manual or automated is to record the arrival of each part or issue of serials in the library. Other tasks are to:

- Renew existing titles (if approved by the authority);
- Place subscription for new titles;
- Cancel subscriptions;
- Follow up delayed and missing parts;
- Prepare lists of newly arrived issues/parts;
- Prepare lists of serials holdings;

- Provide user services; and
- Arrange binding and accessioning of back issues etc.

In an automated environment all the above mentioned works are performed with higher degree of accuracy and better control. A computer-based serials control subsystem may be predictive or non-predictive. Predictive systems predict the arrival of individual parts and generate reminders automatically in case of non-receipted issues or parts within a stated time interval. Non-predictive systems do not predict the arrival of individual parts and therefore, generate claims for non-receipted issues against specific input or instruction. However, prediction can never be completely reliable because of many irregularities in serials publications. The whole array of operations in computerised serials control system may be grouped into four subdivisions—creation and maintenance of master database, subscription and acquisition, cataloguing and article indexing and circulation and binding. Let us discuss them one by one.

## NOTES

### **Creation and Maintenance of Master Database**

Modern LMSs are based on relational data model. Each module of these packages centres on a master database. Any number of addition, modification and deletion is possible in the master database and these changes are automatically reflected in all the sub-modules under that module. Serials control module also follows the same architecture. Records created in the master database are available to all the sub-modules under serials control. It reduces data entry work and ensures standardisation. Generally, this database includes following master entries:

#### ***Title Master***

In this file, bibliographical details of new serials are entered after the selection and approval process. This file also includes details of all the existing serials on the basis of some standard bibliographic format. Any number of addition and deletion in the title master file is possible through appropriate options.

#### ***Country Master***

This file contains name of countries and their corresponding codes for entering country of publication data in sub-modules of serials control. Country code is generally based on ISO-3166 where each country is represented by two unique characters *e.g.*, the code of India is in as per ISO-3166.

#### ***Language Master***

This file contains entries for languages and their three digit codes as per the ISDS manual and CCF manual.

## **Supplier/Publisher/Binder Master**

This master file contains details of all local and foreign subscription agents, publisher of serials and binders along with their corresponding codes. These codes are generally created locally.

### NOTES

### **Subject Master**

This master file holds lists of subject descriptors and are available in various sub-modules for entering subject group(s) of serials. This file also includes class numbers, corresponding to that particular subject.

### **Frequency Master**

Individual parts of serials are published in different time intervals *e.g.*, weekly, monthly, semi-monthly, bi-monthly, quarterly etc. This master file enables one to record serials frequencies and also to create codes for new serials frequencies.

### **Budget Master**

This file enables to record various financial data necessary for serials acquisition. This is required for fund accounting of serials budget and includes entries for main budget head, sub-head, period, account name and code, dates, opening balance, grant amount, amount spent, credit note amount, closing balance etc.

### **Currency Master**

This file contains currency description, codes and exchange rate for foreign currencies in terms of Indian Rupees. The conversion will be calculated automatically on the basis of data given in this file. The frequent update of conversion rate is necessary (if required, daily) on the basis of bank notification.

### **Delivering Mode Master**

Publisher and vendors arrange supply of serials by different modes of delivery. These are:

- Air freight
- Surface Air Lifted
- Shipment/postal
- Hand delivery
- Generally publishers charging extra payment for air mail. This file includes different mode of supply for making these information available for data entry work in various sub-modules.

### **Physical Media Master**

Serials are available in many forms, formats and media. The available physical media are print, CD-ROM, magnetic tape, online, DVD-ROM etc. This file includes description and corresponding codes for different physical media.

### ***Binding Type Master***

Individual parts or issues of journals are bounded volume wise at the end of year or subscription period. Binding of back volumes of journals is a regular work of serials control section of a library. This file contains different modes of binding (e.g., standard, leather binding, cloth and rexin binding etc.) and their corresponding codes.

### ***Letter Master***

This file includes formats for every type of letters required for the generation of outputs such as order letter, cancellation of order letter, reminder letters, claim letters, etc. The terms and conditions formulated by the library may be included in the order letter format.

### ***Subscription and Acquisition***

This group of activities involves the following basic tasks:

- Selection of serials for new subscription;
- Renewal or discontinuation of existing journals/serials;
- Selection of delivery mode;
- Selection of subscription mode;
- Formulation of terms of procurement;
- Selection of vendors;
- Approval from authority;
- Ordering and renewal;
- Payment;
- Receiving and registration;
- Reminder generation; and
- Adjustment of advance payment for non-receipted issues.

### ***Renewal and Discontinuation***

The appropriate authority also decides the continuation or discontinuation of existing serials received by the library against subscription or on exchange or by gift.

### ***Selection for Approval***

The list of journals containing the new titles requested or suggested by experts or authority along with the titles selected for renewal is prepared in this stage after entering necessary input to the database. This list includes bibliographical details of journals along with status (new/renewal), subscription mode, delivery mode and name of vendor or supplier. Finally, list is sent to the appropriate authority or library committee for consideration.

NOTES

### **Approval process**

After receiving a list of approved serials, library staff incorporates recommendations by checking the appropriate options in the database. This is known as approval process. This process helps to merge all the bibliographical details of the approved titles into the master database.

#### NOTES

### **Ordering**

This is typically a group of three activities—order processing, order generation and order cancellation. This step involves the work of creating orders for new serials and for renewed serials. The necessary datasets for order processing include order number, date of order, name of publisher/supplier, type of order, start date of subscription, start volume, date of expiry of subscription, end volume, number of issues etc. A typical set of data elements is displayed below. After entering and saving of necessary datasets, the next step is to proceed for order generation. An order can be cancelled anytime by entering necessary input into the database *e.g.*, order number, remarks etc. Selecting an appropriate option can also generate letter of cancellation of order. Payment Renewal and subscription to journal require release of payment in advance. Supply of journal will only start after the receipt of payment by publisher. The supplier after having remitted the subscription to the publishers will submit the adjustment bills claiming the difference (if the amount remitted is more than the advance payment), along with the proof of payment. The payment sub-module manages this type of payment through adjustment of advance payment option. As per the agreement, the vendor/publisher has to replace all the missing issues without any extra charge. In case missing issues cannot be replaced, the vendor has to refund the proportionate cost of the missing issues. At the end of the subscription period, library claims the proportionate costs for all the missing issues. Supplier or publishers in the form of credit note refund proportionate costs of the missing issues. The payment sub-module should support all these payment-processing activities.

### **Acquisition**

As issues or individual parts of serials appear regularly, acquisition is one of the major functions in any serials control subsystem. The acquisition functions may be divided into two groups—receiving and claiming.

### **Receiving**

Receiving involves a large volume of checking. When an issue arrives it must be checked to ensure that it is the correct item and then the master records must be amended to reflect new receipts. The receiving sub-module generally includes the facilities for schedule generation, check-in, processing of non-receipted issues and report generation. Schedule generation involves the entering of lead-time in terms of days. It is required in a predictive serials control subsystem to generate schedules for expected arrival of serials publications. The schedule is utilised for

entering check-in details of serials. The check-in process will first retrieve all the details of a selected title from the master database. In the next step, library staff will change the status of the issue as received and enter date of receipt before saving of records. If an issue is not received in the library by the expected date, then the status of that particular issue is to be changed as non-receipted issue for the generation of reminder to respective suppliers/publishers.

## NOTES

### ***Claiming***

The claiming process involves a small set of functions, but possibly it is a more complex activity. The final output of claiming process is the generation of reminder letters for non-receipted issues to the respective suppliers/publishers. If prediction can be carried out successfully, then serials parts which are predicted to arrive within a specific time span yet fail to do so, can be assumed to have failed to reach the library through the normal pattern of distribution, and claims may be made for replacement. Claims can be generated manually or automatically. Modern systems support automatic generation of claims as text files or in other formats and can send them to respective vendors/publishers as e-mail notice.

## **Cataloguing and Article Indexing**

### ***Cataloguing***

Cataloguing formats for serials are fundamentally similar to those of monographs. But the content and format of serials bibliographic records varies considerably between systems. Some catalogues are based on ISBD(s) and others on ISDS formats. Some cataloguing systems use local formats. Any such format should take care of changed and variant titles along with references. For example, CCF prescribes following minimum data elements for the bibliographical format of serials.

### ***Article Indexing***

Indexing of articles (also called papers) from journal issues is an optional facility of serials control subsystem. Generally, publishers of primary periodicals produce annual and other sorts of indexes regularly. Apart from such products, libraries also subscribe to number of indexing and abstracting journals related to the areas of their interest. As a result, article indexing is only necessary when available indexing and abstracting services do not cover the core journals on discipline of interest. The absence of publisher produced indexing services is reason enough for a library to consider producing one itself. The first step of article indexing is to determine the bibliographic format of articles to be entered into the system. As per CCF, a bibliographic format for component part in a serial such as article of a periodical may include the following data elements.

A library must follow a subject thesaurus for the standardisation of keywords or descriptors.

## 3.3 CIRCULATION AND BINDING

### NOTES

### **Circulation**

Circulation pattern of serials differ largely from that of books. But if serials are available for ordinary loan, then the same circulation control system will suffice as for monographs. However, serials are generally reserved for reference use only. In special libraries, the short time loan options for journals are common because of the specific need of users. If the number of transactions per day is large enough then such transaction system may be computerised. Such computerised facility must have a list of serials taken, a list of users and their addresses, and transaction interface with options for the generation of required output.

### **Binding**

A further valuable feature of computer based serials control subsystems is their ability to inform the library staff of volumes that have been completed and are now ready for binding. It is a very helpful feature to assist in work scheduling and to spread the binding load to give an even distribution of work in the binding throughout the year. The binding sub-module includes the following functions:

#### ***Binding Set Selection***

This step involves the work of selecting the relevant issues meant to be sent for binding. The selection follows entering of binding type, colour and embossing type by using drop down menu or as per the provisions.

#### ***Order Process***

In this step library has to provide various datasets to generate order letter such as order number, cost of binding, expected date of delivery etc. Binders name and address, budget head etc., may be selected from drop down menu or as per the provision originated from the master database.

#### ***Receiving***

This function includes the receiving of bound volumes against order numbers (after manual checking of the quality of binding) and accessioning of bound volumes. It also includes the work of entering receipt date, classification number, location etc. manually.

#### ***Payment***

Payment work starts with invoice processing, generates payment release note or order and finally records all the details of payment e.g., cheque/DD number, date, amount, budget head, payment acknowledgement receipt number and date etc.

## **Reminder**

It is meant for generating reminder letters to be sent to the binders in case of any delay in receiving the bound volumes. Binding sub-module should generate list of overdue items.

## **NOTES**

## **Output of Computerised Serials Control Subsystem**

Computerised serials control subsystems are designed to manage group of routine clerical chores, and act as MIS tool for the library authority. Such systems are able to answer to various user queries in online mode and also generate information products on demand. As a result outputs of such systems are quite helpful and may be grouped into three basic categories—OPAC, Reports and Off-line information products.

### **OPAC**

Generally, the OPAC module of automation packages provides separate option for online searching of serials database of library. OPAC for serials supports searching by:

- Title
  - Current titles
  - Complete holdings
  - Key title
  - Linked title
  - Variant title
- Subject
  - Broad subject heading
  - Subject divisions, descriptors and class number
- Publisher
- Title history
  - Title split
  - Title merge
  - Title change
  - Title holdings
- ISSN
- Free text OPAC supports both simple and composite searching along with the application of Boolean operators (AND, OR, XOR, NOT), positional operators (SAME, WITH, NEAR, ADJ) and relational operators (less than, greater than, equal to, less than or equal to, greater than or equal to, not equal to). OPAC also enables users to specify the amount of bibliographic data to be displayed during retrieval.

## Reports and Lists

Several reports, letters and statistics can be generated by the automated serials control system. A representative list of such products is as follows:

### NOTES

- List of suggestions
- List of approved titles
- List of titles order
- Subscription overdue list
- List of titles cancelled
- Order letter
- Payment report
- Title report by
  - Alphabetical order
  - Language
  - Subscription
  - Physical media
  - Budget head
  - Department
  - Class number
  - Supplier
  - Receipt mode
  - Delivery mode
  - Location
  - Country
  - Publisher
  - Status
  - Frequency
- List of issues received
- List of non-receipted issues
- List of missing issues
- List of issues by status
- Reminder letters
- Binding reports
- Binding details
  - Binder wise
  - Title wise
  - Binding received
- Payment details

- Accession register
- Title history reports by
  - Title changed
  - Title split
  - Title merged
  - Title holdings
- General reports by
  - Publisher
  - Vendor
  - Binder
- Statistical reports
  - Yearly and monthly reports
  - Subject wise statement
  - User's usage statement

These reports can be consulted online or printouts can be generated in offline mode.

### **Information Products**

Information products of serials control section of a library are generated in offline mode and form the base of various user services. These products *are originated* either from article indexing activities or serials catalogue database and produced on demand.

#### **Information products from serials catalogue database:**

- List of recent arrival for issues of a group of journals (as selected by users);
- List of journals available on a particular discipline;
- Discipline-wise holding list of serials;
- Title wise holding list of serials;
- Union catalogue;
- List of current subscription;
- List of currently available indexing/abstracting journals; and
- List of serials status.

#### **Information products from article indexing:**

- Table of contents service of a group of journals (as per user selection);
- Compilation of subject bibliographies;
- CAS and SDI services in online and offline mode;
- Online database searching and retrieval;
- Altering services on specific topics (online/offline);

NOTES

- Compilation of abstract bulletin on any given topic;
- News compilation services; and
- Product compilation services.

## NOTES

### **Advantages of Computerised Serials Control Subsystem**

Serials management is possibly the most complex process in libraries. It involves a lot of time bound routine activities that requires accuracy and precision. In special libraries, periodicals are the most important resources and take the lion's share from the library budget. The optimum utilisation of these resources demands application of ICT as a management tool. Following are multifaceted features of automated serials control subsystems:

- It reduces workload of library staff and ensures more productive use of manpower especially in the development of information products and services;
- It acts as a decision support tool for the library authority;
- It can generate various reports in required formats for MIS activities;
- It ensures timely reminders generation and better binding control;
- It offers easy and simple solutions for fund accounting, payment management and budget control;
- It offers easy creation and maintenance of article indexing database and thereby generates number of user services on demand;
- It helps library staff in quick production of serials holdings and list of recent arrivals in many forms;
- It facilitates online access to the serials database from anywhere at anytime in any format;
- It predicts the arrival of journal issues and generates schedules for receiving journal issues;
- It can manage export and import of cataloguing data for serials received by the library;
- In an integrated serials control module, the master database supplies most of the bibliographical, financial and administrative data at the time of data entry work. This approach ensures standardisation of entries and also eliminates data redundancy; and
- It supports generation of a variety of information products and user services in the desired form and formats.

### **Basic Functions of Computerised Circulation Subsystem**

Computerised circulation subsystems generally perform a group of functions utilising three basic categories of information (as listed in the previous section) as a base. Apart from the issue and return provision, circulation module performs the following functions:

NOTES

- To locate circulating:
  - Items on loan
  - Items reserved by user
  - Item at binding
  - Items being reprocessed;
- To identify items on loan to a:
  - particular borrower
  - specific/category class of borrowers;
- To record 'personal reserves' for items on loan but desired by another borrower and to issue alerting notice to the library staff on return of the reserved item by a borrower;
- To print recall notices for:
  - returning overdue items
  - renewing of items
- To arrange renewal of loan;
- To notify to the library staff of overdue items and printing of overdue notices;
- To calculate fines or overdue charges for generating:
  - Printout of fine notices
  - Receipts of fine records
  - Printout of fine receipts
- To calculate and print statistical reports;
- To extend provision for handling special categories of borrowers and special types of materials;
- To generate and print gate pass and due date slips;
- To act as decision support system for better circulation management;
- To support various data capturing devices e.g., barcode readers, smart card and RFID equipment; and
- To extend facilities for ILL and maintenance activities.

In summary we can specify a total of nine functional groups of activities performed by a circulation module and these are:

- Circulation control function to perform all charging and discharging activities at a workstation in real time;
- Hold processing function to enable placement of reservation on any title or item in the database;
- Loan period calculation function to determine automatically the loan period for each item at the time it is charged;
- Overdue processing function to automatically produce the appropriate notice when the selected criterion is reached;
- Cash management function to allow users to be billed automatically for recovering overdue charges and value of lost materials;

## NOTES

- Fiscal report generation function to record daily financial transactions including fines and fees levied, fines and fees waived and fines and fees collected;
- Transaction control function to block users who exceed overdue threshold (e.g., too many overdue items) set by library;
- User registration function to assign track and maintain borrower profiles; and
- Report generation function to produce lists, reports and statistics in relation with various facets of circulation management.

### Framework of Computerised Circulation Subsystem

Library materials should be made available to users readily or as soon after the demand arises as feasible. Circulation systems are designed to support this primary activity of a library *i.e.*, document availability. A circulation system records loan transactions to specify:

- What material is in the library stock or readily accessible on ILL;
- Which material is on loan, and from whom or where it can be retrieved; and
- When materials on loan will next be available in library for other users.

A computer based circulation system supports all these primary activities of circulation section and is quite simple in its basic concept. The transaction or loan database is the core of circulation system. This database comprises a series of records, one for each transaction. Each record includes a brief dataset that specifies details of the document (through document number), details of the user (through membership code) and transaction details (e.g., date of issue and date of return are extracted from the system date, and due date is calculated automatically). In an integrated setup, the bibliographical details (e.g., author, title, edition, place and year of publication) of documents on loan are extracted from the catalogue database and the membership database is utilised for collecting user information. Accession numbers of documents are used as the key data elements in first case, whereas membership codes act as pointer to the member database in the second instance.

It is quite clear from the following figures that accession number and member ID are the most important data elements for a circulation control system.

Therefore, accession numbers and member codes must be assigned in such a way that they identify the documents and the borrowers uniquely. Error free capturing of these two important data elements is another essential requirement in computer based circulation subsystem. Data-capture is generally based on barcodes, which are used to encode both accession numbers and member codes. These bar-coded data elements can be read through a portable light-scanning device (e.g., barcode reader).

Some automation packages have also started utilising modern data capturing equipments like RFID (Radio Frequency Identification), smart card etc.

## **Computerised Circulation Operations**

NOTES

Automated circulation management has been very successful since the beginning of library automation. There are obvious reasons for this:

- The operations to be performed are repetitive;
- The procedures to be followed can be described systematically;
- Circulation can be separated from other library operations;
- The bibliographic information used in such systems need not be extensive; and
- Information may be captured in a variety of ways.

The broad groups of activities necessary to manage automated circulation are discussed in the following sections.

### **Membership Management**

This sub-module is basically meant to create and update membership records in a library. The work of this sub-module starts with the creation of master database, which contains details about member categories, institutes, departments, courses/designations and so on. Member enrollment and generation of outputs follow this step. A systematic organisation of activities related to this sub-module will give us three subgroups of facilities:

#### **Master Database Creation and Maintenance Facility**

It includes the creation of various master files necessary for membership management in particular and circulation management in general. The tasks of this subgroup are:

- **Category and privileges determination:** This will enable to assign category code, duration, membership charges if any, and to determine various privileges associated with the respective membership categories. The privileges include types of borrowing items, their issue and reserve periods along with overdue charges.
- **Institute file creation:** It enables to create a master file of institutions by entering institute code, name of the institute and address.
- **Departments/Divisions file creation:** It allows adding a department under institute by entering department code and its name, address etc. Some packages also allow creating course/designation master file under each department. This option is particularly helpful for academic libraries.
- **Calendar:** It is meant for developing library calendar by entering

## NOTES

list of holidays as per the calendar of events of the parent institute. It helps to automatically shift the date of return on a working day, if it falls on a closed holiday. This will also take care of calculating overdue charges accordingly.

- **Member enrollment facility:** This facility is meant to create and update membership records in a library. It includes:
  - **New membership record creation:** It is to enroll members to the library by entering members personal information viz. last name, first name and middle name, type of membership (selected from master database), membership status, category of member, name of department/division, institute and course (if applicable) and contact number, address etc. Membership code may be generated automatically or entered manually into the database. Member ID will be used at various places like identity card, reminder and document transactions.
  - **Modification of membership record:** It is required for editing or updating a member's record using members ID or code.
  - **Deletion of membership:** This facility will enable the library staff to delete membership either group wise or single at the end of membership period. This step requires the clearance of all dues for each individual member of the group.
  - **Renewal of membership:** This provision is required for renewing the date, when the validity of membership expires.

### **Output Generation Facility**

This facility allows:

- searching of membership details against name or member ID;
- generation of 'No-Dues' certificate group-wise or single at the time of membership termination; and
- generation of membership ID card and printing before distribution. It includes the production of barcoded member ID required for barcode based circulation works.

### **Transaction Management**

Transaction sub-module includes all the day-to-day activities of circulation section of a library viz. issue, return, renewal, reservation, reminders for overdue books, searching document availability and listing of items issued to a member.

