COMPUTER FOUNDATION

Foundation Course

BAFNDC

[ENGLISH EDITION]

Directorate of Distance Education

TRIPURA UNIVERSITY

Reviewer

Richa Bhargava

Associate Professor, IT and Operations, IILM Institute for Business & Management, Gurgaon

Authors

Daminni Grover, (Units: 1.2-1.4, 3) © Reserved, 2017

Sanjay Saxena, (Units: 1.5, 1.7, 4, 5.2) © Sanjay Saxena, 2017

Rajneesh Agrawal, (Unit: 1.6) © Rajneesh Agrawal, 2017

B. Basavaraj, (Unit: 2.2-2.5), © B. Basavaraj, 2017

Paulomi M Jindal, (Unit: 5.3) © Reserved, 2017

Vikas Publishing House, (Units: 1.0-1.1, 1.8-1.12, 2.0-2.1, 2.6-2.13, 5.0-5.1, 5.4, 5.5-5.10) © Reserved, 2017

Books are developed, printed and published on behalf of Directorate of Distance Education, Tripura University by Vikas Publishing House Pvt. Ltd.

VIKAS® PUBLISHING HOUSE PVT. LTD.
E-28, Sector-8, Noida - 201301 (UP)
Phone: 0120-4078900 • Fax: 0120-4078999
Regd. Office: 7361, Ravindra Mansion, Ram Nagar, New Delhi 110 055
• Website: www.vikaspublishing.com • Email: helpline@vikaspublishing.com

SYLLABI-BOOK MAPPING TABLE

Computer Foundation

Syllabi	Mapping in Book
Unit - I **Introduction:** Introduction to Computer System, Characteristics of Computers, Uses of Computers, Types and Generations of Computers, Basic Applications of Computer.	**Unit 1:** Introduction to Computer System **(Pages 3-35)**
Unit - II **Data Represenation:** Number Systems and Character Representation, Decimal, Binary, Octal and Hexadecimal System, Conversion from One Number System to Another Number System, Binary Arithmetic (Addtion, Subtraction Using 1's Complement & 2's Complement, Multiplication). Representation of Data/Information Concepts of Data Processing. Definition of Information and Data, Basic Data Types, Storage of Data/Information as Files.	**Unit 2:** Data Representation and Number System **(Pages 37-87)**
Unit - III **Human Computer Interface:** Types of Software, Operating System as User Interface, Utility Programs.	**Unit 3:** Human Computer Interface **(Pages 89-98)**
Unit - IV **Devices:** Input and Output Devices (with Connections and Practical Demo), Keyboard, Mouse, Joystick, Scanner, OCR, OMR, Bar Code Reader, Web Camera, Monitor, Printer, Plotter. **Memory:** Primary, Secondary, Auxiliary Memory, RAM, ROM, Cache Memory, Hard Disks, Optical Disks.	**Unit 4:** I/O Devices and Memory **(Pages 99-134)**
Unit - V **Use of Computers in Education and Research:** Data Analysis, Heterogenous Storage, e-Library, Google Scholar, Domain Specific Packages Such as Office, SPSS, Mathematica etc.	**Unit 5:** Use of Computers in Education and Research **(Pages 135-182)**

CONTENTS

INTRODUCTION

NOTES

Computers have brought about major changes in all spheres of life. Today it is extremely difficult to imagine the world without computers. Computers help us to communicate using modems, telephone and Wi-Fi facilities and it seems as if you are sitting side by side and communicating directly with each other. This modern way of communication has been made possible by computers. Through the Internet and e-mail, we now have the facility to communicate with anybody in any part of the world in a fraction of minutes. Conveniences like ATM bank services, the Internet, video conferencing, wireless telephony and electronic mail could not have been possible without data communication and computer networks. Communication facilities available with an organization or with an individual measure the level of standard for them.

The fact that computers have made a big impact on many aspects of our lives can hardly be questioned. They have opened up an entire world of knowledge and information that is readily accessible. Today, we are using the fifth generation of computers. The term 'generation' is used to distinguish between varying hardware and software technologies. The hardware by itself cannot do any calculation or manipulation of data without being instructed what to do and how to do it. Thus, there is a need of software in a computer system. The software used in a computer system is grouped into application software, system software and utility software.

This book, *Computer Foundation,* attempts to provide a sound theoretical as well as practical basis for understanding the subject. We have tried to present a clear conceptual understanding of computers and have used an easy-to-follow visual style for the practical usage of various packages to facilitate understanding. The material in this book has been provided in the self-instruction or SIM format. Each unit begins with an 'Introduction' to the topic that gives a brief outline of the concept to be dealt with. It is followed by the 'Unit Objectives' and then the details of the topic of each unit. 'Summary' and 'Key Terms' are given after every unit to help students recapitulate the concepts. The 'Check Your Progress' and 'Questions and Exercises' sections in each unit help in better understanding the subject through rigorous exercises given. The 'Further Reading' section creates a research interest in students for further exploration of the topics covered.

UNIT 1 INTRODUCTION TO COMPUTER SYSTEM

NOTES

Structure

1.0 INTRODUCTION

In this unit, you will be introduced to computers. Computers are electronic devices that perform the basic operations of input, process, output and storage under the direction and control of a program. Computers have become an integral part of our lives. Most of the work that is done these days is performed by computers in some way or other.

Computers are used to educate students, obtain any information needed, manage finances and accounts, and for social networking with friends and family members. It plays a vital role in our personal and professional lives. This unit will help you understand the definition, characteristics, generations and classifications of computers and it will help you become computer literate.

1.1 UNIT OBJECTIVES

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

- Define a computer and know its characteristics
- Explain the different generations of computer systems
- Discuss the types of computers
- List some uses of computers
- Identify certain application areas of computers

NOTES

1.2 DEFINITION OF COMPUTER

Defined in the simplest way, 'A computer is an electronic device that operates under the control of a set of instructions that is stored in its memory unit' (see Figure 1.1). It accepts data from the user through an input device and processes the data into useful information. The processed data is displayed on its monitor. In fact, a computer is a collection of hardware and software parts that help you complete various tasks. Hardware is tangible in nature and consists of the computer itself and the peripherals connected to it. Software programs are intangible in nature. These are the set of instructions that the computer follows in performing a task.

Fig. 1.1 *A Computer System*

A complete computer system includes four distinct parts, namely hardware, software, data and a user.

1.3 CHARACTERISTICS OF COMPUTERS

Computers are becoming popular day by day because of their characteristics that make them very useful. The characteristics that make computers indispensable are:

(i) Speed: A computer is able to process data given by a user and gives the output in fractions of seconds. It provides information to the user on time, thus enabling the user to take right decisions on the right time.

(ii) Accuracy: In spite of its high-speed processing, errors seldom occur as a computer's accuracy is always high, which avoids any errors. Since the computer is capable of doing only what it is instructed to do, faulty instructions for data processing may lead to inaccurate and inconsistent data or defective programs as well as defective program design. This is termed as Garbage In Garbage Out (GIGO).

(iii) Large storage capacity: Computers can store high volumes of data in small and compact storage devices.

(iv) Constant efficiency: Computers do not feel tired and distracted. They perform multiple tasks simultaneously with the same ease.

(v) Versatility: It is a very versatile kind of machine. It can work upon various data types, such as numbers, characters, graphics, audio, video, etc.

1.4 COMPUTER GENERATIONS

NOTES

In the development of computers, different steps and different technologies were used. These steps of technological differences are called 'Generation of Computers'. One generation of computers is marked by various technological developments and used in the advancements of the computer technology. Over the years, the circuit has become small and powerful than the earlier ones. This resulted in the growth of miniaturization, high speed and powerful computers. With the advent of new discoveries it profoundly affects our way of living. As of today, the evolution of computer systems has been divided into five generations.

Computers of the First Generation (1950s): Vacuum Tubes

The first generation computers like the ENIAC were based on vacuum tubes, which use to overheat and blow up. Hence, they were very unreliable. Tube replacement was a continuous process. The major focus of this era was to keep the hardware operating. Little attention was paid to any of the other four components of IS, i.e., people, procedures, software and data. The first generation computers were mainly designed for scientific computations but a small number of units were also being used for accounting purpose.

The first generation computers used vacuum tubes which were fragile glass devices for circuitry and magnetic drums for memory. They were often enormous in size and took up the size of entire rooms. Their basic technology comprised circuitry consisting of wires and thermionic valves which look somewhat similar to the pre-transistorized type of radio system. They were very slow and operated in millisecond speed. Extensive air-conditioning was required because of the heat generated by the vacuum tubes. They use modest magnetic tape and could only support assembly (low-level) languages or FORTRAN. Some examples of this generation computers were IBM 701 and 650 systems, ENIAC, EDVAC, EDSAC, UNIVAC I, IBM 701, etc.

These first generation computers had many distinctive features, such as they used the vacuum tube technology. During the 1950s, these computers were the fastest computers, as well as very large and bulky in size. They needed spacious rooms to be installed. The rooms had to be air conditioned, as these computers used hundreds of vacuum tubes and, hence, generated much heat. The components and equipment were non-portable and worked very slowly, lacking in adaptability and speed. In addition, these first generation computers were very expensive and needed a large amount of electricity. These computers were not consistent and are prone to repeated hardware failures, thus leading to their continuous maintenance. At the same time, each computer parts had to be assembled individually.

Second Generation Computers (1960s): Transistors

By the mid to late 1950s, vacuum tubes were replaced by transistors. Machine reliability increased dramatically. These were the second generation computers. The primary

NOTES

focus thus shifted from hardware to programming. People thought of ways to make programming more efficient and reliable. Hence, a new type of programs called assemblers and compilers developed. The purpose of the software was to translate symbolic code that people can understand to machine code, i.e., 1s and 0s.

They were smaller in size, generated less amount of heat and had higher capacity of internal storage and their processors operated in microsecond speed range. The second generation computers made substantial use of magnetic tape and had memory capacity of 32 kB and 64 kB. They supported high-level languages, such as ALGOL, COBOL and FORTRAN II. They had multi-programming capabilities and could communicate data over telephone lines. Some examples of this generation computers were Burrough 5000, IBM 1401, 7080 and 7090, Philco S-2000, Honeywell 400, etc.

The second generation computers had various distinctive features, such as these machines were based on transistor technology, were smaller than the first generation computers. The computation time of these machines increased and these computers were more reliable and less prone to hardware crash, thus requiring less maintenance. These machines were potable and gave out less heat, but still air-conditioning was required. In the second generation computers, the Assembly language was used to program computers, due to which programming became more time efficient and less cumbersome. It requires manual assembly of individual components into a functioning unit.

Third Generation Computers (1970s): Integrated Circuits

In third generation computers, transistors and other electronic components were combined on a single silicon chip called integrated circuits (IC's). These ICs are popularly known as CMIPs. With this technology, computers became smaller, faster and even more reliable. Powerful computer languages and computer programs were developed. From the mid-1960s to 1970s the focus turned towards the data component of IS. This era saw the birth of database management system (DBMS) on mainframe computers. Higher level languages, Beginners All Purpose Symbolic Instruction Code, such as (BASIC) were also developed during this period. Some of the examples of this generation computers were IBM 360, IBM 370, ICL-1900 and VAX-750.

The third generation computers, based on integrated circuit (IC) technology, further reduced the computation time from microseconds to nanoseconds. Their distinct features included easy portability; as these computers were relatively smaller in size as compared to previous computers. The other features were—increased reliability, less heat generation (requiring air conditioning only in some cases), thus, less electricity was consumed. The hardware failure was rare which led to low maintenance cost. These third generation computers used high-level languages.

Fourth Generation Computers (1980s): Microprocessors

In this generation electronic components were further miniaturized and condensed into very large scale integrated circuits (VLSI). One result of VLSI was that it became possible for an entire computer to be put on a single chip. This led to the creation of the personal computers (PCs) in the late 1970s. Thus, the computers

NOTES

which occupied very large rooms earlier could now be placed on a table. Consequently, computer technology was now available to the common people for communication and business needs. This generation represents an upgradation of the third generation brought about by the introduction of advanced micro technology. This led to the possibility of extremely large internal and external storage capacities as well as extensive computer networks and widely distributed databases. Some examples of this generation computers were Apple II, IBM PCs, TRS 80, VAX 9000, etc.

The main features of these generations were that these machines were based on microprocessors; these machines were smaller in size than the earlier generation computers. These computers were quite economical as compared to the earlier generations. They were portable and reliable. They generations very low heat, hence they hardly require air-conditioning. The hardware failure was reduced ever further, thus bringing down the cost of maintenance and production. The graphical user interface (GUI) enabled users to understand the use of these computers easily. And finally, due to networking of these computers, better communication and resource sharing was possible.

Fifth Generation Computers (1990s): Artificial Intelligence (AI)

Computers that can 'think' and are capable of taking decisions like human beings have been characterized as fifth generation computers. They are also termed as *Thinking Machines*. The speed of this generation computers are very high. They use the concept of 'Artificial Intelligence' and possess voice recognition capabilities. Artificial intelligence (AI), a branch of computer science, is concerned with making computers perform functions associated with human intelligence, such as reasoning, learning, self-improvement, etc. In 1956 John McCarthy coined the term.

AI is used to play games, such as chess and checkers, etc. It is also used as expert systems, i.e., programming computers to make decisions in real-life situations, for example, many expert systems help doctors to diagnose diseases based on symptoms. AI is also employed to program computers to comprehend natural languages. It is even used in neural networks, i.e., systems that simulate intelligence by attempting to reproduce the types of physical connections that occur in animal brains. It is also used in robotics, to program computers to see and hear and react to other sensory stimuli. At present, there is not a single computer that has full artificial intelligence.

These computers use parallel processing of semiconductors for advanced computing. These computers have the capabilities of learning and self organization. Some examples of this generation computers are IBM notebooks, Pentium PCs, SUN Workstations, IBM SP/2, etc.

The three main features of the fifth generation of computers are (i) based on mega chips, (ii) parallel processing and (iii) AI.

(i) These computers use super large scale integrated (SLSI) chips. There are millions of electronic components on a single chip. These machine need to store a great amount of storage capacity to store instructions

NOTES

and information. Thus, these chips enable the computer to approximate the memory capacity of the human mind.

(ii) Computers using parallel processing can access many instructions at a time and process them simultaneously through use of multiple CPUs.

(iii) The idea of AI points to a series of technologies that try to simulate and reproduce human intelligence, such as learning, reasoning, self-improvement, etc. AI consists of a group of related technologies, such as natural language processing (NLP), expert systems (ES), speech recognition, vision recognition and robotics, etc.

CHECK YOUR PROGRESS

1. Define computer.
2. Give some examples of first generation computers.
3. Define the term Artificial Intelligence.

1.5 CLASSIFICATION OF COMPUTERS

Computers can be classified on the basis of their size, processing speed and cost. The various types of computers are as follows:

- Personal computers
- Workstations
- Notebook/Laptop computers
- Tablet PC
- Personal Digital Assistant or PDA
- Mainframe computers
- Supercomputers

Analog

Analog computers are generally used in industrial process controls and to measure physical quantities, such as pressure, temperature, etc. An analog computer does not operate on binary digits to compute. It works on continuous electrical signal inputs and the output is displayed continuously. Its memory capacity is less and can perform only certain type of calculations. However, its operating speed is faster than the digital computer as it works in a totally different mode.

Analog computers perform computations using electrical resistance, voltage, etc. The use of electrical properties signifies that the calculations can be performed in real time or even faster at a significant fraction of the speed of light. Typically, an analog computer can integrate a voltage waveform using a capacitor which ultimately accumulates the charge. The basic mathematical operations performed in an electric analog computer are summation, inversion, exponentiation, logarithm, integration

NOTES

with respect to time, differentiation with respect to time, multiplication and division. Hence in the analog computers, an analog signal is produced which is composed of Direct Current (DC) and Alternating Current (AC) magnitudes, frequencies and phases. The starting operations in an analog computer are done in parallel. Data is represented as a voltage that is a compact form of storage.

Digital

Digital computers are commonly used for data processing and problem solving using specific programs. A digital computer stores data in the form of digits (numbers) and processes. It is in the discrete form from one state to the next. These processing states involve binary digits which acquire the form of the existence or nonexistence of magnetic markers in a standard storage devices, ON/OFF switches or relays. In a digital computer, letters, words, symbols and complete texts are digitally represented, i.e., using only two digits 0 and 1. It processes data in discrete form and has a large memory to store huge quantity of data.

The functional components of a typical digital computer system are input/output devices, main memory, control unit and arithmetic logic unit. The processing of data in a digital computer is done with the help of logical circuits, which are also termed as digital circuits. All the circuits processing data in side a computer function in an extremely synchronized mode, which is further controlled using a steady oscillator acting as the computer's 'clock'. The clock rate of a typical digital computer ranges from several million cycles per second to several hundred million cycles whereas the clock rate of fastest digital computers are about a billion cycles per second. Hence, the digital computers operate on very high speed and are able to perform trillions of logical or arithmetic operations per second to provide quick solution to problems, which is not possible for a human being to do manually.

Hybrid

Hybrid computers are the combination of digital and analog computers. A hybrid computer uses the best features of digital and analog computers. It helps the user to process both continuous and discrete data. Hybrid computers are generally used for weather forecasting and industrial process control.

The digital component basically functions as a controller to provide logical operations, whereas the analog component functions as a solver to provide solutions of differential equations. Remember that the hybrid computers are different from hybrid systems. The hybrid system is a digital computer equipped with an analog-to-digital converter for input and a digital-to-analog converter for output. The term 'hybrid computer' signifies a mixture of different digital technologies to process specific applications with the help of various specific processor technologies.

General Purpose

Workstations are high end, general purpose computers designed to meet the computing needs of engineers, architects and other professionals who need computers with greater processing power, larger storage and better graphic display facilities. These are commonly used for Computer Aided Design (CAD) and for multimedia applications such as creating special audio/visual effects for television programmes

NOTES

and movies. A workstation looks like a PC and can be used by only one person at a time. The characteristics of a workstation, which are often used to differentiate it from a PC are as follows:

- **Display Facility:** Most workstations have a large screen monitor (21 inches or more) capable of displaying high resolution graphics as compared to PCs, which have a small screen monitor (19 inches or less).
- **Storage Capacity:** Workstations have a larger main memory than PCs, which have only a few hundred MB of main memory. The hard disk capacity of workstations is also more than that of PCs.
- **Processing Power:** The processing power of workstations is several times greater than that of PCs.
- **Operating System:** PCs can run any of the five major operating systems namely MS DOS, Microsoft Windows, Windows NT, Linux and UNIX, but all workstations generally run the Unix operating system or a variation of it, such as AIX (used in IBM workstations), Solaris (used in SUN workstations) and HPUX (used in HP workstations).
- **Processor Design:** PCs normally use CPUs based on the Complex Instruction Set Computer (CISC) technology whereas workstation CPUs are based on the Reduced Instruction Set Computer (RISC) technology.

Special Purpose

A special purpose computer is a digital or an analog computer specifically designed to perform desired specific task. These are high performance computing systems with special hardware architecture, which is dedicated to solve a specific problem. This is performed with the help of specially programmed Field Programmable Gate Array (FPGA) chips or custom Very Large Scale Integration (VLSI) chips. They are used for special applications, for example, astrophysics computations, Very Large Scale Integration or GRAvity PipE (GRAPE) 6 (for astrophysics and molecular dynamics), Hydra (for playing chess), MDGRAPE-3 (for protein structure computations), etc.

Micro, Mini, Mainframe and Supercomputers

Microcomputers

Microcomputers are developed from advanced computer technology. They are commonly used at home, classroom and in the workplace. Microcomputers are called home computers, personal computers, laptops, personal digital assistants, etc. They are powerful and easy to operate. In recent years, computers were made portable and affordable. The major characteristics of a microcomputer are as follows:

- Microcomputers are capable of performing data processing jobs and solving numerical programs. Microcomputers work rapidly like minicomputers.
- Microcomputers have reasonable memory capacity which can be measured in megabytes.

NOTES

- Microcomputers are reasonably priced. Varieties of microcomputers are available in the market which can be as per the requirement of smaller business companies and educational institutions.
- Processing speed of microcomputers is measured in MHz. A microcomputer running at 90MHz works approximately at 90 MIPS.
- Microcomputers have drives for floppy disk, compact disk and hard disks.
- Only one user can operate a microcomputer at a time.
- Microcomputers are usually dedicated to one job. Millions of people use microcomputers to increase their personal productivity.
- Useful accessory tools, such as clock, calendar, calculator, daily schedule reminders, scratch pads, etc., are available in a microcomputer.
- Laptop computers, also called notebook computers are microcomputers. They use the battery power source. Laptop computers have a keyboard, mouse, floppy disc drive, CD drive, hard disk drive and monitor. Laptop computers are expensive in comparison to personal computers.

Minicomputers

Minicomputers are a scaled down version of mainframe computers. The processing power and cost of a minicomputer are less than that of the mainframe. The minicomputers have big memory sizes and faster processing speed compared to the microcomputer. Minicomputers are also called workgroup systems because they are well suited to the requirements of the minor workgroups within an organization. The major characteristics of a minicomputer are as follows:

- Minicomputers have extensive problem solving capabilities.
- Minicomputers have reasonable memory capacity which can be measured in MB or GB.
- Minicomputers have quick processing speeds and operating systems facilitated with multitasking and network capabilities.
- Minicomputers have drives for floppy disk, magnetic tape, compact disk, hard disks, etc.
- Minicomputers can serve as network servers.
- Minicomputers are used as a substitute of one mainframe by big organizations.

Mainframe Computers

Mainframe computers are generally used for handling the needs of information processing of organizations like banks, insurance companies, hospitals and railways. This type of system is placed in a central location with several user terminals connected to it. The user terminals act as access stations and may be located in the same building Figure 1.2.

NOTES

Fig. 1.2 *Mainframe Computer*

Mainframe computers are bigger and more expensive than workstations. They look like a row of large file cabinets and need a large room with closely monitored humidity and temperature levels. A mainframe system of lower configuration is often referred to as a minicomputer system. The various components of a mainframe computer are as follows:

- **Host, Front End and Back End Computers:** A mainframe system consists of several computers, such as a host computer that carries out most of the computations and has direct control over all other computers. The front end portion is used for handling communications to and from all the user terminals connected to the mainframe computer. The back end portion is used to handle data input/output operations. The host computer and other computers are located in the systems room, to which entry is restricted to system administrators and maintenance staff only.
- **Consoles:** Console terminals are directly connected to the host computer and are mainly used by the system administrator to perform certain administrative tasks like installing new software on the system, taking system backups and changing the configuration of the system.
- **Storage Devices:** A mainframe computer has several magnetic disk drives directly connected to the back end computer. The host computer, via the back end computer, accesses all data to and from these magnetic disks. In addition, a mainframe computer also has a few tape drives and a magnetic tape library (located in the systems room) for restoration and backup of data. The tape drives are located in the users' room, so that users' tapes can be used for input and output.
- **User Terminals:** User terminals are used to access the required stations, which may be located at different locations. Since mainframe computers support multiprogramming with time sharing, they can run different operating systems and can be accessed by multiple users simultaneously.
- **Output Devices:** A mainframe computer has several output devices like printers and plotters, connected to the back end computer, so that these devices are accessible to the user for taking their outputs. A plotter is a device that prints vector graphics on paper using ink pens and pencils on mechanical arms mainly used for large size printouts of architectural and engineering drawings.

Supercomputers

Supercomputers are the most powerful and expensive computers available today. They are primarily used for processing complex scientific applications that involve tasks with highly complex calculations and solving problems with mechanical physics, such as weather forecasting and climate research systems, nuclear weapon simulation and simulation of automated aircrafts. Military organizations, major research and development centres, universities and chemical laboratories are major users of supercomputers.

Supercomputers use multiprocessing and parallel processing technologies to solve complex problems promptly. They use multiprocessors, which enable the user to divide a complex problem into smaller problems. A parallel program is written in a manner that can break up the original problem into smaller computational modules. Supercomputers also support multiprogramming, which allows simultaneous access to the computer by multiple users. Some of the manufacturers of supercomputers are IBM, Silicon Graphics, Fujitsu and Intel.

Personal Computers

A PC is a small single user microprocessor based computer that sits on your desktop and is generally used at homes, offices and schools. As the name implies, PCs were mainly designed to meet the personal computing needs of individuals. Personal computers are used for preparing normal text documents, spreadsheets with predefined calculations and business analysis charts, database management systems, accounting systems and also for designing office stationary, banners, bills and handouts. Children and youth love to play games and surf the Internet, communicate with friends via e-mail and Internet telephony and do many other entertaining and useful tasks.

The configuration varies from one PC to another depending on its usage. However, it consists of a CPU or system unit, a monitor, a keyboard and a mouse. It has a main circuit board or motherboard (consisting of the CPU and the memory), hard disk storage, floppy disk drive, CD-ROM drive and some special add-on cards (like Network Interface Card or NIC) and ports for connecting peripheral devices like printers.

PCs are available in two models namely desktop and tower. In the desktop model, the monitor is positioned on top of the system unit whereas in the tower model the system unit is designed to stand by the side of the monitor or even on the floor to save desktop space. Due to this feature, the tower model is very popular.

Some popular operating systems for PCs are MS DOS, Microsoft Windows, Windows NT, Linux and UNIX. Most of these operating systems have the capability of multitasking, which eases operation and saves time when a user has to switch between two or more applications while performing a job. Some leading PC manufacturers are IBM, Apple, Compaq, Dell, Toshiba and Siemens.

NOTES

Types of Personal Computers

Notebook/Laptop Computers

Notebook computers are battery operated personal computers. Smaller than the size of a briefcase, these are portable computers and can be used in places like libraries, in meetings or even while travelling. Popularly known as laptop computers, or simply laptops, they weigh less than 2.5 kg and can be only 3 inches thick (refer Figure 1.3). Notebook computers are usually more expensive as compared to desktop computers though they have almost the same functions, but since they are sleeker and portable they have a complex design and are more difficult to manufacture. These computers have large storage space and other peripherals, such as serial port, PC card, modem or network interface card, CD-ROM drive and printer. They can also be connected to a network to download data from other computers or to the Internet. A notebook computer has a keyboard, a flat screen with Liquid Crystal Display (LCD) display and can also have a trackball and a pointing stick.

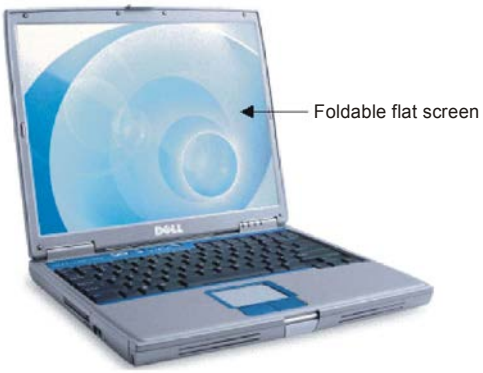


Fig. 1.3 *Laptop Computer*

A notebook computer uses the MS DOS or Windows operating system. It is used for making presentations as it can be plugged into an LCD projection system. The data processing capability of a notebook computer is as good as an ordinary PC because both use the same type of processor, such as an Intel Pentium processor. However, a notebook computer generally has lesser hard disk storage than a PC.

Tablet PC

Tablet PC is a mobile computer that looks like a notebook or a small writing slate but uses a stylus pen or your fingertip to write on the touch screen. It saves whatever you scribble on the screen with the pen as shown in picture in the same way as you have written it. The same picture can than be converted to text with the help of a HR (Hand Recognition) software.

PDA

A Personal Digital Assistant (PDA) is a small palm sized hand held computer which has a small color touch screen with audio and video features. They are nowadays

used as smart phones, Web enabled palmtop computers, portable media players or gaming devices.

Most PDAs today typically have a touch screen for data entry, a data storage/ memory card, Bluetooth, Wireless Fidelity (Wi-Fi) or an infrared connectivity and can be used to access the Internet and other networks.

1.6 USES OF COMPUTERS

Information Technology (or IT as it is popularly called) has changed our life in many ways. Like electricity, IT has impacted all parts of life and, in fact, its usage is so universal that it is difficult for today's generation to lived without computers. It is used in different industries to achieve the following function:

(i) Decreasing the cost of operations by increasing operational efficiency and staff productivity.

(ii) Improving revenues by helping management in informed decision-making and focusing on priority areas.

(iii) Improving customer satisfaction by providing better, faster and value-added services.

IT has opened up several allied industries and employment opportunities which never existed before like Business Process Outsourcing or BPO or Web-enabled services (medical transcription, call centres, etc.). The Internet has brought the world closer.

1. Railways

The railway industry is the backbone of the country's economy. So, for better economic growth, the rail network has to be effective, efficient and timely. In the case of the Indian Railways, which is one of the largest rail networks of the world, it has to manage 11,000 trains every day covering around 108,706 kms, connecting 6,853 stations and a 1.54 million workforce. Since it was very difficult and complex to manage and operate on such a large scale, Indian Railways decided to go for computerization to simplify their operations and have a better control over the management. Some of the major IT initiatives taken by Indian Railways are as follows:

- All India centralized reservation system provides the facility for the passenger to book tickets from any destination and is one of the most successful examples of computerization in the country.
- IRCTC is an online railway ticket booking portal which enables the passenger to book railway tickets for any destination in India from anywhere in the world. It also provides very useful information like computerized reservation-related enquiries about booking status, train schedule and trains between pairs of stations. Booking a railway ticket is now easy with an option to print your tickets from your printer or get it couriered within twenty-four hours. Apart

NOTES

from this, the railway timetables, network maps and freight charges are also available on the Indian Railways Website.

- Indian Railways has also launched the online computerized system for unreserved passengers. This service caters to almost 14 to 15 million people travelling with unreserved tickets. These unreserved tickets being available from locations other than boarding stations reduce the long queues and chaos at the railway station ticket counters.
- Indian Railways has also started a pilot project for issuing periodical season tickets through ATMs (Automated Teller Machines) and another pilot project for buying tickets through smart cards.
- The Interactive Voice Response System (IVRS) has also been introduced to update passengers with railway inquiry and other related information. This national enquiry system is capable of providing train running positions in a real-time system through the IVRS and other output devices.

2. Airlines

The airline business is one of the largest users of computers. Computers have been deployed in almost all aspects of the airline business for increasing revenues, reducing cost and enhancing customer satisfaction.

Imagine the trouble airline companies will have booking air tickets across thousands of booking counters spread all over the world without computerized booking systems and interconnectivity of these systems. The airline industry is using a vast Web-based online system that can be accessed by anyone from anywhere in the world. Some of the major IT initiatives taken by the airline industry are as follows:

- **Online Ticket Booking Through the Internet:** Almost all the airline companies, may it be domestic or international, sell air tickets online. Air tickets can be booked online by paying through credit cards and e-tickets can be printed on your printer.
- **Flight and Seat Availability Information:** Flight and seat availability information, along with the cost of the ticket, is now easily available online with an option to compare it with other available airlines, making it easier to choose the airlines according to the time and price that best suits an individual. The facility of choosing the seat position and the meals (vegetarian or non-vegetarian) if offered, is also available. Yatra.com, makemytrip.com and travelguru.com are a few Websites available for online booking other than the official Website of each airline.
- **Last Minute Deals and Auctions:** To recover the lost revenue opportunity on unsold seats, most airlines have started bidding for last minute tickets in online auction. Several specific airline ticket auctions sites like razorfinish.com are also available. This option is beneficial both for the passenger as well as the airline company since the airline company gets some revenue on unbooked seats and the passenger gets a good deal by paying much less than the regular price.

All these facilities would not have been possible for airline companies without the use of computers.

3. Banking

NOTES

In the 1960s, with the increasing workload, rise in the customer database and bank branches coming up all over, it was getting extremely difficult for banks to fight the competition in providing good customer service. With the increasing number of branches, banks had to simply recruit more manpower to cope up with the additional workload which meant more operational cost and cost was the main basis for competition. To reduce this operational cost, banks computerized all the customer accounts. However, computerization meant reduction in back office operational cost only, and banks still needed manpower as customer relation officers for the front office. To deal with this problem, banks played a smart move by introducing the Automated Teller Machines (ATMs). With the introduction of the ATMs, customer interaction with the bank staff was drastically reduced and customers had the privilege of 24×7 banking. This meant they could take advantage of certain banking facilities round the clock like depositing and withdrawing cash, checking available balances, requesting for statements and requesting for cheque books and drafts without even entering the bank.

Today the banking industry has come a long way since then. It is now one of the largest users of IT. Some of the major IT initiatives taken by the banking industry are as follows:

- **Back Office Computerization:** Not only international banks, but also national banks today run on a fully integrated online system where all back office operations like account transaction postings, bank reconciliation, cheque clearings and other back office banking operations are fully computerized.
- **Front Office Computerization:** With the introduction of the computerized systems, all banks today are in a position to provide certain facilities likes instant bank account statement, electronic fund transfer, direct debit facility etc., to their customers.
- **Automated Teller Machines:** ATMs are a wonderful invention that enabled customers to do their regular bank transactions without visiting their bank branch, instead, visiting their nearest bank ATM which are much more in number and easily accessible. All these transactions, such as withdrawing cash, checking account balance and viewing the bank statement are possible on swiping the card and clicking a button on the computerized machine installed at many locations easily accessible to the customer. ATMs drastically reduced infrastructure and operational costs and enabled the banks to provide a hassle free and a better service 24×7.
- **Internet Banking:** Almost all banks today have an extremely user-friendly Website for e-banking, where the customer can carry out the typical banking transactions, such as making request for cash and cheque pickup, cash delivery, generating account statements, requests for cheque books

NOTES

and drafts online without the hassle of physically visiting the bank. The innovative use of Internet banking enables convenient, cost-effective, easy and quick banking for the customer and hassle-free operation for the banks.

- **Credit Card Operations:** Credit cards provide one with the facility of cashless purchases with an added advantage of credit from the bank at a certain percentage charged as interest. The customer simply swipes his credit card at the time of clearing the bill at the merchant's Point Of Sale (POS) machine that instantly connects with the bank's database through the telephone line and authenticates the customer for credit, simultaneously authorizing the merchant to carry on the transaction and receive the payment from the bank. The customer gets the advantage of not carrying huge sums of money for shopping and the banks earn money on every transaction from the merchant and interest from the customer on the credit given to the customer.

All these operations would not have been possible without the use of IT.

4. Insurance

The insurance industry also involves substantial paperwork, such as maintaining policyholder databases, clearing insurance claims filed, maintaining survey and investigation reports, premium payment receipts, premium overdue, list of policies lapsed, and so on.

Today, with the increasing number of insurance policyholders, it will be very difficult for an insurance company to function without the use of computers. With the help of these computerized systems, insurance companies are able to provide quicker and more efficient service to its customers. The computerized database system also has an option to get information about the customers' financial, economic and demographic details enabling the insurance company to minimize risk and maximize profits.

5. Financial Accounting

In most cases, computerization of business organizations started with the computerization of financial accounting systems.

Computerization of accounts made life much more easier for accountants as they no longer had to maintain manual books, filled vouchers, update registers, maintain long ledgers and then spend days cross-checking the manual entries. Now they can concentrate more on analysing information.

Features like ledger database, automatic calculation, figures tally systems, checks and validations, automatic posting of entries from the voucher to the profit and loss account and the balance sheet, inventory systems, invoicing, creation of challans and purchase orders, creation of relevant reports, interconnectivity between users at different geographical locations and many more, make the computerized accounting systems an error-free, time saving and a fairly simple system to use.

NOTES

According to the difficulty of the operation and size of the organization, various ready-made financial accounting software are available in the market. Tally, EX and Busy are financial accounting software which are good enough for most small- and medium-scale organizations.

On the other hand, larger organizations with offices in multiple locations and multiple operations use Enterprise Resource Planning (ERP) software like ORACLE Financials, BAAN and SAP. Though ERP software cost millions of rupees and are comparatively much more difficult to implement, they provide an excellent platform for ensuring that the company's system and procedures are consistently followed at multiple locations at the same time. Since ERP software provide complete integrated solutions for all functions of business such as financial accounting, payroll management and inventory control, they make it easy for the companies with office at multiple locations to consolidate disparate information at a centralized place, thus enabling the top management to get effective real-time management information.

6. Inventory Control

Most large manufacturing units usually need a variety of raw material for production. Manually managing such a large number of raw materials and then keeping track of the finished products is not an easy task. IT plays a useful role here; a computerized inventory management system provides the facility of specifying the right amount of inventory in hand and determines the time at which, and the amount of inventory required.

The computerized inventory control system maintains a date-wise list of all items along with the maximum, minimum, reorder and inventory in hand and automatically updates the list according to the available material. The computerized inventory control system is also capable of preparing many other useful MIS reports such as aging analysis, goods movement analysis, slow and fast moving stock report, valuation report and it assists the storekeeper and accountants.

ERP software like ORACLE Financials, BAAN, SAP are some examples of sophisticated inventory control packages which can also generate purchase orders automatically whenever the minimum level of stocks is reached, provide automatic posting of accounting entries as soon as any purchase or sale is carried out and generate analytical reports which (itemize) show the previous and future trends in inventory consumption.

Many organizations nowadays have inventory control systems connected through the Internet or the intranet whereby the request for purchase or the purchase order is instantly delivered to the department or the vendor through e-mails, the moment they receive an order or request for an item which is out of or low in stock.

Some interesting IT-based innovations used to simplify and improve inventory management are as follows:

- **Barcode Readers:** Barcodes are a series of black and white parallel and adjacent bars with spaces which represent a string of characters. Bar coding is a quick and easy method for automatically reading barcodes from the products identifying their batch numbers, manufacturing and

NOTES

expiry dates, etc., without having to manually read and type it in the computer to generate bills and track inventory.

- **HandHeld Terminals (HHTs):** HHTs are microprocessor-based simple devices used to communicate with any type of microprocessor-based device. HHTs' standard input device is basically a calculator-like device with a small LCD display output. It is a compact, easy to use device designed for collecting data from large warehouses.

7. Hotel Management

The hotel industry is an important aspect of tourism, which in turn is an source of revenue for any country, especially India. The use of IT in the sector has revolutionized both the hotel and tourism industries. Over the Internet, one can get information about tourist spots, hotel locations, room availability and price details as well as pictures of hotels and locations. IT can play a crucial role for the hotel industry because of its potential of creating customer relationships and the flow of information between the industry and its customers.

Information technology is widely used in the hotel industry; some common instances of use of IT in hotel management are as follows:

- **Computerized Records:** A hotel can track and record all the interactions between a guest and the hotel, from the booking of the room to the check-in and all transactions such as meals, drinks and services provided during their stay. This data can be used to calculate bills and receipts or to help identify particular client types in order to cater to their needs better in future. The data regarding room bookings can be used in a larger hotel management system to track the number of guests in the hotel at a particular time and can also be used in an online booking facility.
- **Online Booking:** Data regarding room availability can be used by the hotel in an online booking system on the hotel's Website. This allows a potential guest to see the availability of different types of rooms as well as the pricing schemes of the hotel. The user can then book a room using this system which will automatically update the hotel management system once the booking is confirmed. The hotel staff will be made aware of the bookings and can make preparations for the guest's arrival.

8. Education

In a traditional teaching model, a teacher would present the study material through verbal presentation, the use of a black or white board to write and illustrate. There would generally be a text book to accompany this material which the students and teacher alike could refer to. Information technology has the potential to enhance and complement traditional teaching methods by providing additional tools to a teacher to display and explain ideas to their students and providing students with innovative but practical learning tools to help study.

For a teacher, the presentation of study material can be enhanced through multimedia presentations where a slide show, animation, video with sound is projected

NOTES

for the students to view. This makes the material more interesting and accessible than if it was merely presented verbally and through a textbook. Furthermore, concepts and ideas presented in a visual format like these as well as verbally are more easily understood and retained longer by students.

Some interesting developments and uses of IT in education are as follows:

- **Computer Based Training (CBT):** Advanced educational institutions can conduct classroom sessions using computer-based training or CBT. Each student will sit at a computer terminal which operates software that presents course material in interactive sessions. It includes refreshers and quizzes of the material presented to reinforce the students' understanding. The benefits of such learning software are that the student can learn at their own pace, and it allows the student to explore and discover ideas and concepts within the material.
- **Internet:** The Internet is a huge source of academic information; a student can use the Web to help his research and study from textbooks and libraries. Search engines allow students to locate relevant and accurate material for study.
- **Distance Learning:** IT applications, such as e-mail, videoconferencing, Web-based study has made distance learning available to many students who are not able to study on location at a university. The students receive and submit assignments, get course material, course information over the Internet and are able to contact their tutors through videoconference or e-mails.

Note that it is important to understand that information technology is not a replacement for real teaching or learning but a tool to improve these things.

9. Telephone Exchanges

In its early stages after invention by Alexander Graham Bell, the telephone system worked on a 'point-to-point' basis. That is, all telephones had to be directly connected by telephone wire in order to make a call. This system was adequate when the number of phone users was very few and the scale of the 'telephone network' was very small but the consequences of using this system on a larger scale became apparent very quickly. A 'point-to-point' telephone network was unworkable for the following reasons: the difficulty of connecting individual phones on a larger network without a centralized system to control this activity and the telephone wire requirements increased as the network grew.