The details are as follows:

- **Issue/Check-in.** To issue a document, accession number of that document is to be entered either manually or by using a data capture device. It will display bibliographical information of the document along with its status (whether available for issue or reserved). The next step is to enter member code of the borrower.

It will show details of the member along with category, number of documents issued against the member and overdue charges etc. The library staff can then issue the document by selecting appropriate option. The system after checking various parameters settled by the library, will generate appropriate message along with *date of issue and due date of return*. This sub-module generally supports extension of due date, if required.

## NOTES

- **Return/Check-out.** This facility is meant for receiving back the documents lent out to users. The work starts with entering accession numbers of returned document. It will display the status of the document and the details of member who borrowed that particular book. The library staff may then mark return the document by clicking necessary option. If the book has been returned within due date a successful return message will be displayed. In case of late return (after due date) the system will be asking for fine collection by generating a fine receipt form.
- **Renewal.** This is meant for issuing a document to the same member again by renewing the transaction record. Due date gets changed automatically in the process. The renewal process will be successful, if someone has not reserved the document.
- **Reservation.** It supports booking of already issued document for a member. Reservation facility is programmed to generate message after returning a reserved document. The reservation process also allows change of reservation priority and cancellation of reservation, if required.
- **Recall.** This facility is useful if the document issued needs to be recalled and a reminder is to be sent to the member. It allows printing of recall letters and storing of letters for record.
- **List of document borrowed.** This facility helps to display details of documents borrowed by a member and information about overdue charges etc.
- **Document availability.** Circulation staff has to search sometimes details of the document for checking of bibliographical information or document availability. This facility provides a link to the OPAC module from the transaction sub-module in an integrated architecture.

### ***Reminder Generation***

This facility is meant for generating reminders for overdue documents:

- To a group of members;
- To individual members;
- For a particular due date; and
- To all the format and text of reminder letter may be modified by using this facility or by using the master database.

## Fiscal Management

It provides option to manage outstanding dues against a member. This facility allows entering of overdue charges or amount against a member and generation of payment receipt. The change of payment if any, is generally admissible by the use of authorised code. This facility should also allow printing of fine statement if a member wants to have a statement of fines.

### NOTES

### *Inter Library Loan (ILL)*

Inter library loan method simply means that documents of a library can be issued to the members of other libraries. No libraries in the world can procure all the published documents on all subject areas. As a result libraries have to depend on each other for the satisfaction of users. The success of ILL clearly depends on the availability of union catalogues. The circulation module of LMSs should support ILL to help library staff in carrying out activities related to ILL method. This sub-module generally includes three main facilities:

- ILL Membership;
- ILL Borrowing; and
- ILL Lending.
- **ILL membership.** It allows enrolling a new member for ILL through entering details about the proposed member such as name, department and institute, member privileges (maximum items allowed, loan period and overdue charges). The cancellation of ILL membership or any modification in the membership profile is also admissible through this facility.
- **ILL borrowing.** It involves tasks to procure document(s) for a member of the library who has asked for an ILL from some other library. The tasks are:
  - Receiving of ILL request and entering of relevant information into the database (member code, item details, type of item etc.);
  - Generation and printing of request letter addressed to the library from where the document is to be borrowed;
  - Receiving of document(s) on ILL by selecting the received status and entering date of arrival in the module;
  - Sending intimation about arrival of document(s) on ILL to the requester;
  - Issuing document(s) to the member who has requested it;
  - Returning the item(s) borrowed on ILL;
  - Generating reminder letter in case of overdue document(s); and
  - Sending ILL document back to the library.
- **ILL lending.** It is meant for lending of library documents to member libraries on ILL through the use of ILL sub-module. The tasks to be performed for the service are:

NOTES

- Issuing of item(s) to the member library through entering necessary details (accession number of document and name and code of the member library);
- Receiving back the item(s) on ILL;
- Sending reminder to the library concerned in case of delay in returning; and
- Generating list of documents which have been loaned to other libraries on ILL.

### ***Maintenance***

This sub-module is generally attached with circulation module for recording information about lost documents, documents sent for binding, damaged documents, missing documents and documents withdrawn from library. The activities of this sub-module include following groups of tasks:

- **Lost documents.** Loss of a document is managed through two options- replacement of document and amount recover. In the first case replacement is confirmed after entering code of the concerned member and accession number of the lost document. If amount is to be recovered from a member for the lost document, the 'amount recover' option is to be activated by entering member code and price to be recovered. Payment receipt may also be generated by selecting appropriate option.
- **Binding.** The reasons for binding documents are two-fold *i.e.*, their preservation for posterior and to keep them in a usable state for users of the day. Binding facility of maintenance sub-module is a combination of five basic steps:
  - (i) Entering accession numbers of the documents to be sent for binding;
  - (ii) Order generation for binding of selected documents through inputting of order number, expected date of delivery, price, name of binder, budget head and binding type;
  - (iii) Receiving of documents from binder and processing of invoice;
  - (iv) Release of payment; and
  - (v) Generation of reminder for overdue items.
- **Withdrawal.** Stock revision or weeding out of documents is essential in order to make enough space for the usable stock. The catalogue database is also to be modified through recording details of weeded out documents. In computerised circulation, the task of weeding out is performed by entering accession numbers of withdrawn documents in appropriate sub-module. Such a system has the provision to reintroduce an already weeded out document by simply checking/clicking the necessary button.

## Advantages of Computerised Circulation Subsystem

### NOTES

Computerised circulation subsystems apart from performing all the basic activities related to circulation, offer extensive control of stock. Transaction records can be entered and saved into the main database through a terminal. The central transaction database is updated immediately and subsequent consultation of the database will communicate the current situation. Some of the important issues may be enumerated as:

- fines can be calculated on demand;
- reservation and other modification to document records can be made instantly;
- automatic identification of over borrowing and problem borrowers;
- error-free data capturing through barcode, RFID and smart card technology; and
- back up provision to cover telecommunication or computer system failure; The circulation modules of modern LMSs provide many advantages over the manual system. Such a system can:
  - identify and trap reserved documents on their return from loan;
  - generate (print and electronic) overdue and recall notices;
  - collect and display issue statistics;
  - print checkout and check-in slips;
  - manage a variety of categories of stock and borrowers, each with its own library defined circulation parameters;
  - keep records of fine payment in detail;
  - manage multi-site provision, with each site having its own stock, borrowers and associated circulation parameters;
  - change status of an item in the OPAC after it has been issued or returned;
  - circulate un-catalogued items on the fly and to add such records to the catalogue database, not to the circulation file;
  - support multiple items renewals in one step;
  - alert operator if an item being charged is already issued to another user; and
  - allow users to place reservation on an available or charged/issued item without library intervention;

A recent trend of LMSs is to incorporate self-checking or self-issue option in the circulation module. RFID (Radio Frequency IDentification) is being used to manage un-manned self-service counters for the issue and the return of documents. An RFID system comprises three components: a tag, a reader and an antenna. The tag is paper-thin chip, which stores necessary bibliographic data. The tag is to be placed on the inside cover of the corresponding document. RFID reader and antenna are often integrated

into patron self-checkout machines or inventory readers. The reader powers the antenna to generate RF field to decode information stored on the chip. Reader sends information to the central server, which can communicate with the library automation software. RFID, apart from self-issue facility, also supports stock verification, theft detection, and identification of misplaced books and inventory counts.

NOTES

---

### 3.4 USE OF COMPUTERS IN INFORMATION SERVICES

---

We are utilising ICT as a tool (such as Library Management Software (LMS), Internet, Telecommunication etc.) to serve our users right information at the right time, and the father of reference and information services, suggested in 1876 four basic functions in the area of library services:

- **Instructing the User:** It is intended to help users to learn how a library is organised so that they could take advantage of knowledge contained in information resources.
- **Assisting the user with his/her queries:** It is intended to provide reference and information services that involve either finding the required information on behalf of the users, or assisting the users in finding information.
- **Aiding the users in selection of good works:** It is the link between librarian's knowledge of the collections and needs of the users. It is intended to guide users in selecting the most appropriate information sources and services.
- **Promoting the library within the community:** It is intended to relate the activities of the library to the needs of its parent community. It helps to make the library visible in the community. The success of any library depends on the recognition by the parent community.

Although over a century has passed, these four functions remain the core of reference and information services in today's digital environment.

ICT is a medium and a tool that enables the organisation and dissemination of information. It is a conduit for the delivery of information from providers to users. The collaboration between reference librarians and technology experts provides software solutions for online reference and information services (such as LSSI's electronic reference software and Endeavor's Encompass) to support:

- **Immediacy:** To help users to receive quick library services at any time of day or night, any and everyday of the week.
- **Interactivity:** To help users to follow-up responses.
- **Personalisation:** To help users to receive information package in the way they want.

- **Mobility:** To help users to access library services from any Internet connected computer at anywhere and any time. The wireless telephony and wireless computing will increase the value of mobility in future.

## NOTES

### Types of Library Services

Libraries provide a variety of information services to satisfy different information requirements of users. The major and important library services are:

**Reference Service:** Reference services help users to locate and obtain specific pieces of information from information sources such as reference books, catalogues, directories, files, abstracting and indexing periodicals, databases (online and CD-ROM) and other reference materials. Library personnel may either help users in searching (direct search) or they themselves do the search for users (delegated search) in online or offline mode.

**Referral Service:** Referral services aim to refer users to the sources of information such as secondary publications, information units, professional organisations, research institutions and specialists/experts. Such services do not provide the documents or information required by the user for his/her query. Librarians utilise directories and databases on sources, specially designed and developed for rendering referral services.

**Current Awareness Service (CAS):** CAS satisfies users' current approach to information and thereby keeps them up-to-date in the field of their work. The important characteristics of CAS are as follows:

- It is a technique of communicating current information to users.
- It provides latest developments in a subject field and does not provide answer to any specific query.
- generally covers a broad subject area and supplements the user's own channel/media of obtaining information.
- It is known for the speed and timeliness.
- It is meant for use before its contents are absorbed by secondary publications like abstracting and indexing journals.

CAS may be provided through variety of media and channels such as current awareness lists, current contents, routing of periodicals, list of research in progress and forthcoming meetings/seminars/conferences, newspaper clippings etc. Some commercial publishers and database vendors provide free online CAS through Internet. These are:

- Contents Direct service (Elsevier)
- IDEAL Alert (Academic press)
- Wiley book notification service (Wiley)
- Book information (Amazon.com)
- Current contents and ISI alerting service (Institute of Scientific Information, Philadelphia - Fee based)

**Selective Dissemination of Information (SDI) Service:** SDI is a special type of current awareness service. It supplies each user with the references of documents to their predefined areas of interest, selected from document published recently or received during a particular span of time. H.P. Luhn first coined the concept of SDI as a computer mediated information services. The workflow of SDI service is based on the following steps:

NOTES

- **Step I-Users' profile:** In the first step needs and interests of each user or a group of users having similar requirements are ascertained and carefully analyzed. These are then expressed in terms of some keywords, collected from an accepted thesaurus. User profile may be stored as a database file in case of computerised SDI.
- **Step II-Document profile:** In this step contents of selected documents are analyzed and expressed in terms of keywords selected from the same accepted thesaurus. It is necessary for precise matching. This may also be stored as a database file.
- **Step III-Matching:** The first two steps are the work of library professionals. This step *i.e.*, matching of two profiles is conducted by computer at regular intervals. The result of matching is then saved as a file in the required format such as text, html etc.
- **Step IV-Notification:** This step involves communication of result to the users. Notification may be sent to users through e-mail by attaching the result file or as a hard copy by obtaining printout of the result file(s).
- **Step V-Feedback:** SDI includes a mechanism of feedback from the side of users. Generally, a feedback form is sent to each user along with the notification in which he/she has to indicate whether the document really interests him/her, whether he/she needs a copy or why it is of no interest to him/her. On the basis of feedback user's profile is updated regularly.

**Literature Search Service:** It is an extension of reference service. This service includes the following steps:

- Analysis of the search parameters of a query.
- Formulation of a suitable strategy for searching different information sources.
- Identification and choosing of most appropriate sources to be searched and the order of searching them.

**Document Delivery Service (DDS):** Document delivery is a key element in access to information. Unless the documents required by the user are available to him/her, all the other services are of no use. DDS is a complex process and is concerned with supply of documents to users

## NOTES

on demand in required format. All the information services we discussed so far (*i.e.*, CAS, SDI, Literature search etc.) are aimed at guiding users to the documents of their interests. DDS is the last point in this chain of information services that actually locates the required document and supplies it to users in required format. Electronic DDS supports delivery of documents in digitised form at anytime from anywhere.

**Translation Service:** In the area of science and technology about half of the world's literature is published in languages other than English. Access to non-English literature by people who know English is possible through translations. Translation services thus help in the global access of information. In India, DESIDOC, NISCAIR, IASLIC, ONGC, BARC, BHEL, DRDO laboratories and several wings of the Ministry of Defence and Ministry of Science and Technology provide translation facilities.

---

### 3.5 SUMMARY

---

- Serial control is one of the most complex subsystems in libraries. This is mainly because of the continuing nature of subscriptions to serials. Computerised serials control subsystems are able to generate a variety of outputs and these are very useful for the design and development of information products and user services in the required forms and formats.

---

### 3.6 REVIEW QUESTIONS

---

1. Explain 'Check-in' operation of serials control.
2. What is the role of article indexing in serials control?
3. Enumerate important information products of computerised serials control subsystem.
4. List out various reports required for serials management.
5. Discuss advantages of automated serials control over the manual system.
6. Write down the advantages of predictive serials control.
7. What are the primary functions of a computerised circulation subsystem?
8. Write down the functions related to overdue management.
9. Explain the barcoded circulation system.
10. What do you mean by ILL?
11. What is RFID based circulation system?
12. Discuss advantages of computerised circulation over the manual one.

---

### 3.7 FURTHER READINGS

---

NOTES

1. Arms, William Y. (2000), *Digital Libraries*, The MIT Press: Cambridge, MA.
2. Arms, W.Y. (1995). *Key Concepts in the Architecture of the Digital Library*. D-lib Magazine.
3. Haigh, S.(1996). *Optical Character Recognition (OCR) as a Digitization Technology*. Network Notes. no.37.
4. IMLS: A Framework of Guidance for Building Good Digital Collections. (<http://www.ims.gov/pubs/forumframework.htm>)
5. Jantz, Ronald. (2001). *Technological Discontinuities in the Library: Digital Projects that Illustrate New Opportunities for the Librarian and the Library*. IFLA Journal 27,74-77.
6. Kenney, Anne R. and Stephen Chapman. (1996). *Digital Imaging for Libraries and Archives*. Ithaca: Dept. of Preservation and Conservation, *Cornell University Library*.
7. Kessler, Jack. (1996). *Internet Digital Libraries: The International Dimension*. Boston: Artech House Publishers.

### NOTES

### ★ STRUCTURE ★

- 4.1 Introduction
- 4.2 Database Approach
- 4.3 Search Techniques
- 4.4 Computer Networks
- 4.5 Library and Information Networks
- 4.6 Summary
- 4.7 Review Questions
- 4.8 Further Readings

---

### LEARNING OBJECTIVES

---

After going through this unit, you should be able to:

- know about database approach
- know about different types of database
- define databases
- define various search techniques
- explain the computer networks
- describe library and information networks.

---

### 4.1 INTRODUCTION

---

Data is an important resource for any organisation. Many organisations/enterprises irrespective of their nature are concerned with collection and manipulation of data. Data can be defined as the raw numbers or letters stored in a computer, which represents facts about things. Data has little meaning in its stored state, whether we think of data in the physical form in which it is stored or in its logical form, since facts without context are unusable. A store of data is referred to as a database.

Ideally, data must be accessible to any authorised person in an organisation for whatever application in whatever format desired. But, traditional computer files maintained in different organisations do not facilitate the access of this nature. In other words, computer files are designed to be of use for just one application and the set of programs associated with that application. They generally have little or no relation to other

data used and required within the same organisation because they are initiated and developed independently. Such files are often unrelated to user's requests for information, particularly when these requests change in un-predictable ways.

When files are designed for a specific application, and when an organisation has many applications, the files often contain redundant data. This means the same data is stored in more than one file. For example, an employee's name and address might be contained in both personnel and payroll files. In addition to requiring more storage space, redundant data presents updating problems, since updating data in one file will not update the same data in another file. As a result, data in one file may not correspond to data in another file, although, it is supposed to. Further, since files are designed for application specific programs, a change in the format of data can affect many applications and many programs. If a new field has to be added to a file, then the associated program must be changed. In other words, it might be said that the programs are data dependent.

To overcome the problems of data unrelated to user's needs, data redundancy, and program dependency on data, databases have been developed. The basic purpose of a database is to enable access to stored data so that it can provide information that users require. This Unit is intended to furnish detailed information relating to database concept and explain the database components.

## NOTES

---

## 4.2 DATABASE APPROACH

---

Traditionally, data accessed through computers has been stored on different storage media in the form of individual files. Files proved to be quite satisfactory so long as computerisation was limited to a few application areas and use of computers is restricted. However, as the actual users grew in number, with the advent of online sharing systems, the file systems gave rise to many serious problems. The discipline of database systems evolved in response to these problems.

It may be emphasised that the database approach is more than merely a different computer technique involving the storage of data and the use of additional generalised software. It involves a new approach to designing and operating information systems throughout the enterprise and can have far reaching effects beyond the data processing department. Stated differently, database approach regards data as a resource to be managed along with more generally recognised resources of an organisation (such as staff, finance, equipment and premises) so as to be available to a variety of applications and users. The integrated database is intended to provide a consistent view of the institution's/enterprise's data for all user departments. Although, each department has responsibility for specific data, several departments using the same type of data can operate on the same data

## NOTES

values. In other words, a parochial view of data would be discouraged and the uncoordinated exchange of information among departments would not be necessary. Instead, all basic data would be input to the database by each department responsible for it and retrieved (possibly in summarised form) by those departments needing access to it.

The database approach is not dependent on any particular structure of data; many of these concepts can be applied to files as well. The database approach cannot be practically implemented in the absence of a computer. In many respects database approach is characterised by the word efficiency. In principle, one can do any thing with non-database systems as with database systems. In practice, the cost of doing certain things without the benefit of the database approach is so prohibitive that the possibility is dismissed. The two main contributors to prohibitive cost are the data redundancy and the uncontrolled searching time. The database approach is beneficial when online access to data is provided. With online access, the collection of data and extracting of timely reports becomes fairly straightforward as the restrictions or delays of batch processing are not imposed on the user. However, it may be emphasised that the database approach is rooted in the attitude of: sharing valued data resources; releasing control of those resources to a common responsible authority; and

### **Database Definition**

The major problem one faces in the study of database technology relates to the determination as to what precisely constitutes a database. The casual use of the term database tends to refer to any organised collection of data capable of being accessed by a computer. This could be applied to a couple of reels of magnetic tape or a few boxes of punched cards (obsolete now) or a collection of floppies containing data. As such, this interpretation does not constitute a precise definition for the concept.

The literature published on the subject of databases in the field of library and information science tends to focus on the use of online search services and searching techniques. There is not much clearly identifiable literature on the creation and management of textual databases covering the procedural aspects of building and managing information packages consisting largely bibliographic information. This section outlines the concept of database, issues related to its need and its composition.

### **Different Approaches to Database**

In the literature of computer science, one encounters a number of definitions for the term database. Let us examine some of the definitions so as to understand different approaches to this concept.

- (i) "A collection of data on a defined range of subjects together with all the information needed to access that data".

- (ii) "A named collection of units of physical data which are related to each other in a specific manner".
- (iii) "A generalised, common integrated collection of (company or installation owned) data which fulfils the data requirements of all applications which exist in an enterprise".
- (iv) "The term database refers to just the information file. Database software is the set of programs whose function is to manage the data and programs that operate on it. The database system is the entire hierarchy of elements, files and application programs that result in efficient management of information".
- (v) "A database is a collection of data organised in a manner which allows retrieval and uses that data by anyone needing it. A database is organised and designed to allow a large number of users to draw information from it for many purposes in many different formats".
- (vi) "In its most basic form, a database consists of a number of data elements, each of which is a unit of data that is complete in itself. A part number for example, is a typical data element. These elements are organised into logically related groups called data structures. The data files in database system are organised in a fashion that permits their use in several applications rather than a single application. Thus, in a database system the focus shifts from a particular application and its specific input and output needs to a more general requirement for the data files to serve a number of applications'.

## NOTES

It may be noted that each of the above definitions regards the 'database' from a different view point, its access, purpose, description, contents and integration. Yet, each refers to a specific collection of organised data rather than any data on computer-readable media. The last mentioned definition is a comprehensive one, which clearly explains the concept of a database and also brings out the essential difference between a database and a traditional computer file. This definition also refers to a database system. It is important to distinguish the term database from the term database system, all the components of which cooperate to collect, manipulate, manage and deliver information.