The solution to both these problems was the telephone exchange. A telephone user would be connected directly to their local telephone exchange (as would all other telephone users) where the operator would connect a user to his desired destination by physically connecting the incoming caller's wire to the destination user's telephone wire — known as switching. This process was by no means perfect because as the telephone network grew, the number of exchanges increased and the number of connections at each exchange grew. The whole telephone network required a lot of labor and the process of connecting calls was prone to errors due to the large number of connections at each exchange.

NOTES

The telephone network in its conceptual operation has not changed much since it was first introduced; a telephone user picks up the phone at one end and makes a connection to a user at the other. The complexity, efficiency and reliability of the network has been increased a lot through evolution and developments in technology used in its implementation. Now telephone exchanges are automated and computerized not only for switching calls but also for passing information to a user regarding the status of their call. They can switch and connect many calls at a time with no delays or mistakes. Thus, human beings are no longer required to perform the function of connecting calls.

Computerization and digitization of telephone exchanges have allowed integration of other systems into the telephone network making a variety of services and functionality available to users. These include the following:

- Call waiting.
- Caller identification information.
- Digital answer phones.
- Automated billing.
- Customer information services.

10. Mobile Phones

In this era of technology, using IT tools of communication has become an important aspect of all types of business and personal communication. Until a few years ago, a majority of the population in developing nations remained disconnected from the outside world primarily due to demographics and economic factors. This was especially true in less populated areas where landlines or any other wired modes of communication were never installed due to the low benefit over cost ratio. Even the maintenance of these telephone cables across difficult regions poses a huge challenge to network providers.

The perfect solution for providing connectivity is a mobile phone. It works on the fundamental concept of a radio. Increased coverage can be provided while keeping cost and power expenditure at a minimum by dividing coverage area into cells. Each cell contains a base station within its area of coverage. The base station hosts an antenna and other radio equipment which wirelessly connects with the mobile phones located in its proximity. Mobility beyond the coverage of a cell is achieved by allowing inter-cell communication and transfer of connection from one cell to another.

Mobile technology has made great advances over a short period, starting from the first generation mobile phones, such as the Motorola, using which the subscriber could only make or receive phone calls, down to the latest multipurpose sophisticated mobile phones which can be used to send/receive SMSs, browse the Internet, make calendar entries or watch TV, etc. With the addition of these features, the cost-effectiveness and the efficiency of mobile technology has further improved, thus increasing the popularity of mobile services by making them more affordable. Another reason for the rapid acceptance of mobile phones in our society is due to

the vast functions they offer, such as video recording, music, radio, office document editing, GPS, etc., all integrated into a pocket size device.

Some of the popular functions of mobile phones which are based on IT are as follows:

- **Short Messaging Service (SMS):** Messages containing plain text can be exchanged between any two mobile devices for exchanging information. SMSs are limited by the number of characters each message can contain.
- **Address Book:** It is an electronic medium for storing contact information inside the mobile phone. An address book can contain fields such as full name, phone number, e-mail, fax number, and other information.
- **Calendar Schedules or To-do Lists:** A mobile phone user can fix reminders for upcoming events by putting notes and other relevant information against a particular day and date. A to-do list is a basic text file which stores brief information about unfinished tasks in the chronological order.
- **Send or Receive E-mail:** Mobile phones using WAP, GPRS, EDGE or 3G can send and receive e-mails from the Internet by combining mobile phone features to network services. By using the onboard memory, a user can not only send and receive e-mail but also store them for offline viewing.

11. Weather Forecasting

Predicting the condition or state of the atmosphere after a period of time and over a certain region(s) is known as weather forecasting. The professionals involved in the study and prediction of weather are called *meteorologists*. The atmosphere depends on various factors, such as temperature, humidity, wind speed, etc.

With the development in IT, weather forecasting has become a science. Weather forecasting requires processing and analysing huge amounts of data quickly. This makes it suitable for the application of IT. The volume of data to be processed and the complexity of calculations that must be made in order to forecast weather with a certain degree of accuracy can be gauged by the fact that this task can only be performed by supercomputers which work at very high speeds and can process huge amounts of data very quickly.

The software and hardware tools provided by IT help in making accurate weather forecasts over longer time intervals. Large amounts of data are collected by weather balloons, satellites, sensors and radar instruments and fed into computers with huge processing power and data storage capacity for quantitative analysis and weather modelling. Some examples of software used in this process would be Digital Atmosphere, Forecast Laboratory and RAOB. Accurate assessments of weather conditions over a period of three to six days can be made using hydrological forecasts and warnings of extreme events can be issued five to ten days in advance.

There is still a huge scope for development in the field of weather forecasting and IT is helping it by developing better software for computer modelling, building and designing weather monitoring sensors. Software is also being developed for data collection, analysis and growing channels of weather forecasting services and making systems with huge computing power and storage space available. The screened displays of a weather forecasting is made by using software named METLAB (see Figure 1.4).

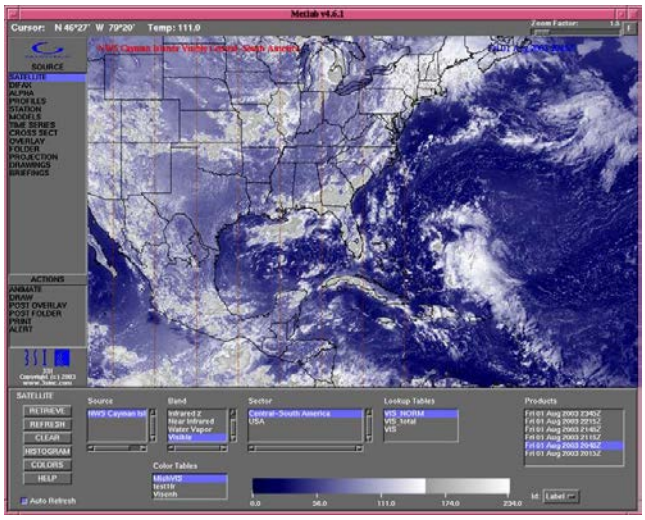


Fig. 1.4 *A Screenshot of a Weather Forecast*

Remote Sensing

The retrieval of data and information regarding an object or phenomenon without coming into physical contact with it is known as remote sensing. The devices used for recording such data are known as sensors, and depending upon the method of retrieval there can be either *recording* or *real-time* sensors. The technique of remote sensing determines if it is *active remote sensing* or *passive remote sensing*.

In active remote sensing, artificial radiation focuses over a particular region of interest and the reflected rays are detected by the sensors to collect data and relevant information. An example of active remote sensing is radar technology.

Passive remote sensing only detects natural radiations of an object or that reflected from its surrounding area. Remote sensors do not emit radiation for measuring values of the object. A remote camera set up to observe wildlife and natural phenomenon is a good example of passive remote sensing.

With the help of IT remote sensors, computer systems and software were developed to monitor and collect geographic or spatially referenced data. Figure 1.5 shows the process of remote sensing.

The various applications of IT in the field of remote sensing are as follows:

(i) **Software:** Embedded software are used to process data from remote sensors and turn it into relevant information. They also control the functions of a remote sensor by judging the data returned from it. Image

enhancement and grouping applications, for example, help in clearing interference from raw images (captured images from camera with minimally processed data and huge detail) and can be used to transform multiple images into one high-resolution continuous image.

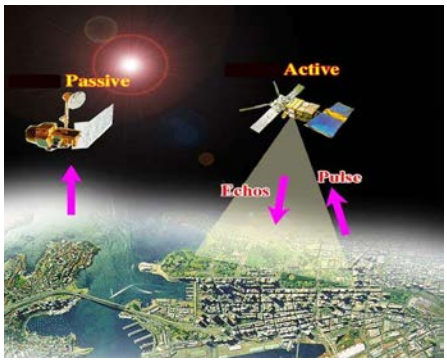


Fig. 1.5 *The Process of Remote Sensing*

(ii) **Hardware**: IT helps in designing customized hardware components for the purpose of remote sensing. The capabilities of a sensor can be optimized if they are redesigned for each application.

(iii) **Telecommunication:** Advancements in the communication between the sensor and the base station have helped in increasing the remote distance. Global environmental mapping, for example, would not have been possible without worldwide telecommunication.

12. E-Commerce

E-commerce is the exchange of goods and services involving financial transaction over an electronic medium by utilizing information and communication technologies, such as Electronic Data Interchange (EDI), Electronic Funds Transfer (EFT), online shopping portal, etc.

An e-commerce system replaces almost all the paperwork in an organization with cheap and more efficient electronic exchange of data which is reliable and secure. The general way of doing business in e-commerce and traditional commerce is the same—a buyer and seller indulge in swapping products and/or services for money. Instead of shops and stores, mail order catalogue or telephone, e-commerce is conducted over a network, such as the Internet. IT tools, such as networking, software development, data mining, data warehousing and Enterprise Resource Planning (ERP) have provided sellers a way to conduct business without the need of building physical infrastructure, printing advertisement pamphlets or recruiting a large number of staff. Additionally, automation of billing and tracking systems has further cut labor costs and time. For distribution of digital content, the distribution channels and delivery time have been minimized as they can readily be downloaded after payment through online payment gateways. Due to the World Wide Web, sellers are not restricted to local markets. They can target customers across the world and provide the same quality of service irrespective of distance and country.

NOTES

E-commerce, by a combination of facilitates such as online payment gateways, customer analysis and report generation, advertisement channels and security coded access to goods and services, provides a sales and distribution channel which is extremely customer-friendly.

The Internet revolution and availability of cheap computers and Internet bandwidth have been the main reasons for the proliferation of e-commerce as a viable alternative to the 'brick-and-mortar' businesses.

'Brick and mortar' is a term used since the development of e-commerce to refer to a traditional model of a business that is based at a commercial address made up of brick and mortar. Before the advent of the Internet and e-commerce, a commercial address or shop front was required for any business that wanted to sell goods directly to the public.

A brick and mortar business can offer some features that an e-commerce business cannot. They are as follows:

- **Customer Security:** Many people feel more comfortable buying goods or services from a business in the real world rather than a virtual business where the customer cannot visit in person.
- **Increased Customer Relations:** A business operating at a commercial address with sales staff can offer a more personal experience to the customer, increasing satisfaction and the possibility of repeat business.
- **Many Businesses Simply Require a Physical Location in Order to Provide Services:** Such businesses could include healthcare services or motor vehicle repair shops.

A brick and mortar business also has disadvantages compared to businesses that operate only online. These are:

- **Increased Cost:** The cost of setting up and running a brick and mortar business is far more than those in e-commerce. The basic reason behind this is the cost of operating in a commercial property. Property cost, maintenance charges, tax, insurance and employees are all costs that e-commerce can avoid but a brick and mortar business cannot.
- **Smaller Customer Base:** A brick and mortar business with no online presence has a customer base that is limited by geography and local population. An online store has no such limitations.

Therefore, many brick and mortar businesses now recognize the potential for business expansion through e-commerce and so simultaneously run online and shop front operations. In that case, there is no competition amd both business can harmoniously coexist and complement each other. So for every example of 'pure' online stores (like e-bay.com and Amazon.com) which sell only on the Web, there are many examples of standard businesses using the online business to complement its brick and mortar operation (like HP and Sony).

NOTES

13. Web Publishing

Traditionally, when we talk about the term publishing then printers, paper, distribution, expensive infrastructure and static content comes to mind. The drawbacks of these are that they require a huge amount of investment, the productivity is low as a lot of manual and machine work is involved, the content published cannot be changed easily and the scope of marketing the product is very limited. All these drawbacks have been overcome by the development of Web publishing.

Web publishing involves putting content on the World Wide Web and includes all the support arrangements required for it. It includes custom Web designs for Web development, Website hosting and e-commerce. Originally, Web publishing simply meant putting selected content on paper into HTML over a Website for public access; this is also known as *ipaper*. This method of publishing is not widely used any more as professional Web publishers now use modern software, such as content management systems for rearranging the structure of a Website and modifying its content.

The most important tool of information technology used in the process of Web publishing is the World Wide Web. This makes content available twenty-four hours a day, seven days a week, to anybody in the world who is connected to the Internet. The only requirement for publishing and viewing the content online is a computer or a handheld device which has an Internet connection and a Web browser. The scope of Web publishing in terms of penetration is very high, with an estimated 1.5 billion Internet users worldwide, as of 2007. The relatively low cost of buying a domain name and hosting a Website is another major driver behind the large amount of online data available over Websites.

14. Management Information Systems

Management Information Systems (MIS) consist of people, technology and procedures to collect, process, store and disseminate data and information required by a business organization for planning, controlling, monitoring, decision-making and other functions of management.

Some of the most commonly used applications of management information systems are accounting, financial management, production planning and control, sales and marketing, human resource management, project management, etc.

The two most important components of an MIS system are database management system and software which allows users to work on these databases for performing various business functions like sales, accounting, etc. The first component is typically covered by a Relational Database Management Systems, also called a RDBMS, and the second by Enterprise Resource Planning, also called ERP.

Although MIS systems need not always be based on standard ERP software since businesses can use custom-built MIS software to serve one or more specific areas of a business, such as inventory control or human resource management.

NOTES

However, the use of standard Commercial Off-The-Shelf (COTS) packages, such as ERP packages, are increasing.

Before the advent of computers and computerized MIS, it was very difficult to manage huge databases of customers, suppliers, shareholders, etc. Since these databases could not be maintained and managed properly, the data therein could not be analysed easily. This meant that managers had to make decisions on either partial data or rely more on 'hunch' or 'gut feel' rather than on facts and figures.

Nowadays, large RDBMSs like Oracle and Ingress, along with ERP packages (which are sophisticated MIS packages covering the entire gamut of functions of an organization) like Oracle, SAP and BAAN and MySQL help organizations to manage vast databases more efficiently, accurately, quickly and easily as compared to past methods of manual data base management. The use of these sophisticated databases and packages has meant that managers can now be better informed and base their decisions on factual analysis rather than timates. While nobody can still deny the role of 'gut-feel' and intuitive decision-making, these tools, when supported by hard facts and figures, provide the basis for making better decisions.

The use of the Internet and Web-based management information systems have made MIS simpler and faster, making it possible to get data and information instantly at the click of a mouse from anywhere across the globe.

15. Decision Support Systems

Decision Support Systems (DSS) is a computerized information system that supports organizational and business decision-making tasks. It is basically an interactive computer software developed to enable decision makers to compile information from basic data, documents, knowledge and business models to enhance the ability to make decisions. A DSS may present information graphically and may support or be complemented by an Artificial Intelligence (AI) or expert system.

A decision support application mainly collects and presents the following information:

- Inventory of all present information assets like data warehouses, data marts, etc.
- Comparison of periodic data, e.g., inventory for this month with past months.
- Projected data based on assumptions, e.g., projected sales figures based on sales assumptions.
- Consequences of different decision alternatives based on past experience.

16. Medical

IT has completely transformed the way modern medical systems work—from storing information about a patient's history to developing new ways of diagnosing patients and educating students in medicine. IT has become such an integral part of the modern medical system that nowadays it is inconceivable to think how this industry worked without it.

Developments in medicine due to IT have offered important benefits to patients and healthcare systems. Research in hi-tech medicine, such as genetic research, DNA modification, hospital infrastructure, rapid ambulance services, etc., have been facilitated by IT. Medical scientists can now use computers to check the effectiveness of a drug against a disease by modelling their genetic structure on computer-based software and using high-speed processors to support the process (see Figure 1.6).

Fig. 1.6 *Use of IT in Medical Science*

The storage and rapid access to electronic medical records and its instant transmission over the Internet in large amounts is called *teleconsulting* where practitioners share patients' data across the world to diagnose patients cooperatively without experiencing their medical history. Videoconferencing between surgeons allows the sharing of expertise so that complicated procedures can be carried out by sharing knowledge in real time. This allows doctors to develop expertise without the need for supervising surgeons to travel. Operations can be performed in areas in which they would not ordinarily be accessible, potentially saving or improving many lives, with the help of IT.

Medical images are sometimes so complicated that they cannot be effectively analysed without using computers. They can not only improve the image quality but also adapt images to fit in accordance to the doctor's desire.

Check Your Progress

4. What are mainframe computers?
5. Define PDA.
6. What is remote sensing?
7. Define E-commerce.

NOTES

1.7 APPLICATIONS OF COMPUTER

Computers are useful in various ways. With the increasing availability of more complex and dynamic operating systems, the primary use of a computer is only limited to the imagination and technical know-how of the user. Everything from your cell phone, DVD player, to your TV has some sort of microprocessor in it, giving it computer-like abilities. Computers have the ability to help the society by various means. They have a number of applications in science as well. In space aeronautics, computers are used in space shuttles for data collection as well as control of flights. In the medical industry, computers are being used in conjunction with robotics to create a new breed of machines that can perform operations on a minimally invasive scale, thereby, increasing the patient's survival rate and reducing the healing time. Computers are also used in agriculture for controlling complex irrigation systems and sensors that detect soil pH among others things, to give the crops a higher yield and faster grow times.

The various applications of computers are as follows:

- **Word processing:** Word processing software automatically corrects spelling and grammar mistakes. If you want some content of your document to be repeated, you do not have to type it each time. You can use the copy and paste features. Images can also be added to your document.
- **Internet:** It is a network of almost all the computers in the world. You can browse through much more information than you could do in a library. That is because computers can store enormous amounts of information. You can also have very fast and convenient access to information. Through e-mail, you can communicate with a person sitting thousands of miles away in seconds. The chat software enables you to chat with another person on a real-time basis. Videoconferencing tools are becoming readily available to the common man.
- **Digital video or audio composition:** Audio or video composition and editing have been made much easier by computers. It no longer costs thousands of dollars of equipment to compose music or make a film. Graphics engineers can use computers to generate short or full-length films or even create three-dimensional models. Anybody owning a computer can now enter the field of media production. Special effects in science fiction and action movies are created using computers.
- **Desktop publishing:** Page layouts for books can also be created on your personal computer.
- **Medicine and health care:** Software is used in magnetic resonance imaging to examine the internal organs of the human body and also used for performing surgery. Computers are used to store patient data.
- **Mathematical calculations:** Computers have computing speeds of over a million calculations per second, using which we can perform various mathematical calculations.

NOTES

- **Banks:** All financial transactions these days are done by computer software. They provide security, speed and convenience.
- **Travel**: One can book air tickets or railway tickets and make hotel reservations online.
- **Communication:** Software is widely used in this field through which you can interact with people around the world.
- **Military:** There is software embedded in almost every weapon. Military software is used for controlling flights and for marking target in ballistic missiles. Software is used to control access to atomic bombs.
- **E-learning:** For a student, it is easier to learn from e-learning software instead of a book.
- **Examinations:** You can give online exams and get instant results.
- **Certificates:** Different types of certificates can be generated and it is very easy to create and change layouts.
- **ATM machines:** The computer software authenticates the user and dispenses cash for banks.
- **Marriage:** There are matrimonial sites through which one can search for a suitable groom or bride.
- **News:** There are many websites through which you can read the latest or old news.
- **Planning and management:** Software can be used to store contact information, generate plans and schedule appointments and deadlines.
- **Plagiarism:** Software can examine content for plagiarism.
- **Sports:** It is used for making umpiring decisions. There are simulation software using which a sportsperson can practice his skills. Computers are also to identify flaws in technique.
- **Airplanes:** Pilots train on software, which simulates flying.
- **Weather analysis:** Supercomputers are used to analyse and predict weather.
- **Research:** Computers are widely used for research purposes in various fields such as follows:
  - Network-attached storage (Linux distribution named FreeNAS)
  - Media Server (Hewlett-Packard makes a dedicated version)
  - Graphics design (Adobe is the forefront in design software)
  - Architectural design (AutoCAD/CAM)
  - Online banking (savings, loans, insurance, credit, mutual funds, etc.)
  - Gaming (computer 3D games, etc.)
  - Social networking (Myspace, Facebook, Twitter)
  - Knowledge sharing (WikiAnswers, Wikipedia, Lifehacker, Gizmodo, etc.).

NOTES

1.8 SUMMARY

- A computer is an electronic device that operates under the control of a set of instructions that is stored in its memory unit. It accepts data from the user through an input device and processes the data into useful information. The processed data is displayed on its monitor.
- The first generation computers like the ENIAC were based on vacuum tubes, which use to overheat and blow up. Hence, they were very unreliable.
- By the mid to late 1950s, vacuum tubes replaced by transistors. Machine reliability increased dramatically. These were the second generation computers. They were smaller in size, generated less amount of heat and had higher capacity of internal storage and their processors operated in microsecond speed range.
- In third generation computers, transistors and other electronic components were combined on a single silicon chip called integrated circuits (IC's). These ICs are popularly known as CMIPs. With this technology, computers became smaller, faster and even more reliable.
- In fourth generation electronic components were further miniaturized and condensed into very large scale integrated circuits (VLSI). One result of VLSI was that it became possible for an entire computer to be put on a single chip. The main features of these generations were that these machines were based on microprocessors; these machines were smaller in size than the earlier generation computers. These computers were quite economical as compared to the earlier generations. They were portable and reliable.
- Computers that can 'think' and are capable of taking decisions like human being have been characterized as the fifth generation computers. They are also termed as *Thinking Machines*. The speed of this generation computers are very high. They use the concept of 'Artificial Intelligence' and possess voice recognition capabilities.
- Artificial intelligence (AI), a branch of computer science, is concerned with making computers perform functions associated with human intelligence, such as reasoning, learning, self-improvement, etc.
- Analog computers are generally used in industrial process controls and to measure physical quantities, such as pressure, temperature, etc. It works on continuous electrical signal inputs and the output is displayed continuously. Its memory capacity is less and can perform only certain type of calculations.
- Digital computers are commonly used for data processing and problem solving using specific programs. A digital computer stores data in the form of digits (numbers) and processes.
- Microcomputers are developed from advanced computer technology. They are commonly used at home, classroom and in the workplace. Microcomputers are called home computers, personal computers, laptops, personal digital assistants, etc.

NOTES

- Minicomputers are a scaled down version of mainframe computers. The processing power and cost of a minicomputer are less than that of the mainframe. The minicomputers have big memory sizes and faster processing speed compared to the microcomputer.
- Mainframe computers are generally used for handling the needs of information processing of organizations like banks, insurance companies, hospitals and railways.
- Supercomputers are the most powerful and expensive computers available today. They are primarily used for processing complex scientific applications that involve tasks with highly complex calculations and solving problems with mechanical physics, such as weather forecasting and climate research systems, nuclear weapon simulation and simulation of automated aircrafts.
- Computer is used in different industries to achieve the following function:
  (i) Decreasing the cost of operations by increasing operational efficiency and staff productivity.
  (ii) Improving revenues by helping management in informed decision-making and focusing on priority areas.
  (iii) Improving customer satisfaction by providing better, faster and value-added services.
- E-commerce is the exchange of goods and services involving financial transaction over an electronic medium by utilizing information and communication technologies, such as Electronic Data Interchange (EDI), Electronic Funds Transfer (EFT), online shopping portal, etc.
- Decision Support Systems (DSS) is a computerized information system that supports organizational and business decision-making tasks.

1.9 KEY TERMS

- **Computer:** An electronic device that operates under the control of a set of instructions that is stored in its memory unit.
- **Main memory:** Also known as the primary memory, it is a part of the CPU and the combination of both RAM and read- only memory (ROM).
- **Information system:** A set of interrelated components that work together to achieve a common goal.
- **Artificial intelligence:** A branch of computer science that is concerned with making computers perform functions associated with human intelligence, such as reasoning, learning, self-improvement, etc.
- **Special-purpose computers:** These computers are chiefly designed to perform specific operation and usually satisfy the needs of a particular type of problem.
- **General-purpose computers:** These computers are designed to perform a large array of functions and operations.

NOTES

1.10 ANSWERS TO 'CHECK YOUR PROGRESS'

1. A computer is an electronic device that operates under the control of a set of instructions that is stored in its memory unit. It accepts data from the user through an input device and processes the data into useful information. The processed data is displayed on its monitor.
2. Some examples of first generation computers were IBM 701 and 650 systems, ENIAC, EDVAC, EDSAC, UNIVAC I, IBM 701, etc.
3. Artificial intelligence (AI), a branch of computer science, is concerned with making computers perform functions associated with human intelligence, such as reasoning, learning, self-improvement, etc.
4. Mainframe computers are generally used for handling the needs of information processing of organizations like banks, insurance companies, hospitals and railways. This type of system is placed in a central location with several user terminals connected to it.
5. A Personal Digital Assistant (PDA) is a small palm sized handheld computer which has a small colour touch screen with audio and video features. They are nowadays used as smart phones, Web enabled palmtop computers, portable media players or gaming devices.
6. The retrieval of data and information regarding an object or phenomenon without coming into physical contact with it is known as remote sensing.
7. E-commerce is the exchange of goods and services involving financial transaction over an electronic medium by utilizing information and communication technologies, such as Electronic Data Interchange (EDI), Electronic Funds Transfer (EFT), online shopping portal, etc.

1.11 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is a computer?
2. What do you understand by the term artificial intelligence?
3. What are the different types of computers?
4. Define general and special purpose computers.

Long-Answer Questions

1. What are the characteristics of computers?
2. Briefly describe the various generations of the computer.
3. Explain the different types of personal computers.
4. What are the uses and application areas of computers?

1.12 FURTHER READING

NOTES

William, Brin K. Stacey C and Sawyer. 2007. *Using Information Technology: A Practical Introduction to Computers and Communications.* Ohio, US: McGraw-Hill Irwin.

Long, Larry and Nancy Long. 2004. *Computers: Information Technology in Perspective*. Upper Saddle River, New Jersey: Prentice Hall.

Sinha, P. K. 2004. *Computer Fundamentals: Concepts Systems and Applications*. New Delhi: BPB.

Norton, Peter. 2008. *Introduction to Computers*. Ohio, US: Tata McGraw.

Basandra, Suresh K. 2007. *Computers Today*. New Delhi: Galgotia.

UNIT 2 DATA REPRESENTATION AND NUMBER SYSTEM

NOTES

Structure

2.0 INTRODUCTION

A mathematical tool used to count and measure is known as a number. The moment we hear the 'term' numbers, our mind begins to picturize digits like 0, 1, 2, 3, etc. Numbers are used in everyday life for calculating even the smallest of things, and has brought system to the world as a whole. Numbers are used everywhere by everyone, in shops, offices, businesses, homes, etc. This unit will discuss number systems, bits and bytes, binary number system, decimal number system, octal number system and hexadecimal number system. You will also learn about conversion from one number system to another, representation of characters and binary arithmetic.

The data plays an important role in programming and all computer programs involve applying operations on the data. The data may be a value or a set of values, such as name and age of a person, grade of a student, salary of an employee, and so on. The data is just a collection of values and no conclusion can be drawn from it; however, after processing it becomes information that can be helpful in making some decisions. You will learn the difference between data and information, and how data can be classified into structured and unstructured data. You will then learn the logical and physical concept of data.

2.1 UNIT OBJECTIVES

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

- Understand the various number systems

NOTES

- Perform conversions from one number system to another
- Differentiate between data and information
- Classify data into structured and unstructured data
- Describe various methods of data processing
- Outline the cycle of data processing
- Explain the meaning and usage of data files

2.2 NUMBER SYSTEM

A number represents a thought that refers to a precise amount of something. Numbers can be expressed in words, gestures and symbols. When expressed in words, numbers are spoken out. Numbers are expressed through gestures using (usually) our hands. Numbers are expressed in symbols that can be written down. A number symbol is known as *numeral*. Hence, a number is a precise idea about an amount, which we form in our minds when we look at a numeral, hear it when it is spoken or see it when it is signalled by hands.

On hearing the word number, we immediately think of the familiar decimal number system with its 10 digits; 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9. These numerals are called Arabic numerals. Our present number system provides modern mathematicians and scientists with great advantages over those of previous civilizations, and is an important factor in our advancement. Since fingers are the most convenient tools nature has provided, human beings use them in counting. So the decimal number system followed naturally from this usage.

A number having base, or radix r, is a system that makes use of distinct symbols of r digits. A string of digit symbols represent numbers. To find out the quantity represented by a number, it is essential that each digit be multiplied by an integer power of r, and then the sum of all the weighted digits be formed. Any whole number greater than one can be used as a base in building a numeration system. The number of digits in use will always be equal to the base.

There are four systems of arithmetic, which are often used in digital systems. These systems are:

1. Decimal
2. Binary
3. Hexadecimal
4. Octal

In any number system, there is an ordered set of symbols known as digits. Collection of these digits makes a number which in general has two parts, integer and fractional, and are set apart by a radix point (.). Hence, a number system can be represented as,

$$N_{\hat{b}} = \underbrace{a_{n-1}a_{n-2}a_{n-3} \ldots a_1a_0}_{\text{Integer portion}} \cdot \underbrace{a_{-1}a_{-2}a_{-3} \ldots a_{-m}}_{\text{Fractional portion}}$$

NOTES

where, N = A number

b = Radix or base of the number system

n = Number of digits in integer portion

m = Number of digits in fractional portion

a_{n-1} = Most significant digit (MSD)

a_{-m} = Least significant digit (LSD)

and $0 \leq (a_i \text{ or } a_{-f}) \leq \cdot\, b_{-1}$

Base or Radix: The base or radix of a number is defined as the number of different digits which can occur in each position in the number system.

Bits and Bytes

All data to be stored and processed in computers are transformed or coded as strings of two symbols, one symbol to represent each state. The two symbols normally used are 0 and 1. These are known as BITS, an abbreviation for BInary digiTS.

Let us now understand some commonly used terms.

BITS The smallest component used by a computer is the bit that holds one of the two likely values.

Value	Meaning
0	Off
1	On

A bit which is OFF is also considered to be FALSE or NOT SET; a bit which is ON is also considered to be TRUE or SET.

Since a single bit can only store two values, there could possibly be only four unique combinations as follows,

00 01 10 11

Bits are therefore, combined together into larger units so that they can hold greater range of values.

NIBBLE A nibble is a group of 4 bits. This gives a maximum number of sixteen possible different values.

$2^4 = 16$ (2 to the power of the number of bits)

BYTES Bytes are a grouping of 8 bits (two nibbles) and are often used to store characters. They can also be used to store numeric values.

$2^8 = 256$ (2 to the power of the number of bits)

Binary Number System

A number system that makes use of just two digits, 0 and 1 is called the **binary number system**. The binary number system is also known as a **base two system**. The two symbols 0 and 1 are known as **bits** (binary digits).

NOTES

The binary system groups numbers by twos and by powers of two, shown in Figure 2.1. The word binary is derived from a Latin word which means two at a time.

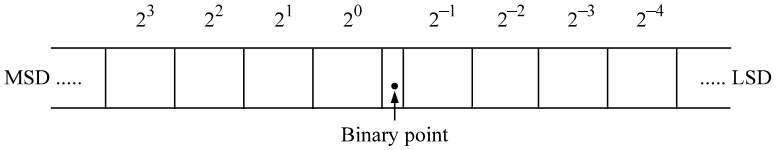


Fig. 2.1 *Binary Position Values as a Power of 2*

The weight or place value of each position can be expressed in terms of 2, and is represented as 2^0, 2^1, 2^2, etc. The least significant digit has a weight of 2^0 (= 1). The second position to the left of the least significant digit is multiplied by 2^1 (= 2). The third position has a weight equal to 2^2 (= 4). Thus, the weights are in the ascending powers of 2 or 1, 2, 4, 8, 16, 32, 64, 128, etc.

The numeral 10_{two} (one, zero, base two) stands for **two**, the base of the system.

In binary counting, single digits are used for **none** and **one**. Two-digit numbers are used for 10_{two} and 11_{two} [2 and 3 in decimal numerals]. For the next counting number, 100_{two} (4 in decimal numerals) three digits are essential. After 111_{two} (7 in decimal numerals) four-digit numerals are used until 1111_{two} (15 in decimal numerals) is reached, and so on. In a binary numeral, every position has a value 2 times the value of the position to its right.

A binary number with 4 bits, is called a **nibble** and a binary number with 8 bits is known as **a byte**.

For example, the number 1011_2 actually stands for the following representation:

$$1011_2 = 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$$
$$= 1 \times 8 + 0 \times 4 + 1 \times 2 + 1 \times 1$$
$$\therefore \quad 1011_2 = 8 + 0 + 2 + 1 = 11_{10}$$

In general,

$$[b_n b_{n-1} \ldots b_2, b_1, b_0]_2 = b_n 2^n + b_{n-1} 2^{n-1} + \ldots + b_2 2^2 + b_1 2^1 + b_0 2^0$$

Similarly, the binary number 10101.011 can be written as

1	0	1	0	1	.	0	1	1
2^4	2^3	2^2	2^1	2^0	.	2^{-1}	2^{-2}	2^{-3}
(MSD)						(LSD)		

$$\therefore \quad 10101.011_2 = 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 + 0 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-3}$$
$$= 16 + 0 + 4 + 0 + 1 + 0 + 0.25 + 0.125 = \mathbf{21.375_{10}}$$

In each binary digit, the value increases in powers of two starting with 0 to the left of the binary point and decreases to the right of the binary point starting with power –1.

Why is Binary Number System used in Digital Computers?

Binary number system is used in digital computers because all electrical and electronic circuits can be made to respond to the two states concept. A switch, for instance, can be either opened or closed, only two possible states exist. A transistor can be made to operate either in cut-off or saturation; a magnetic tape can be either magnetized or non-magnetized; a signal can be either HIGH or LOW; a punched tape can have a hole or no hole. In all of the above illustrations, each device is operated in any one of the two possible states and the intermediate condition does not exist. Thus, zero can represent one of the states and one can represent the other. Hence, binary numbers are convenient to use in analysing or designing digital circuits.

Decimal Number System

The number system which utilizes ten distinct digits, i.e., 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9 is known as decimal number system. It represents numbers in terms of groups of ten, as shown in Figure 2.2.

We would be forced to stop at 9 or to invent more symbols if it were not for the use of positional notation. It is necessary to learn only 10 basic numbers and positional notational system in order to count any desired figure.

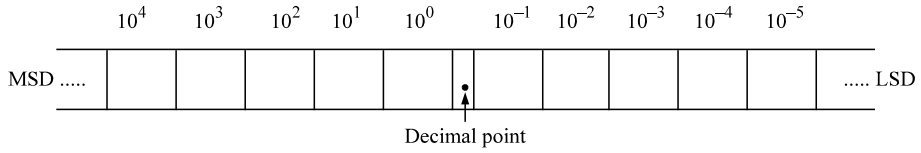


Fig. 2.2 *Decimal Position Values as Powers of 10*

The decimal number system has a base or radix of 10. Each of the ten decimal digits zero through nine, has a place value or weight depending on its position. The weights are units, tens, hundreds and so on. The same can be written as the power of its base as 10^0, 10^1, 10^2, 10^3... etc. Thus, the number 1993 represents quantity equal to 1000 + 900 + 90 + 3. Actually, this should be written as $\{1 \times 10^3 + 9 \times 10^2 + 9 \times 10^1 + 3 \times 10^0\}$. Hence, 1993 is the sum of all digits multiplied by their weights. Each position has a value 10 times greater than the position to its right.

For example, the number 379 actually stands for the following representation.

100	10	1
10^2	10^1	10^0
3	7	9

$3 \times 100 + 7 \times 10 + 9 \times 1$

$$\therefore \quad 379_{10} = 3 \times 100 + 7 \times 10 + 9 \times 1$$
$$= \mathbf{3 \times 10^2 + 7 \times 10^1 + 9 \times 10^0}$$

In this example, 9 is the least significant digit (LSD) and 3 is the most significant digit (MSD).

NOTES

Example 1. Write the number 1936.469 using decimal representation.

Solution: $1936.469_{10} = 1 \times 10^3 + 9 \times 10^2 + 3 \times 10^1 + 6 \times 10^0 + 4 \times 10^{-1} + 6 \times 10^{-2} + 9 \times 10^{-3}$

$= 1000 + 900 + 30 + 6 + 0.4 + 0.06 + 0.009 = \mathbf{1936.469}$

It is seen that powers are numbered to the left of the decimal point starting with 0 and to the right of the decimal point starting with –1.

The general rule for representing numbers in the decimal system by using positional notation is as follows:

$$a_n a_{n-1} \ldots a_2 a_1 a_0 = a_n 10^n + a_{n-1} 10^{n-1} + \ldots a_2 10^2 + a_1 10^1 + a_0 10^0$$

Where n is the number of digits to the left of the decimal point.

Octal Number System

The octal number system was used extensively by early minicomputers. However, for both large and small systems, it has largely been supplanted by the hexadecimal system. Sets of 3 bit binary numbers can be represented by octal numbers and this can conveniently be used for the entire data in the computer.

A number system that uses eight digits, 0, 1, 2, 3, 4, 5, 6 and 7 is called an **octal number system.**

It has a base of **eight**. The digits, 0 through 7 have exactly the same physical meaning as decimal symbols. In this system, each digit has a weight corresponding to its position as shown below:

$$a_n 8^n + \ldots a_3 8^3 + a_2 8^2 + a_1 8^1 + a_{-1} 8^{-1} + a_{-2} 8^{-2} + \ldots + a_{-n} 8^{-n}$$

Octal Odometer

Octal odometer is a hypothetical device similar to the odometer of a car. Each display wheel of this odometer contains only eight digits (teeth), numbered 0 to 7. When a wheel turns from 7 back to 0 after one rotation, it sends a carry to the next higher wheel. Table 2.1 shows equivalent numbers in decimal, binary and octal systems.

Table 2.1 *Equivalent Numbers in Decimal, Binary and Octal Systems*

Decimal (Radix 10)	*Binary (Radix 2)*	*Octal (Radix 8)*
0	000 000	0
1	000 001	1
2	000 010	2
3	000 011	3

4	000 100	4
5	000 101	5
6	000 110	6
7	000 111	7
8	001 000	10
9	001 001	11
10	001 010	12
11	001 011	13
12	001 100	14
13	001 101	15
14	001 110	16
15	001 111	17
16	010 000	20

Consider an octal number $[567.3]_8$. It is pronounced as five, six, seven octal point three and not five hundred sixty seven point three. The co-efficients of the integer part are $a_0 = 7$, $a_1 = 6$, $a_2 = 5$ and the co-efficient of the fractional part is $a_{-1} = 3$.

Hexadecimal Number System

The hexadecimal system groups numbers by sixteen and powers of sixteen. Hexadecimal numbers are used extensively in microprocessor work. Most minicomputers and microcomputers have their memories organized into sets of bytes, each consisting of eight binary digits. Each byte either is used as a single entity to represent a single alphanumeric character or broken into two 4 bit pieces. When the bytes are handled in two 4 bit pieces, the programmer is given the option of declaring each 4 bit character as a piece of a binary number or as two BCD numbers.

The hexadecimal number is formed from a binary number by grouping bits in groups of 4 bits each, starting at the binary point. This is a logical way of grouping, since computer words come in 8 bits, 16 bits, 32 bits and so on. In a group of 4 bits, the decimal numbers 0 to 15 can be represented as shown in Table 2.1.

The hexadecimal number system has a base of 16. Thus, it has 16 distinct digit symbols. It uses the digits 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9 plus the letters A, B, C, D, E and F as 16 digit symbols. The relationship among octal, hexadecimal, and binary is shown in Table 2.2. Each hexadecimal number represents a group of four binary digits.

Table 2.2 *Equivalent numbers in Decimal, Binary, Octal and Hexadecimal Number Systems*

Decimal (Radix 10)	*Binary (Radix 2)*	*Octal (Radix 8)*	*Hexadecimal (Radix 16)*
0	0000	0	0
1	0001	1	1
2	0010	2	2
3	0011	3	3
4	0100	4	4
5	0101	5	5

NOTES

6	0110	6	6
7	0111	7	7
8	1000	10	8
9	1001	11	9
10	1010	12	A
11	1011	13	B
12	1100	14	C
13	1101	15	D
14	1110	16	E
15	1111	17	F
16	0001 0000	20	10
17	0001 0001	21	11
18	0001 0010	22	12
19	0001 0011	23	13
20	0001 0100	24	14

Counting in hexadecimal

When counting in **hex**, each digit can be incremented from 0 to *F*. Once it reaches *F*, the next count causes it to recycle to 0 and the next higher digit is incremented. This is illustrated in the following counting sequences: 0038, 0039, 003A, 003B, 003C, 003D, 003E, 003F, 0040; 06B8, 06B9, 06BA, 06BB, 06BC, 06BD, 06BE, 06BF, 06C0, 06C1.

CHECK YOUR PROGRESS

1. Name the systems of arithmetic used in digital system.
2. What is a nibble?
3. What is an octal number system?

2.3 CONVERSION FROM ONE NUMBER SYSTEM TO THE OTHER

Binary to Decimal Conversion

A binary number can be converted into decimal number by multiplying the binary 1 or 0 by the weight corresponding to its position and adding all the values.