### Database Features

One of the main purposes of a database is that the data in the database should be used for a variety of different applications. To achieve this it is important for a database to possess the following features:

it must be substantially non-redundant (that is to say that the database should not have duplication of data) because duplication of data leads to difficulty in ensuring data consistency, and results in the wastage of storage space;

**NOTES**

it must be program-independent so that the data can be moved or restructured without the need to make alterations to programs. This concept is known as 'data independence';

it must be capable of being used by all programs;

it must include all necessary data relationships, to support the variety of different uses to which data is put;

it must have common approach to retrieval, additions and deletions and amendments to data.

A database can be analysed from two viewpoints: the physical storage of the data or logical or conceptual view of data. Files are used to physically store data in a database. Most databases use either direct files or indexed files or a combination of the two to physically store data on disk. Users and applications do not need to know anything about the physical data storage. Stored with actual data will be a description of the database, which enables the DBMS to retrieve information from the database and to store new data in appropriate places in the database establishing relationships with other data if relevant.

The logical or conceptual database is concerned with how the data is logically organised and how the data can be retrieved for information purposes. In case access is required to a series of linked files, it is necessary to have guidelines regarding allocation of data to specific files within the database system, and defining the optimum links between files. Based on the model followed for the structuring of data there are three basic types of databases and the associated DBMS namely: hierarchical, network and relational.

Hierarchical databases are structured in such a way that the relationships between data items follow a branching tree-type arrangement. In other words, the database consists of elements, which act, in a parent-child relationship. The relationships within the database are established when the database is created, that is to say that the database designer defines which is a child element of the parent element. An element within a database can have only one parent element. The data stored in the lower levels of a hierarchy database can only be accessed through the parent element.

The network database approach is based on explicit links or pointers between related entities. In a network model of database, there is more direct link between the data items at various levels. This is achieved by the use of pointers linking data at different levels. This approach requires a large number of links established between data elements, which occupy a large amount of storage space.

Relational databases use a type of data structure, which has been commonly adopted in database systems. In relational database systems, information is held in a set of relations or in the form of tables. Rows in such tables correspond to records while columns in these tables are equivalent to

fields. The data items in various relations are linked through a series of keys. Relational databases are designed using a technique known as 'normalisation'. 'Normalisation' is used to break data into tables so that the fields in each table are dependent only on one key field and not linked to any other key. This process ensures that insertions, deletions and amendments may be made on to the data without any difficulty.

In addition to the above-discussed types of databases, other types of database structures such as multimedia databases and object oriented databases are also in existence. Multimedia structures are used to manage such databases, which deal with pictures, animation, sound and text as well as tables. The storage needs of these materials are certainly different from the types discussed earlier.

The Multi Media DBMS (MM-DBMS) attempts to use a range of technologies like relational technology for tables, image storage devices for graphics and animation, and provide facilities to the users.

## NOTES

### **Types of Databases**

There are many ways of categorising databases. One of the categorisations might be a numerical and textual database. Another way of looking at them is by their coverage *i.e.*, local, regional and global. Databases are generally stored on magnetic or optical media such as disks and accessed either locally or remotely. They may include access to a particular organisation's database covering transactions and financial records, or to other databases that might be accessed remotely. Some of these databases may hold publicly accessible information such as abstracting and indexing databases, full texts of reports or directories, etc., on the other hand, their might be databases which are shared within an organisation or group of organisations.

Databases available to information users in the public domain and which might be accessed remotely or online via some online search services, or locally on CD-ROM, can be categorised as Reference or Source databases.

Reference databases direct the user to another source where the information sought by the user is available. This might be a document, an organisation or an individual. Some of the examples include: bibliographic databases, catalogue databases and referral databases. Let us try to know what these categories of databases are.

### **Catalogue Databases**

These types of databases indicate the collection of a given library or a group of libraries constituting a library network. These databases list the type of collection, namely, monographs, journal titles and other items possessed by the library. They merely provide citations to the documents along with their call numbers to enable easy location of the documents.

## **Referral Databases**

This type of databases offers references to information or data such as names and addresses of organisations, and other directory type of data.

It may be mentioned here that source databases contain the original source data, and are considered as one type of electronic document. After successful consultation of a source database, the user should have the information that is required by him and should not need to seek information in another original source. Source databases may be categorised into:

### NOTES

- numeric databases, which contain numerical data of different types such as statistics and survey data;
- full text databases of news paper items, technical specifications and computer software;
- hybrid databases which contain a mixture of textual and numeric data;
- multi-media databases, which include information stored in a mixture of different
- types of media including sound, video, pictures, text and animation.

## **Bibliographic Databases**

These types of databases contain citations or bibliographic references sometimes along with brief abstracts of literature. They indicate to the user the content of the full text, the source where it can be located (*e.g.*, journal title, conference proceedings) and whether they contain mere citations or provide abstracts or summaries of the original documents covered.

Bibliographic databases, contain a series of bibliographic records, with each record containing some combination of the under mentioned components:

- document number
- author
- title
- source reference
- abstracts
- full text
- indexing terms or keywords or phrases
- citations including the total number of references
- language of the document
- call number or location

Each of these items is known as a data element and is represented by a field. There are different bibliographic record formats and there is considerable variation between them. It may be mentioned here that the

components listed above do not generally (except in case of those containing abstracts) give information of the text of the document but only indicate where the information might be found. Of course, a good informative abstract, if provided for each reference may furnish valuable information to the user and enhance the utility of the database.

### Database Architecture

The architecture of a database is commonly viewed in terms of three separate levels of description: conceptual, external and internal.

The overall logical description of the entire database is the conceptual level. This overall description is commonly known as a schema. It may also be called a community user view. Subsets of the schema that contain only the data needed for particular applications may be defined. These are called sub-schemas or user views. The sub-schemas provide a description at the external level. The description of physical storage structures used to store database on a specific computer system is the internal description.

Explained in simple language, it might be stated that a database can be analysed from two view-points—the physical storage of the data and the logical or conceptual view of data. Files are used to physically store data in a database. Most databases use either direct files or indexed files or a combination of the two to physically store data on a disk.

The logical or conceptual view of a database is concerned with how data is logically organised and how data can be retrieved for information purposes. There are three different methods (architectures) of logically organising data in a database. They are hierarchy model, network model and relational model.

### Advantages of Database Systems (DBMS's)

The Database Systems provide the following advantages over the traditional file system.

1. *Controlled redundancy*: In a traditional file system, each application program has its own data, which causes duplication of common data items in more than one file. This duplication/redundancy requires multiple updations for a single transaction and wastes a lot of storage space. We cannot eliminate all redundancy due to technical reasons. But in a database, this duplication can be carefully controlled, that means the database system is aware of the redundancy and it assumes the responsibility for propagating updates.
2. *Data consistency*: The problem of updating multiple files in traditional file system leads to inaccurate data as different files may contain different information of the same data item at a

NOTES

NOTES

- given point of time. This causes incorrect or contradictory information to its users. In database systems, this problem of inconsistent data is automatically solved by controlling the redundancy.
3. *Program data independence:* The traditional file systems are generally data dependent, which implies that the data organization and access strategies are dictated by the needs of the specific application and the application programs are developed accordingly. However, the database systems provide an independence between the file system and application program, that allows for changes at one level of the data without affecting others. This property of database systems allow to change data without changing the application programs that process the data.
  4. *Sharing of data:* In database systems, the data is centrally controlled and can be shared by all authorized users. The sharing of data means not only the existing applications programs can also share the data in the database but new application programs can be developed to operate on the existing data. Furthermore, the requirements of the new application programs may be satisfied 'without creating any new file.
  5. *Enforcement of standards:* In database systems, data being stored at one central place, standards can easily be enforced, by the DBA. This ensures standardised data formats to facilitate data transfers between systems. Applicable standards might include any or all of the following—departmental, installation, organizational, industry, corporate, national or international.
  6. *Improved data integrity:* Data integrity means that the data contained in the database is both accurate and consistent. The centralized control property allow adequate checks can be incorporated to provide data integrity. One integrity check that should be incorporated in the database is to ensure that if there is a reference to certain object, that object must exist.
  7. *Improved security:* Database security means protecting the data contained in the database from unauthorised users. The DBA ensures that proper access procedures are followed, including proper authenticational schemes for access to the DBMS and additional checks before permitting access to sensitive data. The level of security could be different for various types of data and operations.
  8. *Data access is efficient:* The database system utilizes different sophisticated techniques to access the stored data very efficiently.
  9. *Conflicting requirements can be balanced:* The DBA resolves the conflicting requirements of various users and applications by knowing the overall requirements of the organization. The DBA can structure the system to provide an overall service that is best for the organization.
  10. *Improved backup and recovery facility:* Through its backup and

recovery subsystem, the database system provides the facilities for recovering from hardware or software failures. The recovery subsystem of the database system ensures that the database is restored to the state it was in before the program started executing, in case of system crash.

## NOTES

11. *Minimal program maintenance:* In a traditional file system, the application programs with the description of data and the logic for accessing the data are built individually. Thus, changes to the data formats or access methods results in the need to modify the application programs. Therefore, high maintenance effort are required. These are reduced to minimal in database systems due to independence of data and application programs.
12. *Data quality is high:* The quality of data in database systems are very high as compared to traditional file systems. This is possible due to the presence of tools and processes in the database system.
13. *Good data accessibility and responsiveness:* The database systems provide query languages or report writers that allow the users to ask ad hoc queries to obtain the needed information immediately, without the requirement to write application programs (as in case of file system), that access the information from the database. This is possible due to integration in database systems.
14. *Concurrency control:* The database systems are designed to manage simultaneous (concurrent) access of the database by many users. They also prevents any loss of information or loss of integrity due to these concurrent accesses.
15. *Economical to scale:* In database systems, the operational data of an organization is stored in a central database. The application programs that work on this data can be built with very less cost as compared to traditional file system. This reduces overall costs of operation and management of the database that leads to an economical scaling.
16. *Increased programmer productivity:* The database system provides many standard functions that the programmer would generally have to write in file system, The availability of these functions allow the programmers to concentrate on the specific functionality required by the users without worrying about the implementation details. This increases the overall productivity of the programmer and also reduces the development time and cost.

### Disadvantages of Database Systems

In contrast to many advantages of the database systems, there are some disadvantages as well. The disadvantages of a database system are as follows:

## NOTES

1. *Complexity increases:* The data structure may become more complex because of the centralised database supporting many applications in an organization. This may lead to difficulties in its management and may require professionals for management.
2. *Requirement of more disk space:* The wide functionality and more complexity increase the size of DBMS. Thus, it requires much more space to store and run than the traditional file system.
3. *Additional cost of hardware:* The cost of database system's installation is much more. It depends on environment and functionality, size of the hardware and maintenance costs of hardware.
4. *Cost of conversion:* The cost of conversion from old file-system to new database system is very high. In some cases the cost of conversion is so high that the cost of DBMS and extra hardware becomes insignificant. It also includes the cost of training manpower and hiring the specialized manpower to convert and run the system.
5. *Need of additional and specialized manpower:* Any organization having database systems, need to be hire and train its manpower on regular basis to design and implement databases and to provide database administration services.
6. *Need for backup and recovery:* For a database system to be accurate and available all times, a procedure is required to be developed and used for providing backup copies to all its users when damage occurs.
7. *Organizational conflict:* A centralised and shared database system requires a consensus on data definitions and ownership as well as responsibilities for accurate data maintenance.
8. *More installational and management cost:* The big and complete database systems are more costly. They require trained manpower to operate the system and has additional annual maintenance and support costs.

## Databases

A user seeks information to satisfy his or her information need which may be available in varieties of documents stored a library/information centre. Library catalogues (both card catalogue and OPAC) and other bibliographic databases act as an interface between users and their documents of interest. Users approach to the documents, that can satisfy his/her information need, can be categorised into two groups [Bhattacharya, 1979]—Known document approach and Unknown document approach. In the first case, user can specify a document by using the name(s) of its author(s), or collaborator(s), or series, or by its title. When the query for the document is formulated by using the name-of-subject contained in it, the approach is termed as unknown document approach. The treatment of a document as unknown document in catalogues or bibliographic databases

requires subject description and bibliographic description. Similarly, the treatment of a document to satisfy known document approach calls for bibliographic description and headings or access points derived from bibliographic description. Bibliographic description is the process of recording details for describing and identifying bibliographic items. The process of bibliographic description requires standardisation, for making bibliographic file to be consistent. The development of national, international and regional systems for the exchange of machine readable bibliographic information will not be possible unless a high level of standardisation is attained. A standard related to bibliographic description is a set of general rules agreed to internationally or nationally. The practice of standard bibliographic description in the development of library catalogues and bibliographic databases requires a series of operations. The identification of these unit operations leads to the formulation of principles of bibliographic description and general rules of procedures. The principles and general rules of bibliographic description provide a scientific basis to the field of cataloguing and to the creation of bibliographic records. The application of ICT to the cataloguing process has brought with it demands for precision and logic in the recording of bibliographic data. For bibliographic control, the first principle is standardisation, compatibility and integration. Within the same principle standards may vary. Easy interchange of bibliographic records requires agreement on standards governing the medium of exchange.

Bibliographic data are generated by bibliographers, cataloguers, abstractors and indexers, publishers and booksellers and appear in a range of products, including but not limited to library catalogues, online databases, publishers and booksellers lists, abstracting and indexing services and bibliographies. The nature of bibliographic and cataloguing data appears to be identical, but bibliographic data have an independent existence and need separate consideration. Cataloguing, in fact, is one of the many applications of bibliographic data management. Haggler and Simmons [1982] define bibliographic data as "elements of information, which help to identify a piece of recorded communication as a physical object". They identified three types of functional data groups:

- Data that uniquely identify a particular document and distinguish it from others;
- Data, which reveal an association of two or more documents (e.g., common authorship, continuation or reprint); and
- Data, which describe some characteristics of the intellectual content of a document (e.g., statement of subject).

### **Bibliographic Description: An Overview**

Organisation of bibliographic data elements leads to the creation of bibliographic records. Bibliographic record has been defined as the sum of all the areas and elements, which may be used to describe,

### NOTES

## NOTES

identify or retrieve any physical item of information content. Bibliographic description is the assemblage of data elements sufficient to identify a bibliographic item and to distinguish it from others. In manual systems (e.g., card catalogue), a collection of bibliographic data elements are grouped under the main access points or headings as per the cataloguing code in use. Such record of an item in a catalogue is called an 'Entry'. Entries are usually identified by the kind of access they provide e.g., 'author entry' or 'subject entry'. The distinction between bibliographic record and entry is most visible in computerised environment where the master bibliographic record is stored in the machine and computer programmes generate entries from it. Dempsey [1989] identified three groups of bibliographic dataset—bibliographic description and control data (data describing, identifying and providing controlled access to items), subject data and content description. The first two groups of data generally appear in library catalogues and bibliographic databases. They include:

- data naming an item (e.g., title, alternative title);
- data naming persons or bodies connected with the creation of an item (e.g., author, artist, cartographic agency);
- data describing hierarchical, lateral or lineal relationships between items (e.g., component part, host item, numbering in series, companion item, name of earlier edition or version);
- data indicating intellectual content (e.g., subject heading, abstract);
- data naming persons or bodies connected with the production of an item as a physical object (e.g., publisher, designer);
- data indicating form or nature of item (e.g., bibliography, documentary, novel);
- data indicating mode of expression or communication (e.g., verbal, pictorial);
- data describing the physical appearance, characteristics and constituents of an item (e.g., map, film, dimensions, number of volumes or parts, technical information needed for use); and
- data assigned by a bibliographic or other agency for purpose of identification and control (e.g., ISBN).

The above list shows that bibliographic description deals with two categories of data—data providing access and data describing items. The level and extent of bibliographic description depends on the application and purpose of bibliographic records. The major application domains are the production or creation of:

- authoritative national records and national bibliography;
- bibliographic records for international exchange;
- bibliographic records for cooperative systems;
- records for use in individual libraries;
- records for abstracting and indexing services;

- records for subject bibliographies/ authors' bibliographies;
- records for use in online information retrieval systems (including WWW); and
- records in book trade.

Bibliographic record may be viewed as a package of data, the content of which varies according to the different needs and purposes for which it is intended. The selection and inclusion of data elements for the bibliographic record must be based on user needs. The aggregate of data in a bibliographic record are broadly divided into following groups:

## NOTES

- Descriptive data elements (as defined in the ISBDs);
- Data elements used in headings for persons, corporate bodies, titles and subjects. They function as filing devices or index entries;
- Data elements used to organise a file or file of records (such as classification numbers, abstracts, summaries or annotations); and
- Data specific to the copies of the library collections (such as accession numbers and call numbers).

Bibliographic record should be constructed according to the agreed rules and standards. There are many widely used standards for constructing bibliographic records (*e.g.*, AACR2 for national bibliographies or library catalogues) but the most striking contribution has been made by IFLA, with its programme of ISBDs. ISBD(G) [General International Standard Bibliographic Description] is intended to provide the generalised framework for descriptive information required in a range of different bibliographic activities. The bibliographic data elements which are required for this purpose are set out in eight areas: Title and statement of responsibility –Edition–Material specific data–Publication, distribution data–Physical description–Series–Notes–Standard number and terms of availability. Each of these areas is further divided into discrete elements. The elements are cited in given order and separated by the punctuation prescribed. The complete set of ISBD data is sufficient to ensure identification of bibliographic item and many cataloguing codes (including AACR2) have adopted ISBD(G) as a basis for their own rules for description. The family of ISBDs (includes standard for cartographic materials, non-book materials, printed music, antiquarian books, monographs, serials and other continuing materials and electronic resources) is utilised for the purpose of bibliographical description but the choice and form of access points are based on the Paris Principles (the statement of principles adopted at the International Conference on Cataloguing Principles held in Paris in 1961) adopted in national cataloguing rules. Bibliographic formats (such as MARC family, CCF, UNIMARC, etc.) have also applied ISBDs as base format. But cataloguing codes and bibliographic formats cannot represent all the characteristics of different digital information resources. As a result various general and domain specific metadata schemas (such as Dublin Core, FGDC, ONIX, GILS, etc.) have been developed for description of electronic resources.

## 4.3 SEARCH TECHNIQUES

### NOTES

You must have understood that the searching subsystem is one of the major subsystems of an information retrieval system. In this subsystem, users' queries are received and interpreted, appropriate search statements are formulated, and the actual search (*i.e.*, matching queries with the surrogates of information resources file) is conducted with a view to retrieving the required information. All these tasks can be performed manually, as is used to be done in the earlier systems, or these can be automated.

The information retrieval may be of different types—reference, document, fact and knowledge. In the reference retrieval, information related to specific questions are retrieved. In document retrieval, retrieval by surrogates of documents such as author, title, subject, etc., are made which may provide the information on a bibliographic citation with or without abstracts. Nowadays complete texts are also retrieved (text retrieval system). In the fact retrieval system, specific data or facts are retrieved (*viz.*, numerical databases). The knowledge retrieval system is a rule based system in which there is a knowledge base with capability for knowledge acquisition and inference engine. It processes facts, information into knowledge, based on rules.

The advancements in Information and Communication Technologies (ICT) brought out the possibility of more dynamic searching via online methods. The concept of online searching has occupied a large and significant area in the study and research of modern information retrieval. Online information retrieval systems appeared over four decades ago, and over the years they have developed significantly in terms of number, features and facilities. However, a user often faces difficulties in approaching an online information retrieval system, especially in formulating an appropriate search statement. The cost of searching a database, whether in-house or external, can be reduced significantly if an appropriate strategy for searching is followed. The search strategy helps the user select the optimum path for searching a file or a database. This involves a number of measures that are to be taken before and during a search. This Unit covers the basic concepts of the search strategy and describes the actual searching process in the context of online and web information retrieval systems.

### Search File—An Essential Component

A search file is an essential component of a database which describes the document collection of IR system. Generally the search for information is through subjects or keywords. This implies that every aspect or characteristics of documents are to be properly identified and indexed which form the essential component of a search file. The steps in creating a search file are:

- (a) deciding the subject characteristics of documents;

- (b) translating these characteristics into indexing terms;
- (c) recording the terms with citation/location; and
- (d) arranging term in searchable order.

### **Search Strategies and Pre-requisites**

Search strategy is the action plan which is drawn to conduct a search. It encompasses several steps and levels of work in information retrieval. There are many issues that need to be considered while formulating an appropriate search statement. These are:

- (i) the concepts or facets to be searched and their order;
- (ii) the term(s) that appropriately represent(s) the search concept;
- (iii) the feature(s) of the retrieval system concerned; and
- (iv) the measures to be taken in revising a search statement.

Developing a good search strategy requires knowledge about the nature and organization of target database(s) and also the exact needs of the user. Knowledge of the user's exact requirement can greatly affect the actual search and retrieval process. In some cases, the user may want only a few relevant items on a given topic, in which case the task of searching will obviously be limited. Conversely, the user may wish to obtain all the relevant items (obviously with as small a number of non-relevant items as possible), in which case the search must be exhaustive. Thus, an information search may fall in one of the following three categories.

- (a) *High recall search*: when the user needs to find out all the relevant items on the stated topic. Recall is a parameter used to measure the performance of information retrieval systems; it is measured as the proportion of relevant items retrieved from a collection in a given search session;
- (b) *High precision search*: when the user needs only relevant items, i.e., as small a number of non-relevant items as possible. Precision is a parameter used to measure the performance of information retrieval systems; it is measured as the proportion of the retrieved items that are relevant in a given search session; and
- (c) *Brief search*: when the user wants only a few relevant items as opposed to all the relevant items.

### **The Pre-Search Interview**

The results of a search depend heavily upon the correct understanding of the users' precise needs. This understanding can be developed through a pre-search interview. A pre-search interview is a conversation that takes place between a user and member of the information staff regarding the actual information requirement of the user. In order to conduct a successful pre-search interview, the interviewer should possess the following qualities:

NOTES

## NOTES

- ability to conduct personal communication;
- conceptual skills;
- analytical skills;
- knowledge of file organisation;
- understanding of indexing policy and vocabulary control; and
- subject knowledge.

It may be noted that the concept of pre-search interview presupposes the existence of a search intermediary. This concept was developed to get a better understanding of the search requirements of a user in an online search environment. The concept of pre-search interview, although is very important in a mediated search process, has very little relevance in the context of information retrieval from the World Wide Web and digital libraries, since these systems are designed to be used by the end users without any direct involvement of human intermediaries. Nevertheless, an understanding of the activities involved in a search interview process may be quite useful for designers of information retrieval systems in the web and digital library environment.

The involvement of intermediaries reduces the uncertainty of retrieval to a great extent, thus, providing effectiveness in retrieval. The functions of the intermediaries are to ascertain exact information needs of the users as far as possible, identifying the concepts that describe the needs, translating these concepts into the terms of IR system and delineating the relationships among them, stipulating the parameters of the search. This, they do by interpersonal communication and participative discussion. The other functions are selection of database(s) for search, identification of alternative sources, formulation of alternative search strategies, matching of formulated search expression with the features of database(s) and conducting past search review. Many times, these are done by users interface. According to Taylor [1968], queries of a searcher shift from actual need to unexpressed need. The need is refined to the conscious description need or conscious need which is finally formalized as formalized need. The searcher's actual query formulation is done with interactive interface, in case of online systems and reference interviews.

Different techniques are available to conduct searches effectively so that maximum relevant information are retrieved according to users need. Such techniques are:

### **Keyword and Phrase Search**

A search can be conducted by entering a single search term or a phrase comprising more than one term. Keyword search is the simplest form of search facility offered by a search system. In the keyword search mode, the system searches the inverted file (the index) for each keyword/term forming the search expression. The search terms can be entered through the keyboard or can be selected from an index or vocabulary control tool, such as subject headings lists or thesauri. More than one keywords,

forming a search expression, have to be combined using the Boolean or proximity operators (discussed below).

In a phrase search, the system searches for the entire phrase rather than each individual keyword forming the phrase. Phrase searches can be conducted only on those fields that are phrase indexed. If the index file comprises only single terms, then phrase search cannot be conducted, except of course by using the proximity operators where the system searches for each constituent keyword in the search expression separately, and retrieves only those records where the keywords occur consecutively (as prescribed by the proximity operator) like in a phrase. Usually only a few fields in a database are phrase indexed, and phrase search can be conducted only on those fields. A search phrase can simply be entered through the keyboard, or through the index file or vocabulary control tools like subject headings lists and thesauri.

Different search system provides different facilities for conducting keyword and phrases searches. For example, in DIALOG search one can simply enter a keyword or a phrase preceded by the search command (s or select or find) and can click on the search button to conduct the search. User can also restrict the search to one or more fields (for further details on DIALOG search). Some search engines, for example, in AltaVista a phrase search is conducted when a search expression is entered within double quotes.

### **Keyword and Subject Search**

Many bibliographic information retrieval systems provide two types of search facilities for conducting an unknown item search: keyword search and subject search. A keyword search allows users to enter one or more keywords pertaining to their query. These keywords can be chosen by the user in any combination depending on the requirements and there are several search operators that can be used to combine several keywords to formulate a search expression. The search keywords can appear anywhere, or in one or more chosen fields, in the database records. A subject search is the one that allows user to submit a subject expression that reflects his/her information requirement. Such a search is conducted on the subject field that contains one or more assigned subject headings entered by the indexer while creating the record in the database. Thus, a record will be retrieved only when user's subject search expression exactly matches the subject heading assigned by the indexer. In order to standardize the process, and also to help the user identify the appropriate subject heading, information retrieval systems use certain tools, called vocabulary control tools.

### **Boolean Search**

This is a very common search technique that combines search terms according to the Boolean logic. Three types of Boolean search are possible: AND search, OR search and NOT search.

## NOTES

Boolean AND search allows users to combine two or more search terms using the Boolean AND operator. A Boolean AND search will retrieve all those items where all the constituent terms occur. For example, the following search expression "Internet and WWW" will retrieve all those records where both the terms occur. Boolean AND search adds more restrictions to a search expression by adding more search terms. Therefore, the more search terms are ANDed, the more restricted, or specific, will be the search, and as a result the less will be the search output. Sometime, a search may not produce any result if too many search terms are added.

Boolean OR search allows users to combine two or more search terms such that the system retrieves all those items that contain either one or all of the constituent terms. Thus, the following search expression "Colleges or Universities" will retrieve all those records (1) where the term Colleges occurs, (2) where the term Universities occurs, and (3) where both the terms occur. Note that this is contrary to the use of the term 'or' in normal English. Boolean OR search, though adds more terms to a search expression, adds less restrictions to a given search expression, because the search is conducted for occurrence of each single ORed term irrespective of whether the other term(s) occurs or not. Consequently, the output of OR searches will be more. When too many search terms are ORed, the search output may be too big to handle.

Boolean NOT search allows users to specify those terms that they do not want to occur in the retrieved records. For example, the following search expression "Search engines NOT Hotbot" will retrieve all the records on search engines except those where the term 'Hotbot' occurs. Boolean NOT searches add restrictions to a search by forcing the search system to discard those items where the NOT term(s) occur. Hence the search output will decrease with increase in the NOT terms. The operators used for conducting a Boolean search vary from one search system to the other.

In some search engines there are two approaches to conduct a Boolean search. The usual approach is by combining the search terms using the Boolean AND, OR, NOT, or any corresponding operators. For example, in AltaVista and HotBot, Boolean searches can be conducted using Boolean AND (or &), OR (or |) and NOT (or !), and combining terms and operators using parentheses. However, most search engines have another way to conduct search that implies Boolean search. For example, the plus operator (+) placed before a word or phrase means that all returned pages should contain that search term, and thus if the '+' symbol is used before two search terms, then the result will be as good as a Boolean AND search. Similarly the minus operator (-) can be placed before a word or phrase to exclude all documents containing that search term, and this implies the Boolean NOT search.

### Truncation Search

Truncation is a search facility whereby a search can be conducted for all

the different forms of a word having the same common root. As an example, the truncated word COMPUT\* will retrieve items on COMPUTER, COMPUTING, COMPUTATION, COMPUTE, etc. A number of different options are available for truncation, viz., right truncation, left truncation, and masking of letters in the middle of the word. Left truncation retrieves all words having the same characters at the right hand part, e.g., 'hyl' will retrieve words like 'methyl', 'ethyl', etc. Similarly, middle truncation retrieves all words having the same characters at the left and right hand part. For example, a middle truncated search term 'colr' will retrieve both the terms 'colour' and 'color'. A 'wild card' is used to allow any letter to appear in a specific location within a word.

Right truncation and character masking or wild card are the most common truncation search facilities available in search systems. Operators used for truncation search vary from one information retrieval system to another; the most commonly used truncation operators include: \*, \$, !, and ?.

### Proximity Search

This search facility allows user to specify (1) whether two search terms should occur adjacent to each other, (2) whether one or more words occur in between the search terms, (3) whether the search terms should occur in the same paragraph irrespective of the intervening words, and so on. The operators used for proximity search and their meaning differ from one search system to the other. A proximity search is as good as a Boolean and search in the sense that it searches for the occurrence of two or more search terms in the documents. However, it adds more constraints by specifying the distance between the search terms, and therefore the search output becomes more specific.

#### Example

sun within 4 words after moon  
 sun within 3 words before moon  
 sun within 2 words of moon  
 ice within 2 words after fire  
 ice within 4 words before fire  
 fire within 5 words of ice.

### Field-Specific Search

A search can be conducted on all the fields in a database, or it may be restricted to one or more chosen fields to produce more specific results. Specific fields and codes vary according to the search systems and database. The following examples show some valid DIALOG searches that have been restricted to some specific fields. The general format for using suffix codes is: "Syntax: SELECT <term>/xx,xx... where xx is

NOTES

NOTES

a Basic Index field code(s)"

Select computer?/TI Terms searched in the Title (/TI) field only.  
S (information OR Terms searched in either the Descriptor (/DE)  
Communication)/DE, ID or Identifier(/ID) field.  
S S12/TI, AB Restricts set S12 to either the Title (/TI) or Abstract (/AB) field.

In some cases one can use some prefix codes to restrict a search in a specific field. For example, in DIALOG one can enter the following search expressions to restrict the search in author or corporate source:

Select AU = Chowdhury, G

### Limiting Search

Sometimes a user may want to limit a given search by using certain criteria, such as language, year of publication, type of information sources, and so on. These are called limiting searches. Parameters that can be used to limit a search are decided by the database concerned.

---

## 4.4 COMPUTER NETWORKS

---

With the increasing use of computers it was observed that the use of the computer could not be restricted to a particular place. A need was felt to link computers located at different places, e.g., either in the same room or scattered through a building or at distant places for exchange of data/information. A major development, in the early 1950s, was the use of communication links to connect central computers to remote terminals and other peripheral devices. The number of such devices expanded rapidly in the 1960s with the development of time-shared computer systems and with the increasing power of central computers. With the proliferation of remote peripheral devices, it became uneconomical to provide a separate long-distance communication link to each peripheral. Finally to free the central processor from handling all this communication, special processors called front ends were developed to control the communication to and from all the peripherals. Such connectivity of the computer with remote peripherals was referred to as Data Network or Computer Communication Network.

### Network Components

The following are the essential components for computer networking.

### Network Software

Software is the first part of the network. Before selecting a suitable software, the network card, cable and the topology to be used must be decided. The main categories of network software are:

## **Peer-to-Peer**

Peer-to-peer network operating systems allow users to share resources and files located on their computers and to access shared resources found on other computers. However, they do not have a file server or a centralized management source. In a peer-to-peer network, all computers are considered equal; they all have the same abilities to use the resources available on the network.

## **Client/Server-based**

Client server operating systems allow the network to centralize functions and applications in one or more dedicated file servers. The file servers become the heart of the system, providing access to resources and providing security. Individual workstations (clients) have access to the resources available on the file servers. The network operating system allows multiple users to simultaneously share the same resources irrespective of physical location. Some server-based networking softwares are: Novell Netware, Windows NT, Unix, Windows 2000, etc.

## **Network Card**

The network card in the computer represents the middle and most important part of the connection. The network software dictates the network protocol, which in turn dictates the speed at which the network can operate and the kind of cable used. Network cards must match the system bus structure. The bus is physically represented on the bottom card edge that slides into the main system board called motherboard.

Once the card is inserted into the slot of the motherboard it becomes a part of the computer. The cards are physically different. Depending upon the architecture the cards are of three types. They are 8-bit cards, 16-bit cards and 32-bit cards. Each card has its own method of sending information (network protocol) through the cable. The most commonly used protocol is Ethernet protocol. It's a standard for Media Access Control (MAC) sub-layer. This is preferred for smaller networks. It provides a decent transmission speed when compared to the cost of the card and cabling. Xerox Corporation developed it in the 1970's. The *transmission* speed is rated at 10 megabits per second called 10BaseT or 100 megabits per second called 100BaseT. The maximum length of the cable between two computers ranges from 500 to 1500 feet. Gigabit Ethernet built on top of the Ethernet protocol provides high bandwidth capability for backbone design, *i.e.*, 1 Gbps (1000 Mbps). Another type of card eliminates the need to run cable from system to system. Wireless LAN is the method for doing so. The signal used in this is radio or infrared.

- Simplex transmission refers to one-way data exchanges. This requires only a single pair of wires.
- Duplex transmission refers to two-way data exchanges between two parties.

NOTES

- In half duplex transmission, signals are sent in only one direction at a time.
- In full-duplex transmission, signals are sent in both directions simultaneously.

## NOTES

### **Transmission Media**

Communication of data propagation and processing of signals is called transmission. Transmission systems allow transport of signal from one point to another. At the transmitter end, the data is encoded as energy and the energy is transmitted through some sort of medium. At the receiver end, the energy is decoded back into data.

The energy can be electrical, light, radio, sound, etc. Each form of energy has different properties and requirements of transmission. Electrical/ Radio transmission from transmitter to receiver goes over some transmission medium using electromagnetic (EM) waves. The transmission media are as under:

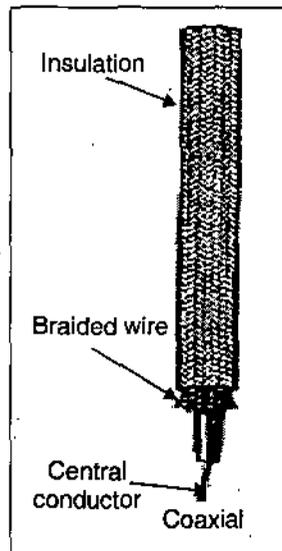
### **Guided Media**

Waves are guided along a physical path; Coaxial cable, Twisted pair, Optical fibre.

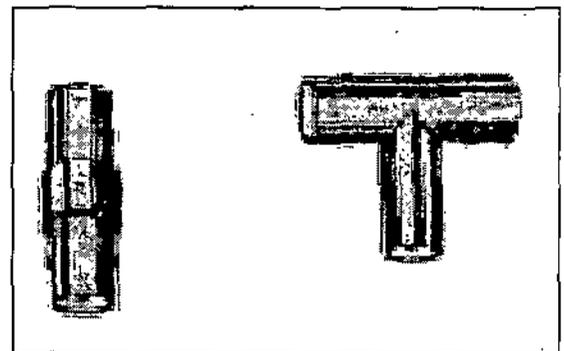
### **Baseband Coaxial cable and BNC Connections**

Coaxial cable, abbreviated "coax" looks like the cable used to bring the cable TV signal to television. One strand (a solid core wire) runs down the middle of the cable. Around the strand is insulation. Covering the insulation are braided wire and possible metal foil, which shield against electromagnetic interference. A final layer of insulation covers the braided wire.

Two kinds of coaxial cables are widely used. The 50-ohm cable is used for digital transmission. The other kind, 75-ohm cable, is used for analogue transmission.



Coaxial Cable



BNC Connector

Coaxial Cable for Ethernet networks is either Thick Ethernet which can carry signals up to 500 metres or Thin Ethernet which can carry signals up to 180 meters. Thin Ethernet is mostly used with Ethernet cards. This cable uses BNC connectors and sleeves of 0.2 inch diameter. These connectors use a twist-to-lock sleeve that attaches to the T-connector on the back of the network card. Cable can be attached to the connector only after carefully stripping the insulation from each layer. Once the cable is inserted the sleeve must be crimped. Many network problems can be traced to poor cable connections. Coaxial cable is not easy to install. It cannot be pulled into a sharp 90 degree angle and requires careful attachment to the BNC connectors.

## NOTES

### ***Broadband Coaxial Cable***

Cable uses analogue transmission on standard cable television cabling. It is called broadband. Although the term "broadband" comes from the telephone world, where it refers to anything wider than 4 KHz, in the computer networking world "broadband" means any cable network using analogue transmission.

Since broadband networks use standard cable television technology. The cables can be used upto 300 MHz (and sometimes up to 450 MHz) and can run for nearly 100 kms due to the analogue signaling, which is much more critical than digital signaling. Typically, a 300 MHz cable will support a total rate of 150 Mbps.

One key difference between baseband and broadband is that broadband systems need analogue amplifiers to strengthen the signal periodically. These amplifiers can only transmit signals in one direction. Therefore Dual Cable Systems have two identical cables running next to each other. All computers transmit on cable 1 and receive on cable 2.

### ***Twisted Pair Copper Wire***

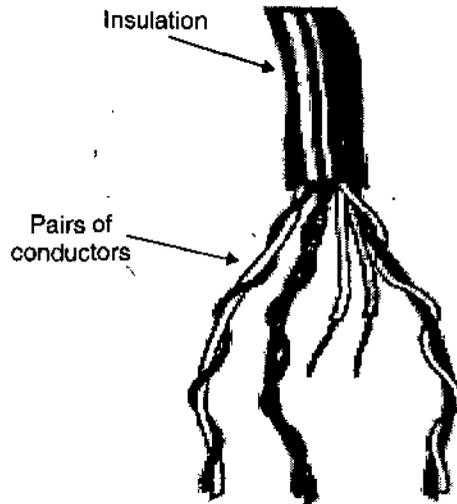
Unshielded Twisted-Pair (UTP) cable and RJ-45 connectors. Unshielded Twisted-Pair (UTP) cables look like phone cables. Each of the pair of wires contained in the cable is twisted with the other. The typical twisted-pair cable rated for network use contains three or four pairs of wires. The arrangement helps shield against electromagnetic interference. There are generally three categories of UTP cabling.

- Category 1:** UTP cables and associated connecting hardware whose transmission characteristics are specified up to 16 MHz.
- Category 2:** UTP cables and associated connecting hardware whose transmission characteristics are specified up to 20 MHz
- Category 3:** UTP cables and associated connecting hardware whose transmission characteristics are specified up to 100 MHz
- Category 4:** UTP cables and associated connecting hardware whose transmission characteristics are specified up to 250 MHz.

The higher the category, the greater the level of protection from any outside (electrical) interference, and of course, the higher the price.

The twisted-pair cable uses small plastic connectors designated as RJ-45. These are similar to the phone connectors except that instead of the four wires, RJ-45 uses eight wire contact.

**NOTES**



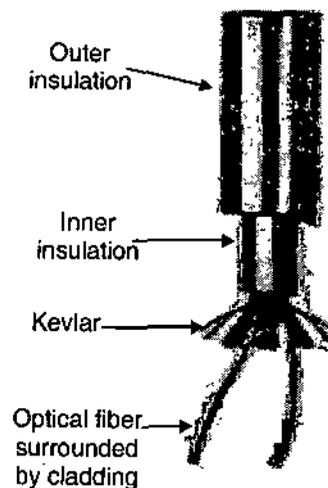
RJ - 45 Connector

Twisted pairs  
UTP Cable

***Fibre Optic Cable***

These cables contain a hair-thin strand of optically pure plastic fibre, surrounded by special shielding and insulation. An optical fibre cable carries pulses of light as opposed to bursts of electricity carried by the twisted-pair or coaxial cables. A fibre optic network uses either a laser or a light emitting diode (LED) to translate electrical signals into light pulses, and then transmits those pulses through the core of the cable.

**All fibres are made up of three general parts:**



Fibre Optic Cable

**Core:** The innermost section, where the light travels, is made from doped silica, i.e., chemicals are added to pure silica (glass) to provide various light transmission properties.

**Cladding Layer:** The cladding layer is made of pure silica and has a lower refractive index than the core. This difference in refractive index is what allows the light to travel within the core. The core/cladding interface acts like a mirror to keep the light from refracting out of the fibre. Optical fibre works because of principle of total internal reflection.

**Coating:** The outer layer of the fibre is coating. Generally made of an acrylate material, the coating has no effect on the transmission properties. It protects the fibre when it is being processed into a cable and when being handled during installation and termination

**Types of Fibre:** There are two general types of fibres available for use in data communication networks today - Multi mode and Single mode.

**Multimode:** At presently backbones in buildings or campuses use multi-mode fibres which meet both the distance and data rate demands of most LAN networks. Multimode cables cost far less than single mode cables since the optoelectronics used with multimode fibres are much cheaper than those used with single mode fibre. This cost advantage is the main reason for the popularity of multimode fibre over single mode fibre in LANs. Because of larger core diameter, multimode fibre LED sources are cheap.

**Single Mode:** As the name suggests, single mode fibre is designed to support one mode of light. It is used with a very narrow laser source and carries large amount of information over long distances. Its core size is typically around 8.5 um and cladding diameters are the standard 125 um. Single mode fibre cables are preferred where long distance and higher information carrying capacity is required. Because of small core diameters, single made fibres require lasers for effectively launching light into the fibre. Lasers are costlier than LEDs.

**Connectors:** There are many types of fibre optic connectors available today. They differ in their installation method, materials and shape, but perform the same basic function. Some of the most popular connectors are SC, ST, SMA, D4 connectors etc.

### **Unguided media: Waves are not Guided; Air Waves, Radio**

Data communications through space are predominantly sent by radio communications microwave links. The space media can be infrared, light, microwave and radio carrier.