Example 2. Convert the binary number 110111 to decimal number.

Solution:
$$110111_2 = 1 \times 2^5 + 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$$
$$= 1 \times 32 + 1 \times 16 + 0 \times 8 + 1 \times 4 + 1 \times 2 + 1 \times 1$$
$$= 32 + 16 + 0 + 4 + 2 + 1$$
$$= \mathbf{55_{10}}$$

We can streamline binary to decimal conversion by the following procedure:

Step 1. Write the binary, i.e., all its bits in a row.

NOTES

Step 2. Write 1, 2, 4, 8, 16, 32, ..., directly under the binary number working from right to left.

Step 3. Omit the decimal weight which lies under zero bits.

Step 4. Add the remaining weights to obtain the decimal equivalent.

The same method is used for binary fractional number.

Example 3. Convert the binary number 11101.1011 into its decimal quivalent.

Solution:

Step 1:	1	1	1	0	1	.	1	0	1	1
					↑					
					Binary Point					
Step 2:	16	8	4	2	1	.	0.5	0.25	0.125	0.0625
Step 3:	16	8	4	0	1	.	0.5	0	0.125	0.0625

Step 4: $16 + 8 + 4 + 1 + 0.5 + 0.125 + 0.0625 = [29.6875]_{10}$

Hence, $[11101.1011]_2 = \mathbf{[29.6875]_{10}}$

Table 2.3 lists the binary numbers from 0000 to 10000. Table 2.4 lists powers of 2 and their decimal equivalents and the number of K. The abbreviation K stands for $2^{10} = 1024$. Therefore, 1K = 1024, 2K = 2048, 3K = 3072, 4K = 4096, and so on. Many personal computers have 64K memory, this means that computers can store up to 65,536 bytes in the memory section.

Table 2.3 *Binary Numbers*

Decimal	*Binary*
0	0
1	01
2	10
3	11
4	100
5	101
6	110
7	111
8	1000
9	1001
10	1010
11	1011
12	1100
13	1101
14	1110
15	1111
16	10000

Table 2.4 *Powers of 2*

Powers of 2	*Equivalent*	*Abbreviation*
2^0	1	
2^1	2	
2^2	4	
2^3	8	
2^4	16	
2^5	32	
2^6	64	
2^7	128	
2^8	256	
2^9	512	
2^{10}	1024	1K
2^{11}	2048	2K
2^{12}	4096	4K
2^{13}	8192	8K
2^{14}	16384	16K
2^{15}	32768	32K
2^{16}	65536	64K

NOTES

Decimal to Binary Conversion

There are several methods for converting a decimal number to a binary number. The first method is simply to subtract values of powers of 2 which can be subtracted from the decimal number until nothing remains. The value of the highest power of 2 is subtracted first, then the second highest and so on.

Example 4. Convert the decimal integer 29 to the binary number system.

Solution: First the value of the highest power of 2 which can be subtracted from 29 is found. This is $2^4 = 16$.

Then, $29 - 16 = 13$

The value of the highest power of 2 which can be subtracted from 13, is 2^3, then $13 - 2^3 = 13 - 8 = 5$. The value of the highest power of 2 which can be subtracted from 5, is 2^2. Then $5 - 2^2 = 5 - 4 = 1$. The remainder after subtraction is 1^0 or 2^0. Therefore, the binary representation for 29 is given by,

$$29_{10} = 2^4 + 2^3 + 2^2 + 2^0 = 16 + 8 + 4 + 0 \times 2 + 1$$
$$= 1 \quad 1 \quad 1 \quad 0 \quad 1$$
$$[29]_{10} = [11101]_2$$

Similarly, $[25.375]_{10} = 16 + 8 + 1 + 0.25 + 0.125$

$$= 2^4 + 2^3 + 0 + 0 + 2^0 + 0 + 2^{-2} + 2^{-3}$$

$$[25.375]_{10} = \mathbf{[11011.011]_2}$$

This is a laborious method for converting numbers. It is convenient for small numbers and can be performed mentally, but is less used for larger numbers.

Double-Dabble Method

A popular method known as **double-dabble method,** also known as divide-by-two method, is used to convert a large decimal number into its binary equivalent. In this method, the decimal number is repeatedly divided by 2, and the remainder after each division is used to indicate the co-efficient of the binary number to be formed. Notice that the binary number derived is written from the bottom up.

Example 5. Convert 199_{10} into its binary equivalent.

Solution:

$199 \div 2 = 99 +$ remainder	1	(LSB)
$99 \div 2 = 49 +$ remainder	1	
$49 \div 2 = 24 +$ remainder	1	
$24 \div 2 = 12 +$ remainder	0	
$12 \div 2 = 6 +$ remainder	0	
$6 \div 2 = 3 +$ remainder	0	
$3 \div 2 = 1 +$ remainder	1	
$1 \div 2 = 0 +$ remainder	1	(MSB)

NOTES

The binary representation of 199 is, therefore, 11000111. Checking the result we have,

$$[11000111]_2 = 1 \times 2^7 + 1 \times 2^6 + 0 \times 2^5 + 0 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$$

$$= 128 + 64 + 0 + 0 + 0 + 4 + 2 + 1$$

$$\therefore \quad [11000111]_2 = \mathbf{[199]_{10}}$$

Notice that the first remainder is the LSB and last remainder is the MSB. This method will not work for mixed numbers.

Decimal Fraction to Binary

The conversion of decimal fraction to binary fractions may be accomplished by using several techniques. Again, the most obvious method is to subtract the highest value of the negative power of 2, which may be subtracted from the decimal fraction. Then, the next highest value of the negative power of 2 is subtracted from the remainder of the first subtraction, and this process is continued until there is no remainder or to the desired precision.

Example 6. Convert decimal 0.875 to a binary number.

Solution:

$$0.875 - 1 \times 2^{-1} = 0.875 - 0.5 = 0.375$$

$$0.375 - 1 \times 2^{-2} = 0.375 - 0.25 = 0.125$$

$$0.125 - 1 \times 2^{-3} = 0.125 - 0.125 = 0$$

$$\therefore \quad [0.875]_{10} = \mathbf{[0.111]_2}$$

A much simpler method of converting longer decimal fractions to binary consists of repeatedly multiplying by 2 and recording any carriers in the integer position.

Example 7. Convert 0.6940_{10} to a binary number.

Solution:

$$0.6940 \times 2 = 1.3880 = 0.3880 \text{ with a carry of } 1$$

$$0.3880 \times 2 = 0.7760 = 0.7760 \text{ with a carry of } 0$$

$$0.7760 \times 2 = 1.5520 = 0.5520 \text{ with a carry of } 1$$

$$0.5520 \times 2 = 1.1040 = 0.1040 \text{ with a carry of } 1$$

$$0.1040 \times 2 = 0.2080 = 0.2080 \text{ with a carry of } 0$$

$$0.2080 \times 2 = 0.4160 = 0.4160 \text{ with a carry of } 0$$

$$0.4160 \times 2 = 0.8320 = 0.8320 \text{ with a carry of } 0$$

$$0.8320 \times 2 = 1.6640 = 0.6640 \text{ with a carry of } 1$$

$$0.6640 \times 2 = 1.3280 = 0.3280 \text{ with a carry of } 1$$

We may stop here as the answer would be approximate.

$$\therefore \quad [0.6940]_{10} = \mathbf{[0.101100011]_2}$$

If more accuracy is needed, continue multiplying by 2 until you have as many digits as necessary for your application.

NOTES

Example 8. Convert 14.625_{10} to binary number.

Solution: First the integer part 14 is converted into binary and then, the fractional part 0.625 is converted into binary as shown below:

Integer part	*Fractional part*
$14 \div 2 = 7 + 0$	$0.625 \times 2 = 1.250$ with a carry of 1
$7 \div 2 = 3 + 1$	$0.250 \times 2 = 0.500$ with a carry of 0
$3 \div 2 = 1 + 1$	$0.500 \times 2 = 1.000$ with a carry of 1
$1 \div 2 = 0 + 1$	

$\therefore$ The binary equivalent is $\mathbf{[1110.101]_2}$

Octal to Decimal Conversion

An octal number can be easily converted to its decimal equivalent by multiplying each octal digit by its positional weight.

Example 9. Convert $(376)_8$ to decimal number.

Solution: The process is similar to binary to decimal conversion except that the base here is 8.

$$[376]_8 = 3 \times 8^2 + 7 \times 8^1 + 6 \times 8^0$$
$$= 3 \times 64 + 7 \times 8 + 6 \times 1 = 192 + 56 + 6 = \mathbf{[254]_{10}}$$

The fractional part can be converted into decimal by multiplying it by the negative powers of 8.

Example 10. Convert $(0.4051)_8$ to decimal number.

Solution:
$$[0.4051]_8 = 4 \times 8^{-1} + 0 \times 8^{-2} + 5 \times 8^{-3} + 1 \times 8^{-4}$$
$$= 4 \times \frac{1}{8} + 0 \times \frac{1}{64} + 5 \times \frac{1}{512} + 1 \times \frac{1}{4096}$$
$$\therefore \quad [0.4051]_8 = \mathbf{[0.5100098]_{10}}$$

Example 11. Convert (6327.45_8) to its decimal number.

Solution:
$$[6327.45]_8 = 6 \times 8^3 + 3 \times 8^2 + 2 \times 8^1 + 7 \times 8^0 + 4 \times 8^{-1} + 5 \times 8^{-2}$$
$$= 3072 + 192 + 16 + 7 + 0.5 + 0.078125$$
$$[6327.45]_8 = \mathbf{[3287.578125]_{10}}$$

Decimal to Octal Conversion

The methods used for converting a decimal number to its octal equivalent are the same as those used to convert from decimal to binary. To convert a decimal number to octal, we progressively divide the decimal number by 8, writing down the remainders after each division. This process is continued until zero is obtained as the quotient, the first remainder being the LSD.

The fractional part is multiplied by 8 to get a carry and a fraction. The new fraction obtained is again multiplied by 8 to get a new carry and a new fraction. This process is continued until the number of digits have sufficient accuracy.

NOTES

Example 12. Convert $[416.12]_{10}$ to octal number.

Solution: Integer part

$416 \div 8 = 52 + \text{remainder}\ 0$ (LSD)

$52 \div 8 = 6 + \text{remainder}\ 4$ ↑

$6 \div 8 = 0 + \text{remainder}\ 6$ (MSD)

Fractional part

$0.12 \times 8 = 0.96 = 0.96$ with a carry of 0

$0.96 \times 8 = 7.68 = 0.68$ with a carry of 7

$0.68 \times 8 = 5.44 = 0.44$ with a carry of 5

$0.44 \times 8 = 3.52 = 0.52$ with a carry of 3

$0.52 \times 8 = 4.16 = 0.16$ with a carry of 4

$0.16 \times 8 = 1.28 = 0.28$ with a carry of 1

$0.28 \times 8 = 2.24 = 0.24$ with a carry of 2

$0.24 \times 8 = 1.92 = 0.92$ with a carry of 1 ↓

$\therefore [416.12]_{10} = \mathbf{[640.07534121]_8}$

Example 13. Convert $[3964.63]_{10}$ to octal number.

Solution: Integer part

$3964 \div 8 = 495$ with a remainder of 4 (LSD) ↑

$495 \div 8 = 61$ with a remainder of 7

$61 \div 8 = 7$ with a remainder of 5

$7 \div 8 = 0$ with a remainder of 7 (MSD)

$\therefore [3964]_{10} = \mathbf{[7574]_8}$

Fractional part

$0.63 \times 8 = 5.04 = 0.04$ with a carry of 5

$0.04 \times 8 = 0.32 = 0.32$ with a carry of 0

$0.32 \times 8 = 2.56 = 0.56$ with a carry of 2

$0.56 \times 8 = 4.48 = 0.48$ with a carry of 4 ↓

$0.48 \times 8 = 3.84 = 0.84$ with a carry of 3 [LSD]

$\therefore [3964.63]_{10} = \mathbf{[7574.50243]_8}$

Note that the first carry is the MSD of the fraction. More accuracy can be obtained by continuing the process to obtain octal digits.

Octal to Binary Conversion

Since eight is the third power of two, we can convert each octal digit into its 3 bit binary form and from binary to octal form. All 3 bit binary numbers are required to

represent the eight octal digits of the octal form. The octal number system is often used in digital systems, especially for input/output applications. Each octal digit that is represented by 3 bits is shown in Table 2.5.

NOTES

Table 2.5 *Octal to Binary Conversion*

Octal digit	*Binary equivalent*
0	000
1	001
2	010
3	011
4	100
5	101
6	110
7	111
10	001 000
11	001 001
12	001 010
13	001 011
14	001 100
15	001 101
16	001 110
17	001 111

Example 14. Convert $[675]_8$ to binary number.

Solution:

Octal digit	6	7	5	
	↓	↓	↓	
Binary	110	111	101	
$\therefore$ $[675]_8$ =	[110	111	101]$_2$	

Example 15. Convert $[246.71]_8$ to binary number.

Solution:

Octal digit	2	4	6	.	7	1	
	↓	↓	↓		↓	↓	
Binary	010	100	110		111	001	
$\therefore$ $[246.71]_8$ =	[010	100	110	.	111	001]$_2$	

Binary to Octal Conversion

The simplest procedure is to use the **binary-triplet method**. The binary digits are grouped into groups of three on each side of the binary point with zeros added on either side if needed to complete a group of three. Then, each group of 3 bits is converted to its octal equivalent. Note that the highest digit in the octal system is 7.

NOTES

Example 16. Convert $[11001.101011]_2$ to octal number.

Solution: Binary 11001.101011

Divide into groups of 3 bits

011	001	.	101	011
↓	↓		↓	↓
3	1		5	3

Note that a zero is added to the left-most group of the integer part. Thus, the desired octal conversion is $\mathbf{[31.53]_8}$.

Example 17. Convert $[11101.101101]_2$ to octal number.

Solution: Binary $[11101.101101]_2$

Divide into groups of 3 bits

011	101	.	101	101
↓	↓		↓	↓
3	5		5	5

$\therefore$ $[11101.101101]_2 = \mathbf{[35.55]_8}$

Hexadecimal to Binary Conversion

Hexadecimal numbers can be converted into binary numbers by converting each hexadecimal digit to 4 bit binary equivalent using the code given in Table 2.5. If the hexadecimal digit is 3, it should not be represented by 2 bits $[11]_2$, but it should be represented by 4 bits as $[0011]_2$.

Example 18. Convert $[EC2]_{16}$ to binary number.

Solution:

	Hexadecimal number	E	C	2
		↓	↓	↓
	Binary Equivalent	1110	1100	0010
$\therefore$	$[EC2]_{16} =$	**[1110**	**1100**	**0010]$_2$**

Example 19. Convert $[2AB.81]_{16}$ to binary number.

Solution: Hexadecimal number

		2	A	B	.	8	1
		↓	↓	↓		↓	↓
		0010	1010	1011		1000	0001
$\therefore$	$[2AB.81]_{16} =$	[0010	1010	1011	.	1000	0001]$_2$

Binary to Hexadecimal Conversion

Conversion from binary to hexadecimal is easily accomplished by partitioning the binary number into groups of four binary digits, starting from the binary point to the left and to the right. It may be necessary to add zero to the last group, if it does not end in exactly 4 bits. Each group of 4 bits binary must be represented by its hexadecimal equivalent.

NOTES

Example 20. Convert $[10011100110]_2$ to hexadecimal number.

Solution: Binary number $[10011100110]_2$

Grouping the above binary number into 4-bits, we have

	0100	1110	0110
Hexadecimal equivalent	↓	↓	↓
	4	E	6

$\therefore \quad [10011100110]_2 = \mathbf{[4E6]_{16}}$

Example 21. Convert $[111101110111.111011]_2$ to hexadecimal number.

Solution: Binary number $[111101110111.111011]_2$

By Grouping into 4 bits we have,	1111	0111	0111	.	1110	1100
	↓	↓	↓	.	↓	↓
Hexadecimal equivalent	F	7	7	.	E	C

$\therefore \quad [111101110111.111011]_2 = \mathbf{[F77.EC]_{16}}$

The conversion between hexadecimal and binary is done in exactly the same manner as octal and binary, except that groups of 4 bits are used.

Hexadecimal to Decimal Conversion

As in octal, each hexadecimal number is multiplied by the powers of sixteen, which represents the weight according to its position and finally adding all the values.

Another way of converting a hexadecimal number into its decimal equivalent is to first convert the hexadecimal number to binary and then convert from binary to decimal.

Example 22. Convert $[B6A]_{16}$ to decimal number.

Solution: Hexadecimal number $[B6A]_{16}$

$$[B6A]_{16} = B \times 16^2 + 6 \times 16^1 + A \times 16^0$$
$$= 11 \times 256 + 6 \times 16 + 10 \times 1 = 2816 + 96 + 10 =$$

$\mathbf{[2922]_{10}}$

Example 23. Convert $[2AB.8]_{16}$ to decimal number.

Solution: Hexadecimal number

$$[2AB.8]_{16} = 2 \times 16^2 + A \times 16^1 + B \times 16^0 + 8 \times 16^{-1}$$
$$= 2 \times 256 + 10 \times 16 + 11 \times 1 + 8 \times 0.0625$$

$\therefore \quad [2AB.8]_{16} = \mathbf{[683.5]_{10}}$

NOTES

Example 24. Convert $[A85]_{16}$ to decimal number.

Solution: Converting the given hexadecimal number into binary, we have

$$[A85]_{16} = \overset{A}{1010}\ \overset{8}{1000}\ \overset{5}{0101}$$

$$[1010\ \ 1000\ \ 0101]_2 = 2^{11} + 2^9 + 2^7 + 2^2 + 2^0 = 2048 + 512 + 128 + 4 + 1$$

$$\therefore \quad [A85]_{16} = \mathbf{[2693]_{10}}$$

Example 25. Convert $[269]_{16}$ to decimal number.

Solution: Hexadecimal number

$$[269]_{16} = \overset{2}{0010}\ \overset{6}{0110}\ \overset{9}{1001}$$

$$[001001101001]_2 = 2^9 + 2^6 + 2^5 + 2^3 + 2^0 = 512 + 64 + 32 + 8 + 1$$

$$\therefore \quad [269]_{16} = [617]_{10}$$

or, $$[269]_{16} = 2 \times 16^2 + 6 \times 16^1 + 9 \times 16^0 = 512 + 96 + 9 = \mathbf{[617]_{10}}$$

Example 26. Convert $[AF.2F]_{16}$ to decimal number.

Solution: Hexadecimal number

$$[AF.2F]_{16} = A \times 16^1 + F \times 16^0 + 2 \times 16^{-1} + F \times 16^{-2}$$
$$= 10 \times 16 + 15 \times 1 + 2 \times 16^{-1} + 15 \times 16^{-2}$$
$$= 160 + 15 + 0.125 + 0.0586$$

$$\therefore \quad [AF.2F]_{16} = \mathbf{[175.1836]_{10}}$$

Decimal to Hexadecimal Conversion

One way to convert from decimal to hexadecimal is the **hex Dabble method**. The conversion is done in a similar fashion, as in the case of binary and octal, taking the factor for division and multiplication as 16.

Any decimal integer number can be converted to hex successively dividing by sixteen until zero is obtained in the quotient. The remainders can then be written from bottom to top to obtain the hexadecimal results.

The fractional part of the decimal number is converted to hexadecimal number by multiplying it by sixteen and writing down the carry and the fraction separately. This process is continued until the fraction is reduced to zero or the required number of significant bits is obtained.

Example 27. Convert $[854]_{10}$ to hexadecimal number.

Solution:
$$854 \div 16 = 53 + \text{with a remainder of } 6$$
$$53 \div 16 = 3 + \text{with a remainder of } 5$$
$$3 \div 16 = 0 + \text{with a remainder of } 3$$

$$\therefore \quad [854]_{10} = \mathbf{[356]_{16}}$$

NOTES

Example 28. Convert $[106.0664]_{10}$ to hexadecimal number

Solution: Integer part

$106 \div 16 = 6 +$ with a remainder of 10 ↑

$6 \div 16 = 0 +$ with a remainder of 6 |

Fractional part

$0.0664 \times 16 = 1.0624 = 0.0624 +$ with a carry of 1

$0.0624 \times 16 = 0.9984 = 0.9984 +$ with a carry of 0

$0.9984 \times 16 = 15.9744 = 0.9744 +$ with a carry of 15

$0.9744 \times 16 = 15.5904 = 0.5904 +$ with a carry of 15 ↓

Fractional part $[0.0664]_{10} = [0.10FF]_{16}$

Thus, the answer is $[106.0664]_{10} = \mathbf{[6A.10FF]_{16}}$

Example 29. Convert $[65, 535]_{10}$ to hexadecimal and binary equivalents.

Solution: (*a*) Conversion of decimal to hexadecimal number

65,535	÷	16	=	4095 + with a remainder of	F ↑
4095	÷	16	=	255 + with a remainder of	F |
255	÷	16	=	15 + with a remainder of	F |
15	÷	16	=	0 + with a remainder of	F |
∴	$[65535]_{10}$		=		**$[FFFF]_{16}$**

(*b*) Conversion of hexadecimal to binary number

F	F	F	F
1111	1111	1111	1111

∴ $[65535]_{10} = [FFFF]_{16} = [1111\ \ 1111\ \ 1111\ \ 1111]_2$

A typical microcomputer can store up to 65,535 bytes. The decimal addresses of these bytes are from 0 to 65,535. The equivalent binary addresses are from

0000 0000 0000 0000 to 1111 1111 1111 1111

The first 8 bits are called the upper byte, and second 8 bits are called lower byte.

When the decimal is greater than 255, we have to use both the upper byte and the lower byte.

Hexadecimal to Octal Conversion

This can be accomplished by first writing down the 4-bit binary equivalent of hexadecimal digit and then partitioning it into groups of 3 bits each. Finally, the 3-bit octal equivalent is written down.

Example 30. Convert $[2AB.9]_{16}$ to octal number.

Solution: Hexadecimal number 2 A B . 9

↓ ↓ ↓ ↓

4 bit numbers 0010 1010 1011 . 1001

3 bit pattern	001	010	101	011	.	100	100	
	↓	↓	↓	↓		↓	↓	
Octal number	1	2	5	3	.	4	4	

$\therefore$ $[2AB.9]_{16} = \mathbf{[1253.44]_8}$

NOTES

Example 31. Convert $[3FC.82]_{16}$ to octal number.

Solution:

Hexadecimal number	3	F	C	.	8		2	
4 bit binary numbers	0011	1111	1100	.	1000		0010	
3 bit pattern	001	111	111	100	.	100	000	100
	↓	↓	↓	↓		↓	↓	↓
Octal number	1	7	7	4	.	4	0	4

$[3FC.82]_{16} = \mathbf{[1774.404]_8}$

Notice that zeros are added to the rightmost bit in the above two examples to make them group of 3 bits.

Octal to Hexadecimal Conversion

It is the reverse of the above procedure. First the 3-bit equivalent of the octal digit is written down and partitioned into groups of 4 bits, then the hexadecimal equivalent of that group is written down.

Example 32. Convert $[16.2]_8$ to hexadecimal number.

Solution:

Octal number	1	6	.	2
	↓	↓		↓
3 bit binary	001	110	.	010
4 bit pattern		1110	.	0100
		↓		↓
Hexadecimal		E	.	4

$\therefore$ $[16.2]_8 = \mathbf{[E.4]_{16}}$

Example 33. Convert $[764.352]_8$ to hexadecimal number.

Solution:

Octal number	7	6	4	.	3	5	2	
3 bit binary	111	110	100	.	011	101	010	
4 bit pattern	0001	1111	0100	.	0111	0101	000	
	↓	↓	↓		↓	↓	↓	
Hexadecimal number	1	F	4	.	7	5	0	

$\therefore$ $[764.352]_8 = \mathbf{[1F4.75]_{16}}$

CHECK YOUR PROGRESS

4. What is a double-dabble method?
5. How is the binary Triplet method used?
6. How can conversion from binary to hexadecimal be accomplished?

NOTES

2.4 REPRESENTATION OF CHARACTERS

Binary data is not the only data handled by the computer. We also need to process alphanumeric data like alphabets (upper and lower case), digits (0 to 9) and special characters like + – * / ( ) space or blank etc. These also must be internally represented as bits.

Binary Coded Decimal (BCD)

Binary Coded Decimal (BCD) is one of the early memory codes. It is based on the concept of converting each digit of a decimal number into its binary equivalent rather than converting the entire decimal value into a pure binary form. It further uses four digits to represent each of the digits. Table 2.6 shows the BCD equivalent of the decimal digits.

Table 2.6 *BCD Equivalent of Decimals*

Decimal Number	Binary Equivalent
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

Converting $(42)_{10}$ into its BCD equivalent, would result in:

$$(42)_{10} = \underset{4}{0100}\ \underset{2}{0100} \text{ or } 01000010 \text{ in BCD}$$

As seen, 4 bit BCD code can be used to represent decimal numbers only. Since 4 bits are insufficient to represent the various other characters used by the computer, instead of using only 4-bits (giving 16 possible combinations), computer designers commonly use 6 bits to represent characters in BCD code. In this, the 4 BCD numeric place positions are retained, but two additional zone positions are added. With 6 bits it is possible to represent 2^6 or sixty-four different characters. This is therefore sufficient to represent the decimal digits (10), alphabetic characters (26), and special characters (28).

Extended Binary Coded Decimal Interchange (EBCDIC)

The major drawback with the BCD code is that it allows only sixty-four different characters to be represented. This is not sufficient to provide for decimal numbers (10), lowercase letters (26), uppercase letters (26), and a fairly large number of special characters (28 plus).

The BCD code was therefore extended from a 6 bit to an 8 bit code. The added 2 bits are used as additional zone bits, expanding the zone bits to four. This

NOTES

resulting code is called the Extended Binary Coded Decimal Interchange Code (EBCDIC). Using the EBCDIC it is possible to represent 2^8 or 256 characters. This takes care of the character requirement along with a large quantity of printable and several non-printable control characters (movement of the cursor on the screen, vertical spacing on printer etc.).

Since EBCDIC is an 8 bit code, it can easily be divided into two 4 bit groups. Each of these groups can be represented by one hexadecimal digit (explained earlier in this unit). Thus, hexadecimal number system is used as a notation for memory dump by computers that use EBCDIC for internal representation of characters.

Developed by IBM, EBCDIC code is used in most IBM models and many other computers.

American Standard Code for Information Interchange (ASCII)

A computer code that is very widely used for data interchange is called the 'American Standard Code for Information Interchange' or ASCII. Several computer manufacturers have adopted it as their computers' internal code. This code uses seven digits to represent 128 characters. Now an advanced ASCII is used having 8 bit character representation code allowing for 256 different characters. This representation is being used in micro computers.

Let us look at the encoding method. Table 2.7 below shows the bit combinations required for each character.

Table 2.7 *Bit Combinations for Each Character*

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	TAB	LF	VT	FF	CR	SO	SI
10	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
20		!	"	#	$	%	&	'	(	)	*	+	,	-	.	/
30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
50	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
60	'	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	DEL

Thus, to code a text string 'Hello.' in ASCII using hexadecimal digits:

H	e	l	l	o	.
48	65	6C	6C	6F	2E

The string is represented by the byte sequence 48 65 6C 6C 6F 2E.

CHECK YOUR PROGRESS

7. Convert $[2AB.9]_{16}$ to octal number.
8. What concept is BCD based on?
9. What is the major drawback with the BCD code?

NOTES

2.5 BINARY ARITHMETIC

Arithmetic operations are done in computer not by using decimal numbers, as we do normally, but by using binary numbers. Arithmetic circuits in computers and calculators perform arithmetic and logic operations. All arithmetic operations take place in the arithmetic unit of a computer. The electronic circuit is capable of doing addition of two or three binary digits at a time and the binary addition alone is sufficient to do subtraction. Thus, a single circuit of a binary adder with suitable shift register can perform all the arithmetic operations.

Binary Addition

Binary addition is performed in the same manner as decimal addition. Binary addition is the key to binary subtraction, multiplication and division. There are only four cases that occur in adding the two binary digits in any position. This is shown in Table 2.8.

(a) 1 + 1 + 1 = 11 (i.e., 1 carry of 1 into next position)

(b) 1 + 1 + 1 + 1 = 100

(c) 10 + 1 = 11

The rules of (1), (2) and (3) in Table 2.8 are just decimal addition. The rule (4) states that adding 1 and 1 gives one zero (meaning decimal 2 and not decimal 10).

There is a carry from the previous position. 'Carry overs' are performed in the same manner as in decimal arithmetic. Since 1 is the larger digit in the binary system, any sum greater than 1 requires that a digit be carried out.

Table 2.8 *Binary Addition*

Sl. No.	*Augend (A)*		*Addend (B)*	*Carry (C)*	*Sum (S)*	*Result*
1	0	+	0	0	0	0
2	0	+	1	0	1	1
3	1	+	0	0	1	1
4	1	+	1	1	0	10

Example 4. Add the binary numbers (*i*) 011 and 101, (*ii*) 1011 and 1110, (*iii*) 10.001 and 11.110, (*iv*) 1111 and 10010, and (*v*) 11.01 and 101.0111.

Solution: (*i*) Binary number — Equivalent decimal number

	11 ← Carry	
	011	3
	+ 101	5
Sum =	1000	8

(*ii*)

Binary		Decimal
11	← Carry	
1011		11
+1110		+ 14
Sum = 11001		25

(*iii*)

Binary		Decimal
1	← Carry	
10.001		2.125
+11.110		+3.750
Sum = 101.111		5.875

NOTES

(*iv*)	Binary		Decimal
	11	← Carry	
	1111		15
	+ 10010		+ 18
Sum =	100001		33

(*v*)	Binary		Decimal
	11	← Carry	
	11.01		3.25
	101.0111		+ 5.4375
Sum =	1000.1011		8.6875

Since the circuit in all digital systems actually performs addition that can handle only two numbers at a time, it is not necessary to consider the addition of more than two binary numbers. When more than two numbers are to be added, the first two are added together and then their sum is added to the third number and so on. Almost all modern digital machines can perform addition operation in less than 1 μs.

Larger Binary Numbers

Column by column addition applies to binary as well as decimal numbers.

Example 35. Add the following binary numbers.

(*i*) 1101101 and 1001110 (*ii*) 1111001 and 1100101

(*iii*) 110011 and 111000 (*iv*) 1111110 and 11100111

Solution:

	1			1	1	carry			
		1	1	0	1	1	0	1	
(*i*)		1	0	0	1	1	1	0	
	1	0	1	1	1	0	1	1	
		1					carry		
		1	1	0	0	1	1		

	1	1					1	carry	
		1	1	1	1	0	0	1	
(*ii*)	1	1	0	0	1	0	1		
	1	1	0	1	1	1	1	0	
		1	1	1	1	1	1	carry	
		0	1	1	1	1	1	1	0

(*iii*)		1	1	1	0	0	0		
	1	1	0	1	0	1	1		

(*iv*)	1	1	1	0	0	1	1	1	
	1	0	1	1	0	0	1	0	1

Example 36. Add these 8-bit numbers : 0110 1011 and 1011 0110. Then, show the same numbers in hexadecimal notation.

Solution:

8 bit binary			Hexadecimal equivalent
1111	11	carry	
0110	1011		6 B H
+ 1011	0110		+ B 6 H
10010	0001		1 2 1 H

Logic equations representing the sum is also known as the exclusive OR function and can be represented also in Boolean ring algebra as $S = \overline{A}B + \overline{B}A = A \oplus B$.

Binary Subtraction

Subtraction is the inverse operation of addition. To subtract, it is necessary to establish procedure for subtracting a large digit from a small digit. The only case in which this

NOTES

occurs with binary numbers is when 1 is subtracted from 0. The remainder is 1, but it is necessary to borrow 1 from the next column to the left. The rules of binary subtraction are shown below in Table 2.9.

1. 0 – 0 = 0
2. 1 – 0 = 1
3. 1 – 1 = 0
4. 0 – 1 = 0 with a borrow of 1
5. 10 – 1 = 01

Table 2.9 *Binary Subtraction*

Sl. No.	*Minuend A*	_	*Subtrahend B*	*Result*	
1	0	–	0	0	
2	0	–	1	0	with a borrow of 1
3	1	–	0	1	
4	1	–	1	0	

Example 37. (*i*)

	Binary	Decimal	(*ii*) Binary	Decimal
Solution:	1001	9	10000	16
	– 101	– 5	– 011	–3
Difference =	100	4	1101	13

(*iii*)	Binary	Decimal	(*iv*) Binary	Decimal
	110.01	6.25	1101	13
	– 100.1	– 4.5	– 1010	– 10
	1.11	1.75	0011	3

Example 38. Show the binary subtraction of $(128)_{10}$ from $(210)_{10}$.

Solution: Converting the given decimal numbers into corresponding hexadecimal number, we have

$$210 \rightarrow \text{D 2 H} \rightarrow 1101\ 0010$$
$$128 \rightarrow 8\ 0\ \text{H} \rightarrow 1000\ 0000$$

1101 0010	D 2 H
– 1000 0000	– 8 0 H
0101 0010	5 2 H

1's and 2's compliments

Subtraction of a number from another can be accomplished by adding the complement of the subtrahend to the minuend. The exact difference can be obtained with minor manipulations.

1's Complement

The 1's complement form of any binary number is obtained simply by changing each 0 in the number to a 1 and each 1 in the number to a 0.

Binary number		1's complement
1011	→	0100
110110	→	001001
1100 1011	→	0011 0100
1011 1010 1011 1001	→	0100 0101 0100 0110

NOTES

1's complement arithmetic

(*a*) Subtrahend is smaller than the minuend.

1. Complement the subtrahend by converting all 1's to 0's and all 0's to 1's.
2. Proceed as in addition.
3. Disregard the carry and add 1 to the total (end-around-carry).

Example 39. Perform the subtractions using 1's complement addition of the following binary numbers:

(*a*) 110010 – 101101 (*b*) 111001010 – 110110101 (*c*) 11010101 – 10101100

Solution: (a) 110010 – 101101 ⇒ 110010 + 010010 [1's of 101101]

1000100 end-around-carry

└→ 1

000101

(b) 111001010 – 110110101 ⇒ 111001010 + 001001010 [1's of 110110101]

1000010100 end-around-carry

└→ 1

00010101

(c) 11010101 – 10101100 ⇒ 11010101 + 01010011 [1's of 10101100]

100101000 end-around-carry

└→ 1

00101001

(*b*) Subtrahend is larger than the minuend.

1. Complement the subtrahend.
2. Proceed as in addition.
3. Complement the result and place a negative sign in front of the result.

Example 40. Perform the subtractions using 1's complement of the following binary numbers:

(*a*) 1011010 – 1101010 (*b*) 1101011 – 1110101 (*c*) 11110011 – 11111010

NOTES

Solution: (a) 1011010 ⇒ 1011010 (1's of 1101010)
– 1101010 + 0010101
1101111

1's complement of 1101111 = – **0010000**

(b) 1101011 ⇒ 1101011 (1's of 1110101)
– 1110101 + 0001010
1110101

1's complement of 1110101 = – **0001010**

(c) 11110011 ⇒ 11110011 (1's of 11111010)
– 11111010 + 00000101
11111000

1's complement of 11111000 = – **00000111**

Example 41. Use the 1's complement to perform the following subtractions:
(*a*) 1111–1011 (*b*) 110011–100101

Solution: (*a*) Minuend 1111 1111
(*b*) Subtrahend – 1011 0100 (1's of 1011)
10011
↳ 1
0100

1111 – 1011 = $\mathbf{0100_2 = 4_{10}}$

(*b*) Minuend 110011 110011
Subtrahend – 100101 + 011010 (1's of 100101)
1001101
↳ 1
1110

110011 – 100101 = $\mathbf{1110_2 = 14_{10}}$

2's Complement Subtraction

(a) Subtrahend is smaller than the minuend

1. Determine the 2's complement of the smaller number.
2. Add this to the larger number.
3. Disregard the carry.

Example 42. Subtract the following using 2's complement method: (*i*) $(1011)_2$ from $(1100)_2$ (*ii*) $(1001)_2$ 1001_2 from 1101_2 (*iii*) 0101_2 from 1001_2.

Solution: (*i*) Direct subtraction 2's complement method

1100 1100
– 1011 + 0101 ← [2's complement of 1011]
0001 Carry → 1 0001

∴ The carry is disregarded. Thus, the answer is $\mathbf{(0001)_2}$.

(*ii*) Direct subtraction 2's complement method

1101		1101	
– 1011		0111	← 2's complement of 1001
0100	Carry → 1	0101	

∴ The carry is disregarded. Thus, the answer is $\mathbf{(0100)_2}$.

(*iii*) Direct subtraction 2's complement method

1001		1001	
– 0101		+ 1011	← 2's complement of 0101
0100	Carry → 1	0100	

∴ The carry is disregarded. Thus, the answer is $\mathbf{(0100)_2}$.

(*b*) Subtrahend is larger than the minuend.

Example 43. Subtract the following using 2's complement method : (*i*) $(1011)_2$ from $(1101)_2$, (*ii*) $(1100)_2$ from $(1000)_2$.

Solution: (*i*) Direct subtraction 2's complement method

1001		1001	
– 1011		+ 0101	← 2's complement of 1011
0010	No carry →	1110	

No carry is obtained. Thus, the difference is negative and the true answer is 2's complement of $(1110)_2$, i.e., $\mathbf{(0010)_2}$.

(*ii*) Direct subtraction 2's complement method

1000		1000	
– 1100		+ 0100	← 2's complement of 1100
0100	No carry →	1100	

Since no carry is obtained, the difference is negative and therefore the true answer is the 2's complement of $(1100)_2$, i.e., $\mathbf{(0100)_2}$.

Binary Multiplication

The multiplication of binary numbers is done in the same manner as the multiplication of decimal numbers. The following are four basic rules for multiplying binary digits:

(1) $0 \times 0 = 0$ (2) $0 \times 1 = 0$ (3) $1 \times 0 = 0$ (4) $1 \times 1 = 1$

In a computer, the multiplication operation is performed by repeated additions, in much the same manner as the addition of all partial products to obtain the full product. Since the multiplier digits are either 0 or 1, so we are always multiplying by 0 or 1 and no other digit.

NOTES

Example 44. Multiply the binary numbers 1011 and 1101.

$1011 \leftarrow$ Multiplicant $= 11_{10}$

$\times 1011 \leftarrow$ Multiplier $= \times 13_{10}$

143_{10}

1011

0000 $\leftarrow$ Partial product = 14310

1011 $\leftarrow$ Final product = 14310

1011

10001111

Example 45. Multiply the following binary numbers.

(i)

1 1 0 0		12
× 1 0 1 0		×10
0 0 0 0		120
1 1 0 0		
0 0 0 0		
1 1 0 0		
1 1 1 1 0 0 0		

(ii)

1 0 1 1		11
× 1 1 0 0		×12
0 0 0 0		132
0 0 0 0		
1 0 1 1		
1 0 1 1		
1 0 0 0 0 1 0 0		

(iii)

1 . 0 1	1.25
×1 0 . 1	×2.5
1 0 1	3.125
0 0 0	
1 0 1	
1 1 . 0 0 1	

Typical 8 bit microprocessor 6502 is used in software multiplication. In other words, multiplication is done with addition instructions.

Complement Arithmetic

Addition in the 2's Complement System

There are four possible cases:

(*i*) Both numbers positive

(*ii*) A positive number and a smaller negative number

(*iii*) A negative number and a smaller positive number

(*iv*) Both numbers negative

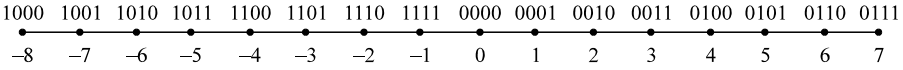


Fig. 2.3 *Binary Odometer Representation in 2's Complement*

The binary odometer is a marvellous way to understand 2's complement representation. There are two important ideas to notice about these odometer readings: (*i*) The MSB is the sign bit : 0 for a + sign and 1 for a – sign (*ii*) The

NOTES

negative numbers shown in Figure 2.3 represents the 2's complements of the positive numbers.

Except for the magnitude, the positive and negative numbers are 2's complements of each other. Hence, we can take the 2's complement of a positive binary number to find the corresponding negative binary number.

Case 1: Two positive numbers

Consider the addition of + 29 and + 19.

	+ 29	0001 1101	(augend)
Adding	+ 19	0001 0011	(addend)
	+ 48	0011 0000	(sum = 48)

Case 2: Positive and smaller negative number

Consider the addition of +39 and –22, remembering that the –22 will be in its 2's complement form. Thus +22 (0001 0110) must be converted to –22 (1110 1010).

	+ 39	0010 0111
Adding	– 22	1110 1010
	17	1 0001 0001
		↑

This carry is disregarded, so the result is 10001.

In this case, the sign bit of addend is 1. The sign bits also participate in the process of addition. In fact a carry is generated in the last position of addition. This carry is always disregarded.