#### ***Microwave***

A microwave link can carry huge traffic/data between two stations without a physical connection. It can connect longer distances with the help of a repeater. It is used for the connectivity between two

NOTES

NOTES

distant stations where legacy connectivity with cable/optical fibre laying is difficult. It requires Parabolic dish antenna, about 3m in diameter, fixed rigidly with a focused beam and its coverage per hop is limited by line-of-sight. Long distance microwave transmission is achieved by a series of microwave relay towers. With no obstacles, maximum distance (D, in km) between antennae can be

$$D = 7.14 \sqrt{Kh}$$

Where 'h' is antenna height and 'K' is an adjustment factor to account for the bend in microwave due to earth's curvature, enabling it to travel further than the line of sight; typically  $K = 4/3$ . For example two microwave antennae at a height of 100 m may be as far as

$$7.14 \times \sqrt{133} = 82 \text{ km}$$

The range of operational frequency is, mainly from 2 GHz to 8 GHz. The microwave link is susceptible to fading in atmosphere specially heavy rain absorbing microwave, resulting in attenuation. The transmission hierarchy can be PDH/ or SDH (Synchronous Digital Hierarchy). It normally uses Quadrature Amplitude Modulation (QAM) technique.

### **Radio Communication**

Radio Channels carry signals in the electromagnetic spectrum. They are an attractive media because they require no physical "wire" to be installed, can penetrate walls, provide connectivity to a mobile user, and can potentially carry a signal for a long distance. The characteristics of a radio channel depend significantly on the propagation environment and the distance over which a signal is to be carried. Environmental considerations determine path loss and shadow fading (which decreases a signal's strength as it travels over a distance and around/through obstructing objects), multiple fading (due to signal reflection off interfering objects), and interference (due to other radio channels or electromagnetic signals). It operates at a lower frequency than microwave mainly hundreds MHz. It may operate in Time Division Multiple Access (TDMA) in Coded Division Multiple Access mode (CDMA).

Terrestrial radio channels can be broadly classified into two groups: those that operate as local area networks (typically spanning 10's to be a few hundred meters) and wide area radio channels that are used for mobile data service (typically operating within a metropolitan region). A number of wireless LAN products are in the market, operating in the 1 to 10's of Mbps range. Mobile data service typically provides channels that operate at 10's of Kbps.

### **Satellite Communication**

The communication satellite is a microwave relay station between two or more ground stations (also called earth stations). It uses different frequency bands for incoming (uplink) and outgoing (downlink) data. Satellite communication is mainly a broad-cast network. It can operate

in C-Band, KU-Band etc. Access of satellite is FDMA, TDMA. Most satellites use 5.925 - 6.425 GHz band for uplink and 3.7 - 4.2 GHz band for downlink (4/6 band). A single satellite can operate on a number of frequency bands, known as transponder channels or transponders. The propagation delay of about a quarter second due to long distance leads to problems in error control and flow control. Geostationary satellites permanently remain above the same spot on the earth. This stationary presence is achieved by placing the satellite in orbit at 36,000 kilometers above the Earth's surface. This huge distance between ground station through satellite back to ground station introduces a signal propagation delay of 250 milliseconds. Nevertheless, satellite links are often used in telephone networks and in the backbone of the Internet. VSATs (Very Small Aperture Terminals) are used to share a satellite capacity for data transmission.

VSATs are basically satellite earth terminals having very small antennae (diameter of 2.4 m or less). They normally use geosynchronous satellites for supporting two way telecommunication and information services such as voice, data and video. Two VSATs cannot directly communicate with each other but they do so via a large earth station called the Hub Station. A VSAT network offers several advantages like better reliability, high availability, modular/easy network growth, network access to / from remote places. They are cost effective and simple to operate, maintain, monitor and control.

A VSAT consists of two units viz. an outdoor unit (ODU) and an indoor unit (IDU). The outdoor unit is basically mounted alongside the antenna. Because of its small size, the antenna and ODU can be installed on rooftops. The ODU contains RF and IF electronics whereas the baseband electronics comprising digital circuits is housed in the IDU. These two units are connected to each other by a cable. In addition to RF/IF signals, this cable carries power supply for the ODU. A personal computer can directly be hooked onto the indoor unit of VSAT. Alternatively, the indoor unit can be interfaced with LAN via a suitable router.

### ***Infrared Communication***

Limited to short distances and highly directional, it cannot penetrate walls. It requires no licensing; and no frequency allocation issues.

### ***Transmission Capacity***

#### ***Bandwidth***

The amount of data that can be transmitted via a given route of lines from server to client and vice versa.

Data transmission rates expressed in bits/second are unambiguous and indicate the rate of data transfer exactly. The baud rate, however refers to the rate of transitions imposed onto a transmission line.

NOTES

## **Measure of Performance**

### NOTES

### **Throughput**

Throughput and Latency are completely independent issues. A channel with high throughput can move large quantities of data rapidly, but the first bit of data can never arrive faster than the latency permits. Connectionless network, such the Internet and other IP-based networks, use throughput capacity efficiently, rather than minimizing latency.

### **Latency**

Latency refers to the delay between the occurrence of two events. Some of the most widely used latency measures for networks are end-to-end trip time, round-trip-time, key-stroke response time, and transaction complete time.

End-to-end trip time is the time it takes a packet or other unit of data to travel from source to destination. Round-trip time adds the time for a return response or acknowledgement to the end-to-end latency.

### **Data Transfer Rate**

Speed at which data can be transferred, once transmission has begun.  
(bit/sec)

Message transfer time = latency + length of message / Data transfer rate  
Shannon's limit (on ideal conditions):

Max. data rate [bit/s] = carrier BW [Hz] · log<sub>2</sub> (1 + ( signal / noise))

ex.: phone line BW = 3 kHz, S/N = 30 dB = 1000

Max. data rate = 30 kbit/s

## **Types of Computer Network**

### **Types of Networking**

There are three types of networking, which may be said to have considerable impact on library and information services. They are: communication networks, computer networks and information networks. Although as information professionals, we are to understand the functioning of each of these types, in the context of this Unit, we shall limit our discussion mainly to communication networks.

### **Communication Networks**

Communication networks are made up of transmission lines, concentrators, switching mechanisms and non-data processing components. Due to increased competition in the communication industry, new and specialised common carriers and value-added networks with data processing equipment, is also linked to communication networks.

The current technological trend setter in communications is "packet switching technology". The greatest advantage of packet switched networks is in

applications where the distance involved is great. Information services involve relatively large quantities of data for full bibliographic or full text retrieval, and therefore, the distance advantage applies more clearly to these applications. Depending upon the area of administrative jurisdiction, networks fall into three categories: Local Area Networks (LANs), Metropolitan Area Networks and Wide Area Networks (WANs).

## NOTES

***Local Area Networks (LANs)***

A Local Area Network connects a large number of different types of equipment, including computer terminals, fax, telex, CD-ROM players, etc., on a single site. The majority of LANs use packet switching technology and offer very high transmission speeds — up to one hundred times faster than those available on public networks at present. Most of these are baseband that is they carry only digital computer data, but broadband LANs can carry video and voice signals as well as data. The topology of LANs can vary greatly (*i.e.*, linear, star, ring) affecting efficiency, reliability and costs.

***Metropolitan Area Networks***

These type of networks generally cover specific metropolitan area and provide to them all the services.

***Wide Area Networks (WANs)***

Wide Area Networks (WANs) are private networks linking equipment at several different sites and have much in common with conventional large scale data networks.

In configuring a network to communicate information the focus is on minimising costs while maintaining the desired performance (or minimum desired quality) standards depend on the nature of the information to be sent. In the case of voice, quality is measured by word intelligibility or articulation index. For video, the Television Allocations Study Organisation has adopted the TASO grades ranging from excellent to poor that specify the degree of subjective acceptability of the picture displayed to viewing audiences. For both voice and video, the corresponding quantified measure of quality, signal-to-noise ratio, is used as a technical design parameter. For digital information, say from one computer to the other the quality of the transmission is expressed as the number of bit 'errors', a appearing as '0' or vice versa, that occur as a fraction of the total number of bits sent.

Along with performance requirements, it is necessary to choose the channel bandwidth for analog transmission. The transmission speed of digital signals is also referred to as data rate or bit rate. These factors partly determine cost. Therefore, depending on whether the transmission is from one point to another, one point to several, or several points to a central point, a network is designed. The simplest 'network' might

**NOTES**

involve just two stations (point to point) linked for example, to transfer data from one computer to another, or it might be a complicated system where several stations are connected to a local node, several local nodes connected to a central node, and so on in hierarchical fashion.

Developments in network technology include the increasing use of computers to control switching and information flow, most significant is the modernisation of the public network so that voice and data can be carried over a single network. This type of single network is referred to as an Integrated Services Digital Network (ISDN).

---

## **4.5 LIBRARY AND INFORMATION NETWORKS**

---

The world is in the information age. It is commonly recognized that information and its associated technologies are playing more and more important roles in socio-economic development, and that a nation that does not keep pace with the latest advances in information and communication technologies will be left behind in world development. Like most countries in the world, India is drawing up its own information super-highway programme, trying to make it part of the international super-highway system and has greatly promoted the computerization and networking of library and information services.

Computer application in the library and information field has made phenomenal progress with the development in hardware, software and communication facilities. In view of these technological advancements, many libraries in the western countries are able to computerise their entire library and information system with great success. Besides improving services and operations for a better performance, libraries are also able to evolve effective computer networks towards optimum utilization of resources and facilities. Such efforts are there in many countries, including India.

The collection of a library could broadly be classified into two groups— one satisfying the core interest of the institution to which the library belongs, and the other serving peripheral interests. With the financial crunch, the library could restrict acquisition of materials in the peripheral areas, but tries its best not to shed anything from its core acquisition list. India is spending more than Rs. 5000 million a year on procuring new library resources to meet its users requirement, and in that around Rs. 2,000 million is spent on acquisition of costly periodicals/journals. The annual increase in the price of these resources means that even financially sound organizations experience difficulties in meeting their user requirements. Therefore, in a collective development situation, it is logical for a library to look up to other institutions for meeting its peripheral interests. This requires networking of libraries to meet the requirements of users. Apart from supplementing the capacity of individual libraries in providing conventional facilities, the collective effort could bring new generation services like the Bulletin Board and Online Access to remote databases.

The activities concerning library automation, inter-library cooperation, generation of new services naturally demand time and energy of the participating libraries. However, it is unlikely that, the libraries would get additional manpower for such activities, One way to handle the situation would be to relieve the librarians of some of the activities such as book processing, cataloguing and classification through the shared efforts, so that they could attend to new services.

A network is developed when a group of libraries and/or information centres decide to exchange information through computer application. UNISIS-II working document defined the term Information Network as 'a set of inter-related information systems associated with communication facilities, which are cooperating through more or less formal agreements in order to jointly implement information handling operations with a view to pooling their resources and to offer better services to the users. They generally follow identical or compatible rules and procedures'.

The term 'Network' is used in the present times in place of 'Resource Sharing' or 'Cooperative systems'. Networking and Modernization are becoming very important in all types of libraries as they enable the users to have access to the resources of many other libraries in addition to their own one. The benefits of networking include: Preparation of union catalogues, retrospective conversion, provision of bibliographies, optimum use of resources including rare collections, cooperative acquisition of documents, resource sharing, time saving, minimizing cost of the library services. With the rapid advancements in the technologies such as telecommunications, satellite communications, computers, etc., several networks are either functional or in the formative stage in India.

### **Objectives of Library and Information Networks**

- Provide reliable access to document collection of libraries, *i.e.*, Union Catalogues etc.
- Provide access to worldwide bibliographical information.
- Provide document delivery service.
- Optimise information resources through resource sharing mechanisms.
- Facilitate computerisation of all the libraries.
- Facilitate communication among teachers, students, scientists and others.
- Provide effective access to library resources.
- Encourage resource sharing.
- Train manpower required by participating libraries.
- Evolve standards, uniform guidelines, methods, and procedures, both for data capturing as well as hardware and software.

NOTES

## Definitions

### NOTES

One comes across several definitions of library networks. Susan Martin defines a network as a "group of individuals or organizations that are interconnected to form a system to accomplish some specified goal. This linkage must include a communications mechanism, and many networks exist for the express purpose of facilitating certain types of communication among members."

Joseph Becker suggests that, "when two or more libraries engage formally in a common pattern of information exchange, through communications, for some functionally interdependent purpose, we have a library network."

**UNISIST II**, main working document defines Information Network as "A set of interrelated information systems associated with communication facilities, which are cooperating through more or less formal agreements and institutional agreements, in order to jointly implement information handling operation, with a view to pooling their resources and to offer better services to the user. They generally follow identical or compatible rules and procedures".

### Existing Data Networks in India

The development of computer communication networks, an important technological transmission. Telecommunication networks form an integral part of accessing, communicating, and transforming information. In India, the Department of Electronics, Department of Telecommunications (DOT), Videsh Sanchar Nigam Limited (VSNL), Government of India, are responsible for providing and maintaining national and international telecommunication facilities. These communication networks are:

#### **INDONET**

The INDONET, a computer-based network commissioned by Computer Maintenance Corporation (CMC), was the first Indian commercial computer communication network. It came into operation in 1986. It is an integrated information management and distributed data processing facility spanning the entire country. The INDONET aims to provide facility for distributed data processing on an all-India basis to large organizations in the network using the CMC computers for their data processing operations. It also plans for provision of data communications between its users in their respective locations in the network, even if the users are not accessing CMC's nodal computers. Distributed databases in various subjects and access to specialized applications software locally, or in remote locations obviate the need for duplication of software and hardware facilities at each location.

INDONET-CMC's Value Added Services network is a data network, which supports both TCP/IP and x.25 Network Protocols. It is approved by the Department of Telecommunications having:

## High speed leased links of 64 kbps connecting 9 locations in India

- Delhi, Mumbai, Kolkata, Chennai, Bangalore, Hyderabad, Ahmedabad, Pune, Vizag.
- Has alternate Datalinks at each location, for better reliability and uptime.
- 2 RISC based servers at each location

### Present services of INDONET include:

- E-mail/E-mail-fax services
  - File Transfer, Credit card Authorization, Database Services
- Hosting User Applications in the Net
  - Providing Network-based connectivity solutions, Corporate Intranet
  - Network Management, Managed Data Network Services, Help Desk
  - EDI Services, Webbased applications
  - Students registration in the Net for admission
  - Students assignment submission grading over the Net.
  - Electronic Reference Library access in the Net
  - Data Management and data processing services
  - Recruitment Services using Newer Technologies (OMR, OCR) etc.
  - ISP Services

### Future services planned are:

- Full fledged E-commerce services, ERP/Application Integration Services
- Voice/Video Conferencing facilities,
- Payment transaction to *Utilities* (Electricity, Water, Telephone etc.)
- Fax-Store and Forward
- Call Centre Services
- VSAT Services

## Ernet (Education and Research Network)

In this era of globalization and hyper-competition, the concept of teaching has under-gone a sea change. Learning and dissemination of information is becoming more important. Internet-based education and e-learning are the trends of the day. The Department of Electronics had initiated a project "ERNET" with funding from UNDP. The objective was to

NOTES

NOTES

create expertise, R&D and education in the country in the area of networking and the Internet in the country.

Today ERNET is the largest nationwide terrestrial and satellite network with a presence in the premier educational and research institutions in major cities in the country. The focus of ERNET is providing network connectivity and meeting the entire needs of educational and research institutions by hosting and providing relevant information to their users. Research and Development and training are integral parts of ERNET activities.

ERNET (Education and Research Network) has made a significant contribution to the emergence of networking in the country. It has built up national capabilities in the area of networking, especially in protocol software engineering. It has not only succeeded in building a large network that provides various facilities to the intellectual segment of Indian society, that is the research and education community, it has over the years become a trendsetter in the field of networking. The Govt. of India has committed itself to further strengthen the project by including it in the 9th Plan with the allocation of funds and by creating of a new organisational set-up in the form of a Society.

ERNET was initiated in 1986 by the Department of Electronics (DoE), with funding support from the Government of India and the United Nations Development Programme (UNDP), involving eight premier institutions as participating agencies-NCST (National Centre for Software Technology) Bombay, IISc (Indian Institute of Science) Bangalore, the five IITs (Indian Institutes of Technology) at Delhi, Bombay, Kanpur, Kharagpur and Madras, and the DoE, New Delhi. ERNET began as a multiprotocol network with both the TCP/IP and the OSI-IP protocol stacks running over the leased-line portion of the backbone. Since 1995, however, almost all traffic is carried over TCP/IP.

**The Objectives of ERNET include:**

- ERNET operations, *i.e.*, providing state-of-the-art communication infrastructure and services to academic and research institutions, Govt. organisations, NGOs, private sector R&D organisations, and various other non-commercial organisations;
- Research and development;
- Training and consultancy;
- Content development.

**Achievements**

- Foundation of a national capability building in the area of computer networking laid through:
- Setting up of a chain of core groups as the participating agencies with a minimal set of lab facilities and creation of skilled manpower to carry out R&D

- Generating manpower at different levels
- Making the world of standards (TCP/IP, OSI etc.) well understood
- Providing an insight into emerging issues such as ATM networks, networked multimedia, and information infrastructure
- Network infrastructure and services set up, including
- Installation, maintenance and operation of large campus LANs
- Design, commissioning and testing of SATWAN hub and the installation of VSATs
- Seamless interconnection of LAN-WAN segments and multi-protocol capability provided
- Provision of the whole range of Internet services
- Deployment of TDM/TDMA based VSAT network for Internet access
- Research and Development
- Research and development in the area of computer networking has been the forte of ERNET.

## NOTES

**NICNET: National Informatics Centre Network**

The satellite based National Informatics Centre Network (NICNET) (Seshagiri et al, 1987) was set up to provide informatics services to the Central and State Government Departments and then organizations. NICNET provides state of the art solutions and decision support for information management and decision support requirements of the Government of India and the corporate sector. The information technology services provided by NIC range from conducting feasibility studies for developing and implementing computer-based information systems, undertaking large turnkey networks and imparting training. It has developed extensive expertise in integrating IT-based systems with the working of user organizations. NIC continue to provide value added network services, viz., E-mail, database access, Internet etc.

The network consists of Master Earth Station, Remote Micro Earth Stations and a Geosynchronous Satellite. The master earth station is located at CGO Complex, New Delhi. It comprises a 13-metre antenna, a network control centre and a packet switch. The micro earth station connects remotely located district computers with the State computer, which, in turn, are connected to the regional headquarters. Specialized services such as computer-aided design and computer-aided management are also offered over NICNET. The network supports X.25 switch and operates at 1200 bps transmission speed and 19.2 Kbps receive speed. The present configuration handles 300 packets (128 bytes each) per second. The host computers are connected to this packet switch.

NICNET is currently using the INTELSAT-V satellite which mainly functions as a relay station between the master earth station and the

**NOTES**

micro earth station. NICNET has used terrestrial communication for distribution of terminals (Local and Remote from NEC-S1000, CEBER-730, CYBER-830, ND-550 or Super AT386 systems) and for development of Local Area Network. Terminals in a building are distributed over RS-232C cable or dedicated lines using line drivers, depending upon the distance involved. The terminals outside the building are connected over data circuits leased from the local telephone authorities. Dial up access using Public Switched Telephone Network (PSTN) is also provided to many users.

The Local Area Network with a gateway to the NICNET has been developed at the NIC Headquarters. A large number of terminals have been distributed from super ATs installed in various buildings. These super ATs are connected to the host machine/packet switch in NIC Headquarters. Initially, it connected 40 interactive local terminals and 11 minicomputer systems to the in-house mainframe CYBER 170/730 system. These are spread out over a radius of 25 km at Delhi. The remote terminals and the minicomputers are connected by point-to-point data circuits in star configuration; 2 concentrators have also been used to support remote terminals. The network facilitates flow of information among 4 national/regional nodes (NEC at Pune, Bhubaneswar and Hyderabad, ND 550's at State capitals), 32 state/union territory nodes and 439 district nodes.

NICNET has expanded as a dedicated network having more than 500 nodes geographically distributed over the country to address the rapidly growing awareness to of computerization in different sectors of the Government. Each district information centre consolidates information for monitoring the socio-economic development of the district. Each district is connected to its State's information centre for flow of information from district level to State level. The State centre in turn sends processed information to the regional and the national centres and is also connected to other States. Hence, any user connected to a remote or master earth station can link to any other remote micro earth station. The national centre at New Delhi is the repository of all information systems and conducts research and development of relevant software and hardware tools.

The various facilities and services of NICNET include:

- Gateway to Internet by providing Internet connectivity
- Electronic mail service
- USENET bulletin board service
- Telnet facility to login to remote host
- File transfer protocol facility
- Research and Education Network of NIC (RENNIC) service to research, education and medical institutions
- Electronic Data Interchange service
- Bibliographic information service

- Geographical Information Systems
- General information Services terminal(GISTNIC) for common public
- Training facilities

### **VIKRAM**

VIKRAM (Lahiri, 1991, pp. 13-14) is the packet switched public data network under development by the Department of Telecommunications. This network will initially have 8 switching nodes in Delhi, Bombay, Kalkata, Madras, Bangalore, Hyderabad, Ahmedabad and Pune and 12 remote access nodes with its network management centre located at Delhi. It will support packet switching interface to CCITTs X.25, X.28, X.29 and X.75 recommendations.

### ***Bibliographic Applications***

Like NICNET and INDONET, Vikram also has bibliographical applications. It has offered its infrastructure to NISSAT for pilot experimentation on library networking in the country.

### **BITSNET**

*Recognizing the importance of information technology for pursuing advanced research in modern biology and biotechnology, a bio-informatics programme, envisaged as a distributed database and network organisation, was launched during 1986-87. The programme has become a very successful vehicle for transfer and exchange of information, scientific knowledge, technology packages, and references in the country involving 10-12 thousand scientific personnel. Ten Distributed Information Centres and an Apex Centre at the Department of Biotechnology, and 44 Sub-Distributed Information Centres, located in universities and research institutes of national importance, are fully engaged in this task. Six national facilities have been set up for interactive graphics-based molecular modeling and other bio-computational needs. Four long-term courses at the level of post MSc Diploma in Bio-informatics, at Poona University, Jawaharlal Nehru University, Kolkata University and Madurai Kamaraj University, are fulfilling the long outstanding need for trained human resources in this inter-disciplinary area.*

### **INET**

INET is India's X.25-based packet switched public data network; it was commissioned by DOT and paved the way for highly reliable, cost effective and flexible ways of national data transfer and information access. Packet switching enables error-free transmission with dynamic rerouting of calls and provides interconnection between computers/terminals at different speeds and protocols. In its first phase, INET had nodes at New Delhi, Mumbai, Kolkata, Chennai, Bangalore, Hyderabad,

### **NOTES**

NOTES

Pune, Kanpur and Ahmedabad; and connected through 9.6 kbps and 64 kbps links. In subsequent phases, this facility was extended to 88 other cities throughout the country. It supports CCITT interfaces/protocols X.25, X.28, X.29, X.3 and X.75. Access is through dedicated leased lines for asynchronous (X.28) or synchronous (X.25) and dial-up mode (X.28). The facilities offered by INET include: reverse charging, closed user groups, fast selects, charge information indication, call redirection, call deflection, abbreviated address calling, hunt groups, multiple packet sizes and network user identification. The typical applications are e-mail services, corporate communications, information retrieval, database services, remote job applications, credit card verifications, travel reservations and electronic fund transfers (DOT, 1999b).

### **SIRNET**

The SIRNET (Scientific and Industrial Network) (SIRNET NETletter., 1990-), a project of INSDOC aims at networking all 40 CSIR laboratories under SIRNET. It was made operational in December 1989. At present, SIRNET provides electronic mail facility as its first application service from the SIRNET servers with a number of user nodes. For transmitting a message, a user has to deposit the message to one of the SIRNET mail service nodes situated at INSDOC, Delhi and at its regional centre at Bangalore from where it can be transmitted to its destination which may be any of the CSIR laboratories at present linked to the mail node. The SIRNET, in turn, is connected to a large network-ERNET (Educational and Research Network) which is connected to the international network UUNET (Unix User Network) through which other international networks like BITNET, CSNET and JANET are accessible. The SIRNET's mail node at the INSDOC also acts as a gateway to ERNET and through ERNET to other networks. Connections between various laboratories of CSIR are established using dial-up telephone lines, while SIRNET is directly connected to DoE mail server VIKRAM which acts as the clearing node in Delhi ERNET.

### **Other Important Library Networks**

#### ***DELNET : Developing Library Network***

DELNET has been in operation since January 1988, and was registered as a society in 1992. It was initially sponsored by the National Information System for Science and Technology (NISSAT), Department of Scientific and Industrial Research, Government of India and is currently being promoted by the National Informatics Centre, Ministry of Information Technology, Government of India, and India International Centre, New Delhi.

DELNET has been established with the prime objective of promoting resource *sharing* among libraries through the development of a network

of libraries. It aims to collect, store, and disseminate information besides offering computerised services to users, to coordinate efforts for suitable collection development and also to reduce unnecessary duplication wherever possible.

DELNET has been actively engaged in the compilation of various Union Catalogues of the resources available in member-libraries. It has created a number of databases listed in the services. All the DELNET databases have been resident on DELSIS, an in-house software developed on BASISPlus, an RDBMS, the product of Information Dimensions Inc. of USA which has been provided to DELNET courtesy National Informatics Centre, New Delhi.

DELNET provides an array of facilities including e-mail to its 243 member-libraries including both institutional and associate institutional members. DELNET'S relentless efforts in resource sharing have proved extremely effective. It has indeed been a big leap towards the modernisation of libraries in India.

### ***Delnet-Service***

#### **Electronic Mail**

DELNET provides RENNIC E-mail facility to its member-libraries which was introduced by the National Informatics Centre. This gives the members access to both national and international E-mail users and also to INTERNET users.

#### **Online Access to different Databases:**

- Union Catalogue of Books of 8,77,772 bibliographic records.
- Union Catalogue of Books : MARC Format has presently 27,231 records.
- Union List of Current Periodicals with 16,497 periodicals and is regularly updated. Union Catalogue of Periodicals contains 10,623 records.
- Database of Periodical Articles has around 2,00,410 records.
- Indian Specialist's Database has 2,000 records of eminent scientists, educationists and writers from all over the country.
- CD-ROM Database has 1,214 Records.
- Union List of Video Recordings has about 2,278 listings.
- Union List of Sound Recordings consists of 708 audio cassette records.
- Union Catalogue of Hindi has Books nearly 3,000 records in Hindi and can be retrieved through the GIST compatible system.
- Urdu Manuscripts' Database contains 210 manuscripts.
- Database of Theses and Dissertations has 16,587 records.
- Retro-conversion facilities provided to the libraries through

NOTES

NOTES

specialised agencies and also facilitates the use of modern tools such as CD-ROMs and online facilities for retro-conversion.

- Referral Services to participating libraries.
- Document Transfer/Copy Facilities for transferring or copying of the documents to its users.
- DELNET conducts training programmes in the use of DELNET services, software, E-mail, AACR2 and LC Subject Headings, Internet, etc., from time to time.

### ***CALIBNET: Calcutta Library Network***

CALIBNET, a Government of India project, has been launched by the National Information Systems for Science and Technology (NISSAT), Department of Scientific & Industrial Research (DSIR); and managed by the CALIBNET Society established under the West Bengal Government's Societies Registration Act 1961. CALIBNET aims to provide individual libraries and their reading members with cost-effective solutions to their information problems.

CALIBNET was planned to link 38 science and technology libraries of various sizes located in the Kolkatta Metropolitan Area in two phases (eight in the first and thirty in the second phase). All these libraries as a pre-requisite, are first to be computerised and then linked to the network. The main features of the network include online/offline output, editing and retrieval of records. Eventually CALIBNET will be connected to other metropolitan networks, viz., DELNET, BONET, MALIBNET, ADINET, etc., through packet-switched data network.

### ***Services***

The network is intended to provide current awareness services, SDI, union catalogue, partial databases and access to national and international networks. Services also include:

- Access to CD-ROM databases
- LCMARC
- Biblio file
- Inside information
- E-mail service

### ***ADINET: Ahmedabad Library Network***

ADINET is a network of libraries in and around Ahmedabad. ADINET was registered as a society in October 1994. It is sponsored by the National Information System for Science and Technology (NISSAT), Department of Scientific and Industrial Research, Government of India.

ADINET aims to bring about a cooperative mode of working amongst the libraries and information centers in and around Ahmedabad. The main

objective of ADINET is to promote sharing of resources and disseminate in of information among member libraries by networking them and creating a centralized Union catalogue of their holdings. It plans to coordinate efforts for suitable collection development and reduce unnecessary duplication wherever possible.

A centralized database of periodicals, books and non-book materials available in the libraries of Ahmedabad is being created by ADINET. Records collected from participating libraries are being suitably formatted and merged to generate a Union Catalogue of Ahmedabad libraries. With the cooperation of participating libraries, this database is being regularly updated.

### **Services**

- Online information—accessing the Union Catalogue.
- Inter-Library Loan—enables a user to find the location of periodical/book/report of his/her interest Books, reports, theses and borrowed on an inter-library loan basis. Inter-Library document delivery service is available.
- Photocopying services
- Current Awareness Services
- Information Service
- Internet Services

### **MYLIBNET: Mysore Library Network**

The National Information System for Science and Technology (NISSAT), Dept. of Scientific and Industrial Research (*DSIR*), took the initiatives design and develop library networks in India in 1985 in order to share the resources available in the libraries located in various parts of the country. The Mysore Library Network was set up during May 1995 in the city of Mysore with financial assistance from NISSAT.

The Mysore Library Network is housed inside the Central Food Technological Research Institute (CFTRI) campus.

### **Objectives of MYLIBNET**

- To share the resources available with all the libraries.
- To provide faster communication to all the libraries through electronic mail facility.
- To develop software tools for better library management.
- To create awareness in the field of latest information technology by conducting seminars/workshops/training programmes.
- To set up an information base in collaboration with industries.
- To conduct surveys.
- To flash the arrival of new books/journals, announcement of events like seminars/ workshops/training programmes.

NOTES

### ***PUNENET: Pune Library Network***

Punenet is planned for implementation in four phases with the objectives of better utilization of funds through sharing of resources by the creation of commonly usable databases and communication between libraries and automating the functions of the individual libraries.

#### **NOTES**

#### ***Services***

The network services to be offered are union catalogue, current awareness, SDI, authority data, acquisition and fund accounting, serials control, books and journals maintenance, circulation, user services, interlibrary loan services, document transfer/ copy, access to national/international databases.

### ***MALIBNET: Madras Library Network***

- Initiated in 1991 by INSDOC
- Feasibility study 1992
- Registered as society in 1993
- Facilities : Computers, Sybase RDBMS, UNIX
- Databases : serials, books, other INSDOC databases
- Membership : Multi type, 15 members
- Services : e-mail, MALIBNET card, CC, DDS, Training
- Location : INSDOC Regional Centre, Chennai.

---

## **4.6 SUMMARY**

---

- This Unit commences with an introduction in which the importance of data and its organisation has been emphasised. The term 'database' has been defined and interpreted in different ways in the literature of computer science. The Unit discusses the concept of 'database' and explains the reasons for its emergence. It also includes a detailed discussion on database approach. It has been emphasised that the database approach developed in an attitude of sharing valued data resources, releasing control of those resources to a common responsible authority cooperating in the maintenance of the shared data resources.
- An information search is a complex process and it involves several stages and several decisions. Traditional online information services were designed as intermediary-based services, in the sense that they were expected to be used by the end users with the help of expert search intermediaries.
- The physical layer is the basis of all networks. Signals travel from transmitter to receiver via a path. This path called the medium can be guided or unguided. The principal guided media are twisted

pair, coaxial cable and fiber optics. The choices for cable is affected by the network card, cable layout, possibility of electromagnetic interference and of course the budget.

- Unguided media include radio, microwaves, infrared, and lasers through air. In LANs there is less emphasis on error recovery and flow control and there are differences in protocol layering arising from differences in topology. The IEEE 802 and CCITT x.25 define standards of LAN and WAN respectively. Optical Networking is the future, fast replacing the traditional media of communications.

NOTES

---

## 4.7 REVIEW QUESTIONS

---

1. Define database in the context of library and information science.
2. Discuss in brief the different levels of database architecture.
3. What is a pre-search interview in the context of online information retrieval?
4. How does a truncated search improve the results of a search?
5. Which of the two search expressions will produce more hits (search results) and why?
  - (a) Classification and indexing
  - (b) Classification or indexing
6. What is the difference between a Boolean search and a proximity search?
7. Describe the type of media access used by Ethernet.
8. Describe the type of media access used by Token Ring.
9. Describe unicast, multicast, and broadcast transmissions.
10. What is the difference between a modem and an ISDN terminal adapter?
11. What do you understand by Library and Information Networks?
12. Discuss the main objectives of library and information networks.
13. What are the efforts made in India for establishing library and information networks?

---

## 4.8 FURTHER READINGS

---

1. Bay ross, Ivan N. (2003). Database Concepts and Systems. Navi Mumbai: Shroff.
2. Chen, peter p. (1976). The Entity-Relationship Model—Tward a United view of Data, ACM TODS 1, No.1, Republished in Michael

NOTES

- Stonebraker (ed.) Readings in Database Systems. San Mateo, California: Morgan Kaufmann (1988).
3. Desai, Bipin C. (1994). An Introduction to Database Systems, New Delhi: Galgotia Publication.
  4. Frank, Lars (1988). Database Theory and Practice. England. New York: Addison-Wesley publishing.
  5. Hernandez, M.J. (2003). Keys in Relational Databases. Boston: Addison-Wesley.
  6. [http://en.wikipedia.org/wiki/Database\\_management\\_system](http://en.wikipedia.org/wiki/Database_management_system).
  7. <http://support.microsoft.com/default.aspx?scid=kb;EN-US;100139#apliesto>.

---

**UNIT V · DIGITAL LIBRARIES**

---

**★ STRUCTURE ★**

- 5.1 Introduction
- 5.2 Digital Library: Concept and Definition
- 5.3 Digitisation of Library-Materials — Metadata
- 5.4 Management of Digital Libraries
- 5.5 Major Elements of the Management
- 5.6 Summary
- 5.7 Review Questions
- 5.8 Further Readings

NOTES

---

**LEARNING OBJECTIVES**

---

After going through this unit, you should be able to:

- define digital libraries
- know about digitization of library-materials
- define metadata
- explain the management of digital libraries.

---

**5.1 INTRODUCTION**

---

The term digital library has a variety of meanings ranging from a digital collection of material that one might find in a traditional library to the collection of all digital information along with the services that make that information useful to all possible users. Digital libraries have evolved as a result of fast technological development in order to cater to the needs of individuals with varying interests in various fields. Although the term digital library has gained popularity in recent years, such libraries they have evolved along the technological ladder for the past 30 years. There is lot of interest in digital libraries today. This is reflected in the fact that an advanced Altar Vista search conducted in early July 1996, on "digital library" or "digital libraries" retrieved about 20000 entries.

NOTES

In spite of the plethora of literature it is not clear what we mean by the term "digital library". The term is rarely defined, or even characterized. It has been applied to an extraordinary range of applications from digital laboratories to collection of electronic journals, software agents that support inquiry based education, collection of e-mail and similar objects, electronic version of a public library, personal library collection and the entire internet among others. It is not easy to see what these have in common except for their digitization. A digital library contains digital representation of the object found in it.

Most understanding of "digital library" probably also assume that it will be accessible via the internet, though not necessarily to everyone. But the idea of digitization is perhaps the only characteristic of a digital library on which there is universal agreement. Digital library is popularly viewed as an electronic version of a library. The term digital library evokes a different impression in various groups. To some it simply means computerization of traditional libraries. But to others who have studied library science, it indicates carrying out the function of libraries in a new way, encompassing new types of information resources, new approaches to acquisition, new methods of storage and preservation, more reliance on electronic systems and networks. But to a computer professional, a digital library is simply a distributed text-based information system, a collection of distributed information service, etc. A digital library is a library of digital documents, artifacts and records. The advantage of having library material in digital form are: (i) the content occupies less space and can be replicated and used electronically, (ii) the content can be made available on networks, (iii) the search for content can be automated.

---

## 5.2 DIGITAL LIBRARY: CONCEPT AND DEFINITION

---

The working group of the US Government's Information Infrastructure Technology and Applications defined the digital library as 'system' providing user with a coherent access to a very large organised depository of information and knowledge. R.R. Larson defined digital library as a global virtual library—the libraries of thousands of "networked electronic libraries". The digital library need not be networked. A digital library is a library which has all the information in electronic form and having electronic devices to have access to the digitized information. Thus digital library is a library which has number of machine-readable publications and facilities for remote access to several databases.

The American Digital Library Federation has defined the digital library as "Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence overtime of collection of digital works so that they are readily and economically

available for use by a defined community or set of communities."

The definition of a digital library can be given as a set of characteristics as follows:

***The digital library is:***

- a collection of services
- a collection of information objects
- a supporting users with information objects
- organization and presentation of those objects
- available directly or indirectly
- electronic/digital availability

A digital library is much more than just the collection of material in its depositories. It provides a variety of services to all of its users. The basis of the digital library is the information objects that provide the content in the form of digital resource. The goal of the digital library is to satisfy user needs for management, access, storage and manipulation of the variety of information stored in the collection of material that represents the holding of the library. The information objects may be digital objects or they may be in other media but represented in the library via digital means (*e.g.*, metadata). They may be available directly over the network or indirectly. Although the object may not even be electronic, and although the objects themselves may not be available directly over the network, they must be represented electronically in some manner.

There are many definitions of a digital library. The terms such as "electronic library" and virtual library are often used synonymously. The elements that have been identified as common to these definitions are :

- The digital library is not a single entity.
- The digital library requires technology to link the resources of many.
- The linkages between the many digital libraries and information service are transparent to the end user.
- Universal access to digital libraries and information services is a goal.
- Digital library collections are not limited to document surrogates, they extend to digital artifacts that cannot be represented or distributed in printed formats.

The aim of a digital library may be to expedite the systematic development of digital resources collection; the means to collect, store and organise information and knowledge in digital form.

NOTES

## Characteristics of Digital Libraries

The structured information in the digital library is called digital object which includes text, audio, video, image, computer programmes, graphics, and multimedia components in digital form. The digital library requires lots of digital technologies. The digital library:

### NOTES

- Provides access to a very large information collection, both primary and secondary
- Support multimedia components
- Provides links to different digital objects
- Supports and provides search and retrieval interface
- Supports the traditional library mission of collection, development, organization, access and preservation.

Important characteristics of a digital library are:

- (i) Digital collection**—In the digital environment a digital library is expected to develop document collection in a digital format.
- (ii) Technology**—It is understood that a digital library will have digital material in its collection. But in the present day context, both digital and non-digital information belonging to a digital library are to be handled using digital technologies.
- (iii) Work and Service**—The professionals supposed to work in a digital library should have necessary training in handling digital information in order to provide the optimum level of effective service.

The most important component of a digital library is the digital collection it holds or has access to. A digital library can have a wide range of resources. It may contain both paper based conventional documents or information contained in computer-processible form. The collection of a digital library may include—a combination of structured/unstructured texts, numerical data, scanned images, graphics, audio and video recordings.

With the assumption that digital libraries are libraries first, some of the important characteristics of a digital library are:

1. Digital libraries are the digital face of traditional libraries that include both the digital collection and the traditional, fixed media collection, so they encompass both electronic and paper materials.
2. Digital libraries will also include digital material that exists outside the physical and administrative bounds of any one digital library.
3. Digital libraries will also include all the processes and services that are the backbone and nervous system of libraries. However, such traditional processes, though forming the basic digital library work, will have to be revised and enhanced to accommodate the differences between new digital media and traditional fixed media.

4. Digital libraries will serve particular communities or constituencies as traditional libraries do now, though these communities may be widely dispersed throughout the network.
5. Digital libraries will require the skills of both librarians as well as computer professional to be viable.

NOTES

---

### 5.3 DIGITISATION OF LIBRARY-MATERIALS— METADATA

---

Digitising a document in print or other physical media (*e.g.*, sound recordings) makes the document more useful as well as more accessible. It is possible for a user to conduct a full-text search on a document that is digitised and OCRed. It is possible to create hyperlinks to lead a reader to related items within the text itself as well as to external resources. Ultimately, digitisation does not mean replacing traditional library collections and services; rather, it serves to enhance them.

A document can be converted into digital format depending on the objective of digitisation, end user, availability of finances, etc. While the objectives of digitisation initiatives differ from organisation to organisation, the primary objective is to improve access. Other objectives include cost savings, preservation, keeping pace with technology, and information sharing. The most significant challenges in planning and execution of a digitisation project relate to technical limitations, budgetary constraints, copyright considerations, lack of policy guidelines and, lastly, the selection of materials for digitisation.

While new and emerging technologies allow digital information to be presented in innovative ways, the majority of potential users are unlikely to have access to sophisticated hardware and software. Sharing of information among various institutions is often restricted by the use of incompatible software.

One of the main benefits of digitisation is to preserve rare and fragile objects with enhancing their access to multiple number of users simultaneously. Very often, when an object is rare and precious, access is only allowed for certain category of people. Going digital could allow more users to enjoy the benefit of access. Although, digitisation offers great advantages for access, allowing users to find, retrieve, study and manipulate material, it cannot be considered as a good alternate for preservation because of ever changing formats, protocols and software used for creating digital objects.

There are several reasons for libraries to go for digitisation and there are as many ways to create the digitised images, depending on the needs and uses. The prime reason for the digitisation is the need of the user for convenient access to high quality information. Other important considerations are:

NOTES

### **Quality Preservation**

The digital information has potential for qualitative preservation of information. The preservation-quality images can be scanned at high resolution and bit depth for best possible quality. The quality remains the same inspite of multiple usage by several users. However, caution need to be exercised while choosing digitized information as preservation media.

### **Multiple Referencing**

Digital information can be used simultaneously by several users at a time.