Case 3: Positive and larger negative number

Consider the addition of –47 and +29.

	– 47	1101 0001
Adding	+ 29	0001 1101
	– 18	1110 1110

The result has a sign bit of 1, indicating a negative number. It is in 2's complement form. The last seven bits 110 1110, naturally represent the 2's complement of the sum. To find the true magnitude of the sum, we must take the 2's complement of 1110 1110; the result is 10010 (+ 18). Thus, 1110 1110 represents – 18.

Case 4: Two negative numbers

Consider the addition of – 32 and – 44.

	– 32	1110 0000	(augend)
Adding	– 44	1101 0100	(addend)

– 76 11011 0100 (sum = – 76)

↑

This carry is disregarded, so the result is 1011 0100.

NOTES

Subtraction in 2's Complement System

As in the case of addition, there are four cases. The subtraction operation using the 2's complement system actually involves the operation of addition.

Case 1: Both positive numbers

Consider the case where 19 is to be subtracted from +28

+ 28 0001 1100

+ 19 0001 0011

To subtract + 19 from + 28, the computer will send + 19 to a 2's complement circuit to produce – 19 1110 1101

The computer will then add + 28 and – 19 as follows:

+ 28 0001 1100

Adding – 19 1110 1101

(Sum = 9) 1 0000 1001 disregard the carry

Case 2: Positive and smaller negative number

Consider that the minuend is + 39 and the subtrahend is – 21. In 2's complement system, they appear as,

+ 39 0010 0111

– 21 1110 1011

The computer sends – 21 to a 2's complement circuit to produce,

+ 21 0001 0101

The computer then adds + 39 and + 21 as follows:

+ 39 0010 0111

Adding + 21 0001 0101

(Sum = 60) 0011 1100

Case 3: Positive and larger negative number

Consider that the minuend is + 19 and the subtrahend is – 43. In 2's complement system, they appear as,

+ 19 → 0001 0011

– 43 → 1101 0101

The computer sends the 2's complement of – 43, i.e.,

+ 43 → 0010 1011

Now the computer adds + 19 and + 43 as shown below:

	+ 19	→	0001 0011
Adding	+ 43	→	0010 1011
	(sum = 62)		0011 1110

NOTES

Signal Binary Number Representation

In digital computers, the binary numbers are represented by a set of binary storage devices, such as flip-flops. Each device represents one bit. A 6 bit FF register, for example, could store binary numbers ranging from 000000 to 111111 (0 to 63 decimal). This represents the magnitude of the number. Since digital computers and calculators handle positive as well as negative numbers, some means is required for representing the sign of the number (+ or –). This is usually done by placing another bit called sign bit to the left of the magnitude bits. A 0 in the sign bit represents a positive number and a 1 in the sign bit represents a negative number. Therefore, –101, –100 and –010 are coded as 1101, 1100 and 1010 respectively.

The sign bit is used to indicate whether a stored number is positive or negative. The numbers in Figure 2.4 consists of a sign bit and seven magnitude bits. The magnitude bits are true binary equivalent of the decimal value being represented. This is called the sign-magnitude system for representing signed binary numbers.

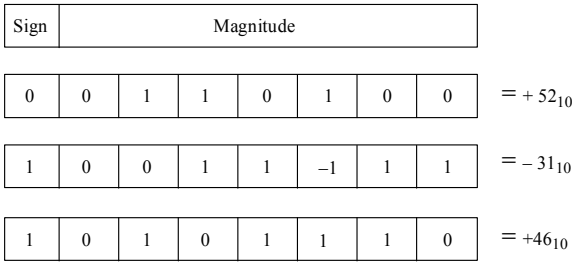


Fig. 2.4 *Representation of Signed Numbers in Sign-magnitude Form*

Three main signed number binary codes are used : (*i*) Sign magnitude (*ii*) 1's complement code, and (*iii*) 2's complement code. The most commonly used system for representing signal binary numbers is the 2's complement system.

CHECK YOUR PROGRESS

10. How are arithmetic operations done in a computer?
11. Add the binary numbers, 1101101 and 1001110.
12. Which is the most commonly used system for representing signal binary numbers?

NOTES

2.6 DATA AND INFORMATION

Data comprises raw facts and/or figures from which meaningful conclusions can be easily drawn. When the data is recorded, classified, organized and related or interpreted within a framework, it provides meaningful information. Information can be defined as 'data that has been transformed into a meaningful and useful form for specific purposes'. Data is represented by the letters of the alphabets or numerals, while the information may be represented in the form of tables, graphs, charts, and so on.

In some cases, data may not require any processing before constituting information. However, data is not useful unless it is subjected to a process through which it is manipulated and organized, and through which its contents are analysed and evaluated. When items are purchased from a departmental store, a number of data items are recorded, such as your name, address, items purchased, the price, the tax, discounts, the amount paid, and so on. If you put these items together and interpret them, they represent information about a business transaction.

Data vs Information

Often the terms data and information are used interchangeably. However, they are different. Data consists of raw facts and figures, such as numbers and text, whereas information is processed data which is presented with some context. For example:

Data	:	261108
Information	:	Joining date of an employee: 26/11/08
	:	Average salary of an entry level software engineer: ₹ 2,61,108
	:	Total number of books available in university library: 2,61,108

The term 'Data Management' refers to all actions related to the management and implementation of storage, updating, retrieval, indexing, security as well as searching of data. Data management is a key process in the functioning of any organization. A well planned and effective data management strategy can make a huge difference to the efficiency and decision-making abilities within an organization. In this unit, we will discuss the need for and applications of data management but first we must understand the concept of 'Data' in the context of business.

Classification of Data

For data management purposes, data is broadly classified into two categories: (i) Structured and (ii) Unstructured data.

Structured Data

Structured data or structured information is the data stored in fixed fields within a file or a record. This form of data representation is also known as 'Tabular Data', where data sets are organized in the form of a table. Structured data is managed by

techniques that work on query and reporting against programmed data types and clear relationships. Databases and spreadsheets are examples of structured data.

Unstructured Data

NOTES

People use and create unstructured data everyday, although they may not be aware, a word processed letter or e-mail, in fact any document and images, such as those captured by a digital camera are all examples of 'Unstructured Data'. Unstructured data primarily consists of 'Textual Data' and 'Image Data'. Textual data being any string of text, this could be a whole book or simply a short note. Images are digital pictures, such as photographs and maps.

Unstructured data in business can take the form of letters, memorandums, reports and legal documents. In order to manage this data effectively, it needs to be organized for storage and retrieval because the information in these documents may be critical to business processes. One technique for organizing or structuring unstructured data is to utilize metadata.

Managing Data

Data is a valuable resource for any organization, large or small. Regardless of the operations and objectives of an organization, it keeps records of its finances, employees, stocks, production, and so on. Whether these records are stored and updated electronically using a computer system or on paper using a filing cabinet, an organization will benefit by managing this data effectively.

Recording and storing data within an organization is only useful if this data is used to benefit the firm. Unused data, apart from the legal requirements of record keeping, is generally considered a wasted resource. Data on stock control and production output in a manufacturing firm can be analysed to identify strengths and weaknesses in the production process, employee records can help identify trends and information regarding salary and demographics to focus development on the workforce. These benefits and many more can be achieved by managing an organization's data.

At its most basic level, managing data is about organizing an environment or system where data can be stored, updated and retrieved. An organization's data management requirements will be greater than this and the specific requirements will be more complex.

Data Management in IT: From the very moment a computer was used to make calculations involving data, the need to store and access this data was identified and the following solutions were developed:

File System was developed in the 1950s followed by hierarchical Database Management System (DBMS) in the 1960s.

Network DBMS, followed by Relational DBMS were developed in the 1970s and later on developed to Object Oriented DBMS in the 1990s.

NOTES

Logical and Physical Concept of Data

The logical architecture of a DBMS is known as three level architecture which was suggested by ANSI/SPARC (American National Standards Institute/Standards Planning And Requirements Committee).

The data can be accessed and manipulated by defining the levels of abstraction. There are three levels of abstraction, which are as follows:

- **Physical or Internal Level:** It is the lowest level of abstraction. It describes how data is actually stored on the physical media.
- **Logical or Conceptual Level:** It is the next higher level of abstraction. It describes what data is stored and how data is interrelated.
- **External or View Level:** It is the lowest level of abstraction as seen by a user. This level of abstraction describes only a part of the entire database or a subset of the database.

Figure 2.5 presents the logical architecture of a typical DBMS.

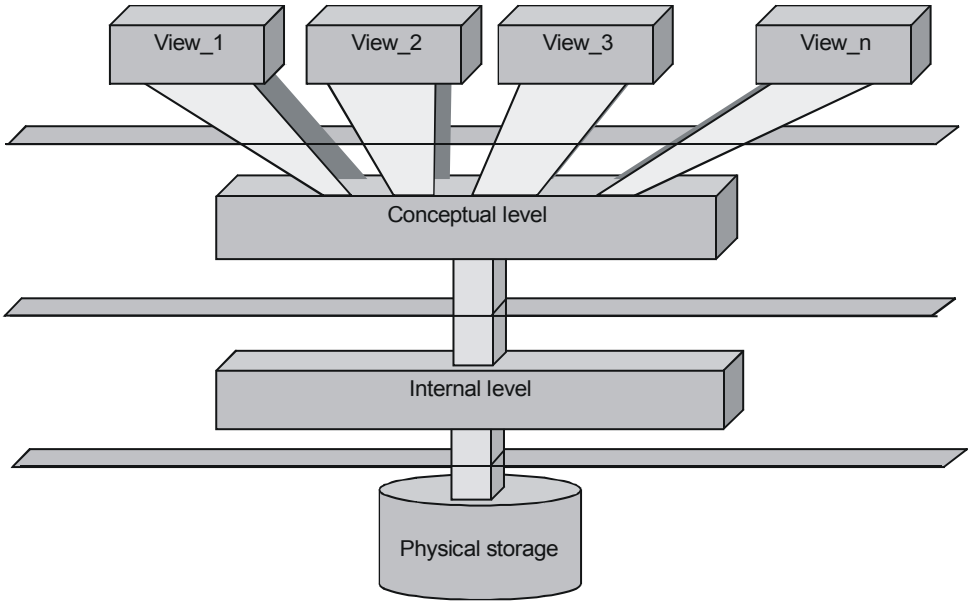


Fig. 2.5 *Logical Architecture of DBMS*

This database architecture clearly separates data representation as seen by the users and the physical data structure layout. This separation has flexibility and adaptability and is known as data independence.

Since a schema defines each view, there exist several schemas in the database partitioned according to the levels of abstraction. The internal view is expressed by the internal schema which contains the definition of the stored record, the method of representing the data fields and the access aids used. The conceptual schema defines

NOTES

this conceptual view. There is only one conceptual schema per database. Each external view is described by means of a schema called an external schema or a subschema. Many external schemas can be defined and even overlap each other.

The concept of level of architecture can be explained with the help of an employee database that contains the details of an employee, such as the employee number, employee name and the department number. The internal level of architecture for the employee database can be represented as follows:

```
Stored_emp  BYTES=20
Prefix      TYPE=BYTE (5), OFFSET=0
Emp         TYPE=BYTE (6), OFFSET=5, INDEX=EMPX
Dept        TYPE=BYTE (2), OFFSET=10
Pay         TYPE=FULLWORD, OFFSET=10
```

`OFFSET` retrieves a portion of the row, which are generated by the query and can skip many rows (not relevant) before returning the rows to the client. `OFFSET 0` is the same as omitting the `OFFSET` clause.

In the internal level, stored record types represent employee, `Stored_emp`, which is 20 bytes long. The `Stored_emp` consists of four stored fields, which are `Prefix`, `Emp`, `Dept` and `Pay`. The `Prefix` contains control information, such as flags or pointers. The data fields represent three properties of employees and also the records in `Stored_emp` are indexed for searching and sorting the records.

In the conceptual level, database contains information about an entity. For example, in an employee database, the conceptual level includes information about employee entity, such as `employee_number`, `dept_number` and `salary`. The conceptual level of architecture for the employee database can be represented as follows:

```
employee
employee_number      CHARACTER (6)
dept_number           CHARACTER (4)
salary               NUMERIC (5)
```

The `employee_number` and `dept_number` are of `CHARACTER` data types having text length equal to 6 and 4, respectively. The salary field is taken as `NUMERIC` having length as 5, which means values up to 5 digits can be entered.

At the external level, the view of the database consists of two fields, `employee_number` and `salary`. The external view shows only the fields that a user needs to view. For example, for an employee database, the external level of architecture consists of two fields, `emp#` and `sal`, which can be represented as:

```
DCL 1 empp,
2 emp# CHAR (6),
2 sal FIXED BIN (30);
```

NOTES

At the external level, the two views can be explained as follows:

View 1

employee_number	FName	Designation	Dept_No

View 2

employee_number	LName	salary

Mapping between the Levels

Mapping is transformation of requests and results between different levels of abstraction. Mapping can be conceptual/internal or external/conceptual. Conceptual/ internal mapping exists between the conceptual and internal levels.

If a modification is made to the structure of the stored database, then accordingly, a change must be made in the conceptual/ internal mapping to ensure that the view from the conceptual level remains the same. In other words, if the physical structure of the database gets modified, the DBMS will be aware of these modifications and continue to provide the same logical view as before the changes. This is physical data independence.

Conceptual/external mapping exists between the conceptual and external levels. Two types of data independence can be defined with respect to the three level architecture, that is, logical data independence and physical data independence. The ability to modify the conceptual scheme without modifying the external schemas or application programs is called logical data interdependence.

The ability to modify the internal schema without changing the conceptual schemas or external schemas is known as physical data interdependence.

If the conceptual view is separated from the internal view, it allows a logical description with no need to specify the physical structures.

Modifications to the internal schema may be required because some physical files need reorganization. This is usually done when the logical database structure is modified.

Physical DBMS Architecture

The related and interconnected software components of a DBMS are described by the physical architecture. At an extremely basic level, this can be split into two parts. These are known as *back end* and *front end*. Back end has the responsibility of managing the physical database supporting and mapping the internal, external and conceptual levels. In addition, the back end is also responsible for the other advantages of a DBMS, such as access control, security and integrity.

The front end is an application running on top of the DBMS and acts as a user interface. The back end is further divided into the functional components of the software, which are as follows:

NOTES

DML Precompiler: This converts the DML (Data Manipulation Language) statements embedded in an application program to normal procedure calls in a host language. Through the DML precompiler, DML commands and application programs written in the host language are separated. DML commands are sent to the DML interpreter for translation into the object code for database access and the rest of the program is sent to the compiler of the host language. Object codes for the commands and the rest of the program are combined together through linking and are sent to the DBMS engine (also called database manager) for execution. The precompiler must interact with the query processor in order to generate the appropriate code. The precompiler interacts with the query processor.

Data Manipulation Language or DML

DML Compiler: This translates DML or Data Manipulation Language statements into low level instructions that a query processor understands.

The manipulation of database is needed if schemas are compiled and database is working with multiple databases. DML, a database language, is used to manipulate data. Data manipulation means accessing and retrieving data, adding new data, manipulating data and deleting data if specific data is not necessary. A query statement in DML is used to retrieve data from database. Query language is posed by a subset of DML. The query language and DML are used as synonymously.

DML is embedded with conventional programming languages, such as C, C++, Pascal, Assembler, COBOL and PL/I. DML statements are called host language if it is embedded with general purpose programming. DML is also known as data sublanguage. The two types of DML are as follows:

- **Low Level or Procedural**

This level facilitates a user to specify what type of data is needed and how to get it. For example, DMLs for hierarchical and network database system.

- **High Level or Non-procedural**

This level facilitates a user to specify what type of data is needed without specifying how to get it, for example, SQL (Sequential Query Language) and QBE (Query By Example).

Data Definition Language or DDL

DDL Interpreter or Compiler: This interprets Data Definition Language or DDL statements and records definitions in the data dictionary.

DDL is also known as database scheme which represents a set of definitions. DDL allows the creation and deletion of structures of database objects as well as provides facilities for defining and altering defined physical data structures. `CREATE`, `DROP` and `ALTER` statements are the most frequently used DDL statements. The definition also includes any constraints that are set of rules to be maintained for integrity of the database.

Database Manager: This is a program module providing an interface for low level data with application programs and queries which are submitted to the database system.

NOTES

The functions of the database manager include:

- Efficient storage, retrieval and updation of data
- Interaction with the file manager
- Ensuring a state of consistency in the database, irrespective of system failures
- Maintenance of integrity controls and user authority for data accession

File Manager: This manages the allocation of disk space and data structures used to represent information on disk.

In addition, several data structures are required for the physical system implementation.

- **Data:** It is stored in data files which store the database itself.
- **Data Dictionary:** Actually, this is a critical element in the DBMS. The result of the compilation of DDL statements is a set of tables that is stored in a special file called data dictionary which documents data in a database. A data dictionary contains metadata (data about data). Metadata is data about the storage details of a database.
- **Indices:** To improve the performance of a DBMS, a set of access aids in the form of indexes are usually provided in the database systems. An index is a data structure that helps access data through a given part of their value. It provides fast access to data. There are several indexes, and for implementing indexes, several techniques are used; however, each technique is specific to a particular database application. A DBMS provides commands to build, maintain and destroy such indexes.
- **Statistical Data File:** The query processor uses statistical data to optimize queries.
- **Log File:** Each log record comprises the values for database items before and after a modification, and it can be utilized for the purpose of recovery.

2.7 DATA PROCESSING

Data processing can be defined as the process of converting raw data into suitable information using a series of operations like classifying, sorting, summarizing and tabulating it for easy storage and retrieval. Processed data is called information.

Data, especially large volumes of it, unless processed properly, are not of much use in the current information driven world. Relevant information can give a definite edge to a business to stay ahead of its competition and plan for the future. In fact, the speed at which information can be extracted from data (a process called data processing) is just as crucial as the information itself. Information usually loses it value if it's not up-to-date. Automatic Data Processing (ADP) applications are gaining wide popularity in the market to solve this very problem. They not only save time but also reduce the cost of data processing. An ADP application, once configured,

NOTES

is ideal for converting similarly structured data into specific sets of information using predefined rules of selection, processing and presentation. Data processing can also include the conversion of one type of information into another for legacy systems transfer.

The data processing activities, including collection, conversion, manipulation, storage and communication, are common to all data processing systems from manual to electronic systems. These activities can be categorized into four groups, namely data input, data processing, data output and storage. These altogether are known as a data processing cycle.

(i) **Input:** The term 'input' denotes the activities that are required for recording data and for making it available for processing. The input can also include the steps that are important to check, verify and validate data contents.

(ii) **Processing:** The term 'processing' refers to the actual data manipulation techniques, such as classifying, sorting, calculating, summarizing, comparing, and so on, which convert data into information.

(iii) **Output:** After the processing of data, the information generated in transmitted through s communication function known as output. The output transmits the information to the persons who need the information. Sometimes output also includes decoding activity, which converts the electronically generated information into human-readable form.

(iv) **Storage:** It is the last stage in the data processing cycle, where data, instruction and information are held for future use. This cycle allows quick access and retrieval of the processed information. It allows to pass the information on to the next stage directly, when needed. Every computer uses storage to hold system and application software.

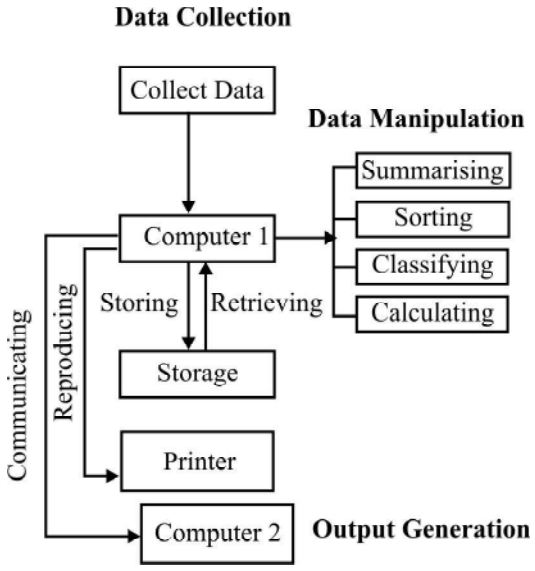


Fig. 2.6 *Data Processing Cycle (DPC)*

NOTES

Typically, a data processing cycle can be broadly divided into three stages:

(i) Data collection

It is the first stage of the cycle, and also the most important. This is because the quality of the data collected can have a strong impact on the output. The collection process should ensure that the data collected are both well-defined and correct, so that the following decisions based on the outcomes are authentic. This stage provides both the baseline from which to measure and a target on what to improve.

Data can be collected in the form of events, transaction or some observations. Following the collection of data, it can be recorded in some usable form. Initially, data may be recorded on paper source documents and then converted into a machine-usable form for processing. On the other hand, they may be recorded by a direct input device in a paperless, machine-readable form. Data collection is also known as data capture.

(ii) Data manipulation

After the collection of data is complete, it is manipulated into a suitable form for further analysis and processing. It is not possible to process raw data, which must also be checked for accuracy. Preparation is about constructing a dataset from one or more data sources, which is to be used for further exploration and processing. The analysis of data, which has not been checked properly for problems, can result in highly misleading results that are heavily dependent on the quality of data prepared.

Once data is collected and converted, it is ready for the manipulation function which converts data into information. Data manipulation includes the following activities:

- **Sorting:** It includes the arrangement of data items in the preferred sequence. Generally, it is easier to work with data if it is arranged in a logical sequence. Usually, the data are arranged in alphabetical order. Sometimes sorting itself will change data into information. Sorting technique is extensively used in business data processing. Almost all the records in business files are maintained in some logical sequence. In computer-based processing systems, numeric sorting is preferred because it is usually faster than alphabetical sorting.
- **Calculating:** It refers to the arithmetic manipulation of data. Items of recorded data can be added to one another, subtracted, divided or multiplied to create new data. Calculation is, therefore, a very important part of data processing.
- **Summarizing:** In this method, large amount of data are condensed or reduced to a more usable and concise form. When the data involved is numbers, you summarize by counting or accumulating the totals of the data in a classification or by selecting strategic data from the mass of data being processed.
- **Comparing:** To compare data means performing an evaluation in relation to some known measure. For example, business managers compare data

to discover how their companies are performing. They may draw a comparison between current sales figures and those of the last year to analyse the performance of the company in the current month.

NOTES

(iii) Output generation

In this stage, the processed information is transmitted to the user. Output is presented to users in different report formats, such as printed report, audio, video or on monitor. Output needs to be understood so that it can provide meaningful information.

Once data has been captured and manipulated, the following activities may be carried out:

- **Storing:** By storing data, the user can utilize the data later or for continued usage. Storage is important for any organized method of processing and re-using data. The storage mechanisms for data processing systems are file cabinets in a manual system and electronic devices, including magnetic disks/magnetic tapes in case of computer based system. The storing activity comprises storing data and information in an organized method for facilitating the retrieval activity. Moreover, data should be stored only if the value of having them in the future exceeds the storage cost.
- **Retrieving:** When the stored data or information is recovered or found again, it is known as retrieval. Retrieval techniques use data storage devices. Therefore, data, whether in file cabinets or in computers, can be recalled for further processing. Retrieval and comparison of old data gives meaning to current information.
- **Communicating:** The process of sharing information is known as communication. Communication is of no use unless the information is made available to the users who require it. Therefore, communication involves the transfer of data and information produced by the data processing system to the potential users of such information or to another data processing system. Accordingly, reports and documents are prepared and delivered to the users. In electronic data processing, results are communicated through display units or terminals.
- **Reproducing:** To reproduce is to copy or duplicate data or information. This reproduction activity may be carried out manually or by machine.

Types of Data Processing

Just as there are different types of data (classified either by usage, attributes or content), there are different methods of processing data. These are as follows:

1. **Manual Data Processing:** This type of processing is used for calculation and processing of data. It is usually simple and operated manually. This type of function is carried out using non-technological tools, such as paper, writing tools and physical filing cabinets.
2. **Electro-Mechanical Data Processing:** It functions by using electric motor, switches and relays for control of processes, for example, desk calculators and punched card processing devices.

NOTES

3. **Electronic Data Processing:** This type of processing is used in the modern age. It includes the modern computer which have evolved into five generations with advancement of processing hardware; the vacuum tubes, transistors, integrated circuits and microprocessors.

   (i) **Batch Processing:** Real-time processing requires high speed broadband connections so that the data inputted from different computers or locations can be used to update a centralized server and database. Setting up such networks is expensive and not always feasible because sometimes the data does not need to be processed immediately. For example, in a BPO (Business Processing Outsourcing) outfit, hundreds of operators may be inputting data, which can be made available to the client only after it is checked and verified by a supervisor(s). Such situations call for batch mode processing, which is used when the conversion of data into information is not required to be done immediately and, therefore, this data processing is done in lots or batches. The advantages of batch processing are that it is cheaper and processing can be done offline (see Figure 2.7).

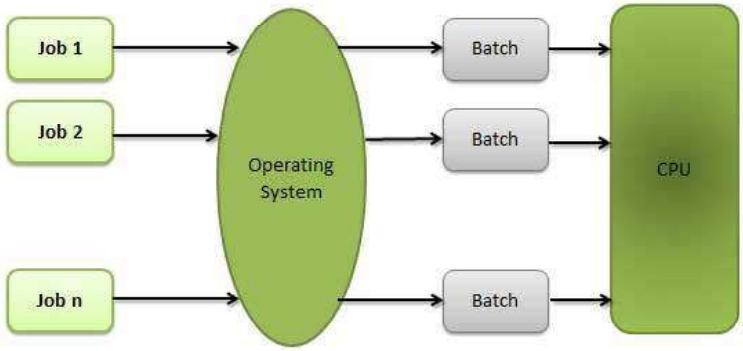


Fig. 2.7 *Batch Processing*

   (ii) **Real-Time Processing:** In this mode, data is processed almost immediately and in a continuous flow (see Figure 2.8). This is of particular advantage when the lifespan of information is small and core business activities are involved. The advantages of real-time processing are that the derived information is up-to-date and so it is more relevant for decision making. For example, in a bank or in an ATM (Automatic Teller Machine), as soon you deposit money in your account, your account status (balance standing to your credit) is updated instantaneously. This enables you as well as the bank to know the exact status of funds, in real time mode, or in other words, as of this minute. Similarly, in a railway reservation system, a train ticket booked from anywhere in the world must update the central database in real-time to ensure that the seat once booked is not sold to anybody else in the world. Real-time processing also requires relatively little storage compared to batch processing.

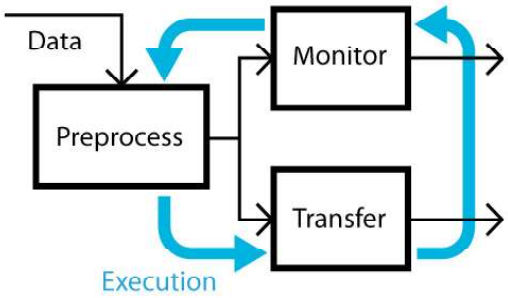


Fig. 2.8 *Real-Time Processing*

(iii) **Online Data Processing:** This is characterized by remote CPU, two-way communication between the CPU and the input or terminal devices, and fast data processing responses. Data is input as soon as it is available with files being update most of the time. Generally, it is expensive to operate; however, the processing speed is faster than offline data processing.

(iv) **Time Sharing:** The basic features of time sharing are multiprogramming, online interaction and real-time response. Figure 2.9 shows time sharing.

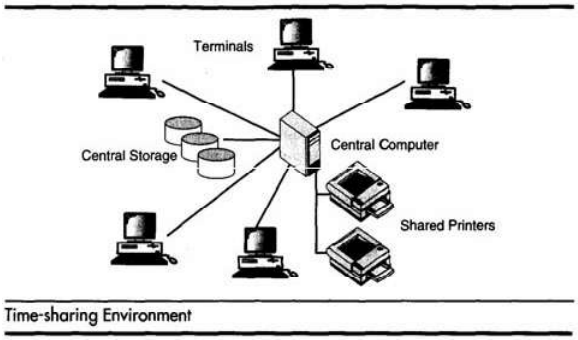


Fig. 2.9 *Time Sharing*

(v) **Distributed Data Processing:** This is characterized by mini- or microcomputers (the satellites) for small scale localized based solution. It can also be used for remote CPU or larger computer for organizational processing or any larger applications for satellite computers; for example, the hospital distributed data processing.

It should be noted that data processing and data conversion are technically quite different; while data conversion only means converting data from one form to another, data processing means conversion of data into information or sometimes vice versa.

Data Processing Cycle

Organizations record the events and transactions that take place in their business activities. The process of converting this data into meaningful information or knowledge is called data processing. Data processing systems are often referred as information processing systems (see Figure 2.10).

NOTES

Fig. 2.10 *Data Processing*

For calculating the salaries of the employees of an organization, a source file is created as input for the program. The accounts assistant collects workers' timecards so he/she knows how many hours each person worked in a particular week and enters the hours from the timecards into a time sheet as shown in Table 2.10. Thus, the generated data is processed to produce information. Table 2.11 is a sample output generated by a data processing application.

Table 2.10 *Time Sheet of an Organization*

Andhra Scientific Works Pvt. Ltd						
Employee Time Sheet						
Work days: 1/12/08 to 7/12/08					**Department Code: Prod.**	
Emp-Name	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Total Hours**
S. Prajit	8	10	8	8	10	44
R. Sujith	8	8	8	8	8	40
K. Manohar	8	8	10	8	10	44
C. Alice	8	8	8	8	8	40

Table 2.11 *A Sample Output Generated by a Data Processing Application*

Andhra Scientific Works Pvt. Ltd					
Employee Weekly Pay Report					
Work days: 1/12/08 to 7/12/08				**Department Code: Prod.**	
Emp-Name	**Total Hrs Worked**	**Pay Rate (₹/hr.)**	**Gross-Pay (₹)**	**Deductions Prof. Tax**	**Net Pay (₹)**
S. Prajit	44	70.00	3,080.00	100.00	2,980.00
R. Sujith	40	75.00	3,000.00	200.00	2,800.00
K. Manohar	44	80.00	3,520.00	400.00	3,120.00
C. Alice	40	80.00	3,200.00	400.00	2,800.00

Data Processing Functions

There are five steps involved in transforming data into information. These are popularly known as *data processing functions* and are defined as follows:

- **Data Collection**: The original data is first collected and recorded into a machine readable format. This can be accomplished by manually entering the data through keyboard.
- **Classification**: The data collected is classified into different classes according to the characteristics of data, such as numeric, alphanumeric or alphabetic.

NOTES

- **Sequencing**: It refers to physical ordering of data. The function sequencing is also referred as sorting.
- **Calculation**: A series of arithmetic and logical operations are performed to validate and manipulate the data so that it is converted into useful information.
- **Transmission**: The information may be moved or transmitted over LAN, WAN or the Internet. It is presented in a variety of formats, such as tables, graphs and charts.

Data Processing Systems

A computer is a programmable machine. Computers act upon executing a series of instructions called software programmes. Data processing systems include computers, application software, printers, scanners, and so on. They are capable of manipulating and storing data.

A Personal Computer (PC) is a desktop computer that runs on a single microprocessor. Typically, a PC consists of a keyboard for entering data; a mouse for interacting with the application software; an output console, called monitor for the display of information; and a storage medium, called hard disk for storing the information. These components of a PC are used for desktop applications, such as Word Processing, Spreadsheets, PowerPoint presentations and other utility software.

The other kind of computer is called mainframe computer which is a bigger computer with high processing and large storage capabilities, usually up to terabytes. Mainframe computers are used for processing large volumes of data. Supercomputers, on the other hand, are extremely fast computers compared to PCs and execute several hundreds of instructions per second.

PC memory and storage cards are the devices that are used to enhance storage capacity of the computers. In order to transmit data files over the Internet network, modems are used. Figure 2.11 illustrates data processing systems.

Fig. 2.11 *Data Processing Systems*

CHECK YOUR PROGRESS

13. Define data and information.
14. What is data processing?
15. List few data processing functions.

NOTES

2.8 DATA FILES

A file contains data/information which are stored permanently in a storage device. Floppy disk and hard disk are commonly used to store file information. When large quantity of data is required to be stored and processed, the concept of file is used. The data entered through a keyboard are stored, retrieved, processed and the results are printed. Consider the Figure 2.12 of file storage and retrieval.

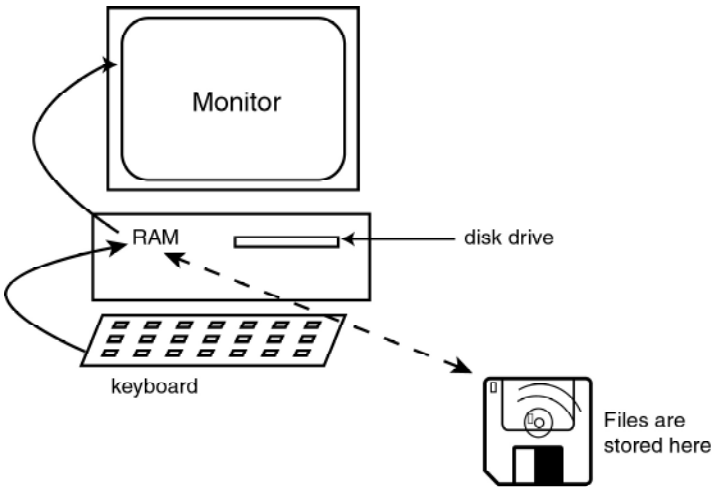
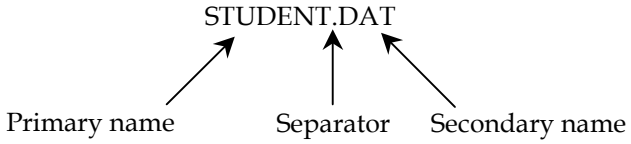


Fig. 2.12 *File Storage and Retrival*

A file stored in a storage device is always identified using a file name (e.g., STUDENT.DAT TEXTINFO.TXT, and so on). Note that a filename normally has a primary name and a secondary name which are separated by a dot (.).



The primary and secondary names are assigned by the programmer. Normally, the secondary name is used to identify the nature of the file. Consider the following secondary names for files:

- DAT is used to refer to a Data file.
- TXT is used to refer to a Text file.

Also note that the secondary names are optional.

Types of Data Files

Files are classified based on the type of data/information stored in them. The two types of files are (i) Data file and (ii) Text file.

A data file contains data stored in the form of records. A record is a collection of data related to a person or an item. For example, a student record may contain data like roll number, students name and marks obtained by him. Note that a

record is called as a structure in C language. A file contains many such records/ structures of data. Imagine that the records are arranged one by one as shown in the Figure 2.13.

NOTES

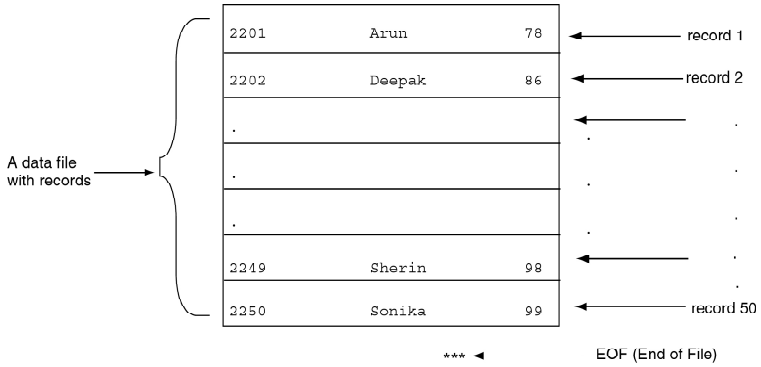


Fig. 2.13 *Records Arranged in a stack attack (data file)*

A text file contains information stored in the form of string of characters. The characters entered through the keyboard are stored continuously as illustrated in Figure 2.14.

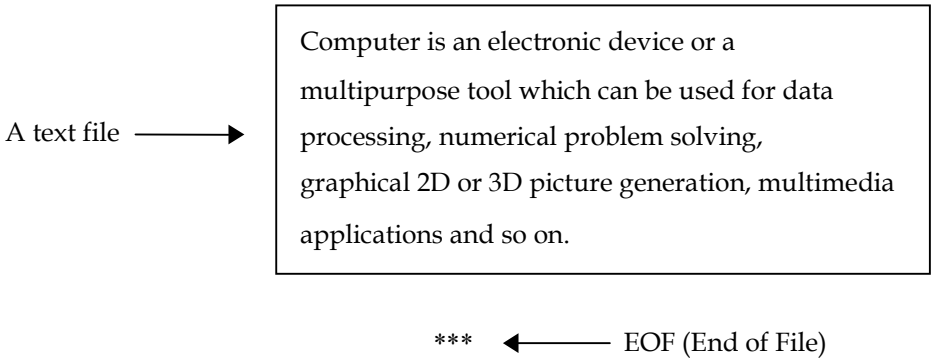


Fig. 2.14 *Text File*

Sequential and Random Files

Depending on the method of accessing the data stored, files can be classified as (i) Sequential files and (ii) Random files.

In a sequential file, information/data is stored sequentially one by one. The data is read in the same order in which they are stored. However, in a random file, information/data can be read randomly. Normally a key data is used to identify the required record in random file accessing. For example, student roll number can be used as a key field in a student data file.

2.9 SUMMARY

- A number represents a thought that refers to a precise amount of something. Numbers can be expressed in words, gestures and symbols. When expressed in words, numbers are spoken out.

NOTES

- A string of digit symbols represent numbers. To find out the quantity represented by a number, it is essential that each digit be multiplied by an integer power of *r*, and then the sum of all the weighted digits be formed. Any whole number greater than one can be used as a base in building a numeration system. The number of digits in use will always be equal to the base.
- There are four systems of arithmetic, which are often used in digital systems. These are decimal, binary, hexadecimal and octal.
- A number system that makes use of two digits, 0 and 1, is known as the binary number system. The binary number system is also known as base two system.
- The octal number system was used extensively by early minicomputers. However, for both large and small systems, it has largely been supplanted by the hexadecimal system.
- The number system which utilizes ten distinct digits, from zero to nine is known as decimal number system.
- The hexadecimal system groups numbers by sixteen and powers of sixteen. Hexadecimal numbers are used extensively in microprocessor work.
- A binary number can be converted into decimal number by multiplying the binary 1 or 0 by the weight corresponding to its position and adding all the values.
- Binary coded decimal (BCD) is one of the early memory codes. It is based on the concept of converting each digit of a decimal number into its binary equivalent rather than converting the entire decimal value into a pure binary form.
- A computer code that is very widely used for data interchange is called the 'American Standard Code for International Exchange' or ASCII. Several computer manufacturers have adopted it as their computers' internal code.
- A binary fraction can be represented by a series of 1 and 0 to the right of a binary point.
- Arithmetic operations are done in computers not by using decimal numbers as we do normally, but by using binary numbers. Arithmetic circuits in computers and calculators perform arithmetic and logic operations. All arithmetic operations take place in the arithmetic unit of the computer.
- Data comprises raw facts and/or figures from which meaningful conclusions can be easily drawn. When the data is recorded, classified, organized and related or interpreted within a framework, it provides meaningful information.
- Structured data or structured information is the data stored in fixed fields within a file or a record. This form of data representation is also known as 'Tabular Data', where data sets are organized in the form of a table.
- Data processing can be defined as the process of converting raw data into suitable information using a series of operations like classifying, sorting, summarizing and tabulating it for easy storage and retrieval.

- A file contains data/information which are stored permanently in a storage device. Files are classified based on the type of data/information stored in them. The two types of files are (i) Data file and (ii) Text file.

NOTES

2.10 KEY TERMS

- **BITS:** It is the smallest element used by a computer.
- **Octal odometer:** It is a hypothetical device similar to the odometer of a car.
- **Binary coded decimal (BCD):** It is based on the concept of converting each digit of a decimal number into its binary equivalent.
- **Data:** It is stored in data files which store the database itself.
- **Statistical Data File:** The query processor uses statistical data to optimize queries.
- **Structured data or structured information:** It is the data stored in fixed fields within a file or a record.
- **Real-Time Processing:** In this mode, data is processed almost immediately and in a continuous flow.
- **Batch Processing**: It is used when the conversion of data into information is not required to be done immediately and, therefore, this data processing is done in lots or batches.

2.11 ANSWERS TO 'CHECK YOUR PROGRESS'

1. There are four systems of arithmetic, which are often used in digital systems.

   These systems are:

   - Decimal
   - Binary
   - Hexadecimal
   - Octal
2. A nibble is a group of 4 bits. This gives a maximum number of sixteen possible different values.
3. A number system that uses eight digits, 0, 1, 2, 3, 4, 5, 6 and 7 is called an octal number system.
4. A popular method also known as divide-by-two method, the double-dabble method is used to convert a large decimal number into its binary equivalent.
5. The simplest procedure is to use the binary-triplet method. The binary digits are grouped into groups of three on each side of the binary point with zero's added on either side if needed to complete a group of three.

NOTES

6. Conversion from binary to hexadecimal is easily accomplished by partitioning the binary number into groups of four binary digits, starting from the binary point to the left and to the right.
7. **Solution:** Hexadecimal number

		2	A	B		9	
		↓	↓	↓		↓	
4 bit numbers	0010	1010	1011		.	1001	
3 bit pattern	001	010	101	011	.	100	100
	↓	↓	↓	↓		↓	↓
Octal number	1	2	5	3	.	4	4

$\therefore$ [2AB.9]16 = **[1253.44]8**

8. Binary Coded Decimal (BCD). It is based on the concept of converting each digit of a decimal number into its binary equivalent rather than converting the entire decimal value into a pure binary form.
9. The major drawback with the BCD code is that it allows only sixty-four different characters to be represented.
10. Arithmetic operations are done in a computer not by using decimal numbers, as we do normally, but by using binary numbers.
11.
```
  1      1 1 carry
    1 1 0 1 1 0 1
+ 1 0 0 1 1 1 0
-----------------
  1 0 1 1 1 0 1 1
```
12. The most commonly used system for representing signal binary numbers is the 2's complement system.
13. Data comprises raw facts and/or figures from which meaningful conclusions can be easily drawn. When the data is recorded, classified, organized and related or interpreted within a framework, it provides meaningful information. Information can be defined as 'data that has been transformed into a meaningful and useful form for specific purposes'. Data is represented by the letters of the alphabets or numerals, while the information may be represented in the form of tables, graphs, charts, and so on.
14. Data processing can be defined as the process of converting raw data into suitable information using a series of operations, such as classifying, sorting, summarizing and tabulating it for easy storage and retrieval. Processed data is called information.
15. Data collection, classification, sequencing, calculation and transmission are some of the main data processing functions.

2.12 QUESTIONS AND EXERCISES

NOTES

Short-Answer Questions

1. Why is binary number system used in digital computers?
2. Write a short note on Octal Number System.
3. Convert [854]10 to hexadecimal number.
4. Write a short note on how binary addition is performed.
5. List the five stages of data processing cycle.

Long-Answer Questions

1. How would you perform a binary to decimal conversion?
2. Explain floating point representation of numbers.
3. Perform the subtraction using 1's complement addition of the following binary numbers: 110010–101101, 111001010–110110101, 11010101–10101100
4. Convert 19910 into its binary equivalent.
5. Convert [65, 535]10 to hexadecimal and binary equivalents.
6. Convert [2AB.9]16 to octal number.
7. Add the following binary numbers:
   (*i*) 011 and 101
   (*ii*) 1011 and 1110
   (*iii*) 10.001 and 11.110
   (*iv*) 1111 and 10010
   (*v*) 11.01 and 101.0111
8. Subtract the following using 2's complement method:
   (*i*) (1011)2 from (1100)2
   (*ii*) (1001)2 10012 from 11012
   (*iii*) 01012 from 10012

2.13 FURTHER READING

William, Brin K. Stacey C and Sawyer. 2007. *Using Information Technology: A Practical Introduction to Computers and Communications.* Ohio, US: McGraw-Hill Irwin.

Long, Larry and Nancy Long. 2004. *Computers: Information Technology in Perspective*. Upper Saddle River, New Jersey: Prentice Hall.

Sinha, P. K. 2004. *Computer Fundamentals: Concepts Systems and Applications*. New Delhi: BPB.

Norton, Peter. 2008. *Introduction to Computers*. Ohio, US: Tata McGraw.

Basandra, Suresh K. 2007. *Computers Today*. New Delhi: Galgotia.

UNIT 3 HUMAN COMPUTER INTERFACE

NOTES

Structure

3.0 INTRODUCTION

In this unit, you will learn how the operating system acts as interface between user and hardware. An operating system is a set of programs that manages computer hardware resources to provide common services for software and application programs. Without an operating system, a user cannot run an application program on their computer unless the application program is executed with self-booting process. For hardware functions, such as input and output, and memory allocation, the operating system acts as an intermediary between application programs and the computer hardware although the application code is usually executed directly by the hardware and will call the operating system or be interrupted by it. An operating system manages the computer's memory, processes and all of its software and hardware. There are many objectives to be met for achieving the ultimate goal of easy to use and human friendly operating system that ensures the computer to be useful, user friendly, acceptable and affordable to everyone in this world.

This unit also explains the characteristics of operating system and types of software.

3.1 UNIT OBJECTIVES

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

- Define operating system
- Understand the role of OS as interface
- Discuss the characteristics of OS
- Explain the types of software

3.2 OPERATING SYSTEM AS USER INTERFACE

NOTES

An OS is a program that acts like an interface between a user of a computer and the computer hardware. The purpose of the operating system is to provide an environment to the user in which he/she can execute programs. The operating system is an integral part of almost every computer system. A computer system can broadly be divided into three components:

(i) The hardware which comprises the memory, the CPU, the ALU, various bulk storage, I/O and peripheral devices.

(ii) Systems programs, such as OS, device drivers, loaders, utilities, etc.

(iii) Application programs such as database systems, business programs, etc.

A computer system is described in Figure 3.1.

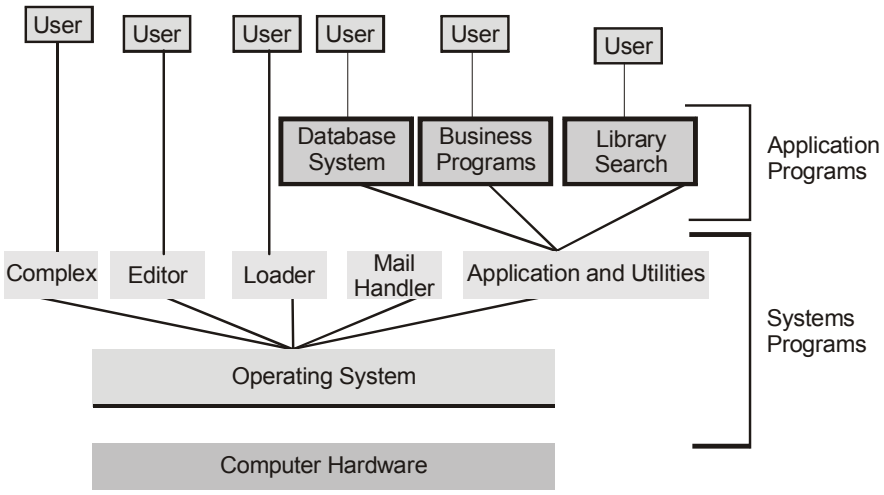


Fig. 3.1 *Conceptual View of a Computer System*

The CPU is located on the chips inside the system unit. The CPU is the brain/heart/soul of the computer system. It is the place where the computer interprets and processes information. The OS is the first component of the systems programs that interests us here.

Systems programs are the programs that are written for direct execution on computer hardware in order to make the power of the computer fully and efficiently accessible to the applications programmers and other computer users. Note that *system programming* is different such as from *application programming*, as it requires an intimate knowledge of the computer hardware as well as the end users' needs.

Systems programs are often large and more complex than application programs, although that is not always the case. Since systems programs provide the foundation upon which application programs are built, it is imperative that systems programs are reliable, efficient and correct.

Definition of Operating System

The OS can be defined as set of programs that are capable to manage resources as well as the operations of a computer. The OS can hide the hardware virtually from the user, i.e., it isolates the hardware from the user (see Figure 3.2). Till the time the functions are performed, user is not concerned about the hardware specifications.

NOTES

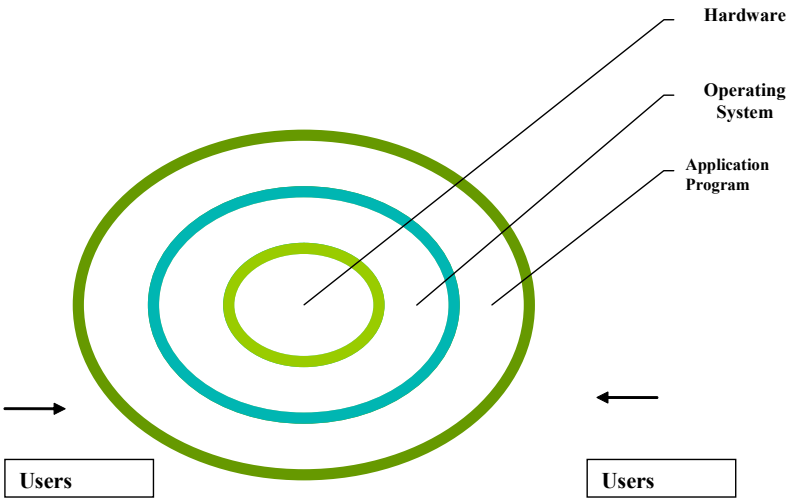


Fig. 3.2 *Isolation of Hardware from the Users by OS*

Characteristics of Operating Systems

It would be useful, at this stage to outline the functions and task commonly performed by the operating systems, irrespective of the makes and models of computers. These are:

- **Storage management**: This involves allocation of space in main storage to meet the processing requirements of the different jobs, transferring data and instructions between the peripherals devices as well as between the peripheral devices and the main storage. Besides this it also keeps track of where the data lies in the main and backing storages and calling programs and subroutines into main storage as and when required.
- **Processing management**: The main activities under the processing management are:
  (a) Equipment control involves activation, deactivation, and synchronization of devices so that two or more processes do not block each other that are getting into what is called a deadly embrace.
  (b) Interrupt handling involves passing of control from one job to another under a system of priorities where more than one application programs is resident in the main memory (multi programming).
  (c) Sequencing involves execution of instructions in their sequence, choosing as necessary between alternative sets of instructions; execution of repeat instructions.

NOTES

(d) Data Transfers between memories, terminals etc.

(e) Arithmetic Operations involves addition, subtraction, multiplication and division of numeric values.

(f) Text Processing involves copying, comparison, extraction, insertion of textual items.

(g) Error Handling deals with error conditions.

- **Device management**: The OS closely manages the I/O job of the computer system. The processor transmits signals which are decoded by the control unit and transmitted to and recognized by the IO devices. Each running process generates the I/O requests which are required to be served by the system. There are differences in the operational speed of the processor and I/O devices. This calls for buffering and blocking of each file which the operating system takes care of. Indeed, the weakest links in the computer system are its I/O subsystems, which can cause errors. Using hardware and software mechanisms, the operating system detects and corrects errors when they occur.
- **File management**: This function is incorporated in the operating system because of the following reasons:
  (a) To relieve the users of the problem of storing files separately from the computer and thus making retrieval a more straightforward operation.
  (b) To allow two or more user to share the same file.

To achieve these objectives, the operating systems are designed to implement policies which provide protection form loss or corrupted files etc. Apart from that, it also provides efficient allocation of space in the secondary storage devices and file sharing besides arranging user computer interface by the use of symbolic file names and attributes.

- **System management**: This may be taken as consisting of the following main sub-systems:
  (a) Security
  (b) Job accounting
  (c) Overall system performance
  (d) Interaction with operators and logging thereof
  - **Security:** This capability includes provision of safeguards against unauthorized access to programs and data resident in the memory, performing validation of inputs, taking action to resolve errors, etc. These measures impose restrictions even on unauthorized users, thereby minimizing the chances of destruction/corruption of data or services.
  - **Job accounting:** This capability implies keeping track of time and resources used by various hobs/users of the system. The accounting of the resources includes keeping count of the following:
    (a) Total system time elapsed

NOTES

(b) Total CPU time elapsed
(c) Total connect time
(d) Units of time for accessing disks
(e) Units of disk storage used
(f) Units of main storage used, i.e., time occupancy
(g) Disk storage when not Online or logged in
(h) I/O transaction time
(i) Time elapsed in reading/writing records

- **Overall system performance**: This includes recording delays between requesting a service and receiving response from the system. Included in system performance is also the recording of percentage of time that the resources like CPU or disk drives were in actual use in relation to the total available time.
- **Interaction with operators**: Interactions between the computer and the operator take place via a control console and/or a CRT with a keyboard. The interaction takes the form of receipt of and acting on operator instructions, acknowledging the same and indicating the status of work and / or any problems encountered. Included in it are keeping of records of communications between the computer and the operator and also the errors and other abnormal conditions encountered during processing.

It may be noted that only such of the above or allied functions can be recognized to the operating system as are carried out by the programs forming part of the operating system software.

The distinction is in the context that in some computer system these functions can as well as be performed by hardware (ROM). Indeed, in the modern computer systems, the hardware and software are so interrelated that only the respective design engineers can tell with certainty which functions are distantly performed by the hardware and which one by the software.

3.3 TYPES OF SOFTWARE

Software is the program used to make the computer usable in terms of accomplishing the users' tasks. Software consists of programs, routines and procedures that can be run on a computer system. The different categories of software are:

- **System software:** System software makes the computer work and gets it to the state where it is ready to run our programs. It consists of all the operating systems, translators and languages to ensure smooth and efficient functioning of the computer system. These programs provide the integration of various components of a computer to work together and provide the way to manipulate these resources according to the need and choice of the user.

NOTES

- **Application software:** Application software refers to all the programs that the user uses in order to complete a particular task. This software allows the user to utilize computers for the tasks which are provided by the software itself, such as data manipulation, documentation, image, and multimedia development and usage, etc. It ranges from playing games to word processing to monitoring the condition of patients in hospitals.
- **Utility packages:** Utility packages are special software which provides a helping hand to the other software in terms of improvements in the systems performances by performing some routine task, such as disk defragmentation, disk scanning and bad sector corrections. The Norton tool kit is one of the utility toolkits that provide a lot of tools to perform tasks that can help in managing and enhancing the systems performances.

System Software

System software comprises different programs that are use to control the computer system and aid a programmer in doing his work efficiently. It consists of all the operating systems, translators and languages to ensure smooth and efficient functioning of the computer system. System software makes very efficient use of the hardware resources. System software can be further categorized as:

- Operating system and controlling programs
- Translators, such as assemblers, compliers and interpreters
- Database management systems (DBMS)
- Drivers and system testing tools

Operating Systems

It is an integrated set of specialized programs that are used to control and manage the resources and overall operations of a computer. Operating system (OS) controls the execution of other computer programs. It provides scheduling, debugging, I/O control, compilation, storage assignment, data management and other related services to all other programs running on the computer. Some examples of popular OS are: MS-DOS (Disk Operating System), OS2, UNIX, Linux, MS-Windows 95/98, MS-Windows NT, MS-Windows XP and MS-Windows Vista.

Translators

Programs that translate a program written in any computer language into machine understandable code are known as translators. Translators are further divided into three categories. These categories are explained in the following:

Assemblers

An assembly language program cannot be directly executed by a computer. To interpret and execute an assembly language program, it has to be converted into its machine language equivalent. An assembler is a program that translates a program written in assembly language into a machine executable code.

An input to the assembler program is termed as *source program* and the output of assembler is a machine language program termed as *object program*. Once an object program is formed, it is translated to the computer's primary memory using the systems loader.

NOTES

Compiler

Compilers are the translators that translate high-level language programs into machine code, and this machine code is executed afterwards. The translated machine code is known as object code. Functions performed by compliers are:

- They allocate addresses for all variables and statements.
- They generate the object program on tape or disk as required.
- They produce a printed listing of the source and object programs, if required.
- They tabulate a list of programming errors found during compilation.

Interpreters

These are used for translating a high-level language program into a machine-level language program line by line and each of the instructions alternately. When a program is to be executed, the interpreter accesses the first interaction, translates it into one or more machine language instructions, and then executes these instructions. The interpreter then accesses the next instruction and repeats these tasks. The process continues until all source language instructions have been translated and executed. The most well known interpreter-based language is BASIC.

Functions of Interpreters

- Interpreters are slow as compared to compilers because of line by line translation and execution.
- Debugging is easier in interpreter-based languages because the interpretation process stops as soon as an error is encountered.

Drivers

A driver is a special type of software created by peripheral device manufacturers to provide facility to the computer to communicate with the peripheral devices, e.g. printer drivers, mouse drivers, etc. Drivers have the ability to convert the data supplied by the computer according to the device setting and than transfer to the device to work over it.

Application Software

Application software enables the computer to produce some useful output, such as specific inventory control reports, attendance accounting, linear programming or medical counting, etc. It is subdivided into three categories: general purpose, special purpose and bespoke (common).

NOTES

General Purpose Software

General purpose or generic software refers to applications that have a number of common uses among a range of different users. The most common types of general purpose software being used in business are:

1. **Word processing applications** are used like typewriters to produce letters and other text based documents including letters, essays, legal documents, contracts, etc.
2. **Spreadsheet applications** are used to handle numbers and calculations, for example, a company accounts, sales figures, etc.
3. **Presentation applications** are used to make multimedia presentations usually to be displayed as a slide show.
4. **Desktop publishing applications** are used to create pages for publications that required more complex layout, for example, magazines, newspapers and brochures.
5. **Database applications** are used to store and search information for a particular purpose, for example, customer records, dental records, etc.
6. **Programming applications** are used for writing computer programs, such as employee attendance, record keeping system, etc.
7. **Multimedia applications** are used to create and display any kind of pictorial information, for example, line drawings, paint packages and animations.

Special Purpose Software

These are the generalized set of programs used to deal with a particular application. The software is normally developed by specialist software developers to solve common problems faced by many users, e.g., MS-Office, WordStar, Lotus and Tally, etc.

Bespoke (Customized) Software

Bespoke (customized) application software is programs written by a use or a programmer in order to perform specific tasks for the user. They are written in a variety of programming languages depending on the task in hand. Normally, these are sets of programs used in conjunction with one another, e.g., payroll system, customized accounting packages for a company, etc.

Check Your Progress

1. What is an assembler?
2. Define drivers.

3.4 SUMMARY

NOTES

- An OS is a program that acts like an interface between a user of a computer and the computer hardware. The purpose of the operating system is to provide an environment to the user in which he/she can execute programs.
- Systems programs are the programs that are written for direct execution on computer hardware in order to make the power of the computer fully and efficiently accessible to the applications programmers and other computer users.
- The various tasks that are generally performed by the operating system are:
  (i) Storage management
  (ii) Processing management
  (iii) Device management
  (iv) File management
- Software is the program used to make the computer usable in terms of accomplishing the users' tasks. Software consists of programs, routines and procedures that can be run on a computer system.
- Application software refers to all the programs that the user uses in order to complete a particular task.
- Utility packages are special software which provides a helping hand to the other software in terms of improvements in the systems performances by performing some routine task, such as disk defragmentation, disk scanning and bad sector corrections.
- General purpose or generic software refers to applications that have a number of common uses among a range of different users.

3.5 KEY TERMS

- **Operating system:** A program that acts like an interface between the user of a computer and the computer hardware.
- **Software:** Programs used to make the computer usable in terms of accomplishing users' tasks.
- **System software:** It consists of all the operating systems, translators and languages to ensure smooth and efficient functioning of the computer system.
- **Application software:** It refers to all the programs that the user uses in order to complete a particular task.

3.6 ANSWERS TO 'CHECK YOUR PROGRESS'

NOTES

1. An assembler is a program that translates a program written in assembly language into a machine executable code.
2. A driver is a special type of software created by peripheral device manufacturers to provide facility to the computer to communicate with the peripheral devices, e.g. printer drivers, mouse drivers, etc.

3.7 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define operating system.
2. A computer system is divided into three components. List them.
3. What do you understand by the term software?
4. What are the different categories of software?

Long-Answer Questions

1. Explain how the OS acts like an interface between a user and computer hardware?
2. Explain the tasks that are performed by operating systems.
3. What are general and special purpose software? Discuss in detail.

3.8 FURTHER READING

William, Brin K. Stacey C and Sawyer. 2007. *Using Information Technology: A Practical Introduction to Computers and Communications.* Ohio, US: McGraw-Hill Irwin.

Long, Larry and Nancy Long. 2004. *Computers: Information Technology in Perspective*. Upper Saddle River, New Jersey: Prentice Hall.

Sinha, P. K. 2004. *Computer Fundamentals: Concepts Systems and Applications*. New Delhi: BPB.

Norton, Peter. 2008. *Introduction to Computers*. Ohio, US: Tata McGraw.

Basandra, Suresh K. 2007. *Computers Today*. New Delhi: Galgotia.

UNIT 4 I/O DEVICES AND MEMORY

NOTES

Structure

4.0 INTRODUCTION

In this unit, you will learn in detail about different types of input and output devices. Information to the CPU is sent through an input device to carry out several functions. A computer is capable of using several types of input devices so that every user gets a unique experience. Keyboard, mouse, webcam, scanner, microphone, etc., can be grouped as input devices. An input device is any peripheral (piece of computer hardware) equipment used to provide data and control signals to the computer.

Output is information produced by the computer program and perceived by the user. In simple words, output is what you get from the computer. This can be information on a screen, sound, images, or a printed page. The output is given to us through various devices which are called output devices. An output device is a computer hardware equipment used to communicate the results of data processing carried out by an information processing system (such as a computer).

Further in this unit, you will learn about data storage in a computer system. The most commonly used form of auxiliary storage involves a kind of magnetic disk. It comes in various sizes and stores data on a magnetic surface. Magnetic disk is known for its high storage capacity and reliability. It also facilitates direct access to data. This unit will also discuss data storage devices like hard disks, diskette and optical disks.

4.1 UNIT OBJECTIVES

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

- Identify the different types of devices used for inputting information in a computer
- Describe the various types of output devices used to display the result of an operation
- Describe the various storage devices used in computers

NOTES

4.2 INPUT DEVICES

An input device accepts data from the outside world and transforms it into a form the computer can interpret. Keyboards are the most commonly used input devices. Point-and-draw devices are used to point to, and select menu items or icons displayed on screen. They provide a means for graphical user interface (GUI). The mouse, trackball, joystick, light pen, and touch screen are commonly used point-and-draw devices. However, scanning devices that provide direct data input from source documents are of two types —CONTACT and LASER. Electronic card readers read the data encoded on electronic cards and convert it to machine-readable form, for further processing. This unit also describes other voice recognition devices that input data in the form of human voice, thereby providing an easy means of data input.

Consider the following example: The average marks of a student need to be calculated based on his marks obtained in various subjects. The marks would typically be available in the form of a document containing the student's name, roll number, and marks scored in each subject. This data must be first stored in the computer's memory after converting it into machine-readable form. The data will then be processed (average marks calculated) and sent from the memory to the output unit, which will present the data in a form that can be read by users.

Figure 4.1 shows the role of I/O devices in a computer system.

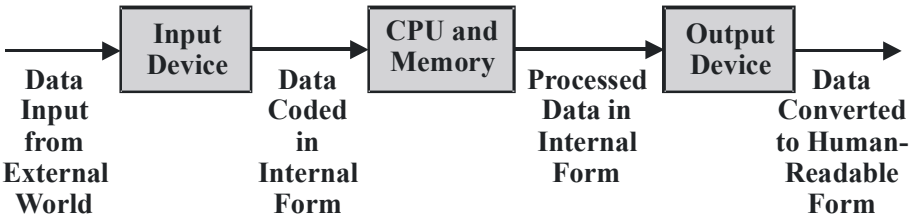


Fig. 4.1 *Roles of I/O Devices*

The I/O devices that provide a means of communication between the computer and the outside world are known as peripheral devices. This is because they surround the CPU and the memory of a computer system. While input devices are used to enter data from the outside world into the primary storage, output devices are used to provide the processed results from the primary storage to the users.

As mentioned earlier in this section, input devices are used to transfer user data and instructions to the computer. The most commonly used input devices can be classified into the following categories:

- Keyboard devices (general and special purpose, key-to-tape, key-to-disk, key-to-diskette)
- Point-and-draw devices (mouse, trackball, joystick, light pen, touch screen)
- Scanning devices (optical mark recognition, magnetic ink character recognition, optical bar code reader, digitizer, electronic-card reader)
- Voice recognition devices
- Vision-input devices (webcam, video camera)

NOTES

Keyboard Devices

Keyboard devices allow input into the computer system by pressing a set of keys which are mounted on a board connected to the computer system. Keyboard devices are typically classified into general-purpose keyboards and special-purpose keyboards.

General-purpose keyboard

The most familiar means of entering information into a computer is through a typewriter-like keyboard that allows a person to enter alphanumeric information directly.

The most popular keyboard used today is the 101-keys with a traditional QWERTY layout, having an alphanumeric keypad, twelve function keys, a variety of special-function keys, numeric keypad, and dedicated cursor-control keys. It is so called because of the arrangement of its alphanumeric keys in the upper left row. Figure 4.2 shows QWERTY keyboard layout.

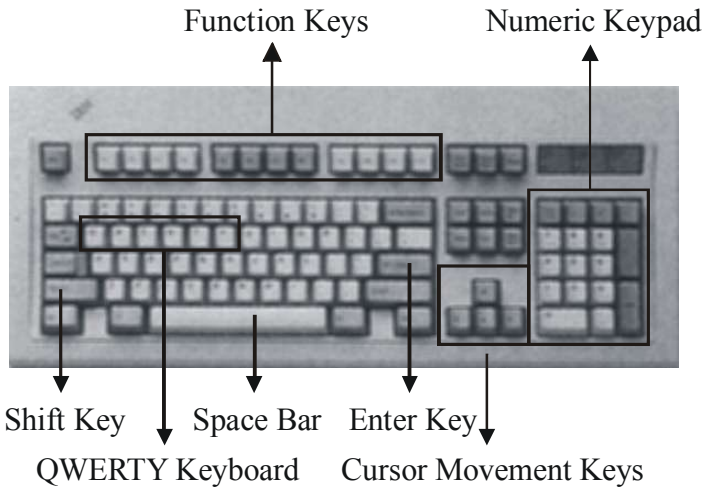


Fig. 4.2 *QWERTY Keyboard Layout*

- **Alphanumeric keypad**: Contains keys for the English alphabets, numbers, 0 to 9, and special characters * + – / [ ].
- **12 function keys**: These are keys labelled F1, F2 ... F12 and are a set of user-programmable function keys. The actual function assigned to a function key differs from one software package to another. These keys are also called soft keys since their functionality can be defined by the software.
- **Special-function keys**: Have special functions assigned to each of these keys. For example, the enter key is used to send the keyed-in data into the memory. Other special keys include:
    - **Shift** – used to enter capital letters or special characters defined above the number keys.
    - **Spacebar** – used to enter a space at the cursor location.

NOTES

- **Ctrl** – used in conjunction with other keys to provide added functionality on the keyboard.
- **Alt** – like Ctrl, it is used to expand the functionality of the keyboard.
- **Tab** – used to move the cursor to the next tab position defined.
- **Backspace** – used to move the cursor one position to the left and also delete the character in that position.
- **Caps Lock** – to toggle between the capital letter lock feature – when 'on', it locks the keypad for capital letters input.
- **Num Lock** – to toggle the number lock feature on and off – when 'on', it inputs numbers when you press the numbers on the numeric keypad.
- **Insert** – used to toggle between the insert and the overwrite mode during data entry – when 'on', entered text is inserted at the cursor location.
- **Delete** – used to delete the character at the cursor location.
- **Home** – used to move the cursor to the beginning of the work area which could be the line, screen or document depending on the software being used.
- **End** – used to move the cursor to the end of the work area.
- **Page Up** – used to display the previous page of the document being currently viewed on screen.
- **Page Down** – used to view the next page of the document being currently viewed on screen.
- **Escape** – usually used to negate the current command.
- **Print Screen** – used to print what is being currently displayed on the screen.
- **Numeric keypad** consists of keys having numbers (0 to 9) and mathematical operators (+ – * /) defined on them. It is usually located on the right side of the keyboard and supports quick entry of numerical data.
- **Cursor-control keys** defined by the arrow keys used to move the cursor in the direction indicated by the arrow (top, down, left, right).

Another well-known key arrangement is the Dvorak system, which was designed to be easier to learn and use. The Dvorak keyboard has the most common consonants on one side of the middle or the home row and the vowels on the other side, so that typing tends to alternate keystrokes back and forth between hands. Although the Dvorak keyboard has never been widely used, it has its adherents.

Special-purpose keyboard

These are standalone data entry systems used for computers deployed for specific applications. These typically have special purpose keyboards to enable faster data entry. A very typical example of such keyboards can be seen at the automatic teller machines or the ATMs, where the keyboard is required for limited functionality (support for some financial transactions) by the customers. Point-of-sale or POS terminals at fast food joints, air/railway reservation counters are some other examples of special-purpose keyboards. These keyboards are specifically designed only for special types of applications.

NOTES

Key-to-tape, key-to-disk, key-to-diskette

These are standalone data entry stations. These units usually have a small processor attached to a keyboard and a visual display unit. The processor checks for the accuracy of data at the time of entry. The screen displays data as it is being entered. These facilities are very useful and desirable during mass data entry and are therefore becoming very popular in data processing centres.

Point-and-Draw Devices

The keyboard facilitates input of data only in the text form. While working with display based packages, we usually point to a display area and select an option from the screen (fundamentals of GUI applications). For such cases, the sheer user-friendliness of input devices that can rapidly point to a particular option displayed on screen and support its selection, resulted in the advent of various point-and-draw devices.

Mouse

A mouse is a small device that a computer user pushes across a desk surface in order to point to a place on a display screen and to select one or more actions possible from that position. The mouse first became a widely used computer tool when Apple Computer made it a standard part of the Apple Macintosh. Today, the mouse is an integral part of the graphical user interface (GUI) of any personal computer. The mouse apparently got its name by being about the same size and colour as a toy mouse.

Figure 4.3 shows a mouse.

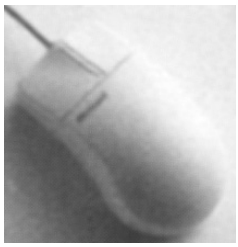


Fig. 4.3 *A Mouse*

NOTES

The most conventional type of mouse has two buttons on top: the left one is used most frequently. In the windows operating systems, it lets the user click once to send a 'Select' indication that provides the user with feedback that a particular position has been selected for further action. The next click on a selected position or two quick clicks on it causes a particular action to take place on the selected object. For example, in Windows operating systems, it causes a program associated with that object to be started. The second button, on the right, usually provides some less-frequently needed capability. For example, when viewing a Web page, you can click on an image to get a pop-up menu that, among other things, lets you save the image on your hard disk. Some models have a third button for additional capabilities. Some mouse manufacturers also provide a version for left-handed people.

Trackball

The trackball is a pointing device that is much like an inverted mouse. It consists of a ball inset in a small external box, or adjacent to–and in the same unit as–the keyboard of some portable computers. Figure 4.4 shows a trackball.

Fig. 4.4 *A Trackball*

It is more convenient and requires much less space than the mouse since here the whole device is not moved (as in the case of a mouse). Trackball comes in various shapes but supports the same functionality. Typical shapes used are a ball, a square, and a button (typically seen in laptops).

Joystick

The joystick is a vertical stick that moves the graphic cursor in the direction the stick is moved. It consists of a spherical ball which moves within a socket, and has a stick mounted on it. The user moves the ball with the help of the stick that can be moved left or right, forward or backward, to move and position the cursor in the desired location. Joysticks typically have a button on top that is used to select the option pointed by the cursor. Video games, training simulators, and control panels of robots are some common uses of a joystick. The following figure shows a joystick:

Fig. 4.5 *A Joystick*

Light pen

The light pen is a pen-shaped device which allows natural movement on the screen. It is made up of a light sensitive cell and a lens assembly designed in such a way that it focuses onto itself any light in its field of view. The pen contains a light receptor and is activated by pressing the pen against the display screen. The receptor is the scanning beam that helps in locating the pen's position (X and Y coordinates on the screen). Suitable system software is provided to initiate the desired action once the area on the display screen is located with the help of the light pen. Light pens are typically used in CAD (Computer Aided Design) applications to directly draw on screen. The following figure shows a light pen:

NOTES

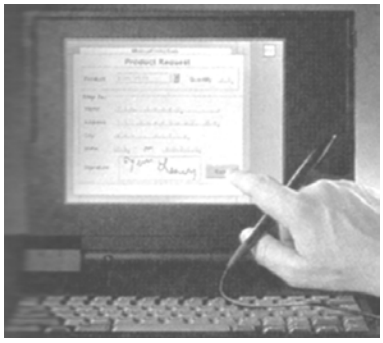


Fig. 4.6 *A Light Pen*

Touch screen

A touch screen is probably one of the simplest and most intuitive of all input devices. It uses optical sensors in or near the computer screen which can detect the touch of a finger on the screen. Once the user touches the screen at a particular position, sensors communicate the position to the computer. This is then interpreted by the computer to understand the user's choice for input. The most common usage of touch screens is in information kiosks, where users can receive information at the touch of a screen. These devices have become very popular today.

Scanning Devices

Scanning devices are input devices used for direct data entry from the source document into the computer system. Scanners facilitate capturing of information and storing it in a graphical format for displaying it back on the graphical screen. They consist of two components, one to illuminate the page so that the optical image can be captured and the other to convert the graphical image into a digital format for storing. The graphical images thus scanned can be seen and processed directly by the computer.

There are two types of scanners, CONTACT and LASER. Both bounce a beam of light off an image, and then measure the reflected light to determine the value of the image. Hand-held contact scanners make contact as they are brushed over the printed matter to be read. Laser-based scanners are more versatile and can read data passed near the scanning area. Hand-held scanners are used where the information to be scanned or the volume of documents to be scanned is very

NOTES

low. They are much cheaper as compared to the flat-bed scanners. The following figures shows hand-held and flatbed scanners:

Fig. 4.7 (a) *Hand-held Scanner*

Fig. 4.7 (b) *Flatbed Scanner*

Capturing information using scanners reduces the possibility of human error typically seen during large data entry. The reduction in human intervention improves the accuracy of data, besides saving time.

Most recent trends for data input are towards source data automation. The equipment used for source data automation, capture data as a by-product of a business activity, thereby completely eliminating manual input of data. Some common examples of these are described as follows:

Optical mark recognition

OMR devices can sense marks on computer readable paper. This type of device is typically used by academic institutions to grade aptitude tests, where candidates need to mark the correct option from a number of alternatives on a special sheet of paper. These answer sheets can be then directly read by the optical mark recognition device and can be used for further processing by the computer.

The actual technique used by an OMR device also involves focusing light on the page being scanned, thereby detecting the reflected light pattern for the marks. Pencil marks made by the user reflect the light determining which responses are marked. The following figure shows an example of a pre-printed answer sheet that can be read by an OMR device:

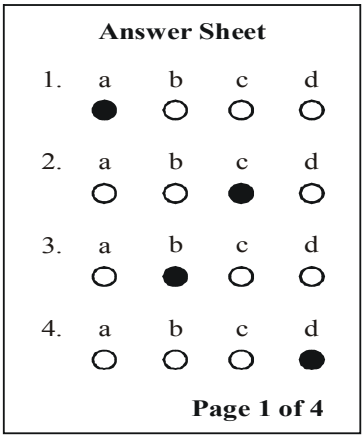


Fig. 4.8 *An Example of a Pre-printed Answer Sheet*

Magnetic ink character recognition (MICR)

NOTES

Magnetic ink character recognition (MICR) is similar to optical mark recognition and is used exclusively by the banking industry. MICR devices are used by the banking industry to read the account numbers on cheques directly and subsequently do the necessary processing.

Banks using the MICR technology print chequebooks on special types of paper. The necessary details of the bank (like the bank's identification code, the relevant account number, and the cheque number) are pre-printed on the cheques using ink that contains iron oxide particles that can be magnetized.

MICR readers are used to read and sort cheques and deposits. An MICR reader-sorter reads the data on the cheques and sorts the cheques for distribution to other banks and customers or for further processing. The following figure shows a bank cheque using MICR technology:

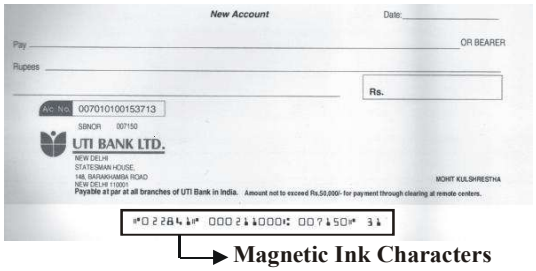


Fig. 4.9 *A Bank Cheque using MICR Technology*

Optical bar code reader (OBR)

Data coded in the form of small vertical lines forms the basis of bar coding. Alphanumeric data is represented using adjacent vertical lines called barcodes. These are of varying widths and spacing between them and is used to uniquely identify books, merchandise in stores, postal packages, etc. The following is an example of a barcode used on one of the books for its unique identification.

A barcode reader uses laser beam technology. The laser beam is moved across the pattern of bars in a bar code. These bars reflect the beam in different ways. The reflected beam is then sensed by a light-sensitive detector, which then converts the light patterns into electrical pulses, thereby transmitting them to logic circuits for further conversion to alphanumeric value.

Barcode devices are available as hand-held devices. The following figures show a barcode reader and a barcode:

NOTES

Fig. 4.10(a) Barcode Reader

Fig. 4.10(b) A Barcode

Digitizer

Digitizers are used to convert drawings or pictures and maps into a digital format for storage in the computer. A digitizer consists of a digitizing or graphics tablet, which is a pressure sensitive tablet, and a pen with the same X and Y coordinates as on the screen. Some digitizing tablets also use a crosshair device instead of a pen. The movement of the pen or the crosshair is reproduced simultaneously on the display screen. When the pen is moved on the tablet, the cursor on the computer's screen moves simultaneously to the corresponding position on the screen (X and Y coordinates). This allows the user to draw sketches directly or input existing sketched drawings easily. Digitizers find most common usage by architects and engineers as a tool for computer aided designing (CAD). The following figure shows a digitizing tablet:

Fig. 4.11 A Digitizing Tablet

Electronic-card reader

Card readers are devices that also allow direct data input into a computer system. The electronic-card reader is connected to a computer system and reads the data encoded on an electronic card and transfers it to the computer system for further processing.

Electronic cards are plastic cards with data encoded on them and meant for a specific application. Typical examples of electronic cards are the plastic cards issued by banks to their customers for use in automatic teller machines or ATMs. Electronic cards are also used by many organizations for controlling access of various types of employees to physically secured areas.

Depending on the manner in which the data is encoded, electronic cards may be either magnetic strip cards or smart cards. Magnetic strip cards have a magnetic strip on the back of the card. Data stored on magnetic strips cannot be read with the naked eye, a useful way to maintain confidential data. The following figure shows an access card security system:

NOTES

Fig. 4.12 *An Access Card Security System*

Smart cards

Smart cards, going a stage further, have a built-in microprocessor chip where data can be permanently stored. They also possess some processing capability making them suitable for a variety of applications. For example, to gain access, an employee inserts a card or a badge in the reader. This device reads and checks the authorization code before permitting the individual to enter a secured area. Since smart cards can hold more information as compared to magnetic strip cards, they are gaining in popularity.

CHECK YOUR PROGRESS

1. What are scanning devices?
2. What are card readers?

4.3 OUTPUT DEVICES

An output device is an electromechanical device that accepts data from the computer and translates it into a form that can be understood by the outside world. The processed data, stored in the memory of the computer, is sent to an output unit, which then transforms the internal representation of data into a form that can be read by the users.

Display Devices

One of the most common and important peripherals in a computer system is the display device. Conventional computers used display terminals known as alphanumeric terminals. These used a form of multi-dot (7×5 or 9×7) array to display characters. These were used to read text information displayed on the screen. The increasing

NOTES

demand for displaying graphs and pictures, for visual presentation of information (more effective for user interaction), brought about the advent of graphic display devices.

Graphic display is typically made up of a series of dots called 'pixels' (picture elements) whose pattern produces the image. Each dot on the screen can be addressed uniquely and directly. Owing to the fact that each dot can be addressed as a separate unit, it provides greater flexibility for drawing pictures. Display screen technology may be one of the following three categories:

Cathode ray tube (CRT)

The main components of a cathode ray terminal are the electron gun, the electron beam controlled by an electromagnetic field, and a display screen, which is phosphor-coated. The screen's phosphor coating is organized into a grid of dots called pixels. An electron gun emits an electron beam, which is directed towards the phosphor-coated display by the electromagnetic field, and this in turn creates the image. There are two types of CRT displays:

- ***Vector CRT display****:* Here, the electron beam is directed only to places where the image is to be created.
- ***Raster scan display****:* In this type, the image is projected on the screen by directing the electron beam across each row of the picture elements from the top to the bottom of the screen. This type of display provides a high dynamic capability since the image is continuously refreshed. It offers full colour display at a relatively low cost and is therefore becoming increasingly popular.

The quality of display is indicated by the resolution of the display device. The number of horizontal and vertical pixels determines the resolution. Typical resolutions in graphic display range from (800 × 600) to (640 × 768) to (1024 × 1024) pixels. Based on the resolution and the number of colours supported, several standards for colour monitors have evolved. The most popular of these include the following:

- Colour graphics adapter (CGA), which has a resolution of (320 × 200) and supports up to sixteen colours.
- Extended graphics adapter (EGA) has a resolution of (640 × 350) and supports up to sixteen colours.
- Video graphics adapter (VGA) has a resolution of (640 × 480) and supports up to 256 colours.
- Super VGA has a resolutions ranging from (800 × 600) to (1280 × 1024) and supports up to 256 or more colours.