### **Wide Area Usage**

Digital information can be made accessible to distant users through the computer networks over the Internet.

### **Archival Storage**

Digitisation is used for restoration of rare material. The rare books, images or archival materials are kept in digitised format as a common practice.

### **Security Measure**

Valuable documents and records are scanned and kept in digital format for safety.

### **Selection of Material for Digitisation**

To begin the process of digitisation, first of all, we need to select documents for digitisation. The process of selection of material for digitisation involves identification, selection and prioritisation of documents that are to be digitized. If an organisation generates contents, strategies may be adopted to capture data that is "born digital". If documents are available in digital form, it can be easily converted into other formats. If the selected material is from the external sources, IPR issues need to be resolved. It is important to obtain permission from the publishers and data suppliers for digitisation, if material being digitized is not available in public domain. Moreover, decision may be taken whether to OCR the digitized images. Documents selected for digitisation may already be available in digital format. It is always more economical to buy e-media, if available, than their conversion. Moreover, over-sized material, deteriorating collections, bound volumes of journals, manuscripts etc. would require highly specialised equipment and highly skilled manpower.

The documents to be digitized may include text, line art, photographs, colour images, etc. The selection of document, need to be reviewed very carefully considering all the factors of utility, quality, security and cost. Rare and highly required documents and images are given first priority in selection without considering the quality. Factors that may be considered before selecting.

**Audio**

The sound quality has to be checked and required corrections made together by the subject expert and computer sound editor.

**Video**

The video clippings are normally edited on Beta max tapes which can be used for transferring it on digital format. While editing colour tone and resolution is checked and corrected.

**Photographs**

The selection of photographs is very crucial process. High resolution is required for photographic images and slides. Especially the quality, future need and the copyright aspects have to be checked.

**Documents**

Documents which are much in demand, too fragile to handle, and rare in availability are reviewed and selected for the process. If the correction of literary value demands much input, then documents are considered for publication rather than digitisation.

**Steps in the Process of Digitisation**

The following four steps are involved in the process of digitisation. Software, variably called document image processing (DIP), Electronic Filing System (EFS) and Document Management Systems (DMS) provides all or most of these functions:

**Scanning**

Electronic scanners are used for acquisition of an electronic image into a computer through its original that may be a photograph, text, manuscript, etc. An image is "read" or scanned at a predefined resolution and dynamic range. The resulting file, called "bit map page image" is formatted (image formats describes elsewhere) and tagged for storage and subsequent retrieval by the software package used for scanning. Acquisition of image through fax card, electronic camera or other imaging devices is also feasible. However, image scanners are most important and most commonly used component of an imaging system for transfer of normal paper-based documents.

**Indexing**

If converting a document into an image or text file is considered as the first step in the process of imaging, indexing these files comprises the second step. The process of indexing scanned image involves linking of database of scanned images to a text database. Scanned images are just like a set of pictures that need to be related to a text database describing them and their contents. An imaging system typically stores a large amount of unstructured data in a two file system for storing

NOTES

## NOTES

and retrieving scanned images. The first is traditional file that has a text description of the image (keywords or descriptors) along with a key to a second file. The second file contains the document location. The user selects a record from the first file using a search algorithm. Once the user selects a record, the application keys into the location index, finds the document and displays it.

Most of the document imaging software packages, through their menu driven or command driven interface facilitate elaborate indexing of documents. While some document management system facilitates selection of indexing terms from the image file, others allow only manual keying in of indexing terms. Further, many DMS packages provide OCR'd capabilities for transforming the images into standard ASCII files. The OCR'd text then serves as a database for full-text search of the stored images.

### *Storing*

The most tenacious problem of a document image relates to its file size and, therefore, to its storage. Every part of an electronic page image is saved regardless of present or absence of ink. The file size varies directly with scanning resolution, the size of the area being digitized and the style of graphic file format used to save the image. The scanned images, therefore, need to be transferred from the hard disc of scanning workstation to an external large capacity storage devices such as an optical disc, CD-ROM/DVD-ROM disc, snap servers, etc. While the smaller document imaging system may use offline media, which need to be reloaded when required, or fixed hard disc drives allocated for image storage, larger document management systems use auto-changers such as optical jukeboxes and tape library systems. The storage required by the scanned image varies and depends upon factors such as scanning resolution, page size, compression ratio and page content. Further, the image storage device may be either remote or local to the retrieval workstation depending upon the imaging systems and document management systems used.

### *Retrieving*

Once scanned images and OCR'd text documents have been saved as a file, a database is needed for selective retrieval of data contained in one or more fields within each record in the database. Typically, a document imaging system uses at least two files to store and retrieve documents. The first is traditional file that has a text description of the image along with a key to the second file. The second file contains the document location. The user selects a record from the first-file using a search algorithm. Once the user selects a record, the application keys into the location index, finds the document and displays it. Most of the document management system provides elaborate search possibilities including use of Boolean and proximity operators (AND, OR, NOT) and wild cards. Users are also allowed to refine their search strategy. Once the required images have been identified their associated document image can quickly be retrieved from the image storage device for display or printed output.

## Digitisation: Input and Output Options

A document can be converted into digital format depending on the objective of digitisation, end user, availability of finance, etc. There are four basic approaches that can be adapted to convert from print to digital:

- Scanned as Image Only
- OCR and Retaining Page Layout
- Retaining Page Layout using Acrobat Capture; and
- Re-keying the Data

### NOTES

### *Scanned as Image Only*

Image only is the lowest cost option in which each page is an exact replica of the original source document. Several digital library projects are concerned with providing digital access to materials that already exists with traditional libraries in printed media. Scanned page images are practically the only reasonable solution for institutions such as libraries and information centres for converting existing paper collection (legacy documents) without having access to the original data in computer processible formats convertible into HTML/SGML or in any other structured or unstructured text. Scanned page images are natural choice for large scale conversions for major digital library initiatives. Printed text, pictures and figures are transformed into computer-accessible forms using a digital scanner or a digital camera in a process called document imaging or scanning. The digitally scanned images are stored in a file as a bit-mapped page image, irrespective of the fact that a scanned page contains a photograph, a line drawing or text. A bit-mapped page image is a type of computer graphic, literally an electronic picture of the page which can very easily be equated to a facsimile image of the page and as such they can be read by humans, but not by the computers, understably "text" in a page image is not searchable on a computer using the present-day technology. An image-based implementation requires a large space for data storage and transmission.

Capturing page image format is comparatively easy and inexpensive, it is a faithful reproduction of its original maintaining page integrity and originality. The scanned textual images, however, are not searchable unless it is OCRed, which in itself, is highly error prone process especially when it involves scientific texts. Options and technology for converting print to digital are given separately.

Since OCR is not carried out, the document may not be searchable. Most of the scanning softwares generate TIFF format by default, which, can be converted into PDF using a number of software tools. Scan to TIFF/PDF format is recommended only when the requirement of project is to make documents portable and accessible from any computing platform. The image can be browsed through a table of contents file composed in HTML that provides link to scanned image objects.

## **OCR and Retaining Page Layout**

The latest versions of both Xerox's TextBridge and Caere's Omnipage incorporate technology that allow the option of maintaining text and graphics in their original layout as well as plain ASCII and word-processing formats. Output can also include HTML with attributes like bold, underline, and italic retained.

### ***Retaining Layout after OCR***

A scanned document is nothing more than a picture of a printed page. It cannot be edited or manipulated or managed based on their contents. In other words, scanned documents have to be referred to by their labels rather than characters in the documents. OCR (Optical Character Recognition) programs are software tools used to transform scanned textual page images into word processing file. OCR or text recognition is the process of electronically identifying text in a bit-mapped page image or set of images and generate a file containing text in ASCII code or in a specified word processing format leaving the image intact in the process.

### ***Retaining Page Layout using Acrobat Capture***

The Acrobat Capture 2.0 provides several options for retaining not only the page layout but also the fonts, and to fit text into the exact space occupied in the original, so that the scanned and OCR'd copy never over or under-shoots the page. Accordingly, it treats unrecognisable text as images that are pasted in its place. Such images are perfectly readable by anyone looking at the PDF file, but which will be absent from the editable and searchable text file. In contrast, ordinary OCR programs treat unrecognised text as tildes or some other special character in the ASCII output. Acrobat Capture can be used to scan pages as images, image +text and as normal PDF, all the three options retain page layout.

**Image Only:** Image only option has already been described in option 1.

**Image + Text:** In image + text solutions, a OCR'd text is generated for each image where each page is an exact replica of the original and left untouched, however, the OCR'd text sits behind the image and is used for searching. The OCR'd text is generally not corrected for errors since it is used only for searching. The cost involved is much less than PDF Normal. However, the entire page is a bitmap and neither fonts nor line drawings are vectorised, so the file size of Image + Text PDFs is considerably larger than the corresponding PDF Normal files and pages will not display as quickly or cleanly on screen.

**PDF Normal:** PDF normal gives the clearest on-screen display, is searchable, and yet with significantly smaller file size than Image + Text. The result is not, however, an exact replica of the scanned page. While all graphics and formatting are preserved, substitute fonts may be used where direct matches are not possible. It is a good choice when files need to be posted to the web or otherwise delivered online. If, during the Capture and OCR process, a word cannot be recognised to the specified confidence

level, Capture, by default, substitutes a small portion of the original bitmap image. Capture "best guess" of the suspect word lies behind the bitmap so that searching and indexing are still possible. However, one cannot guarantee that these bitmapped words are correctly guessed. In addition, the bitmap is somewhat obtrusive, detracting from the 'look' of the page. Further, Capture provides option to correct suspected errors left as bit-mapped image or leave them untouched.

### ***Re-keying the Data***

A classic solution of this kind would comprise keying-in the data and its verification.

This involves a complete keying of the text, followed by a full rekeying by a different operator, the two keying-in operation might take place simultaneously. The two keyed files are compared and any errors or inconsistencies are corrected. This would guarantee at least 99.9% accuracy, but to reach 99.955% accuracy level, it would normally require full proofreading of the keyed files, plus table lookups and dictionary spell checks.

### **Digital Libraries and their Uses**

The important functions and uses of the digital library in context of users are that it:

- (i) provides access to a very large information collection in a digital form
- (ii) supports multimedia content
- (iii) is network accessible
- (iv) provides user friendly interface
- (v) offers links to local/external objects
- (vi) support advance search and retrieval
- (vii) supports the traditional library mission of collection, development, organization, access and presentation
- (viii) supports publishing, annotation and integration of new information
- (ix) brings together people with formal, informal and professional learning missions
- (x) provides faster access to information resources
- (xi) provides an easy mechanism for resource sharing with other libraries. Sharing of digital files is much easier.

### **Major Issues/Challenges**

Creating effective digital libraries poses serious challenges. Some of the more serious issues facing the development of digital libraries are:

- (i) **Technical Architecture**—Libraries need to enhance and upgrade

NOTES

NOTES

current technical architecture such as:

- High speed local network and fast connection to internet
- Relational database that supports a variety of digital formats
- Full text search engines to index and provide access to resources
- A variety of servers such as web services and FTP servers
- Electronic document management system

(ii) **Building Digital Collections** — One of the most important issues in creating a digital library is building of the digital collection. One of the major issue is the degree to which libraries will digitise existing material and acquire original digital works. This is the old access versus ownership issue. How is the specific material to be digitised/to be acquired to be identified by a given library. Who collects and/or digitises which material could be based on factors such as—collection strength, unique collections, the priorities of user groups, manageable portions of collection, technological resources and skills of the staff.

(iii) **Digitisation**—Another aspect is what portion of collection to digitise. Digitisation is conversion of any fixed or analogue media—such as books, journal articles, photos, paintings, microfilm into electronic form either through scanning or rekeying. There are several approaches available, at least theoretically.

- Retrospective conversion of collections
- Digitisation of a particular special collection or a portion of one
- Highlight a diverse collection
- High use materials
- An ad hoc approach -(one digitises and stores material as they are requested)

(iv) **Metadata**—Metadata is the data that describe the content and attributes of any particular item in a digital library. Metadata is important for digital libraries because it is the key to resources, discovery and use of any document like a library catalogue. The "Dublin Core" is one of the prominent schemes. There is the problem of naming, identifiers and persistence. Naming is required to uniquely identify digital objects. Any system of naming should be permanent, lasting indefinitely. The names cannot be bound with specific locations. A global scheme of unique identifier is required. Three schemes proposed to get over the problems of persistent naming are PURLs, URNs, and digital object identifiers:

**PURLs**—are persistent URLs, a scheme developed by OCLC to separate document name from its a location.

**URN**—Uniform Resource Name have been developed by Internet Engineering Task Force (IETF).

**Digital Object Identifier (DOI)**—Developed by Association of American Publishers and Corporation for National Research Initiatives to provide a method by which digital object can be reliably identified and accessed.

(v) **Copyright/Rights Management**—Copyright is one of the most important barriers to digital library development. The current paper-based concept of copyright breaks down in the digital environment because the control of copies is lost. Digital objects are less fixed, easily copied and remotely accessible by multiple users simultaneously. The problems of libraries are that they are for the most part simply caretakers of information, they do not own the copyright of the material they hold. So libraries will never be able to freely digitise and provide access to the copyrighted material in their collection. They have to develop a mechanism for managing copyright.

(vi) **Preservation**—Another important issue is preservation. In the preservation of digital material, the real issue is technical obsolescence. There are three issues of preservation:

- Preservation of the storage medium — tapes, hard drives, floppy discs have a short life span when considered in terms of obsolescence.
- Preservation of access to content—this form of preservation involves preserving access to the content of the document regardless of the format.
- While files can be moved from one storage medium to another, what happens when the formats (*e.g.*, Acrobat PDF) containing the information becomes obsolete?

NOTES

---

## 5.4 MANAGEMENT OF DIGITAL LIBRARIES

---

Computer networks are essential to the operation of libraries. Network, allow library staff to work for cataloguing of books, providing various services, and also allow access to library catalogue and web resources within the library or elsewhere. Without the functioning of a computer network, these things are not possible. Users cannot access network resources within the library or on the Internet. Since the networks were first implemented to connect computer terminals to Online Public Access Catalogue (OPACs), their importance to libraries has continued to grow. The continuous development of information technology will make library networks even more crucial in the near future. During the next few years, as part of rapidly developing information technology, digital audio and video will have a significant impact. The increasing widespread availability of inexpensive products enable audio and video communication through Internet and create new ways to deliver information and new opportunities to enrich communication.

**NOTES**

To support library services, a network must do more than function; it must also provide a certain level of performance. Performance degrades if the network is poorly implemented, or because it lacks the capacity needed to support all its users. These technical issues have a human impact. The Library is the place where patrons come and use Internet resources. Patrons who come to the library to access the resources may stop doing so if they find accessing web pages takes a long time. Similarly, patrons who use electronic resources stored in the library will not be satisfied if retrieving that information takes a long time. The diminished quality of service from inadequate networks poorly serves library patrons.

This unit is not written for technical networking experts or the systems librarian. This only helps the librarians to understand the principal components of their network and how they affect the ability to delivery library services. It also helps librarians become technically literate in this key area of information technology to provide improved services.

**Network Applications and their Requirements**

Collectively, the performance of a network is referred to as its Quality of Service, which has two key measures; latency and jitter. Another term for latency is delay which refers to the length of time for information to cross a network or a piece of that network. Jitter is variation in latency. A network with large amount of jitter delivers information inconsistently, speeding up and slowing down unpredictably. Latency is the product of multiple factors and the most obvious is bandwidth, which is the amount of information that can pass a point in a network in a given amount of time.

Insufficient bandwidth increases latency, because it causes information to take more time to pass through the network. Since bandwidth can have so much impact on latency, the bandwidth requirement of an application is a useful way to state the performance it requires from a network.

Bandwidth is not the only factor in the network that can cause increased latency. The receipt of information, its processing, and then retransmission by each of these devices all add to the total network latency. The delay in getting information through the network is also increased by the time needed to physically transmit information over any segment of the network. More bandwidth means more information can pass through a point in the network at the same time.

Jitter, the variation in latency, occurs because the load on the network increases and decreases. When network load increases there may be contention for network resources. If a network goes from supporting a few users needing limited bandwidth to having a large number of users demanding a lot of bandwidth, latency grows.

**Library Applications and their Network Requirements**

Libraries use more than one type of network. They will have local area networks (LANs); networks connecting computers in a small area. One

or multiple LANs are possible. Where libraries have multiple LANs, or where libraries are in one building connected to other buildings of the library's parent organization, libraries may also connect to campus or backbone networks. The LAN or the backbone of the network is connected to the Internet.

## NOTES

Let us first understand the networking protocol and its role in the network before we get on to Network. A protocol is an agreed way to respond to a situation or to communicate with particular kinds of requirement. A protocol is an agreed way to manage interaction between computers. But the computer network is not so simple that a single protocol can be used. Multiple tasks must be accomplished for one computer to communicate with another. These tasks are not independent, each happening in parallel. Rather, one task depends on another. These different communication tasks appear layered. This set of layered tasks that must occur for two devices to communicate is a protocol stack.

An international standard model exists for the protocol layers involved in computer networks. This model, an ISO standard, is called OSI (Open System Interconnect) Reference Model. This model specifies seven protocol layers. It is the theoretical model of how different networking tasks are handled that is frequently used in explanations of networking.

Let's understand the protocol stack for the first three layers called the bottom three layers.

Network Layer — Layer - 3

Data Link Layer — Layer - 2

Physical Layer — Layer - 1

The bottom layer is the physical layer, Layer 1. The protocol at this layer has to set a standard for use of the physical media to send electrical signals between computers, delivering a stream of bits to another device.

The second layer is the data link layer. It must be able to receive the stream of bits from Layer 1 and interpret it as meaningful chunks of data. Layer 2 protocols organize data into frames. They must also support an addressing system, a way to specify the computer to which frames are being sent. With these two protocol layers specified, there is agreement on how to use cables to communicate and how to package data and send them to specific computers over a network.

Data link layer addresses offer a way to address communications to other computer on a network, but their addressing systems are flat. There is no way to create a route for data traffic among networks based on these addresses. This requires an addressing scheme with addresses that are hierarchical, providing the assurance that a group of addresses are used by one network. The third layer provides a way to cope with more complicated networks.

NOTES

The network layer specifies how to address, format, and route communications that must go among multiple networks.

The OSI protocol model is a theoretical model. It does not specify how these different layers are implemented. Many alternative protocols actually implement these protocol layers. The checklist identifies the protocols used in LANs and on the Internet to perform these layers of tasks. It describes how devices on the network interact in terms of the functional requirements these different layers represent.

**Library Local Area Network Checklists:**

- **Local Area Network Protocol.** The networking standard that is used on the LAN to govern communications
- **Cabling.** The kind of wiring to be used through the LAN
- **Network Interface Cards.** The hardware installed into computers to allow them to communicate over the LAN
- **Structured Cabling System.** The way in which network cabling is installed in a building
- **Network Topology and Connecting Devices.** The logical structure of the network and the hardware to which the computers are connected by the cabling and through which communications flows
- **Network Operating System and Server.** The software and hardware that provides the computer with access to shared services like printers and file servers
- **Wireless LAN.** Means for staff and patrons to work without being tied down by network cables
- **Campus or Backbone Networks.** A high speed data network running through the campus, to which the library LAN may be connected.

Actually LAN does not need all the items mentioned above. A wireless LAN is optional. Some libraries may not have a large enough infrastructure to require a backbone network.

**Policy and Planning for Network Management**

1. Each country should develop its own national policy for libraries and information services with due emphasis on the networking of libraries.
2. A national centre should be identified or established in each country to monitor and coordinate the development of library networking in the country and in the region.
3. International and local professional associations should promote the development of professional library communities through networking-such as the creation and local moderation of electronic discussion lists.
4. All types of library services, especially public and academic libraries,

need to be enhanced and modernised to keep pace with the development, and to meet the challenges, of the Information Society.

5. A National agency or any other body, should convene meetings of interested parties to consider the creation of a regional consortium as an interested community emerges. This should act as a catalyst to develop leadership champions for:
  - creating a culture of sharing;
  - developing policies and guidelines;
  - encouraging collaboration with industry;
  - developing regional consortia for licensing and procurement.

## NOTES

**Technology**

6. International standards should be adopted by libraries to facilitate the exchange of information within the country and globally and to encourage the development of inter-operable systems.
7. In the modernisation of libraries international bibliographic standards, in particular MARC, AACR2, specialised thesauri and national authority databases should be adopted.
8. Libraries can effectively function as centres to provide for public access to electronic information and therefore should be given priority in getting Internet and other communication facilities.

**Human Resources Development**

9. Organisations operating libraries and information services must be encouraged to develop education and training to enable librarians to undertake their professional development continuously throughout their careers, *e.g.*, in information management, library networking, library standards, content creation, effective communication, information entrepreneurship etc.
10. Library associations, academic institutions and other organisations should develop Continuing Professional Development [CPD] programmes to allow librarians to enhance skills and develop core professional competences for the management of electronic resources *e.g.*, leadership in information infrastructure building, data management, electronic library management, needs assessment, etc.
11. Library authorities should create a suitable environment and provide incentives to motivate library professionals to take up continuing professional education in order to become effective change agents in the information society.
12. The appropriate authorities must recognise that the teaching of librarianship and information science [LIS] requires the provision of IT workshops equipped with up-to-date hardware and software and Internet connectivity.