Note that each one of these is implemented by installing an add-on card in the computer, commonly known as graphics adapter or the video card. This card is then connected to the appropriate monitor.

Liquid crystal display (LCD)

Introduced in watches and clocks in the 1970s, LCD is now applied in display terminals. In this, the cathode ray tube was replaced by liquid crystal to produce the

image. It does not have colour capability and the image quality is relatively poor. The main advantage of LCD is its low energy consumption.

It finds most common usage in portable devices where compactness and low energy requirements are of prime importance.

NOTES

Monitors

Monitors use a cathode ray tube (CRT) to display information. It resembles a television screen and is similar to it in other respects. The monitor is typically associated with a keyboard for manual input of characters. The screen displays information as it is keyed in, enabling a visual check of input before it is transferred to the computer. It is also used to display the output from the computer and hence, serves as both an input and an output device. The monitor along with the keyboard is called a visual display unit (VDU). This is the most commonly used input/output device and is also known as the soft copy terminal. A printing device is usually required to provide a hard copy of the output.

Printers

Printers are used for producing output on paper. There are a large variety of commercially available printers today (estimated to be 1500 different types). These printers can be classified into categories based on:

- Printing technology
- Printing speed
- Printing quality

Dot matrix printer

Dot matrix printers were the most popular impact printers used in personal computing. These printers use a print head consisting of a series of small pins to strike a ribbon coated with ink, causing the ink to transfer to the paper at the point of impact. Characters thus produced are in a matrix format. The shape of each character, i.e., the dot pattern, is obtained from the information held electronically.

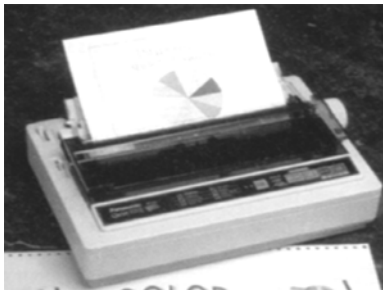


Fig. 4.13(a) *Dot Matrix Printer*

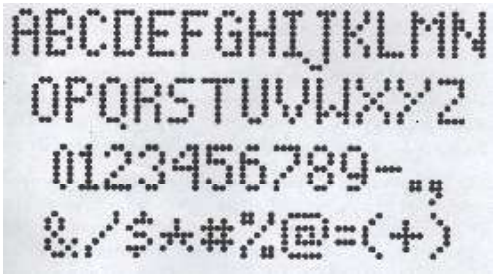


Fig. 4.13(b) *Characters Formed Using Dots*

The speed, versatility and ruggedness, combined with low cost, tend to make such printers particularly attractive in the personal computer market. Typical printing

speeds in case of dot matrix printers range between 40 – 1000 cps (characters-per-second). One major disadvantage of this technology is that the print quality is low.

Inkjet printer

NOTES

Inkjet printers use a series of nozzles to spray drops of ink directly on the paper. These therefore fall under the category of non-impact printers. The printhead of an inkjet printer consists of a number of tiny nozzles that can be selectively heated up in a few microseconds by an IC register. When this happens, the ink near it vapourizes and is ejected through the nozzle to make a dot on the paper placed in front of the printhead. The character is printed by selectively heating the appropriate set of nozzles as the printhead moves horizontally.

If you have ever looked at a piece of paper that has come out of an inkjet printer, you would have noticed the following:

- The dots are extremely small (usually between 50 and 60 microns in diameter), so small that they are tinier than the diameter of a human hair (70 microns).
- The dots are positioned very precisely, with resolutions of up to 1440×720 dots per inch (dpi).
- The dots can have different colours combined together to create photo-quality images.

Inkjet printers are slower than dot-matrix printers (40–300 cps), cheaper to buy but are more expensive in running costs (the ink cartridge cost is considerably higher than that of the DMP ribbon) and are used by people/organizations, where speed of printing is not the most important factor. The following figure shows an inkjet printer:

Fig. 4.14 *Inkjet Printer*

Laser

Laser printers use dry ink (toner), static electricity, and heat to place and bond the ink onto the paper. They use a combination of laser and photocopier technology. Printing is achieved by deflecting laser beam onto the photosensitive surface of a drum after which the latent image attracts the toner to the image. The toner is then electrostatically transferred to the paper and fixed into a permanent image. The following figure shows a laser printer:

NOTES

Fig. 4.15 *Laser Printer*

Laser printers are capable of converting computer output into print, page by page. Since characters are formed by very tiny ink particles, they produce very high quality images (text and graphics), generally offer a wide variety of character fonts, and are silent and fast in use. Laser printers are faster in printing speed than the other printers discussed here. Their speeds can range from ten pages a minute to about 200 pages per minute, depending upon the make/model.

Laser is high quality, high speed, high volume, and non-impact technology that work on plain paper or pre-printed stationery. This technology is relatively expensive but has become very popular because of the quality, speed and noiseless operations.

Plotters

Plotters are used to produce graphical output on paper. It is a device capable of producing charts, drawings, graphics, maps, etc. It is much like a printer but is designed to print graphs instead of alphanumeric characters.

Based on the technology used, plotters may be pen plotters or electrostatic plotters. While pen plotters have an ink pen attached to draw the images, electrostatic plotters work similar to a laser printer. Image is produced by charging the paper with a high voltage. This voltage attracts the toner, which is then melted on the paper with heat. Electrostatic plotters are fast, but the quality is generally considered to be poor when compared to pen plotters. This is why pen plotters are more extensively used as compared to electrostatic plotters. Flatbed plotters and drum plotters constitute the most commonly used plotters.

Flatbed plotters

Flatbed plotters have a flat base like a drawing board on which the paper is laid (as shown in Figure 4.16(a). One or more arms, each of them carrying an ink pen, moves across the paper to draw. The arm movement is controlled by a microprocessor (chip). The arm can move in two directions, one parallel to the plotter and the other perpendicular to it (called the x and y directions). With this type of movement, it can move very precisely to any point on the paper placed below.

NOTES

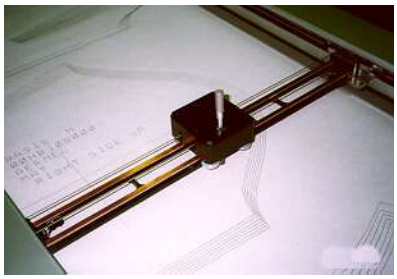


Fig. 4.16(a) *Top View of a Flatbed Plotter*

The computer sends the commands to the plotter which are translated into x and y movements. The arm moves in very small steps to produce continuous and smooth graphics. The size of the plot in a flatbed plotter is limited only by the size of the plotter's bed.

The advantage of flatbed plotters is that the user can easily control the graphics. He can manually pick up the arm anytime during the production of graphics and place it on any position on the paper to alter the position of graphics to his choice. The disadvantage here is that flatbed plotters occupy a large amount of space.

Drum plotters

Drum plotters use a drum revolver to move the paper during printing (as shown in Figure 4.16(b)). The arm carrying a pen moves only in one direction, perpendicular to the direction of motion of the paper. Thus, in drum plotters the pen is moved in a single axis track and the paper itself moves on a cylindrical drum to add the other axis or dimension. The combination of the pen and paper movement creates the graphics.

The size of the graph is therefore limited only by the width of the drum and can be of any length. Drum plotters are very compact and lightweight as compared to flatbed plotters. This is one of the advantages of such plotters. The disadvantage, however, is that the user cannot freely control the graphics when they are being created.

Fig. 4.16(b) *Drum Plotter*

Plotters are more expensive when compared to printers. Typical application areas for plotters include: CAE (computer-aided engineering) applications like CAD (computer-aided design) and CAM (computer-aided manufacturing), architectural drawing and map drawing.

NOTES

CHECK YOUR PROGRESS

3. List the two types of CRT displays.
4. What are plotters?

4.4 STORAGE DEVICES

The most common properties used for characterizing and evaluating the storage unit of the computer system are the following:

1. **Storage Capacity:** Represents the size of the memory. It is the amount of data that can be stored in the storage unit. Primary storage units have less storage capacity as compared to secondary storage units. While the capacity of internal memory and main memory can be expressed in terms of the number of words or bytes, the capacity of external or secondary storage is measured in terms of bytes.
2. **Storage Cost:** Another key factor that is of prime concern in a memory system is cost. It is normally expressed per bit. It is obvious that lower costs are desirable. It is worth noting that as the access time for memories increases, the cost decreases.
3. **Access Time:** The time required to locate and retrieve the data from the storage unit. It is dependant on the physical characteristics and the access mode used for that device.

   Primary storage units have faster access time as compared to secondary storage units.
4. **Access Mode:** Memory is considered to be consisting of various memory locations. Access mode refers to the mode in which information is accessed from the memory. Memory devices can be accessed in any of the following ways:
   (a) *Random access memory (RAM):* It is the mode in which any memory location can be accessed in any order in the same amount of time. Ferrite and semiconductor memories, which generally constitute the primary storage or main memory, are of this nature.
   (b) *Sequential access:* Memories that can be accessed only in a predefined sequence are sequential access memories. Since sequencing through other locations precedes the arrival at a desired location, the access time varies according to the location. Information on a sequential device can be retrieved in the same sequence in which it was stored. Songs

NOTES

stored on a cassette, that can be accessed only one by one, is an example of sequential access. Typically, magnetic tapes are sequential access memory.

(c) *Direct access:* In certain cases, the information is neither accessed randomly nor in sequence but something in between. In this type of access, a separate read/write head exists for each track, and on a track the information can be accessed serially. This semi-random mode of access exists in magnetic disks.

5. **Permanence of Storage:** If the storage unit can retain the data even after the power is turned off or interrupted, it is termed as non-volatile storage. Whereas, if the data is lost once the power is turned off or interrupted, it is called volatile storage. It is obvious from these properties that the primary storage units of the computer systems are volatile, while the secondary storage units are non-volatile. A non-volatile storage is definitely more desirable and feasible for storage of large volumes of data.

Memory Capacity

Capacity, in a computer system, is defined in terms of the number of bytes that it can store in its main memory. This is usually stated in terms of kilobytes (kB). The rapidly increasing memory capacity of computer systems has resulted in defining the capacity in terms of Gigabytes (GB) which is 1024 MB (1,07,37,41,824 bytes).

Thus, a computer system with a memory of 256 MB is capable of storing (256 × 1024 × 1024) 26,84,35,456 bytes or characters.

Main Memory

Static and Dynamic RAM

The main memory is the central storage unit in a computer system. It is a relatively large and fast memory and is used to store programs and data during computer operations. The principal technology used for the main memory is based on semiconductor integrated circuits. Integrated circuit RAM chips are available in two possible modes, static and dynamic.

The static RAM (SRAM) stores binary information using clocked sequential circuits. The stored information remains valid only as long as power is applied to the unit. On the other hand, dynamic RAM (DRAM) stores binary information in the form of electric charges that are applied to capacitors inside the chip. The stored charge on the capacitors tends to discharge with time and so must be periodically recharged by refreshing the dynamic memory. The dynamic RAM offers larger storage capacity and reduced power consumption. Therefore, large memories use dynamic RAM, while static RAM is mainly used for specialized applications.

The different types of memory discussed here are both of the read/write type. What about a memory where only one of the operations is possible, e.g., if we

allow only reading from the memory (cannot change the information in the memory)? The memory might have some major importance, like an important bit of the computer's operating system which normally does not change, can be stored in this type of memory. Such a memory is called ROM (Read Only Memory).

NOTES

Read Only Memory (ROM)

Most of the memory in a general-purpose computer is made of RAM integrated circuit chips, but a portion of the memory may be constructed using ROM chips. Originally, RAM was used to refer to random access memory, but now we use the term read/write memory to distinguish it from read only memory (since ROM is also random access). RAM is used for storing bulk of the programs and data that are subject to change, while ROM is used to store programs that are permanently resident in the computer and do not change once the production of the computer is completed.

Among other things, the ROM portion of the main memory is used for storing an initial program called the bootstrap loader. The bootstrap loader is a program whose function is to start operating the computer software when power is turned on. Since RAM is volatile, its contents are destroyed when power is turned off. The contents of ROM remain unchanged even after the power is turned off and on again.

Read only memories can be manufacturer-programmed or user-programmed. When the data is burnt into the circuitry of the computer by the manufacturer, it is called manufacturer-programmed ROM. For example, a personal computer manufacturer may store the boot program permanently in the ROM chip of the computers manufactured by it. Note that such chips are supplied by the manufacturer and are not modifiable by users. This is an inflexible process and requires mass production. Therefore, a new type of ROM called PROM (Programmable Read only Memory) was designed. This is also non-volatile in nature and can be written only once using some special equipment. The writing process in PROM can be performed electrically by the supplier or the customer.

In both ROM and PROM, the write operation can be performed only once and whatever is written cannot be changed. But what about the cases where you mostly read but also write a few times? Another type of memory chip called EPROM (Erasable Programmable Read only Memory) was developed to take care of such situations. EPROMs are typically used by R&D personnel who experiment by changing micro-programs on the computer system to test their efficiency.

Further, EPROM chips are of two types: EEPROMs (Electrically EPROM) in which high voltage electric pulses are used to erase stored information, and UVEPROM (Ultra Violet EPROM) in which stored information is erased by exposing the chip for some time to ultraviolet light.

NOTES

The following figure summarizes the various types of random access memories.

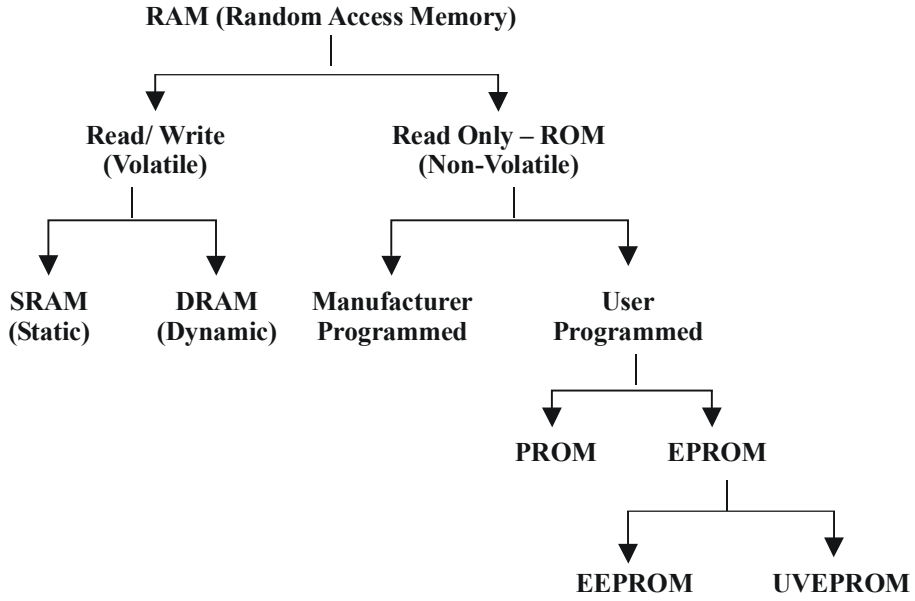


Fig. 4.17 *Types of Random Access Memories*

Cache Memory

Cache memories are small, fast memories placed between the CPU and the main memory. They are faster than the main memory with access times closer to the speed of the CPU. Although caches are fast, they are also very expensive and so are used only in small quantities. For example, caches of size 64K, 128K are normally used in PC-386 and PC-486, where can have 1 to 8 MB of RAM or even more. Cache memories are therefore intended to provide fast speed memory retrieval without sacrificing the size of the memory.

If the memory is so small, how can it be advantageous in increasing the overall speed of memory? The answer to this lies in the phenomenon known as locality of reference. Let us examine what this means.

Locality of reference. Analysis of a large number of typical programs has shown that memory references at any given interval of time tend to be confined to a few localized areas in the memory. This phenomenon is known as the property of locality of reference. This is true because most of the programs typically contain iterative loops (like 'for' or 'while' loops). During the execution of such programs, the same set of instructions (within the loop) are executed many times. The CPU repeatedly refers to the set of instructions in the memory that constitute the loop. Everytime a specific subroutine is called, its set of instructions is fetched from the memory. Thus, loops and subroutines tend to localize the references to memory for fetching instructions.

NOTES

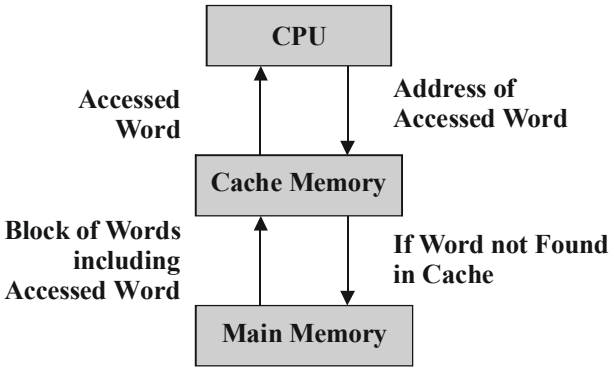


Fig. 4.18 *Functioning of the Cache Memory*

The figure explains the function of the cache memory.

Based on the locality of reference, we understand that the cache contains a copy of certain portions of main memory. The memory read or write operation is first checked with the cache and if the desired data is available in the cache it is used by the CPU directly. Otherwise, a block of words is read from main memory to cache and the word is then used by the CPU from cache.

Secondary Storage Devices

As discussed earlier, RAM is a volatile memory having limited storage capacity. The cost of RAM is also relatively higher as compared to secondary memory. Logic dictates that a relatively cheaper medium, showing some sort of permanence of storage, be used. As a result, additional memory called *external* or *auxiliary memory* or *secondary storage* is used in most computers.

The magnetic medium was found to be long lasting and fairly inexpensive, therefore, became an ideal choice for large storage requirements. Magnetic tapes and disks are commonly used as storage media. With the advancements in optical technology, optical disks are making inroads as one of the major secondary storage devices. The characteristics of all these are discussed in detail in this section.

Magnetic Tapes

Magnetic tapes are used for storing files of data that are sequentially accessed or not used very often and are stored offline. They are typically used as backup storage for archiving of data.

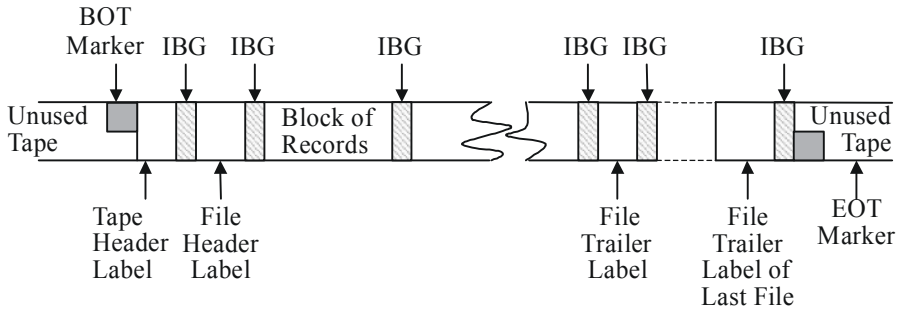


Fig. 4.19(a) *Data Organization on a Magnetic Tape*

NOTES

In case of magnetic tapes, a tape (plastic ribbon usually 1/2 inch or 1/4 inch wide and 50 to 2400 feet long) is wound on a spool and its other end is threaded manually on a take-up spool. The beginning of the tape (BOT) is indicated by a metal foil called a *marker*. When a write command is given, a block of data (records are usually grouped in blocks of two or more) is written on the tape. The next block is then written after a gap (called Inter Block Gap or IBG). A series of blocks are written in this manner. The end of tape (EOT) is indicated by an end-of-tape marker which is a metal foil stuck in the tape. After the data is written, the tape is rewound and kept ready for reading.

Fig. 4.19(b) *Magnetic Tape Reel*

Fig. 4.19(c) *Magnetic Tape Cartridge*

The tape is read sequentially, i.e., data can be read in the order in which the data has been written. This implies that if the desired record is at the end of the tape, all the earlier records have to be read before it is reached. A typical example of a tape can be seen in a music tape cassette where to listen to the fifth song one must listen to, or traverse, the earlier four songs. The access time of information stored on tape is therefore, very high as compared to that stored on a disk.

The storage capacity of the tape depends on its data recording density and the length of the tape. Data recording density refers to the amount of data that can be stored or the number of bytes that can be stored per linear inch of tape. The data recording density is measured in BPI (Bytes per inch).

Thus,

Storage capacity of a tape = Data recording density × Length of tape

It is worth noting that the actual storage capacity for storing user data, is much less owing to the file header labels, file trailer labels, BOT and EOT markers, and the use of IBGs.

Some commonly used magnetic tapes are the following:

- 1/2 inch tape reel
- 1/2 inch tape cartridge
- 1/4 inch streamer tape
- 4 mm DAT (Digital Audio Tape) – typical capacity of 4GB to 14 GB

Magnetic Disks

NOTES

Magnetic disks are direct-access medium and hence they are the most popular online secondary storage devices. Direct-access devices are also called random-access devices because information is literally available at random or in any order. Access to any location on the device is direct and so approximately equal access time is required for each location. An example of this is a music CD, where if you wish to listen to the fifth song, you can directly select the fifth track. It does not require you to fast forward the previous four songs.

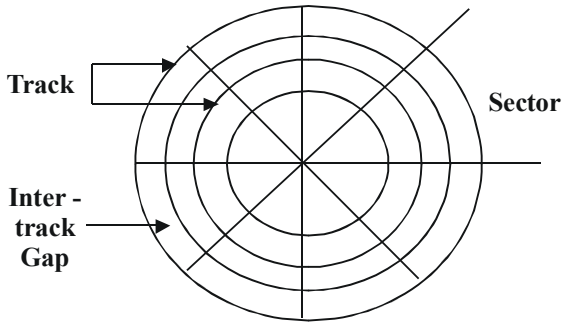


Fig. 4.20 *Logical Layout of a Magnetic Disk*

A magnetic disk is a circular plate made of metal or plastic, coated with magnetized material. Often both sides of the disk are used. Data is recorded on the disk in the form of magnetized and non-magnetized spots (not visible to the naked eye) representing 1s and 0 s.

Data is stored in concentric rings or tracks. To minimize the interference of magnetic fields, the adjacent tracks are separated by inter-track gaps. Tracks are commonly divided into sections called sectors. In most systems, the minimum quantity of information that can be transferred is a sector. Usually, eight or more sectors per track are found.

A track in a given sector near the circumference is longer than the track near the centre of the disk. If bits are recorded with equal density, some tracks would contain more bits than the other tracks. To ensure that each sector can store equal amounts of data, some disks use variable recording density with higher density on tracks near the centre than on tracks near the circumference.

Multiple disks are usually stacked and used together to create disk storage systems having large capacities. In this case, multiple disks are fixed on a central shaft, one below the other to form a disk pack. This is then mounted on a disk drive that has a motor which rotates the disk pack about its axis. The disk drive also has an access arm assembly with a separate read/write head for each surface of the disk pack. The access arms for all the disk surfaces move together. A disk system, is thus addressed by the disk number, the disk surface, the sector number and the track within the sector.

NOTES

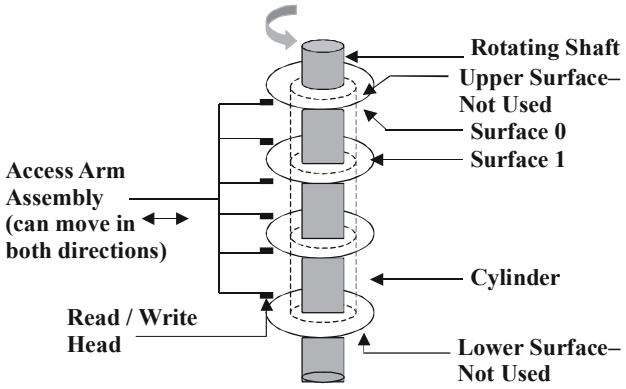


Fig. 4.21 *A Disk Pack with Four Disks*

Usually, the upper surface of the topmost disk and the lower surface of the bottom-most disk are not used since these are prone to getting scratched easily. For faster access of data from disk packs, a concept called cylinders is used. As can be seen in Figure 4.21, a set of corresponding tracks on all the recording surfaces of the disk pack together form a cylinder. Thus, if there are 100 tracks on each disk surface, there are 100 cylinders in the disk pack.

Cylinder-based organization provides faster data access. The related records of a file can be stored on the same cylinder (on multiple disks of a disk pack) and subsequently with one movement of the access arm, all records on, say cylinder 5, (fifth track of every recording surface) can be simultaneously read.

The storage capacity of a disk system can be determined as follows:

Storage capacity = Number of recording surfaces × Number of tracks per surface × Number of sectors per track × Number of bytes per sector

Example: Consider that a disk pack consists of 4 plates each having 2655 tracks with 125 sectors per track. Also, each sector can store 512 bytes. Then,

Storage capacity = $6 \times 2655 \times 125 \times 512 = 1,01,95,20,000$ bytes = 1×10^9 bytes approximately or 1 GB or 1 Gigabyte.

Note: We have six recording surfaces since there are four disk plates.

Access time on disks

As detailed earlier, the disk address is specified in terms of the surface number, the track or cylinder number, and the sector number. The read/write heads need to be first positioned on the track on which the data is to be recorded or from which data needs to be read. Information is always written from the beginning of a sector and can be read only from the beginning of the desired track. Thus, the disk access time depends on the following factors:

- ***Seek time:*** The time taken to position the head on a specific track. The seek time would vary depending upon the position of the access arms at

NOTES

the time the read/write command was received, i.e., if the access arm was positioned on the outermost track and the current read operation required it to be positioned on the fifth track, then the time taken to position the access arm on track 5 is the seek time. It is obvious from this example that moving from the outermost to the innermost track or vice versa would result in the maximum seek time. The average seek time in most systems is 10–100 milliseconds.

- ***Latency time:*** The time required by the desired sector to be positioned under the read/write head, i.e., the time required to spin the desired sector under the head is called latency. Latency is also known as rotational delay and varies depending on the distance of the desired sector from the initial position of the head on the specified track. The rotational speed of a disk is measured in rotations per minute (rpm) and can be anywhere between 300 to 7200 rpm. On an average, latency is equal to half the time taken for a rotation by the disk.

In addition to these two factors, the time taken to read a block of words (Transfer rate) can also be considered. But this is usually too small in comparison to seek time and latency time, and disk access time is generally considered to be a sum of seek time and latency time. Further, since access times to disk are large, a sizeable portion of the data is read in a single go. That is why disks are referenced in blocks.

Based on the size and packaging of the disks, they can be classified into two types – floppy disks and hard disks. Further, disks that are permanently attached to the unit assembly and cannot be removed by the occasional user are called hard disks. A drive using removable disks is called a floppy disk drive.

Floppy Disks

The disks used with a floppy disk drive are small removable disks made of plastic coated with magnetic recording material. Disks of two sizes are commonly used with diameters of 5¼ and 3½ inches.

- The 5¼ inch disk is a floppy disk of diameter 5¼. Earlier such disks recorded data only on one side and were called single-sided (SS) disks. Today, both the surfaces are used for recording and are called double-sided (DS) disks. These are available in two capacities—double density (DD), and high density (HD), where density refers to the number of bits that can be stored per square inch area.
- The 3½ inch disk is a disk of 3½ inch diameter. These record data on both sides and are therefore called double-sided disks. These disks come in three different capacities—double density, high density, and very high density. These are smaller and can store more data than can the 5¼ inch disks.

NOTES

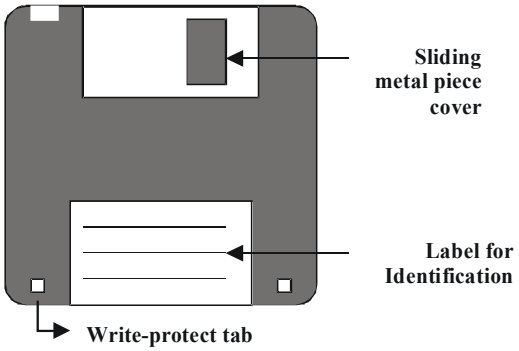


Fig. 4.22 *A 3½ Inch Floppy Disk*

The storage capacity for any disk can be calculated as:

Storage capacity = Number of recording surfaces × Number of tracks per surface × Number of sectors per track × Number of bytes per sector

Thus, for a 3½ inch high density disk which has eighty tracks, eighteen sectors/track, and 512 bytes/sector, the disk storage capacity can be calculated as follows:

$2 \times 80 \times 18 \times 512 = 14,74,560$ bytes or 1.4 MB (approximately)

The following table provides the necessary details and associated storage capacities of various types of floppy disks:

Floppy disks are extensively used in personal computers as a medium for distributing software to computer users.

Table 4.1 *Details of Various Floppy Disks*

Size (diameter in inches)	No. of Recording Surfaces	No. of Tracks	No. of Sectors/ Tracks	No. of Bytes/ Sector	Storage Capacity (approx)
5¼	2	40	9	512	3,68,640 bytes or 360kB
5¼	2	80	15	512	12,28,800 bytes or 1.2 MB
3½	2	40	18	512	7,37,280 bytes or 720 kB
3½	2	80	18	512	14,74,560 bytes or 1.4 MB
3½	2	80	36	512	29,49,120 or 2.8 MB

Hard Disks

Unlike floppy disks, hard disks are made up of rigid metal. The sizes for the disk platters range between 1 to 14 inches in diameter. Depending on the way they are packaged, hard disks can be categorized as disk packs or Winchester disks.

NOTES

- ***Disk packs:*** consist of two or more hard disks mounted on a single central shaft. Because of this, all disks in a disk pack rotate at the same speed. It consists of separate read/write heads for each surface (excluding the upper surface of the topmost disk platter and the lower surface of the bottommost disk platter). Disk packs are removable in the sense that they can be removed and kept offline when not in use (typically stored away in plastic cases). They have to be mounted on the disk drive before they can be used. Thus, different disk packs can be mounted on the same disk drive at different instances, thereby providing virtually unlimited (modular) storage capacity.

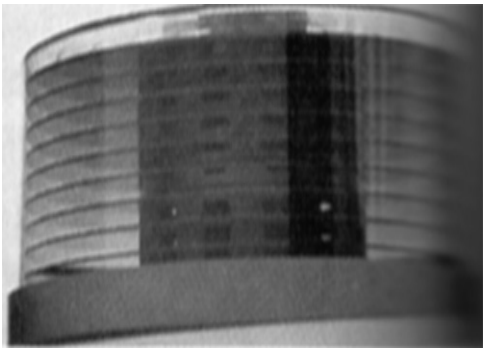


Fig. 4.23 *A Disk Pack*

- ***Winchester disks:*** also consist of two or more hard disk platters mounted on a single central shaft but are of the fixed type. The disk platters are sealed in a contamination-free container. Due to this fact all the disk platters, including the upper surface of the topmost disk platter and the lower surface of the bottommost platter, are used for storing data. So, even though Winchester disks have limited storage capacity as opposed to disk packs, they can store larger amounts of data as compared to the same number of disk platters.

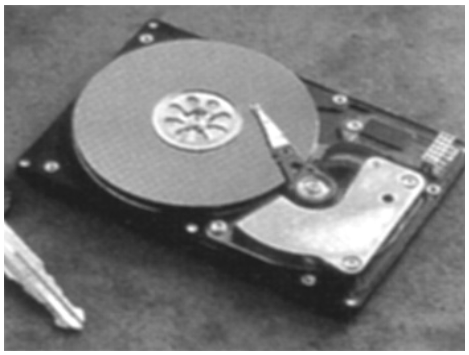


Fig. 4.24 *A Winchester Disk*

Another type of disk called the zip disk is very common today. This consists of a single hard disk platter encased in a plastic cartridge. Such a disk typically has a capacity of about 100 MB. Also, the zip drive can be fixed or portable. The fixed zip drive is permanently connected to the computer system while the portable ones can be carried around and connected to any computer system for the duration of its

use. In both cases however, the zip cartridge (the actual storage medium) is portable just like a floppy, albeit with a nearly 100 times larger storage capacity.

NOTES

Fig. 4.25 *Zip Disks and Zip Drive*

Optical Disks

Optical disks are storage devices with huge storage capacity. It is a relatively new storage medium and uses laser beam technology for writing and reading data.

Optical disks consist of one large track that starts from the outer edge and spirals inward towards the centre (this is unlike the magnetic disk in which tracks are concentric circles on the disk platter). An optical disk is also split up into sectors, but these are of the same length regardless of its location on the track. Data is therefore packed at maximum density over the disk.

The storage capacity of an optical disk is determined as follows:

Storage capacity = Number of sectors × Number of bytes per sector

(Note that we do not consider the number of tracks since there is only one track in this case.)

Thus, a 5.25 inch optical disk with 3,30,000 sectors and storing 2,352 bytes per sector, will have a storage capacity of

3,30,000 × 2352 = 77,61,60,000 bytes or 740 MB (approx.).

The technology used in optical disks uses laser beams to write and read data as opposed to the read /write head used in magnetic disks. Data is recorded by etching microscopic pits (burnt surface) on the disk surface. A high intensity laser beam is used to etch the pits, while a low intensity laser beam is used for data retrieval.

Fig. 4.26 *An Optical Disk and a Disk Drive*

NOTES

Three optical memory devices that are becoming increasingly popular in various computer applications are CD-ROM, WORM, and Erasable optical disks.

CD-ROM: CD-ROM (Compact disk read only memory) is a direct extension of the audio CD. It is usually made from a resin named polycarbonate that is coated with aluminium to form a highly reflective surface. The information on a CD-ROM is stored as a series of microscopic pits on the reflective surface (using a high-intensity laser beam). The process of recording information on these disks is known as 'mastering'. This is so-called because this master disk is then used to make a die, using which copies are made.

Information is retrieved from a CD-ROM using a low-powered laser, which is generated in an optical disk drive unit. The disk is rotated and the laser beam is aimed at the disk. The intensity of the laser beam changes as it encounters a pit. A photosensor detects the change in intensity, thus recognizing the digital signals recorded on the surface of the CD-ROM and converts them into electronic signals of 1s and 0 s.

As the name suggests, information stored in CD-ROM can only be read. It cannot be modified in any way. It is therefore useful for applications in which there is a database of information that is useful as it is and does not need changing in any way, e.g., a directory such as Yellow Pages. CD-ROMs are very useful for distributing large amounts of information to a large number of users. The advantages of CD-ROMs lie in the fact that they provide the following:

- Large storage capacity for information/data
- Fast and inexpensive mass replication
- Suitable for archival storage since they are removable disks

The disadvantages of CD-ROMs are the following:

- They are read-only and cannot be updated
- The access time is greater than that of a magnetic disk

WORM: The drawbacks of CD-ROM were partially resolved by the introduction of WORM ('write-once, read many').

In certain applications, only a few copies of compact disks are required to be made which makes production of CD-ROM economically unviable from a commercial point of view. This is because manufacturers do CD-ROM duplication by using expensive duplication equipment. For such cases, write-once read-many CDs have been developed.

WORM disks allow users to create their own CDs by using a CD-recordable (CD-R) drive. This can be attached as a peripheral device to the computer system. WORM disks recorded in this manner, can be read by any CD-ROM drive.

Erasable optical disk: The most recent development in optical disks is the erasable optical disk. The data in this type of optical disk can be changed repeatedly as in the case of magnetic disks. Erasable optical disks are therefore also known as rewritable optical disks.

NOTES

These disks integrate the magnetic and optical disk technologies to enable rewritable storage with the laser-beam technology and so are also called magneto-optical disks. In such systems, a laser beam is used along with a magnetic field to read or write information on a disk which is coated with magnetic material.

To write, the laser beam is used to heat a specific spot on the magnetic coated material. At this elevated temperature, a magnetic field is applied so that the polarization of that spot can be changed, thereby recording the desired data. This process does not cause any physical changes in the disk and so can be repeated many times. Reading is done by detecting the degree of rotation of the polarized laser beam reflected from the surface. This implies that as the disk spins, the polarized spots pass under the laser beam and depending on their orientation or alignment some of them reflect the light while others scatter it. This produces patterns of 'on' and 'off' that are converted into electronic signals of binary 1s and 0 s.

The capacity of an erasable disk is very high in comparison to that of a magnetic disk. For example, a 5¼ inch optical disk can store around 650 MB of data, while Winchester disks normally can store a maximum capacity of 320 MB. This is why magneto-optical disks are ideal for multimedia applications that require large storage capacities.

Mass Storage Devices

Any physical storage medium has a limit to its capacity and performance. There is a constant effort towards improving such media, and as a result larger capacity secondary storage devices have emerged. These are characterized by using multiple units of the same storage medium, as a single unit, to provide higher storage capacity. Disk arrays (Multiple disks), tape libraries (multiple tapes), and CD-ROM jukebox (multiple CDs) are the three most commonly used mass storage devices.

Mass data storage devices are characterized by relatively slow access time. This is because additional time in terms of first locating the desired disk, tape, or CD-ROM (as the case may be) needs to be accounted for. However, they are more cost effective in case of applications that require huge storage capacity and for which rapid access to data is not the prime consideration.

They can also be used for offline or archival storage of information/data since they can support huge volumes of information/data to be backed up.

Disk array (RAID): RAID (Redundant Array of Inexpensive Disks) is an acronym for a disk array and consists of a number of hard disks and disk drives with a controller in a single box.

The basic idea of RAID was to combine multiple small, inexpensive disk drives into an array of disk drives which yields performance exceeding that of a Single Large Expensive Drive (SLED). Additionally, this array of drives appears to the computer as a single logical storage unit or drive.

NOTES

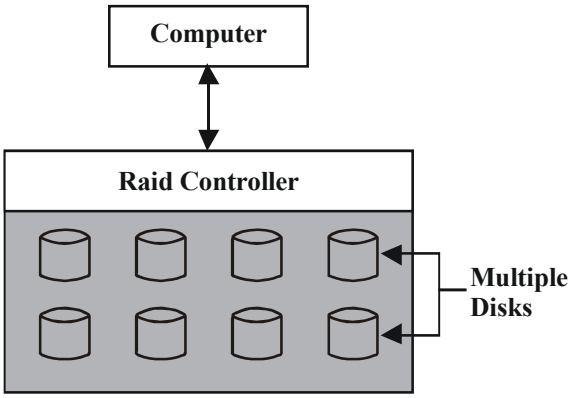


Fig. 4.27 *A RAID consisting of Eight Disks*

The concept was pioneered through academic research funded by Digital Equipment Corporation and has now become a standard in the computing industry for applications requiring fast and reliable storage of large volumes of data.

There are several different types of RAID configurations that are described in terms of 'levels'. The various levels of RAID storage are as follows:

- ***RAID 0:*** Data is split across drives, resulting in higher data throughput. Since no redundant information is stored, performance is very good, but the failure of any disk in the array results in data loss. This level is commonly referred to as striping.
- ***RAID 1:*** It provides redundancy by writing all data to two or more drives. The performance of a level 1 array tends to be faster on reads and slower on writes compared to a single drive, but if either drive fails, no data is lost. This level is commonly referred to as mirroring. Mirroring is the most expensive RAID option (since it doubles storage requirements), but it offers the ultimate in reliability.
- ***RAID 0+1:*** It is a combination of striping and mirroring. This configuration provides optimal speed and reliability, but possesses the same cost problem as RAID1.
- ***RAID 5:*** It employs a combination of striping and parity checking. The use of parity checking provides redundancy without having to double the disk capacity of the overhead. Simply put, parity checking involves determining whether each given block has an odd or even value. These values are summed across the stripe sets to obtain a parity value. With this parity value, the contents of a failed disk can be easily determined and rebuilt on a spare drive.

There are other RAID configurations in addition to the ones described here, but these are the ones most commonly used in the industry.

As can be noticed, RAID configurations result in higher reliability due to the use of multiple disks. In addition to this, both mirroring and striping (techniques used in distributing data across the disks) also result in speeding up the read process since different parts of the same file residing on different disks, can be read at the same time.

NOTES

Tape libraries: Network administrators are hungry for technologies that will allow them to efficiently and economically manage the explosive growth in data stored on networks. As the amount of data increases, the backup process takes longer. Simply adding another tape drive to reduce the backup time does not really help. Further, systems operated in this manner, represent one of the largest operational costs of a data centre and also typically represent the predominant need for human intervention.

The solution to this problem is the multi-drive automated tape libraries. These libraries consist of a set of magnetic tapes with a controller mounted in a single unit. The unit may have one or more tape drives to read and write data on the tapes in the tape library. Automated tape libraries allow random access to large numbers of tape cartridges and concurrent use of two or more drives, rather than manually loading one tape after another. The unit typically has robotic arms to retrieve the appropriate tape from the tape library and mount it on one of the tape drives for processing.

Automated tape libraries can be designed to provide extremely precise control and support for tape drives. Properly implemented, library automation can significantly enhance the operational reliability of tape drives by eliminating the highly variable human/machine interface. In this particular case, the objective of the system design is to avoid failures rather than to tolerate them, as is the case in RAID systems.

Automated tape libraries are typically used for data archiving purposes and as an online data backup device for automated backup.

CD-ROM jukebox: The CD-ROM jukebox is much like the automated tape library but consists of a set of CD-ROM disks instead of the magnetic tapes. The set of CD-ROM disks along with a controller are mounted in a single unit. Here, also the unit can have one or more drives to read data from the disks in the jukebox. The unit has robotic arms to retrieve the appropriate CD and mount it on one of the CD-ROM drives for processing. At the end of processing, the CD is automatically returned to the appropriate slot.

CD-ROM jukeboxes are typically used for archiving read only data that can be accessed online, e.g., online encyclopedias, online directories, etc. A large CD-ROM jukebox may consist of hundreds of disks providing a storage capacity of terabytes.

Data Backup

Data stored on an online storage device, such as a hard disk, can be damaged or lost due to any one of the following reasons:

- Disk crash
- Virus attack
- Accidental deletion by users
- Hardware malfunction
- Natural calamity (e.g., earthquake, fire, floods etc.)

Useful and sensitive data needs to be protected against such eventualities. Data should therefore be copied from online storage devices to secondary storage devices (like magnetic tapes, floppy disks, zip disks) and stored in safe locations. This process is known as backing up of data.

NOTES

Data backup now constitutes an essential part of IT policies in most of the organizations. Different types of backup media may be appropriate for different users and applications depending upon the volumes, periodicity, accessibility, security, sensitivity etc. However, determining the appropriate backup policy which would depend on the unique requirements of each organization, which is outside the scope of this book.

CHECK YOUR PROGRESS

5. List the different ways in which memory devices can be accessed.
6. What do you mean by locality of reference?
7. On what factors does disk access time depend?
8. List any two advantages of CD-ROM.

4.5 SUMMARY

- An input device accepts data from the outside world and transforms it into a form the computer can interpret.
- Point-and-draw devices are used to point to, and select menu items or icons displayed on screen. They provide a means for graphical user interface (GUI).
- The I/O devices that provide a means of communication between the computer and the outside world are known as peripheral devices.
- Keyboard devices allow input into the computer system by pressing a set of keys which are mounted on a board connected to the computer system. The most popular keyboard used today is the 101-keys with a traditional QWERTY layout, having an alphanumeric keypad, twelve function keys, a variety of special-function keys, numeric keypad, and dedicated cursor-control keys.
- A mouse is a small device that a computer user pushes across a desk surface in order to point to a place on a display screen and to select one or more actions possible from that position.
- Scanning devices are input devices used for direct data entry from the source document into the computer system. Scanners facilitate capturing of information and storing it in a graphical format for displaying it back on the graphical screen. There are two types of scanners, CONTACT and LASER.
- Digitizers are used to convert drawings or pictures and maps into a digital format for storage in the computer.

NOTES

- An output device is an electromechanical device that accepts data from the computer and translates it into a form that can be understood by the outside world.
- Dot matrix printers were the most popular impact printers used in personal computing. These printers use a print head consisting of a series of small pins to strike a ribbon coated with ink, causing the ink to transfer to the paper at the point of impact.
- Inkjet printers use a series of nozzles to spray drops of ink directly on the paper. These therefore fall under the category of non-impact printers.
- Laser printers use dry ink (toner), static electricity, and heat to place and bond the ink onto the paper. They use a combination of laser and photocopier technology.
- Plotters are used to produce graphical output on paper. It is a device capable of producing charts, drawings, graphics, maps, etc. It is much like a printer but is designed to print graphs instead of alphanumeric characters.
- The main memory is the central storage unit in a computer system. It is a relatively large and fast memory and is used to store programs and data during computer operations.
- Cache memories are small, fast memories placed between the CPU and the main memory. They are faster than the main memory with access times closer to the speed of the CPU. Although caches are fast, they are also very expensive and so are used only in small quantities.
- Hard disks are made up of rigid metal. The sizes for the disk platters range between 1 to 14 inches in diameter. Depending on the way they are packaged, hard disks can be categorized as disk packs or Winchester disks.
- Optical disks are storage devices with huge storage capacity. It is a relatively new storage medium and uses laser beam technology for writing and reading data.
- RAID (Redundant Array of Inexpensive Disks) is an acronym for a disk array and consists of a number of hard disks and disk drives with a controller in a single box. The basic idea of RAID was to combine multiple small, inexpensive disk drives into an array of disk drives which yields performance exceeding that of a Single Large Expensive Drive (SLED).

4.6 KEY TERMS

- **Secondary storage:** It is the storage other than primary storage. These are peripheral devices connected to and controlled by the computer to enable permanent storage of user data and programs.
- **Storage capacity:** It represents the size of the memory. It is the amount of data that can be stored in the storage unit of the computer.

- **Access time:** It is the time required to locate and retrieve data from the storage unit. It is dependent on the physical characteristics and the access mode used for that device.
- **Sequential access:** Memories that can be accessed only in a predefined sequence are called sequential access memories.
- **Optical disks:** These are storage devices with huge storage capacity. It uses laser beam technology for reading and writing data.
- **Digitizers:** These are used to convert drawings and maps into a digital format for storage in the computer.

4.7 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Scanning devices are input devices used for direct data entry from the source document to the computer system. Scanners facilitate capturing and storing information in a graphical format for displaying it back on the graphical screen.
2. Card readers are devices that allow direct data input into a computer system.
3. Vector CRT display and Raster scan display.
4. Plotters are used to produce graphical output on paper. It is a device capable of producing charts, drawings, graphics, maps, etc. It is much like a printer but is designed to print graphs instead of alphanumeric characters.
5. Random access memory, sequential access and direct access.
6. Analysis of a large number of typical programs has shown that memory references at any given interval of time tend to be confined to a few localized areas in the memory. This phenomenon is known as the property of locality of reference.
7. The disk access time depends on two factors, seek time and latency time.
8. Advantages of CD-ROM are:
   (i) They have large storage capacity.
   (ii) They provide fast and inexpensive replication.

4.8 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are point-and-draw devices? Give examples.
2. What are mass storage devices?
3. What is the need for data backup?
4. Write a short note on the role of monitors as output devices.
5. Find the storage capacity of a magnetic disk which has 80 tracks, 15 sectors per track and stores 512 bytes per sector.

6. Write short notes on plotters.
7. Write a short note on access modes used in memory.
8. Define cache memory and locality of reference.

NOTES

Long-Answer Questions

1. Describe how input devices are used to enter information into the computer.
2. What are scanning devices? Explain in detail.
3. What are display devices? Explain the different display screen technologies.
4. What are printers? Explain the different types with the technology used in them.
5. What are secondary storage devices? Explain.

4.9 FURTHER READING

William, Brin K. Stacey C and Sawyer. 2007. *Using Information Technology: A Practical Introduction to Computers and Communications.* Ohio, US: McGraw-Hill Irwin.

Long, Larry and Nancy Long. 2004. *Computers: Information Technology in Perspective*. Upper Saddle River, New Jersey: Prentice Hall.

Sinha, P. K. 2004. *Computer Fundamentals: Concepts Systems and Applications*. New Delhi: BPB.

Norton, Peter. 2008. *Introduction to Computers*. Ohio, US: Tata McGraw.

Basandra, Suresh K. 2007. *Computers Today*. New Delhi: Galgotia.

UNIT 5 USE OF COMPUTERS IN EDUCATION AND RESEARCH

Structure

5.0 INTRODUCTION

Information Technology (or IT as it is popularly called) has dramatically changed the way of our life. After the discovery of electricity, the computer ranks as one of the most important breakthroughs of the modern era. Like electricity, IT has impacted all facets of life and, in fact, its usage is so ubiquitous that it is hard for today's generation to even visualize how our ancestors lived without computers. From medicine to transportation, from banking to the entertainment industry, there is hardly any industry or sector that does not deploy IT in a fundamental manner to achieve one of the following three goals:

1. Reducing the cost of operations by increasing operational efficiency and staff productivity.
2. Improving revenues and bottom lines by helping management in informed decision making and focusing on priority areas.
3. Improving customer satisfaction by providing better, faster and value-added services.

IT has opened up several allied industries and employment opportunities which never existed before. Whether it is Business Process Outsourcing or BPO remote data processing.

NOTES

Web-enabled services (medical transcription, call centres, etc.), IT has opened up new avenues for jobs. Thanks to the Internet, developing nations today can also participate in the global economy and help bridge the divide between haves and have-nots. In this unit, you will learn about applications of computer in education, research, digital library and domain specific package like MS Office, SPSS etc.

5.1 UNIT OBJECTIVES

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

- Discuss the role of computers in education and research
- Explain the components of digital library
- Identify the need of digital library
- Explain Linking, Crosslinking and Interlinking amongst digital libraries
- Understand what the Google Scholar is
- Discuss the various features and tools in office 2010
- Understand the use of SPSS in data analysis

5.2 APPLICATIONS OF COMPUTERS IN EDUCATION AND RESEARCH

Education

In a traditional teaching model, a teacher would present the study material either through verbal presentation or using a black or whiteboard to write and illustrate. There would generally be a textbook to accompany this material which the students and teacher alike could refer to. Information technology has the potential to enhance and complement traditional teaching methods by providing additional tools to a teacher to display and explain ideas to their students and providing students with innovative but practical learning tools to help study.

The presentation of study material can be enhanced through multimedia presentations where a slide show, animation, video with sound is projected for the students to view. This makes the material more interesting and accessible than if it was merely presented verbally and through a textbook. Furthermore, concepts and ideas presented in a visual format like these are more easily understood and retained longer by students.

Some interesting developments and uses of IT in education are as follows:

- **Computer Based Training (CBT):** Progressive educational institutions can conduct classroom sessions using what is known as CBT. In this case, each student then sits at a computer terminal which operates software that presents course material in interactive sessions. It includes

refreshers and quizzes of the material presented to reinforce the students' understanding. The benefits of such learning software are that the student can learn at their own pace and they can explore and discover ideas and concepts within the material rather than learning by rote.

- **The Internet:** The Internet is a huge source of academic information; a student can use the Web to complement his research and study from text books and libraries. Search engines allow students to locate relevant and accurate material for study.
- **Distance Learning:** IT applications such as e-mail, videoconferencing, Web-based study has made distance learning available to many students who are not able to study on location at a university. The students receive and submit assignments, get course material, course information over the Internet and are able to contact their tutors through videoconference or e-mail.

Note that it is important to understand that information technology is not a replacement for real teaching or learning but a tool to enhance these things.

Research

The use of computer in research is so extensive that it is difficult to conceive today a scientific research project without computers. Many research studies can not be carried out without the use of computer particularly those involving complex computations, data analysis and modelling. Computer in scientific research is used at all stages of study, i.e., from proposal/budget stage to submission or presentation of findings. Physicians can use computers to research diseases, treatments and symptoms. They can use them to help a patient find a particular specialist if one they need is not available where they live. They can use them to record the medical history of patients and keep track of progress or increasing problems. For example, from the research perspective, the Internet is just one of the many pathways to various sites that contain information. Much of the information that can be found by using the Internet can also be found by travelling other paths that may be faster, less expensive, and/or easier to access. It is also important to realize that the information you want may not be available through the Internet.

When you consider using the Internet as an information resource, approach it the same way, as you would in doing any research examine the purpose and goals of the project and use the tools and resources that are appropriate to meet those goals. Keep in mind that the Internet is just one tool that can or should be used. Most of the hard data resources that are on the Internet can be found in libraries, books, CD-ROMs and commercial online databases. Depending on your project, these may be more effective and efficient sources to use. The strength and uniqueness of the Internet is as a communications tool, a way to share ideas with and ask questions from other users which are not covered in the hard or static data resources. The Internet can give you access to experts and specialists on almost any topic imaginable. Figure 5.1 shows a computerized form used for scientific research purposes.

Computerized Form For Scientific Research						
For scientific research	Daily	Weekly	Monthly	Yearly	Never	Intend to in the future
Searching for papers	☐	☐	☐	☐	☐	☐
Researching background information	☐	☐	☐	☐	☐	☐
Instrument control	☐	☐	☐	☐	☐	☐
Programming e.g. Visual Basic etc.	☐	☐	☐	☐	☐	☐
Chemical Drawing packages e.g. ChemDraw	☐	☐	☐	☐	☐	☐
Journal submission	☐	☐	☐	☐	☐	☐
Other *(please specify)*						

Fig. 5.1 *Computerized form for Scientific Research*

In the field of research, computer is being used to carry out the following:

- Access to the Internet has become invaluable as a research tool
- Gather huge amounts of information and store/catalog it
- Search for new information or search the information already acquired
- Interact with other researchers to create/gather more research material
- Access to many remote or obscure locations of the globe and their researchers/experiences/knowledge
- Disseminate results of your own research

5.3 DIGITAL LIBRARY AND GOOGLE SCHOLAR

In this section, we will discuss the concept of a digital library with a focus on its components, technical infrastructure, technical and media type formats, and other relevant concepts.

Components of a Digital Library

Several infrastructural components are needed to set up digital library services and resources, and these are mostly not ready available solutions. What is available is generally sets of different resources and systems from open architecture, which are linked with each other via a network using standard protocols and come together with the same interface which nowadays is the Web. In this way, use can be made of accessories, hardware and software from several different vendors. Generally, the various components of the digital libraries will be internal to the institutions; yet some will lie across the Internet, whose control and ownership will rest with several different independent players.

NOTES

Broadly, the components needed to build a digital library fall into five categories:

(i) Collection infrastructure

(ii) Access infrastructure

(iii) Computer and network infrastructure

(iv) Digital resource organization

(v) Manpower training

Each of the above-mentioned categories is discussed in detail in the following section.

(i) Collection Infrastructure

This generally comprises metadata and digital objects. The role of the metadata is to provide index or bibliographic information for the digital objects, which are the actual documents required by the users. Metadata is responsible for the identification and location of the digital objects via different ways of searching.

Content of different type needs handling specific to the content type. Content could be text which is structured/unstructured or a combination of both, video, audio, graphics, images and numerical data, to name some.

There are four categories into which Chris Rusbridge, Programme Director, Electronic Libraries Programme (University of Warwick) places resources for a digital library:

(i) **Legacy resources:** Mainly non-digital resources that are there in libraries, such as video recordings, audio, maps, slides, prints and manuscripts. Despite all efforts, a large proportion of these will not get digitized for some time.

(ii) **Transition resources:** Mainly those resources that were created for a different media like print have already been or are being digitized. Mostly Optical Character Recognition (OCR) is used in the process of digitization.

(iii) **New digital resources** are either created in digital only or both digital and printable format. Formats, such as XML (Extensible Markup Language) or SGML (Standard Generalized Markup Language) are used more and more by publishers, which allow creation of data files needed for printing. Use is also made of SGML/XML databases to create HTML/PDF/XHTML (Hyper Text Markup Language/Portable Document Format/Extensible Hypertext Markup Language) or Postscript file. Material that has been published in electronic format are available from datasets and databases in various formats, such as numeric and statistical datasets, video, audio, image, full-text and bibliographic format.

(iv) **Future resources**, that is, resources that might come up in the future are like to have such data sets which cannot be formally specified at the present time.

NOTES

The digital constituents of a digital library are as follows:

- Collections available with the library in digital formats
- External digital collections to which the library gains access
- In-house content that was created in digital format (born digital) for printing
- Collection of existing print media that is converted to digital format
- Scanned documents converted to machine readable format via OCR programs
- Organized and structured guide to electronic resources available on the Internet via portal sites or gateways to the electronic collections on the Web

(ii) Access Infrastructure

For a digital resource, the access infrastructure would comprise the following:

- For library Catalogs, WebPACs and/or multi-WebPACs would be used.
- For specialized image-based local collection, specialized collection Websites would be used.
- For Web resources, portals or subject gateways would be used.
- For local collections, search and browse interface would be used.

Network and Computer Infrastructure in a Digital Library

In the setting of a distributed client-server environment, typically, a digital library will require both server side and client side software and hardware.

Server-Side Hardware

Server

Big digital libraries might need to have several specialized servers that perform different tasks, and in the process, also distribute the work workload, removing excess pressure from falling on any single server.

Some library servers will host databases, indices and databases, and object server(s) will host multimedia and digital objects. Small digital libraries use a single server for all activities. Servers should be scalable to meet future requirements.

Input Devices

Image-based implementation needs special equipment for input, such as Photo CD systems, video cameras, digital cameras and scanners.

Storage Devices

Some of the storage devices used in digital libraries will be CDs for large bits of information hierarchical storage mechanisms, fast discs for data used frequently, near line for data used less often, and so on.

Networks that use intelligent storage and snap-servers control the physical storage so that it is available to several servers. To store data for long-term purposes,

use is being made of optical storage devices, such as DVD-ROM and CD-ROM. RAID (Redundant Array of Inexpensive Disks) also provides great performance and security, and is available at a low cost.

NOTES

Communication Devices

Use is made of various communication equipment, such as modems, repeaters, hubs, routers and switches.

Server-Side Software

Software for a digital library will generally be a combination of several open software put together to meet the different requirements of the library and its users. For example, an RDBMS (Relational Database Management System) will be needed to organize the digital objects with associated metadata.

Some of the software that are generally needed in a digital library are as follows:

- Software for scanning an image capturing
- Software for image manipulation and enhancement
- Software to automate the integrated library systems
- Software to run Web servers
- Software to perform information retrieval
- Software for OCR
- Software for database management
- Software for right management

Client-Side Hardware and Software Components

Generally, for client-side, the digital library will need to have an Internet-enabled multimedia computer. Software that the client-side computer will need will be an Internet browser. Other software required could be specifically associated with downloads that are made to support the format of downloads.

(iii) Digital Resource Organization

Since Web and Internet technology are the basis of digital libraries, the libraries make use of the Internet's addressing protocols and objects.

Object Naming and Addressing: URL, PURL, URN and DO1

Full form of **URL** is Uniform Resource Locator. It provides a universal method for locating and accessing information on the Web. URLs form the links between the Web's pages or sites that can be hyperlinked to provide the navigational functionality of the Web.

PURLS or persistent URLs are a scheme developed by OCLC to a document's location from the document itself. This raises the chances of the document being found despite its URL changing since the PURL never changes.

NOTES

Uniform Resource Name (**URN**) is a framework used to define identifiers. URNS need to be resolved to a URL via some database or other such system, and an URN could get resolved to several URL, for example, a URL for every format.

Digital Object Identifier (**DOI**) System has four components:

- **Enumeration:** Assigning an alphanumeric string to the identified digital object
- **Description:** Forming a description of the identified entity
- **Resolution:** Making the identifier actionable
- **Policies:** Rules governing the system's operation

Online Database Connectivity (ODBC)

Some of the methods that aid to connect a database with the Web are as follows:

- ODBC (Open Database Connectivity) Drivers
- Common Gateway Interface (CGI)
- Java Database Connectivity (JDBC)
- Uniform Resource Characteristics or Metadata in Digital Libraries (Descriptive Metadata, Administrative or Technical Metadata and Structural Metadata)

(iv) Manpower Training

As the objects being dealt with in a digital library are different from other physical libraries, the persons working in digital libraries require training for skill and knowledge specifically created to handling digital objects. This area being new, the associated skill are not easy to come by.

Skills required for Digital manpower

There is need for in-depth knowledge in:

- Digital document imaging
- Distributed database management
- Electronic document delivery
- Electronic reference service
- Enforcement of intellectual property rights
- Hypertext
- Information mining
- Information retrieval
- Integration of multimedia information services
- Management and organization of multilingual collection
- Selective dissemination of information

Digital libraries are coming up as an interdisciplinary area for education and research in various areas, such as library science, computer science and information science.

NOTES

Technical Infrastructure of a Digital Library

Internet and the famous World Wide Web can come to your rescue when you need to deliver electronic resources worldwide. These technologies are deployed in a digital library for searching and navigating the electronic content and follow the client-server architecture where both the client and the server are computer programs. Irrespective of the physical location of the data, the server and the client on a particular network receive the requested information.

The server program is generally stored on a powerful server machine that stores all kind of electronic information, while the personal computer of a user is used to store the client program.

A telecommunication network is used by the client and the server for communication purpose over a well-defined protocol. The main functions performed by the server program, which is considered as the main part or the heart of a digital library infrastructure, are as follows:

- It receives the requests from the client.
- It controls access to the information requested by the client.
- It retrieves the desired information after performing the necessary calculations.
- It authenticates the information and sends it to the client.
- It records the usage statistics.

The client programs or the browsers request the information and display it after the request has been processed. These programs are usually in the form of browsers, such as Netscape Navigator and Internet Explorer from Microsoft. These browsers make use of communication protocols to navigate between different servers and are not only easy to use but also have user-friendly interfaces, which are downloaded freely over the Internet.

Some of the different types of servers used by a digital library are as follows:

- **Object Server:** It stores and takes care of the actual digital content stored in the electronic format.
- **Library server:** It performs the main function of maintaining, indexing and supporting digital libraries.
- An appropriate management system that is responsible for intellectual property right issues and unauthorized usage of data.
- An Overview searching feature that makes use of powerful networks to transfer the data stored in a distributed system.

These digital libraries ensure modularity, interoperability, scalability and portability of data by following the Internet standards and protocols. Some of the most commonly used Internet protocols and standards are as follows:

- **Transmission Control Protocol/Internet Protocol (TCPIP):** This protocol creates and controls the transmission paths not only on single networks but also on multiple networks. The protocol, which was originally

NOTES

designed for UNIX operating system, is now widely used for all main types of computer programs, in-house Local Area Networks (LANs) and Public Networks. TCP/IP is a de facto standard for all major kind of networks.

- **HyperText Transfer Protocol (http):** Being the most important protocol used by the World Wide Web, the http protocol forms the basis on which World Wide Web or WWW transmit formats and messages. It carries client requests and returns with the desired information. While on one hand, it requires an HTTP (Hypertext Transfer Protocol) client program or an Internet Browser, on the other hand, it makes use of http server program.
- **File Transfer Protocol (ftp):** This protocol is used to move files and information to and fro from the server and the client or by moving data between tow Internet sites. A user can make use of the FTP to log in into another Internet site, both for sending as well as retrieving information. It is used by many Internet sites to transfer large chunks of information. However, the Internet search engines use hap instead of ftp, which is more complex, comprehensive and powerful when compared to the search operation performed by ftp.

Now let us look at some standards, such as follows:

- **Bibliographic Standards:** These standards are associated with the documents and non-documents as well as the physical attributes in a typical digital library system. These standards not only encompass the most intellectual and difficult part of the object definition but also cover the description of contents in detail, required to process and search the desired material.
- **Machine Readable Catalog (MARC):** This standard is used for recording the bibliographic data at the logical level. MARC, a framework of standards, contains various elements used for physical, content and process description.
- **Dublin Core:** The Dublin Core refers to a set of metadata elements that may be assigned to Web pages in order to facilitate discovery of electronic resources. Originally conceived for author-generated description of Web resources at the OCLC/ NCSA Metadata Workshop held at Dublin, Ohio in 1995, it has attracted the attention of formal resource description communities, such as museums, libraries, government agencies and commercial organizations. It contains elements for physical and process description. MARC is not a single standard but rather a framework within which each country has developed an individual standard.
- **BIB-1** is a type of a record structure that is used for online transmission of data where dialog between two systems takes place. It can be considered as a subset of the MARC standard. The BIB-1 consists of a number of elements that can be mapped to the standards of MARC as well as Dublin Core.

NOTES

- **Text Encoding Initiative (TEI):** This standard is used for marking the beginning and ending of paragraphs, pages, chapters, books, and so on. The main benefit of this standard is its capability to create highly accurate and effective indexes which can be used for efficient search operations. It can also be used to search both textual and linguistic features of text. Here, the techniques, such as SGML encoding, are used.
- **Electronic Archive Description (EAD)** are used for marking up the actual text. This is an example of the encoding scheme which is derived as a part of the SGML framework to constitute the description of contents within the documents and other type of archival objects. The technique makes use of few descriptive elements coded in an extensible manner. The elements so created make use of descriptive records, which help in searching the original material in different ways.

Formats and Media Types in a Digital Library

In digital libraries, the data is saved in digital files, which could be created from the digitization of original record materials or from the most common or standard data formats.

• Still Image File

This is a digital object obtained by digitizing still image, photographs and textual documents in original form. In a still image, the data is a raster or grid of the picture elements (pixels) mapped in such a manner that they create/depict the visual subject. There are many data formats in which one can store raster images, such as it can be stored in a variety of data formats, for example, TIFF (.tif).

o *JPEG: Joint Photographic Experts Group*

JPG/JFIF (Joint Photographic Experts Group/JPEG File Interchange Format) is popular with raster image data storage.

o *TIFF: Tagged Image File Format*

TIFF is a much used choice for raster image data storage. The file extension of a TIFF file is .TIF.

o *PDF: Portable Document Format*

PDF is commonly used to store different data types of data and even raster images. PDF has a number of subtypes, such as PDF-A.

• Audio File

This refers to any file which has been created from original digital or analog audio formats encoded with linear Pulse Code Modulation (PCM). In the case of audio files, one must differentiate between a codec and file format. While codec is responsible for encoding and decoding raw audio data, the data itself gets saved to a specific audio file format, like .wav. Some characteristics of audio files that are important are bit-depth, sampling frequency, and whether they are stereo or monophonic.

NOTES

- ***WAV: Waveform***
- Waveform Audio File Format acronymed as WAVE. It has three character file name extension of WAV.
- **MP3**

  It is an audio coding format for digital audio which uses a form of lossy data compression. It is a common audio format for consumer audio streaming or storage as well as a de facto standard of digital audio compression for the transfer and playback of music on most digital audio players.

• Video File

A video file is a moving image recording, and it has audio synchronized created with original digital video or analog. Some of the characteristics that are important in a video file are pixel array, frame rate per second, aspect ratio, bit rate, field order, color space and the definition being standard or high.

- ***MOV: QuickTime***

  QuickTime is an extensible multimedia framework developed by Apple Inc., capable of handling various formats of digital video, picture, sound, panoramic images and interactivity.

• Motion Picture File

This is a moving image recording of a very high-resolution, and it is mostly synchronized with audio, produced from either original physical or digital formats. Some of the characteristics that are important in a motion picture file are pixel array, frame rate per second, bit depth and color encoding.

- ***DPX: Digital Moving-Picture Exchange***

  The DPX file format is a pixel-based (raster) image format whose every content frame is in a separate data file linked by metadata so that it will play in the correct sequence.

- ***DCP: Digital Cinema Package***

  A DCP is a collection of digital files which help to store and convey digital cinema data streams, image and audio.

Some file formats that are highly recommended for digital libraries are as follows:

For Full Text

• *PDF/A-1b (.pdf)*

An abbreviation for Portable Document Format Archival, it is a subset of Adobe Acrobat's PDF format, and it does not have those features which are detrimental to long-term preservation. This format is an international one and is available from Adobe Acrobat 5 onwards.

For Images

- ***PDF (.pdf)***
- ***PPT (.ppt)***

Before fixing on a format, the original image has to be weighed against the output. Different data works well with different formats. In case the image might need to be printed at some time, it is good to choose a format which will produces a larger file size.

NOTES

Here are some formats and what they best suit:

- **.pdf -** Good for line drawings with searchable text, for example, maps.
- **.jpg -** Better for photographs.
- **.gif -** Better for images other than photos, for example, drawings.
- **.tiff -** For archival images, these files are the largest. More information is stored.
- **.png -** Created to replace gif format and is acceptable for photos also.

Each of the above-mentioned formats are discussed in detail below:

• PDF (.pdf)

PDF is a very good option for storing vector-based graphics since it keep the file size low, provides them a clarity for reading and the text that is included remains searchable. This format is also very apt for storing equations, charts and diagrams that combine text with vector-graphics.

• JPEG (.jpg)

This format is mostly applied to photograph storing. This is a 'lossy' due to which quality of image falls in the process of obtaining small file size. When high quality images are needed, a good choice will be TIFF format. Images that are non-photographic will generate smaller files in the GIF format.

• CompuServe GIF (.gif)

The CompuServe created GIF format is most appropriate for screen-quality images with few colors. A typical GIF file is small in size and is unable to recreate the range of colors needed for reproducing photographic images which is better done in the JPEG format.

• TIFF (.tif)

This is an archival format and will reduce file size with no compromise on quality of image. This is an appropriate format for storage of high quality and detailed images. GIF and JPEG creates smaller file size than TIFF. Also, TIFF files need added plug-ins and software to open and cannot be simply opened in majority of the Web browsers.

• PNG (.png)

PNG works well with low-color images and produces small sized files. This format is good with high-color images and will, therefore, used for storing photographic content well.

NOTES

Video

- **MPEG (.mpg)**

For movies, the most widely supported and the oldest format is MPEG (Moving Picture Experts Group). For all platforms, many MPEG viewers are available. This is the commonest output format from UNIX utilities which create video content.

- **Audio Video Interleaved (.avi)**

This is basically a format specific to Windows and does not have good support on platforms besides Windows.

- **Video Content**

The multimedia of video is an extremely resource intensive one. It is a good idea to keep both a low quality version and a high quality one.

Audio

- **WAV (.wav)**

The WAV format is Windows sound files' default format and majority of the platforms support it.

- **AIFF (.aif)**

AIFF (Audio Interchange File Format) is meant for the Macintosh and like WAV, it does not have support on other platforms.

- **MPEG-3 (.mp3)**

MPEG-3, also known as MP3, is a format which removes that sound data which the human brain and ear cannot perceive strongly. Files in the MP3 format work well when storing large pieces of sound content that do not need to be of high quality.

- **Sound related Suggestions**

Storing speech content on low fidelity recordings of speech content in a very high-quality format, since it would increase the size of the file and not improve the sound quality in any way. It is recommended that recordings that are of high fidelity should be stored at high quality.

Other Formats

When working within a proprietary format, maintain a copy of the original format content and one in a format that people use commonly. In case of large sized multimedia which cannot be Web downloaded, it is a good practice to make available two copies of the content—one in the original quality and one in a low quality whose file size is small.

Zip files work well with files in the digital format. Contents of a zip could be one or more files or folders. A Zip file is an archive file format; therefore, it ensures lossless data compression.

NOTES

Interoperability in Digital Library

The ever-increasing software applications, information resources, users, different types of computer systems and software applications have led the network environments to a critical problem of Interoperability. Digital libraries that contain a large number of traditional print-based resources are facing this problem even more severely as these libraries need to universally access the digital information.

Information sharing in such an environment is only possible when the participants collaborate and share the relevant information over the network and create a required framework for this purpose. So what does the term 'interoperability' mean? Well, the term can be described as follows:

When the various components of a digital library and the services associated with them can be interchanged both functionally and logically, they are said to be interoperable. These components are structured according to a set of well-defined publicly known interfaces. The various services and components along with their clients use open interfaces to communicate with each other. This technique is used to independently develop and create various components of a digital library in an efficient and simple manner.

In a digital library, the technique of interoperability helps in creation of a general framework used for accessing information efficiently along with integrating this information in multiple domains. This standard helps in the creation of digital libraries having large repositories of digital contents, possessing different attributes but used in a similar manner as they share a common interface definition. The concept of interoperability in a digital library setup can be implemented using a large number of approaches, where some of the common ones are as follows:

- **Standardization:** This is one of the most efficient and proven approach to implement interoperability. Certain examples of this technique are data models and protocols, and the famous Schema definition.
- **Mediation:** This technique of interoperability makes use of various mediation machinery and interfaces to translate the different formats of data and usage of different interaction modes between components. Primarily, the network gateways are used for the mediation purpose in the areas of interconnecting different types of networks. In many cases, simple mapping or translation is generally insufficient to achieve interoperability. For example, a lot of work needs to be done before any given two sets of digital libraries can translate by simply mapping each other. This may lead to non-availability of certain data types and operators in both the libraries. In such cases, mediation interfaces can be used to promote functionalities and services that may search two digital libraries and use their own sets of value additions to present the information. These mediation interfaces can then be referred to as proxies or wrappets. You need a proper set of standard to implement the mediation technology.
- **Mobile functionality:** This technology makes use of software agents that travel across the networks to various sites for accessing different

NOTES

types of services required by them. These network agents make use of Java applets and service functionalities that help their clients at run time to gather the required information and then travel back to their original sites. This technique transfers codes, facilitating communication amongst different components.

- **Families of Standards**: In this approach, you can select and implement one or more network standards. This approach was used by the ISO standard for Open Systems Interconnection (OSI) to establish the concept of interoperability framework. Therefore, this seven-layered framework of OSI is based on the Family of Standards. A very popular example of such a protocol is TCP/IP.
- **Specification-Based Interaction:** The semantics and structure of the entire data and operations when described properly helps in achieving interoperability. The requirements of the mediation systems can be circumvented through these specification-based interactions. The Agent Communication Language or ACL is a prevalent example of certain enabling technologies that are well developed.

Preservation in Digital World

For libraries that work with electronic content, a huge matter of concern is the archiving and preservation of content. This relates to content that might have been converted in-house, acquired via subscription services, purchased in digital media, or any other way. Several issues that are serious as well as problematic are associated with archiving and access policies of electronic aggregators and journal publishers. Nevertheless, for an institution's created digital contents, its preservation becomes the responsibility of the institution. In the digital technologies, there is a solution for the preservation of documents that exist in the libraries as also for better and greater access to the digitized documents via electronic networks.

Since digital as well as other Web and Internet technologies are getting better and better every day, they remain in a state of change. Regularly, work is being done to create better and newer technologies that will provide better protocols and standards for storage devices, storage media, network interfaces, hardware components, software, techniques of compression, file formats, to name a few. The situation of transitory standards and 'techno-obsolescence' is a real threat in the digital preservation of data and content by libraries. While new and better hardware for data storage is emerging, there is also a phasing out and disappearance of the once widely used CD-ROM. The problem of backward compatibility for products that were used in the past exists as a reality.

There is a need to continuously migrate the digital content to newer and better media, formats, and so on, to keep it current and prevent it from being on obsolete media that might after a while become unusable and lead to the loss of the data preserved on it.

Content in the digital world must be preserved at two levels, which are as follows:

(i) **Preservation of the storage media:** With technology changing at such a fast pace, the media on which digital data is saved on becomes obsolete very quickly, as fast as between 2–5 years. So data on obsolete media will be unreadable as the hardware and other facilities needed to read will not be available and would have been replaced by more modern technology. So, it is imperative that libraries, to remain current, migrate their digital data form an older media to a more current one on a regular basis.

(ii) **Preservation of access to content:** This requires that access to the content of documents is preserved irrespective of the format that they are stored in. To keep the media current, files can be migrated from one storage device to a more modern one, and in the same way, even the files can be regularly moved from one format to another, whichever is more technologically advanced.

Microfilming is a document preservation technique that is reliable as well as provides longevity (more than 500 years) to preserved documents. If preserved in the correct manner, a microfilm master is the most stable of all reformatting methods. Library scientist and author Don Willis has suggested that digital libraries could preserve data in the form of digital images for the purpose of access and on microfilm for the purpose of preservation. The process would be to first create the master microfilm and use this for the purpose of digitization. This process will enable the circumventing of the problem of software migration. The microfilm master could be used to create new digital files in successive software generations. The current scamming technology does not provide the same high quality as is present on the microfilm and it does not appear that in the new future, this scanning technology will improve to a great extent to reach the quality of the data on the microfilms. Microfilm is a great media for preservation of materials. It provides flexibility, speed and ease of use, and is, therefore, a preferred medium.

Intellectual Property Right (IPR) and Copyright

According to a definition given by library science author Ron Chepesuik in 1977, Copyright can be stated as 'the single most vexing barrier to digital library development'. The paper-based concept of documents so prevalent in traditional libraries has fizzled down in the digital world, as the system loses its control over the copies. As these objects can be accessed by multiple users simultaneously, these objects are available to the masses and less fixed in nature. Also, such documents can be copied easily. All digital libraries act as the caretakers of the information stored therein, unlike the book publishers who own their intellectual property. Even though these libraries are in the ownership of physical or digital material, it is not necessary that they will attain the copyright of this material too. Therefore, this makes it unlikely for the libraries to not only digitize but also distribute this material that they possess freely over the Internet.

The developers of such digital libraries have to take permission from the owners of such information before they can include the copyrighted material. They can also set specific mechanism for including this information in the digitized format without violating the copyright. The constitutions of various countries and international

NOTES

treaties like the Berne Convention govern the Intellectual Property Rights (IPR) as well as the Copyright of the various digitized and other materials. The owners of the digital library need to take permission from the said authorities to include any copyrighted material in the digital format. This is important so that necessary digital information can be provided over the Internet without violating the Copyright Act.

However, certain non-profit educational institutes may use the 'Fair Use' Act to include certain materials without taking explicit permission from the owner only for noncommercial and purposes. Therefore, the 'Fair Use' Act is an exception to the basic Copyright Law. The traditional copyright laws prevalent for data stored in the physical library often tend to be confusing and unclear in the current age of electronic information when performing the function of protecting the intellectual property.

With the digitization of library data in the form of images and its free circulation over the Internet, controlling the copyright data is becoming increasingly difficult as it reaches millions of users simultaneously. Therefore, licenses need to be demonstrated and used to set the copyrights of the digital information. The factors that may affect these type of licenses are as follows:

- The different varieties of the pricing model
- The conditions for each license
- The access limitations

A lot of work is currently happening in the field of model licenses being carried out by the publishers as well as the library associations. A consortium of libraries can negotiate with the publishers to set certain licenses based on the above-mentioned factors.

Linking, Cross-Linking and Inter Linking in Digital Libraries

Linking is a feature that is well supported by digital libraries. Since the Web is a system based on hypermedia, it enables the linking between the various electronic resources that are stored on servers which might even be located at different geographical locations with massive geographical distances between them. DOI or Digital Object Identifier is a unique identifier which is used to support links to bibliographic databases or to full text databases. DOI serves as a system that provides interoperability for exchanging and identifying intellectual property that is available in the digital environment. DOI provides a framework that enables the linking together of content suppliers and customers, making electronic commerce easier and also helping to maintain automated copyright management of media of all kinds.

CrossRef.org (http://www.crossref.org/) is a collaborative reference linking service that employs the use of DOI and associated metadata for citation linking. It is like a digital switchboard since it does not store the content but just provides links to where the content can be accessed. It provides linkages through DOI, which are tagged to article metadata supplied by the participating publishers. While subscribers will generally go straight to the text, non-subscribers will receive information on access via subscription, document delivery or pay-per-view.

Typically, a Web-based electronic document support links at various levels, such as follows:

At Article Level: from Secondary Services to Primary Journals

This linking takes place when research articles indexed in a secondary service link to the full-text digital versions on the site of the publisher.

NOTES

From References in an Article to the Secondary Services

A reference in an article has a link to their full text or to the secondary services.

From References in an Article to the Full Text

References in an article is linked with the full text and more so when it has reference to the same publisher or consortia of publisher.

Electronic Aggregators

Electronic aggregators, such as OCLC Online Journals, have come up with products that are Web-based information providers combining the indexes or table of contents which are linked with full-text articles.

Bibliographic Layer to Full-text Collection

Text collection that are majorly full-text ones are generally bundled together with a bibliographic layer which contains meta information regarding the full-text collection. It is bundled with this meta information and also of other full-text resources. Every bibliographic record in the bibliographic layer will be linked with the Website of the publisher.

Important Digital Library Projects

In this section, we will discuss some important digital library projects in India.

Digital libraries of India

In India, the concept of digital libraries started in mid 1990s with the spread of Internet, information and the support of the Central Government. Some digital libraries that are functioning in India are as follows:

- Archives of Indian Labour
- Digital Library of India
- Digital Library of Art Masterpieces
- Down the Memory Lane
- Indian National Digital Library in Engineering Science & Technology
- Kalasampada
- Khuda Baksh Oriental Public Library
- Mobile Digital Library (Dware Dware Gyan Sampada)
- Mukhtabodha Digital Library and Archiving Project
- National Institute of Advanced Studies (NIAS), Bangalore
- National Mission for Manuscripts
- National Resource Centre for Women
- Parliament Library
- Vidyanidhi

NOTES

Here is a list of digital library projects as compiled and made available by Wikipedia at https://en.wikipedia.org/wiki/List_of_digital_library_projects.