**NOTES**

13. Recognising the increasingly interdisciplinary nature of LIS studies, LIS Departments should work with other disciplines to ensure that the curriculum is broadly based.
14. LIS Departments and national library associations should work towards the establishment of systems of professional accreditation supported, where relevant, by an international professional organisation.
15. The curricula of LIS departments should be revised to integrate IT applications and library networking as significant components.

### **Network Infrastructure**

Telecommunications are playing an increasing role in the exchange of information, flow of ideas, and provision of essential services in education, business, healthcare and government. These technologies will have a profound impact on all citizens in the country. Access to an advanced telecommunications system is essential to ensure long-term strategic development and the quality of life of all its citizens. Computers have changed the way we think about information, the ease and speed with which it can be accessed, and the manner in which it can be assembled to create a new resource for decision makers. The libraries have to realize that the potential rests on the communication network and the readability of the bits and bytes transmitted. The applications that create and store documents and data can seamlessly share the information.

Goals of a Library and Information Networks are:

- To support company-wide connectivity
- To create standardization
- To integrate data
- To implement integrated applications

---

## **5.5 MAJOR ELEMENTS OF THE MANAGEMENT**

---

### **Methods-Technical Architecture of Management of Networks**

An essential element in constructing an integrated environment is an efficient delivery vehicle for processing integrated applications. The design of this delivery system incorporates a complex infrastructure based on the standardization of a Technical Architecture. Technical architecture includes the methods, standards, policies and management directives.

A complex integrated technical architecture requires a highly trained staff to manage resources and provide services. This staff must perform specialized functions: managing daily operations, controlling the systems development environment, maintaining software to operate the environment, and planning for future requirements. Above all else, the technical environment must be continually monitored to ensure that it meets all business and engineering requirements.

It should be enabled that hardware, systems software, communication networks, programme development standards, and data base resources for Network operations are in place and available.

It is on an effective technical architecture that a business's ability to deliver sufficient resources to the systems users depends. The implementation of the architecture will otherwise fail. Providing appropriate types of service delivery and maintaining appropriate performance standards are essential.

The technical architecture incorporates specific services to support the system users. These services range from defining relational database management systems to installing a backbone communications network. Providing these services represents a major cost; however, the return is an effective mechanism for supporting integrated applications and integrated project requirements.

The standardization of a basic technical architecture provides strategic and economic benefits. However, because systems integration requires access to more of the hardware, software, and information resources, it exposes the organization to greater risk by allowing access to sensitive data. Although unauthorized access to information is a risk in any systems environment, the risk is even greater in a decentralized environment because of the increased access to systems and the complexity of controlling information.

The integrated systems environment is characterized by a disciplined approach where the correct information is readily available to quickly and efficiently do the job right the first time. This allows the designers and technicians to fully explore previous operating experience, customer requirements, and total cost implications of alternatives. To achieve this environment, the design function must be carefully evaluated and the identified requirements implemented in an integrated systems environment.

Historically, automation has focused on isolated "islands of automation" due to the paper-based method of information exchange. The systems recommended here will move to an online information exchange among the design groups. As this evolves, further opportunities will occur by linking the databases with new expert systems. Parametric design systems must also be evaluated. These systems have the potential to make conventional software applications, where a "rule based" repetitive design is required, obsolete. Their implications must be evaluated before a final commitment is made to conventional software applications.

- Application Planning/Administration
- Technical Infrastructure
- Service Delivery
- Security and Administration

## NOTES

NOTES

## **Application Planning/Administration**

Application Planning/Administration includes the methods, standards, policies and management directives for maintaining an integrated delivery system. The following recommendations identify the requirements for supporting an integrated environment:

Select a systems development methodology which will fit into the organization's culture and provide the necessary structure for designing and implementing all the required systems in the future. Designate responsibilities for the technical architecture development projects.

## **Technical Infrastructure**

The Technical Infrastructure is the planned configuration of computing hardware and operating systems software, communication networks and databases required to support integrated applications, development functions and administration.

- The first element which must be addressed is the standardization of platforms (*i.e.*, the computing hardware and operating system software).
- The second element is a standard networking technology and communications software to support communications across varying hardware platforms, projects and locations. In order to achieve the integration between the various databases and applications, it is necessary to provide all sites with the capability to communicate either online or through file transfers.
- The third element is a database management system (DBMS).

## **Service Delivery**

A technical environment is configured on the principle that the implementation of new technology is a decision based on cost/benefits and strategic positioning. The following recommendations focus on four key service areas:

Initiate a service management process. The new architecture must provide specific services to support all users. Formalize the process of capacity planning and performance management. This process measures the performance and resource consumption of systems.

Implement a support organization for the resolution of problems. The purpose of problem management is to maintain and protect the integrity of the computing environment.

## **Security and Administration**

Information security and privacy are inextricably intertwined. The former refers to the physical protection of information assets, which includes everything from identifying the asset to be secured to developing relevant policies and procedures, designing a secure environment, implementing the solutions,

training the staff, and testing and monitoring the system. Physical protection is concerned with more than privacy; it also serves to guarantee the integrity and availability of the data.

"Privacy" refers to the goal that secured information can be accessed and seen by authorized and certified personnel only.

## NOTES

### Security Categories

1. Administrative Procedures To Guard Data Integrity, Confidentiality, and Availability.
2. Physical Safeguards to Guard Data Integrity, Confidentiality, and Availability.
3. Technical Security Services to Guard Data Integrity, Confidentiality, and Availability.
4. Technical Security Mechanisms to Guard Data.

The technical environment must be properly managed. The following recommendations focus on four management issues related to data security:

- Implement a security plan to assist management in protecting Library Assets.
- Centralize the management and operations of the backbone network at the corporate level. This requires a centralized function to define, monitor and control network installations, and performance monitoring.
- Initiate a structured approach to evaluate and implement changes into the installed processing environment. This function attempts to minimize the adverse impact on system integrity by planned changes.

This process identifies critical functions and precautions, which must be taken to ensure continual operation in the event of disasters.

### Network Architecture

Computer networks are concerned with linking together various pieces of computer equipment and systems and telecommunication networks. One important aim in a network is to provide end-to-end compatibility so that any terminal or system can be linked to any other. It is useful to examine network architecture. A network architecture consists of:

- The protocols and standards for data encoding and transmission (Protocol, hardware and cabling)
- The topology of the way in which units are distributed through the network (Topology)
- The control and flow of information through the network ( Network operating system)

## Network Protocols

### NOTES

A protocol is a set of rules that governs the communications between computers on a network. These rules include guidelines that regulate the following characteristics of a network: access method, allowed physical topologies, types of cabling, and speed of data transfer. The most common protocols are:

**1. Ethernet.** The Ethernet protocol is by far the most widely used. Ethernet uses an access method called CSMA/CD (Carrier Sense Multiple Access/Collision Detection). This is a system where each computer listens to the cable before sending anything through the network. If the network is clear, the computer will transmit. If some other node is already transmitting on the cable, the computer will wait and try again when the line is clear. Sometimes, two computers attempt to transmit at the same instant. When this happens a collision occurs. Each computer then backs off and waits a random amount of time before attempting to retransmit. With this access method, collisions are minimized. However, the delay caused by collisions and re-transmitting is very small and does not normally effect the speed of transmission on the network. The Ethernet protocol allows for linear bus, star or tree topologies. Data can be transmitted over twisted pair, coaxial, or fiber optic cable at a speed of 10 Mbps.

**2. Local Talk.** Local Talk is a network protocol that was developed by Apple Computer, Inc. for Macintosh computers. The method used by Local Talk is called CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance). It is similar to CSMA/CD except that a computer signals its intent to transmit before it actually does so. Local Talk adapter and special twisted pair cable can be used to connect a series of computers through the serial port. The Macintosh operating system allows the establishment of a peer-to-peer network without the need for additional software. With the addition of the server version of AppleShare software, a client/sever network can be established.

The Local Talk protocol allows for linear bus, star, or tree topologies using twisted pair cable. A primary disadvantage of LocalTalk is speed. Its speed of transmission is only 230kbps.

**3. Token Ring.** The Token Ring protocol was developed by IBM in the mid-1980s. The access method used involves token-passing. In Token Ring, the computers are connected so that the signal travels around the network from one computer to another in a logical ring. A single electronics token moves around the ring from one computer to next. If a computer does not have information to transmit, it simply passes the token on to the next workstation. If a computer wishes to transmit and receive , it attaches data to the token. The Token then proceeds around the ring until it comes to the computer for which the data is meant. At this point, the data is captured by the receiving computer.

The TokenRing protocol requires a wired ring using twisted pair of fiber

optic cable. It can operate at transmission speeds of 4Mbps or 16Mbps. Due to the increasing popularity of Ethernet; the use of TokenRing in library/information center environments has decreased.

**4. FDDI.** Fiber Distributed Data Interface (FDDI) is a network protocol that is used primarily to interconnect two or more local area networks, often over large distances. The access method used by FDDI involves token passing.

FDDI uses a dual ring physical topology. Transmission normally occurs on one of the rings; however, if a break occurs, the system keeps information moving by automatically using portions of the second ring to create a new complete ring. A major advantage of FDDI is speed. It operates over fiber optic cable at 100Mbps.

**5. ATM.** Asynchronous Transfer Mode (ATM) is a network protocol that transmits data at a speed of 155Mbps and higher. ATM works by transmitting all data in small packets of a fixed size; whereas, other protocols transfer variable length packets. ATM supports a variety of media such as video, CD-quality audio, and imaging. ATM employs a star topology, which can work with fiber optics as well as twisted pair cable.

ATM is most often used to interconnect two or more local area networks. It is also frequently used by Internet Service Providers to utilize high-speed access to the Internet for their clients. As ATM technology becomes more cost-effective, it will provide another solution for constructing faster local area networks.

### ***Network Topology***

In network computers are connected to each other. The way in which the connections are made is called the topology of the network. It is important to select the right topology for how the network will be used. Bus, star, ring, and mesh are the four most common topologies.

The term topology or more specifically network topology, refers to the arrangement or physical layout of the computers, cables and other components on the network.

A network topology affects its capabilities. A network topology implies a number of conditions. For examples, a particular topology can determine not only the type of cable used but how the cable is run through floors, ceiling and walls. Topology can also determine how computer communicate on the network. Different topologies require different communication methods and have a great influence on the network.

The type of network installed will depend on a number of factors, including the:

- Network budget
- Network size

NOTES

- Level of security required
- Physical layout
- Type of business
- Amount of network traffic.

## NOTES

### **Network Management**

The terminology "network management" may have many meanings but the definitions given below may be considered for clarifications. Network managements means deploying and coordinating resources in order to plan, operate, administer, analyse, evaluate, design and expand communications networks to meet service-level objectives at all times, at a reasonable cost, and with optimum capacity.

Network management involves planning, organizing, monitoring, accounting, and controlling of network activities and resources. Network management is the process of using hardware and software by trained personnel to monitor the status of network components and line facilities, interacting with end-users and carrier personnel, and implementing or recommending actions to alleviate outages and/or improve communication performance as to well as conduct administrative tasks associated with the operation of the network.

### ***Need for Network Management***

The relentless growth in the information processing needs of organizations has been accompanied by rapid development in computer and data networking technology to support those needs, and explosion in the variety of equipment and networks offered by vendors. Gone are the days when an organization would rely on a single vendor and a relatively straightforward architecture to support its needs. The world is no longer divided into the pure mainframe-based, IBM-computable centralized environment and the PC-based, single-LAN-type, distributed environment. Today's typical organization has a large and growing, but amorphous architecture, with a variety of local area networks (LANs) and wide area networks including PCs, workstations, and servers.

### ***Objectives of Networking***

The very objective of networking is to promote maximum utilization of information through sharing of resources by a group of libraries. It coordinates efforts towards, suitable collection development and reduces unnecessary duplication of costly books, periodicals, etc. The network establishes referral centres to monitor catalogue search and maintain a central online union catalogue of books, serials and non-book materials of all the participating libraries. It is implements computerization of operations and services in the libraries for fast communication of information. It also coordinates with other regional, national and international networks for exchange of information and documents. Another function is to evolve standards for

uniform guidelines in techniques, methods, procedures, hardware, software, services, etc., in order to facilitate sharing and exchange of resources.

### **Benefits of Networking**

Networks can do anything a multi-user system can, and are better at processor and screen intensive tasks, such as spreadsheets and graphics applications, which are run locally (at the workstation). The following may be the advantages of installing a network.

- (i) **Distributed Processing:** Programmes are downloaded from a central point (i.e., the file server) and run locally. The network system does no processing, but merely provides storage space for data and programmes.
- (ii) **Security:** A workstation doesn't need disc drives, so one can stop people from stealing one's data and/or software and one can keep viruses out.
- (iii) **Backing Up:** Where data is centralized, backing up procedures are more convenient and can be more closely controlled.
- (iv) **Shared Resources:** Equipment that would normally be kept idle for long periods can be utilised more effectively when several people share the use of it.
- (v) **Communication:** By networking, sending replies will become easy. Files can be attached to messages, saving papers. Appointments can be arranged between groups of people, so one can use the system like an alarm clock.

An important function of successful networking is to ensure that the right information is accessed by the right person at the right time, meaning thereby that one should get a network that suits one's circumstances rather than change the way you work **Future trends of Network Management.**

Networks are becoming increasingly critical to the majority of corporations for various reasons, such as providing faster and more reliable end-user application services; inter-linking with other corporations and service providers; distributing computing intelligence and database; providing total connectivity across communications forms, network architectures, and network components; supporting brands new applications which require the multiple of presently available bandwidth; and as a combination of all the preceding corporations saying they offer the ultimate advantage over their competitor. For supporting these ambitious goals, new network components and services provided by a considerably high numbers of suppliers have to be carefully considered. But new networking solutions, installation of new components, and implementation of new services will fail if there is no adequate network management solutions.

NOTES

## Networking in India

Forced by the circumstances and by design, there has been a spurt in activities on library network development in the recent past. Some of the factors responsible for the promotion and support of such efforts in India are:

### NOTES

- Increased awareness of the need for resource sharing
- Resource crunch
- Increased installation of computers and access facilities in a library environment and enhancement of computer literacy among the library professionals
- Improvement in computer communication facilities within and across geographical regions, and availability of general data networks like National Informatics Network (NICNET) of MIT, INDONET(CMC Ltd), Education and Research Network -ERNET (MIT) and I-NET(Department of Telecommunications) and VSNL for the Internet.

Considerable study has been conducted as to why and how networking has preceded the efforts to implement and establish library networks in the country. Some of them are:

- (i) The report of the networking group of the Planning Commission on modernization of library services and informatics for the 7th Five Year plan, 1985-90 suggests measures to establish networks of important libraries in the country to enable sharing of resources for the benefit of their user communities.
- (ii) The National Policy on the Library and information System document (1986) accepted by the Ministry of Human Resources Development, Govt. of India, has many recommendations emphasizing the need for the national libraries in the country to form part of one integrated system so that they make a functional whole.
- (iii) The report of the National Policy on University Libraries, Prepared by the Association of Indian Universities (1987), and the seminar on National Policy on University Libraries held on 30th Sept., 1986. recommends a network of university research and other libraries in the country so that the barriers of time, space and language are eliminated.
- (iv) The UGC report on Information and Library Network (INFLIBNET) recommended the establishment of a National Network to link Libraries in universities, colleges and documentation and information processes. Beside these documents many seminars and conferences have also recommended the creation of library and information networks at various levels and coordination of their activities through modernization of the library and information process.
- (v) Various metropolitan networks have been established with the support of NISSAT of DST and the Planning commission like DELNET, CALIBNET, ADINET, BONET, etc.

## Network Selection Criteria and the Challenges in Network Management

Having decided to participate in a network, an intending library/information centre is faced with the questions of selecting the most suitable network. The following factors help in making the choice:

NOTES

- Available network options.
- Criteria to be used in deciding the network.
- Method of evaluating competing networks.
- Selection of functions of the network to be used.
- Reliability of system.
- Quality control standards decided by the network.
- Cost

A network may fail if the basic objectives are not properly formulated. Reasons for network failure can be listed as:

- Emphasis on only a single function (e.g., inter-library loan in library network) which is difficult to justify on the basis of the cost-effect criterion.
- Lack of attention to the use of appropriate technology.
- Lack of attention to formal system design, derived from stated goals and objectives.
- No specification in developing functions.
- Not translating network bylaws into specific operating procedures.
- Unrealistic budgeting and lack of expertise in developing long term funding.

### Challenges

1. Availability—probably the most important concern in any network management services is the easy availability or the uptime of the network resources.
2. Easy manageability—Status indication of network and centralised management of all the resources.
3. Proactive bandwidth planning.
4. Cost effectiveness of the operations
5. Stability of the operations and network resources.

### How much down time can you live with?

A customized support plan is possibly a chief information officer's only true defence against unexpected outage.

## 5.6 SUMMARY

### NOTES

- Digitisation is the process of converting the content of physical media (e.g., periodical articles, books, manuscripts, cards, photographs, vinyl disks, etc.) into digital format. In most library applications, digitisation normally results in a documents that are accessible from the website of a library, and thus on the Internet. Optical scanners and digital cameras are used to digitise images by translating them into bit maps. It is also possible to digitise sound, video, graphics, animations, etc.
- Digitisation is the first step in the process of building digital libraries. Digitisation is also used for achieving preservation and archiving although it is not considered as good option for preservation and archiving. It is highly labour-intensive and cost-intensive process that involves several complexities including copyright and IPR issues. However, digital objects offer numerous benefits in terms of accessibility and search. The documents to be digitised may include text, line art, photographs, colour images, etc. The selection of document, need to be reviewed very carefully considering all the factors of utility, quality, security and cost. Rare and much in demand documents and images are selected as first priority without considering the quality.
- The process of digitisation involves four steps namely, scanning, indexing, storage and retrieval. A scanned document is nothing more than a picture of a printed page. It cannot be edited or manipulated or managed based on their contents. In other words, scanned documents have to be referred to by their labels rather than characters in the documents. OCR (Optical Character Recognition) programs are software tools used to transform scanned textual page images into word processing file. OCR or text recognition is the process of electronically identifying text in a bit-mapped page image or set of images and generates a file containing that text in ASCII code or in a specified word processing format leaving the image intact in the process.
- There is much confusion with regard to the definition of the "digital library". There is such confusion surrounding the terms partly because library professionals have been using different phrases over the years to denote this concept—electronic library, virtual library, library without walls, etc. Another factor adding to the confusion is that digital libraries are at the focal point of many different areas of research. The third confusion arises from the fact that there are many things on the Internet that people are calling "digital libraries" but which differ from the library professionals point of view. Man people consider World Wide Web as a digital library.
- Creating effective digital libraries poses a serious challenge. The integration of digital media into traditional collection will not be straight forward because of the unique nature of digital information—

it is less fixed, easily copied and remotely accessible by multiple users simultaneously.

- As a rule, networking in library and information services is still at the primary stage and is not as well developed as that in other sectors such as banking, civil aviation, customs, public telecommunication and other industrial departments. The major reasons for this, apart from the social and political ones mentioned above, may be that: (1) before networking can be implemented, individual libraries must move to the computerization stage and a range of viable databases must be available; (2) financial constraints make it difficult for libraries to establish their own dedicated networks, and to afford the communication fees for other networks.
- In spite of all these difficulties, Indian librarians still have made efforts and have achievements to their credit.

NOTES

---

## 5.7 REVIEW QUESTIONS

---

1. What are the major issues/challenges in the development of the digital library?
2. Define 'Digitisation'.
3. What are the factors to be considered before selecting different media for digitisation?
4. What are the four basic approaches that can be adapted to translate from print to digital?
5. What are the objectives of the resource sharing network? Explain the components of the resource sharing network.
6. Define Network Infrastructure Management challenges and issues.

---

## 5.8 FURTHER READINGS

---

1. Bay ross, Ivan N. (2003). Database Concepts and Systems. Navi Mumbai: Shroff.
2. Chen, Peter P. (1976). The Entity-Relationship Model-Tward a United view of Data, ACMTODS 1, No.1, Republished in Michael Stonebraker (ed.) Readings in Database Systems. Sun Mateo, California: Morgan Kaufmann (1988).
3. Desai, Bipin C. (1994). An Introduction to Database Systems, New Delhi: Galgotia Publication.
4. Frank, Lars (1988). Database Theory and Practice. England. New York: Addison-Wesley publishing.
5. Hernandez, M.J. (2003). Keys in Relational Databases. Boston: Addison-Wesley.
6. [http://en.wikipedia.org/wiki/Database\\_management\\_system](http://en.wikipedia.org/wiki/Database_management_system).
7. <http://support.microsoft.com/default.aspx?scid=kb;EN-US;100139#apliesto>.