Name	Subject(s)	Volumes	Description
Ahlul Bayt Digital Islamic Library Project (Ahlul Bayt DILP)	Shia Islam		Digitizes and presents resources related to history, law, and society of the Islamic religion and its personalities with particular emphasis on the Twelfth Shi'ah Islamic school of thought.
Aozora Bunko	Japanese language		Digitized Japanese language texts.
Arnetminer	Computer science and information science		A free online service used to index and search academic social networks, hosted at Department of Computer Science and Technology of Tsinghua University.
Arts and Humanities Data Service	Arts and humanities		Images, texts and datasets largely collected from the UK universities. Ceased operation in 2008. Website is still active.
ArXiv	Science		Preprints in mathematics, physics, computer science, quantitative biology and statistics.
L'Association des Bibliophiles Universels	French literature		Digitizes French texts in the public domain.
Atoll Digital Library	General		Open source software for building digital library collections.
Australian Islamic Library	Tafsir, Hadith,Fiqh, Usool and Islamic Scholarship - Sunni School of thought		Library hosts over thirty-five classical and contemporary exegesis of Quran, all major books of hadith, a large number of books on biographies of early Muslims and a lot of other related titles. Library serves international audience and offers books in numerous international languages.
Avalon Project	Legal studies and history		Documents in Law, History and Diplomacy.
Baen Free Library	Science fiction and fantasy		Science fiction and fantasy from publishing house of Baen Books.
Bartleby.com	Literature, reference and verse		Literature, reference and verse.
Biodiversity Heritage Library	natural history, botany, zoology	>112,000 volumes; >40 million pages.	An open access digital library for biodiversity literature. BHL operates as a consortium of natural history and botanical institutions around the world that cooperate to digitize the natural history literature in their collections and make it openly available through their Website at www.biodiversitylibrary.org
Bookshare	General		A library serving accessible books (DAISY and BRF) to people with print disabilities. Currently, over 42,000 popular and educational books in the collection.
Biblioteca Digital Hispánica	General		Literature, pictures, records, memorabilia.
British History Online	History		Core printed primary and secondary sources for the medieval and modern history of the British Isles.
British Library: Online Gallery	General		The British Library's Online Gallery provides online access to a wide variety of its materials.
California Digital Library			
Canadiana Online	History of Canada	60,000,000+ pages	Canadian historical primary sources, including books, newspapers, periodicals and nationally-significant archival materials.
Carrie			An early full text electronic library.
CERN Document Server	Science		Electronic library for CERN related documents.
Center for Research Libraries	General		(CRL): Edward Hunter (1902–1978) collection.

NOTES

Chinese Text Project	Chinese classic texts		Electronic library of early Chinese texts.
Christian Classics Ethereal Library			
Choral Public Domain Library			
CiteSeerX	Computer science and information science		A Scientific Literature Digital Library and Search Engine hosted at Information Sciences and Technology at the Pennsylvania State University.
Civil Rights Litigation Clearinghouse	Civil Rights		
The Collection of Computer Science Bibliographies			
The Complete Works of Charles Darwin Online	Biology		
Cornell University Library Windows on the Past			
D-Scribe Digital Publishing	various academic		Over 100 thematic collections that together contain over 100,000 digital objects of various academic interests, including almost 800 out-of-print titles from the University of Pittsburgh Press.
Digital Accounting Collection	Accounting		
Digital Bibliography and Library Project			
Digital Comic Museum	Comic books		Thousands of freely downloadable public domain Golden Age Comics which are not under copyright.
Digital Himalaya	Himalayas		
Digital library for Dutch literature			Government-funded digital library project.
Digital Library of Georgia	General		
Digital Library of India		325,473	
Digital Media Repository	Arts and humanities, Middletown studies		
Digital Public Library of America	General		
Digital South Asia Library	South Asia		Materials for reference and research on South Asia.
Directory of Open Access Journals	General		
Domínio Público			
EBBA (English Broadside Ballad Archive)	English literature, English broadside ballads	Over 7,000 ballads from libraries worldwide archived.	Online digital archive of primarily 17th century English broadside ballads, digitized into multiple accessible formats for scholars and members of the general public.
EEBO – Early English Books Online			
eGranary Digital Library	Educational resources from over 2,000 Websites and hundreds of CD-ROMs.	An off-line digital library for communities	Over 30 million resources in every format: books, journals, Websites, video, audio, software, multimedia. Includes tools for patrons to create, upload, share and archive their own content. Has

NOTES

		with inadequate Internet bandwidth, i.e., developing countries, rural schools, prisons.	built-in search tools and Web 2.0 services, such as Moodle, WordPress and Drupal.
EuroDocs: Online Sources for European History	European history for over 50 countries		Facsimiles, transcriptions and translations of primary European historical documents.
Europeana		10,000,000+ digital objects	Linking archives, libraries, museums and audio-visual material from across Europe.
The European Library	General		
European NAvigator			First Digital Library on the history of a united Europe.
Gallica			French digital library.
FRASER (Federal Reserve Archival System for Economic Research)	Economics		A digital library of economic, financial and banking materials covering the economic history of the United States.
Greenpilot	Nutrition, environment, agriculture		A digital library focusing on nutritional, environmental and agricultural sciences. Targeting both scientific researchers and the interested public it aims to provide the user with relevant scientific literature which is both easy to access and of high quality.
Greenstone (software)	General		
Google Books	General	30,000,000+	
Harvard University Library digital collections	General		Links to digitized collections from the university's libraries, museums, archives and special collections, covering subjects, such as art, architecture, religion, history, culture, botany, biology, landscape design, music, politics, law and advertising. Projects involved digitizing of analog collections (including images, text, audio files, and music scores), georeferencing maps, or harvesting Web resources.
HathiTrust	General	10,000,000+	
Humanities Text Collection	Humanities		
Humbul			Humanities hub.
Hungarian Electronic Library	Hungarian literature		
Ibiblio	General		
INSPIRE-HEP	High Energy Physics		Scientific documents in the field of high energy physics.
International Dunhuang Project	Manuscripts		International collaborative effort to conserve, catalog and digitise manuscripts, printed texts, paintings, textiles and artefacts from Dunhuang and other archaeological sites at the eastern end of the Silk Road.
International Music Score Library Project			

International Relations and Security Network(ISN), Digital Library	International relations, security		Open-source Digital Library service for material related to international relations (IR) and security. Available content includes PDF-documents (journal articles, books, papers, and reports), a directory of IR- and security-centered organizations, and multimedia material (podcasts, videos).
Internet Archive	General	3,100,000+	
Internet memory	General		Internet Memory Foundation is a non-profit institution. It actively supports the preservation of Internet as a new media for heritage and cultural purpose.
Internet History Sourcebooks Project			
Internet library sub-Saharan Africa (ilissAfrica)			
Internet Sacred Text Archive			
JamJar Story	Social Research		A UK based digital video library containing hundreds of snapshots of daily life across the UK.
JournalServer	Academic journals		An international initiative to create a free digital library of academic journals.
JSTOR	General		Subscription required. Page image archive of important scholarly journals, with searchable OCR text.
The Kurdish Digital Library	General		The Digital Library consists of writings about the Kurds and Kurdistan. Its aim is to make the Kurdish cultural heritage available as digital data. Page image archive of important scholarly journals, with searchable OCR text.
Kujawsko-Pomorska Digital Library	General		
LacusCurtius	Classical studies		
The Latin Library			
Learning Ally	General		A digital library serving accessible audio textbooks and general titles to people with print disabilities. Currently over 75,000 popular and educational books in the collection.
Library of Economics and Liberty			
Librivox	Public Domain books in the United States		An online digital library of free public domain audiobooks, read by volunteers. Around 90 per cent of the collection is in English.
Literary Kicks			
Making of America collections	General		
Ludwig von Mises Institute: Literature	Libertarianism and economics		Libertarian and Austrian School economics resources.
Luna Digital Image Collection	Arts and Humanities	>80,000	Repository of images from the Folger Shakespeare Library Collection, including content related to William Shakespeare, the early modern period, and performance history.
Marefa	Arabic literature		Digitizes Arabic classics and classics of other languages that are written in Arabic script.

NOTES

NOTES

Marx and Engels Internet Archive			
Marxists Internet Archive			
Maryland Digital Cultural Heritage	General		
Memoriademadrid Digital Library	General		A Spanish digital library that provides access to the cultural heritage of the city of Madrid. Memoriademadrid digitizes heritage collections from the archives, libraries and museums belonging to the Local Government of Madrid.
Metelwin Digital Library	General, heritage		Online digital library containing books, magazines, journals, newspapers and manuscripts with index and searchable OCR text. Contains digitized regional heritage collections and collections of different ephemeral printed materials. All content free for use.
Metropolitan Museum of Art	General		Over the last two years, librarians and interns at the Thomas J. Watson Library have digitized over 250 early Museum Publications.
Michael :Multilingual Inventory of Cultural Heritage in Europe	General		A project funded by the European Commission to give access to digital collections in European archives, museums and libraries.
Michigan Digitization Project	General		
Miguel de Cervantes Virtual Library	General		
The Muslim Philanthropy Digital Library (MPDL)			MPDL covers material from 1900 to the present. It encompasses the diverse cultural, political and social factors influencing the practice of philanthropy as well as the significant contributions from across Muslim-majority countries to contemporary global philanthropic practice.
Mutopia project			
NALANDA			
National Academies Press	General	3,600+	
National Digital Information Infrastructure and Preservation Program	General		
National Digital Library Program	General		
National Library for the Blind	General		
New York Public Library NYPL Digital Gallery	General		
New Zealand Electronic Text Centre			
Ohio E-book Project	General		Available through multiple library systems for cardholders in the state of Ohio.
OSU Library Electronic Publishing Center	General		
Online Books Page	General		Provides an extensive list of digital book available online.

Online Text Library of the University of Texas at Austin	General		
On This Day			
Open Content Alliance	General		
Open Library	General	1,000,000+	
Open University Library	General		The Online Library from the Open University is a gateway to a wide range of online information resources. The Library Website provides access to a world-class collection of resources that enhance the learning experience of students and support the learning, teaching, research and personal development of members of staff.
Oxford Text Archive	AHDS Literature, Languages and Linguistics		
PAIL Solicitors Digital Media and Entertainment Resource Library	Digital Media & Entertainment		An index divided into two main categories, digital media, technology and entertainment; and Website and mobile apps projects, consisting of seven subcategories. The site includes more than 100 articles and comments related to law, society and policy.
Pandora Archive	General		
Panjab Digital Library	General		A NGO digitizing and preserving cultural heritage of Punjab. There are many historically significant documents stored and made available online. Its scope covers Sikh and Punjabi culture. The library funded by The Nanakshahi Trust was launched online in August 2009. It is located at Chandigarh.
Perseus Project	General		
Project Diana			Online Human rights library
Project Gutenberg	General	38,000+	Founded in 1971, it was the first project to create a library of freely available online texts.
Project Gutenberg Australia	General		Providing texts under Australian copyright law.
Project Laurens Janszoon Coster	Dutch literature		A collection of Dutch high literature; no longer maintained since 2001.
Project Madurai	Tamil literature		A collection of Tamil literature.
Project Noolaham	General		
Project Rastko	Serbian literature		Publishes Serbian and Serbian-related digital material, both in public domain and copyrighted.
Project Runeberg	Nordic literature		
Project Sugita Genpaku			Translating any text without permission, if there is no copyright trouble. Commercial use of texts is also allowed.
Projekt Gutenberg-DE			
Swiss Foundation Public Domain	Public Domain Music	50,000+ gramophone records and phonograph cylinders	A project for the conservation and utilization of public domain music and film material.
Questia Online Library	General	75,000+	
Rare Book Room	General		
Readme.cc	General		Books translated into ten languages.

NOTES

NOTES

Rubicon Research Repository	Environmental physiology		
Runivers	Russian literature		Website devoted to Russian culture and history.
SciElo			
Scriptorium	Swiss Newspapers	2,500,000+ pages	Digitized and Full text searchable Swiss newspapers published in canton of Vaud in last 250 years.
Shia Islamic Library Archive - Multi-language	Islam		Digitizes and Archive of Islamic Literature (largest collection of Islamic books online) in several languages.
Sophie Project	German women's literature	1,500+	Digitizes and distributed German-speaking women's writing in the public domain.
South Asian American Digital Archive	South Asian Americans		Documents, preserves and provides access to the material history of the South Asian American community.
Texas Digital Library	General		
Text Creation Partnership	General		Three collections (EEBO/TCP, ECCO/TCP, Evans/TCP).
TITUS - Thesaurus of Indo-European Texts and Language Materials database			Aims to prepare all textual material relevant for Indo-European Studies (including Middle Iranian, Tocharian, and so on) in electronic form for analysis.
Traditional Knowledge Digital Library	Indian literature		Repository of the traditional knowledge of India, setup to protect the ancient and traditional knowledge of the country from exploitation such as Unethical patents and monopolization.
UK Web Archiving Consortium (UKWAC)	General		
United States National Agricultural Library	Agriculture		
US National Library of Medicine			
Universal Digital Library	General	1,500,000+	A book digitization project, led by Carnegie Mellon University School of Computer Science and University Libraries. Working with government and research partners in India (Digital Library of India) and China, the project is scanning books in many languages, using OCR to enable full text searching and providing free-to-read access to the books on the Web.
University of Michigan Library Digital Library Production Service	General		
University of Pittsburgh Archives Service Center	• University of Pittsburgh Archives • 20th Century Urban Renewal • Business and Industry • Ethnic Groups in Pittsburgh • Labor and Politics • Social Action and Women's History		

	• Audio-Visual Materials		
University of Wisconsin Digital Collections	General		
Varna Digital Library	General		A project for digitization of periodicals, postcards and photos related to Varna from the end of the 19th, early 20th century
Vascoda	General		German Internet-portal for scientific and scholarly information, offering interdisciplinary and subject-specific search.
Virtually Missouri	General		Digital collections from Missouri libraries, museums and cultural institutions.
Welsh Journals Online	General		
Wordtheque			Word by word multilingual library.
World Wide Web Virtual Library	General		
Wikibooks	General		A digital library of new books edited in a similar way to Wikipedia.
Wikisource	General	3,500,000+	A digital library out of copyright or freely licensed books.
Wired for Books			A project of the WOUB Center for Public Media at Ohio University.
Wisconsin Heritage Online			A collaborative state-wide portal to Wisconsin's history.
Wisconsin Historical Society Digital Collection	History		Wisconsin state, regional and local history books, journals and museum collections.
World Digital Library	General		
Zeno.org			

Building a Digital Library

The activity of digital library creation and development need specialized knowledge stemming from various disciplines, the branches of knowledge, and a clear understanding of the purpose and objective of why the digital library is being set up.

The proposal for a digital library must contain goals, benefits, scope, target users, time needed for establishing the library, deliverables, issues associated with the implementation and the costs that will be incurred.

Planning Digital Library

In the planning stage, one must work out the following:

- Feasibility
- Infrastructure requirements
- Human resources planning
- Managerial planning

The Selection of Material for Digitization and 'Born Digital'

This requires the identification, selection and prioritization of those documents that will be digitized. Documents being available in digital format will make it easier to convert them to other formats. Material that is from external sources will involve issues of IPR. Buying materials in e-media format falls economical compared with performing conversion on other formats. Such issues and various other related will need to be considered and resolved.

NOTES

Technical Specifications

Finalizing technical specifications needs to be done before beginning the process of actual digitization. At this stage, the existing practices, standards and practices need to be reviewed. Then new specifications need to be created and tested via actual demonstrations and then modified if required. Specifications need to be created for metadata creation for digital objects and digital collection.

Implementation

The following is involved during the implementation stage:

- Purchase of hardware and software
- Training or manpower — existing and new
- Contents creation or building of the content
- Setting up of the access infrastructure
- Organizing the digital resources

Testing and Integration

The process of testing and integration needs to be conducted before the data can actually be made available to the users.

Content Dissemination: Extending Digital Library Services

This is the stage at which the content goes live along with other value-added services that a digital library provides to make the content more accessible and usable for the end users.

Evaluation

Post the digital library going live, it needs to be regularly evaluated so that it can be better consolidated, provide improved usability and better value-added facilities, to name a few. This process can make use of potential users, external evaluators and experts. It is important to note that it is a continuous process and not a one-time activity. It needs to be repeated at regular intervals.

Improving Service and Instructions

Post evaluation, it is time to implement the recommendations that come up from the evaluation which has been performed.

5.3.1 Google Scholar

Google Scholar is a freely accessible web search engine that allows you to search across a wide range of academic literature. It is designed to help you discover scholarly sources that exist related to your topic. It acquires information from journal publishers, university repositories, and other websites that it has identified as scholarly.

NOTES

CHECK YOUR PROGRESS

1. List the components needed to build a digital library.
2. Identify the digital constituents of a digital library.
3. State the access infrastructure for a digital resource.
4. Name some of the software that are generally needed in a digital library.
5. List the four components of a DOI System.

5.4 DOMAIN SPECIFIC PACKAGE: MS OFFICE, MATHEMATICA

MS Office 2010 is the powerful application suite of Microsoft. A number of applications like Word, PowerPoint, Excel and Outlook are there that are used for the business or personal tasks. Microsoft Office 2010 is the latest version of MS Office that runs on the Windows XP SP2 operating system or above. It is bundled with many features. It has various components such as Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Microsoft OneNote, Microsoft Outlook, Microsoft Publisher, and Microsoft Access. Microsoft gives all these programs in its Home and Student Editions. In the earlier stages, during its evolution in the beta phase, it was known as Office 14. It was released on April 15, 2010 for volume license customers and it became available to general users on June 15, 2010. This was released on Windows Vista and Windows Server 2008 operating systems. It provided the functionality of using and working with various applications and performing tasks such as creating presentations, reports, managing emails and contacts. Many user interface tools and various applications have been added to make the job easier. Thus, with the help of Microsoft Office 2010, work is done easily in comparison with the earlier version.

5.4.1 Microsoft Office 2010 – Products and Applications

MS Office 2010 is the popular and extensively used application suite of Microsoft with the number of highly beneficial applications and products such as Word, PowerPoint, Excel, Access, Publisher and Outlook.

MS Office 2010 provides a good platform and a reliable Office suite for PC's (Personal Computers) or desktop applications. The applications of MS Office 2010 give more efficiency and flexibility to accomplish tasks.

Following are the applications covering majority of the Microsoft Office 2010 application suite:

- **Word:** Microsoft Word 2010 is a word processing application that has improved features of searching, editing words and pictures and also making additions in a more proper and unambiguous way over its previous editions. It

NOTES

includes the Screenshot feature, Photoshop feature, visual designs and formatting tools that allow you to create professional looking documents and files. New features like artistic effects, clip art, facilities of sharing documents and turning text into diagrams, are also available in this standard word processing application. This tool of Microsoft Office 2010 also makes creating short reports or lengthy documents, an easy and time saving task.

- **Web Apps:** Office Web Apps represent the online version of MS Office. It is a major addition in MS Office 2010. It is available free of cost and is extensively supported in browsers like Windows Internet Explorer version 7 or higher, Chrome and Firefox etc. Using Office Web Apps, users can access and edit documents within the web browser itself using Microsoft's SkyDrive, a cloud-based storage service. You can work on MS Office Web Apps only when you have some idea of how MS Office works. Certain useful tools are included in Office Web Apps for the purpose of creating, editing and manipulating documents like Word, Excel and PowerPoint.
- **Excel:** Microsoft Excel is a useful spreadsheet application of MS Office suite 2010 with some new features and innovations such as Slicers, Sparklines, Pivot Table etc. These features help in visualizing, enhancing and analyzing data trends for home and business tasks.
- **PowerPoint:** Microsoft PowerPoint 2010 is a presentation program, built for the Windows operating system, and is also available for use on other operating systems. It is used for creating and organizing complex, animated and transitive slide shows. Some improvements like the facility of Backstage View, Graphical User Interface (GUI) and Editing are included in this version, over its predecessor, MS PowerPoint 2007. Thus, this tool or application of the MS Office suite, allows you to create and design attractive presentations having text, images, videos, tables, charts and clip art pictures.
- **OneNote:** Microsoft OneNote 2010 was developed and introduced with miscellaneous improvements over its older versions. This is a note-taking and information gathering program. With this, files or documents can be accessed from anywhere: from home, workplace or on-the-go. This organizational tool requires the use of internet for sharing notes with other users using the OneNote application.
- **Publisher:** Publisher 2010 of Microsoft Office is used for working with text, tables, pictures, layouts and templates. It helps to create or produce professional looking applications. A wide range of publications and marketing material, such as brochures, newsletters, invitations and business cards can be produced efficiently using this desktop publishing platform.
- **Outlook:** Microsoft Outlook 2010 is a web application and a personal information manager for e-mail communication using which you can store e-mails on your system. This tool is also well-known for sending, receiving e-mails, managing messages, accounts and contacts. It includes certain advanced and improved features to stay connected with people and manage information and emails in a much better manner as compared to MS Office 2007.

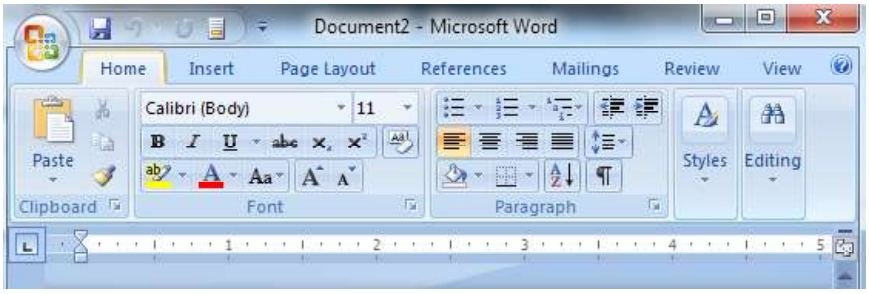
- **Access:** Microsoft Access 2010 has some design modifications over MS Access 2007. It is a relational database application meant for allowing users or programmers to manage, sort and filter large amounts of data. It is a powerful tool for creating and running database solutions and producing meaningful reports.

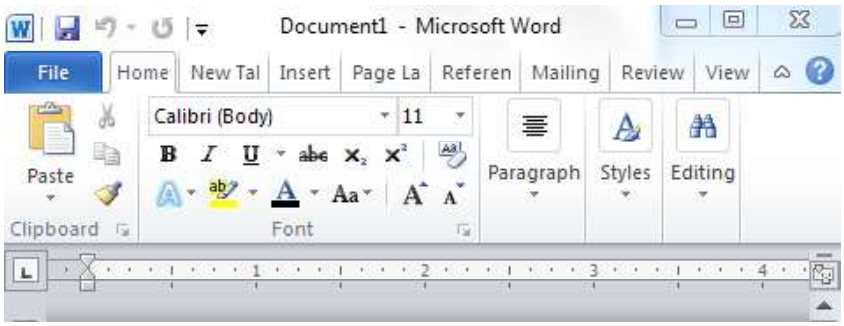
5.4.2 New Features and Tools in Office 2010

There are lots of new features and capabilities which are included in Office 2010. Various features and tools that are included in the suite members of Microsoft Office are as shown:

- **Microsoft Word:** Microsoft Word 2010 has all the functionalities of the previous editions along with some additional features. The "Ribbon User Interface" was introduced in Microsoft Office 2007, and Microsoft Office 2010 maintains this feature along with some additional ones. There are some changes in the UI interface. In the previous version, the background color of Microsoft Office was blue. In the recent edition, the background color of Microsoft Office is grey.

  Screenshots showing these differences in the interface are as shown:





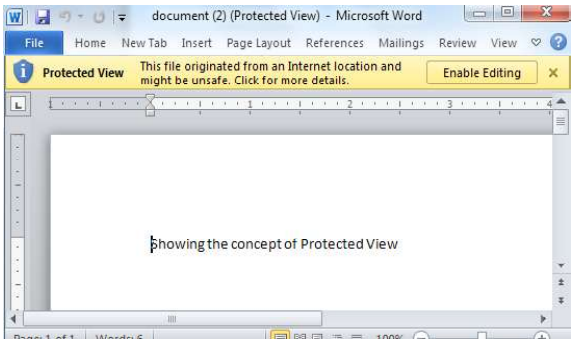
Microsoft Office 2010 is more secure in comparison to its earlier edition. Microsoft Office 2010 provides you with an additional feature to open documents that carry any kind of risks. Risky documents are opened in a sandbox environment of MS Word. Microsoft Office 2010 provides a feature of Protected View. Documents downloaded from the web are opened in protected view. These features make Microsoft Office 2010 more secure.

Unless you click on the 'Enable Editing' button, you cannot make any changes to these Word documents. When you click on the 'Enable Editing' button, it starts

working, like MS Word normally would, but now ensuring that the system on which the Word document is being edited is secure.

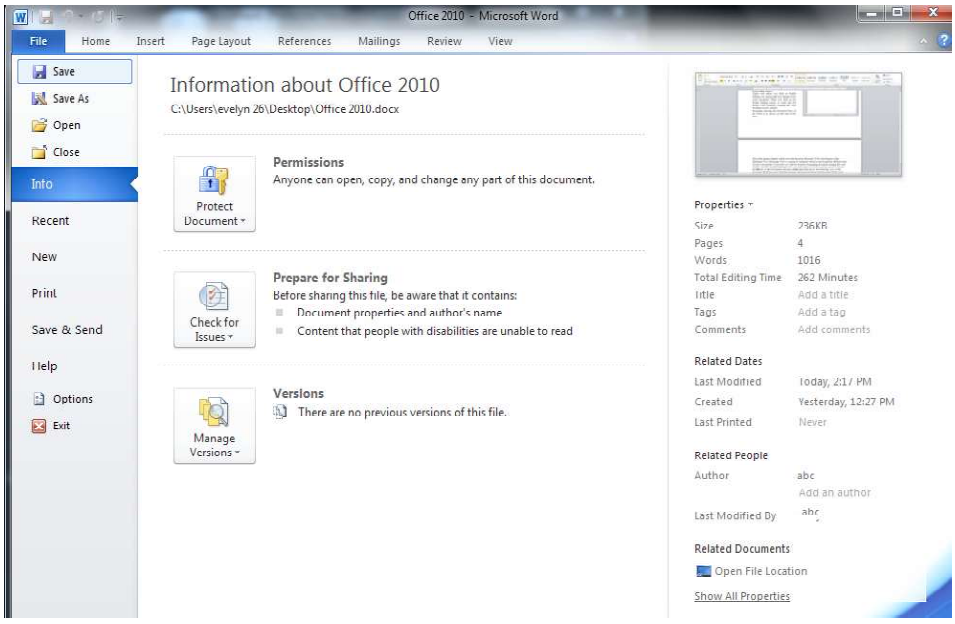
Screenshot showing the Protected View of Word is as shown.

NOTES



One of the unique features which were introduced in Microsoft Office 2010 is the feature of Backstage View. Backstage View is a group of commands which are used to perform different types of tasks on documents. It provides you with the features of managing documents, sharing files and providing greater security. In order to open backstage view, you must open the Word document and then go to the File tab. In this, the default selection is the 'Info option', that shows the backstage view of the document. Backstage view hides the previous document and shows all essential things and permissions associated with the documents. In order to return from the backstage view, you can simply press the "Esc" key or click on the File menu. This will take you to the original document in which you were working.

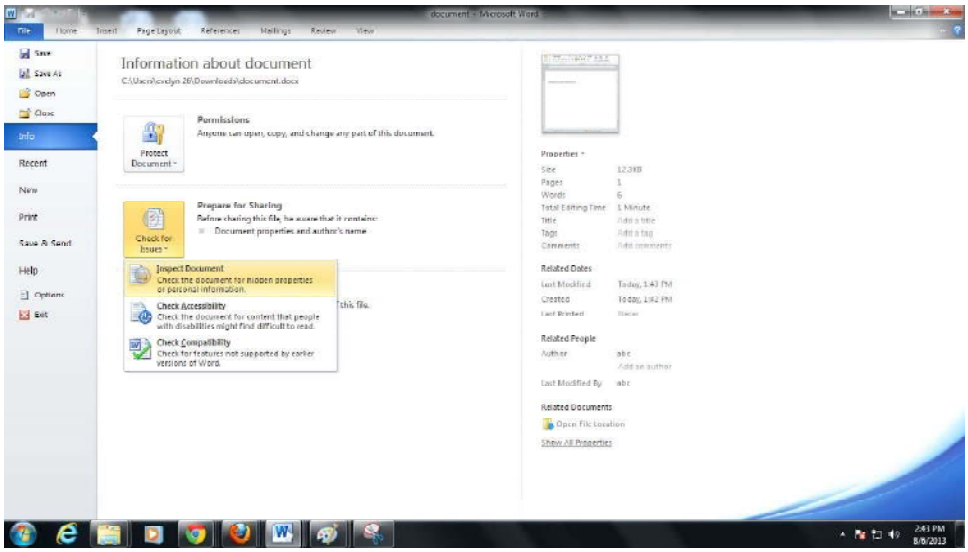
Screenshot showing the Backstage View of MS Word is as shown:



Microsoft Word 2010 provides features of collaboration and sharing of documents. The collaboration and sharing option is opened with the help of the Backstage View. In the Backstage View, there is an option to "Prepare for Sharing".

In this, there are three options, "inspect document", "check Availability", and "check compatibility".

Screenshot showing all these options is as shown:

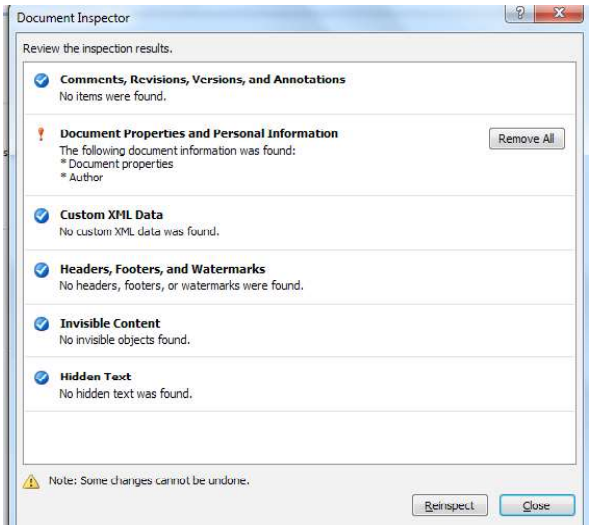


NOTES

The first option is "Inspect Document", which helps in inspecting the document before sending it to clients or colleagues. With the help of "inspect document", you can identify the personal information which gets accumulated in the document over a period of time. Sending personal information by mistake to other organizations will reveal essential details about your organization. That is why inspection of documents is performed before sending them to other users or organizations.

Before performing inspection of the document, the file should be saved on the local computer. This is so because MS Word sometimes is unable to restore documents once inspection of documents is performed.

Screenshot of the "Inspect Document" feature is as shown:



NOTES

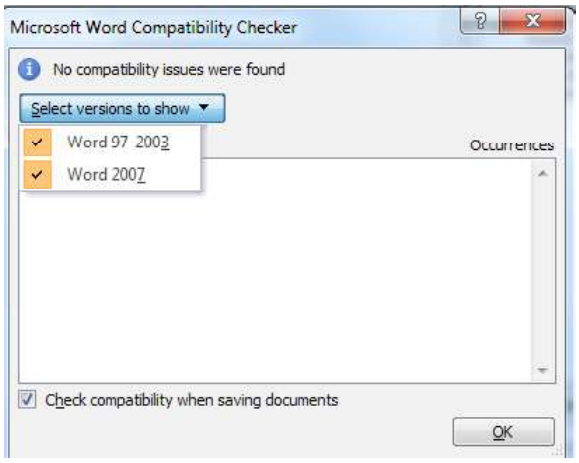
As shown in the earlier screenshot, when inspection is performed, the document property and personal information associated with the document should be removed before sending documents to other users.

The second option is "Check Accessibility" that works in a manner very similar to the Spell Checker. It is recommended to correct the accessibilities so that it is not difficult for anyone to read the content and make changes to your content.

Screenshot shows the state after "Check Accessibility" is performed.

The third option is "Check Compatibility", which checks if the document is compatible with previous versions of Microsoft Office or not. Compatibility is automatically performed by Microsoft Office. Compatibility helps in preserving layout of the document.

The screenshot shows the state after performing "Check Compatibility".



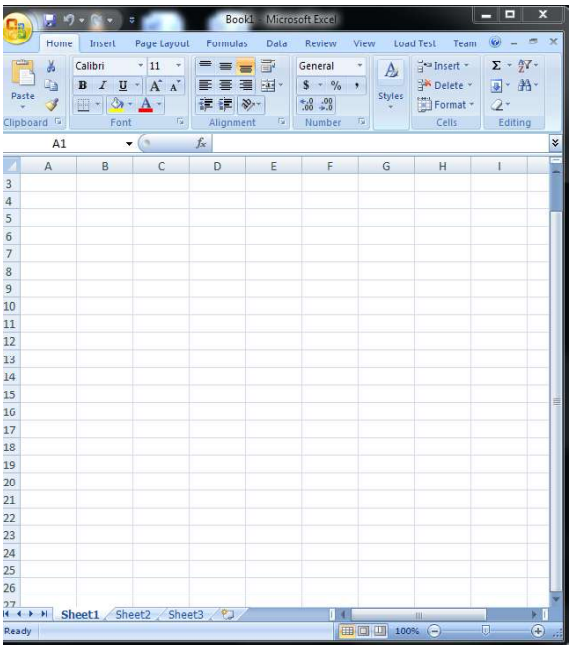
Hence, it can be said that Microsoft Word 2010 is much more powerful application software as compared to its previous versions, not only in terms of its robustness, but also from an aesthetic point of view.

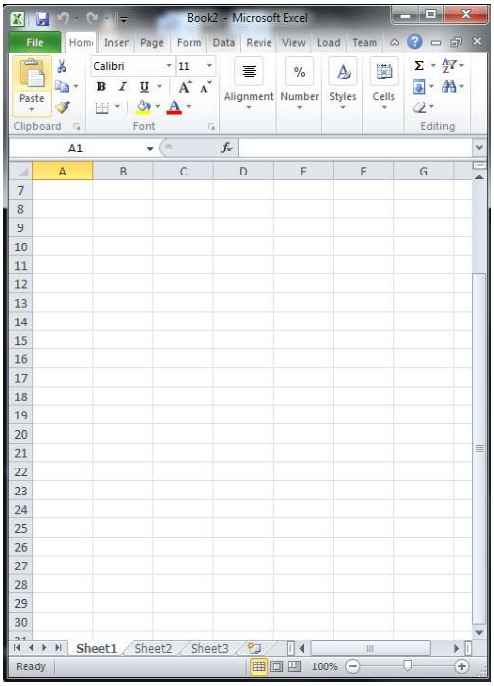
- **Microsoft Excel:** Microsoft Excel 2010 has all the functionalities of the previous edition, and also has some new additional features. Microsoft Excel

NOTES

2010 maintains features of the "Ribbon User Interface" with some additional features. There are also some changes in the UI. In the previous edition, the background of Microsoft Excel was in blue color. In the recent edition, the background color of Microsoft Excel is grey. Microsoft Excel has increased row limits, and thus shows improvement in visualization and analyses of data trends. A lot of features in it have improved over its previous editions, such as formatting, PivotTable and many more.

The screenshot showing differences in interface and row count is as shown:



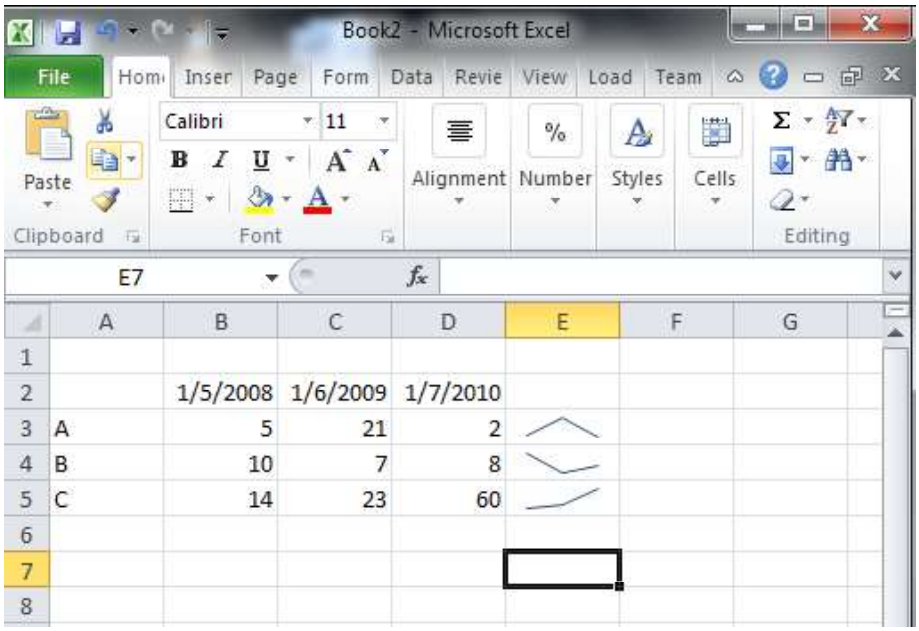


NOTES

It has all the features that are there in Microsoft Word, such as the Protected View, Backstage View, collaboration and sharing.

Sparkline and Slicers are the two important features which are included in MS Excel 2010. Sparkline is small graph that is shown in the cells. It is used to show the flow of data in particular column or row. It needs less amount of space for storage and is useful for the organization where the visualization of the data is necessary. Sparkline is of three types: line, column, and win/loss.

Screenshot showing statistics using line option of Sparkline is as shown:



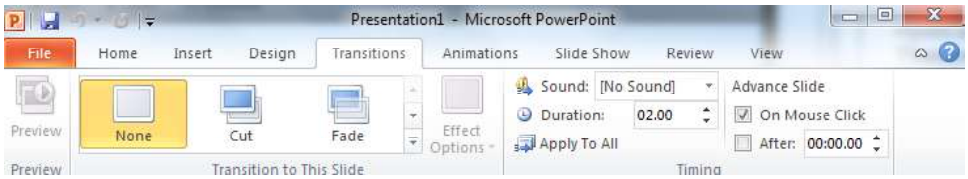
With the help of Slicers, data can be filtered and interpreted in an aesthetic manner. Slicer is inserted by using the 'Insert' option in the Excel Spreadsheet. Slicers help in creating interactive reports as it has improved the features of functions used in Excel.

Thus, you can say that Microsoft Excel 2010 has included new features in order to improve the presentation of the business scenario.

- **Microsoft PowerPoint:** Microsoft PowerPoint 2010 has all the functionalities of the previous editions such as Protected View, Backstage View, Collaboration and sharing, taking screenshot directly. **Transition** tab is the new tab which is introduced in Microsoft Office 2010. It has been separated from the **Animations** tab of 2007 with some more animation effects.

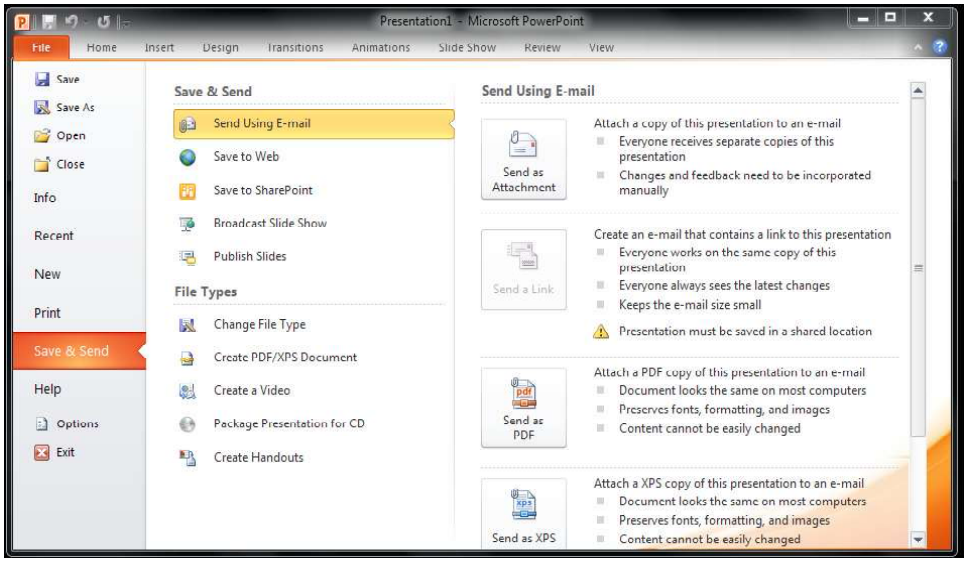
Screenshot showing the **Animation** tab of 2007 is as shown:

Screenshot showing the **Transitions** tab of 2010 is as shown:



In this, one can convert the PPT to PDF or XPS formats. However in the earlier editions, this feature was not present. It also provides the features of sharing a presentation over web and converting presentations to video. It also provides many options of video editing.

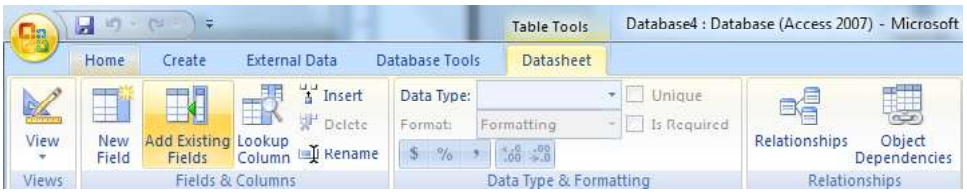
The screenshot shows few additional features that are included in MS PowerPoint 2010:



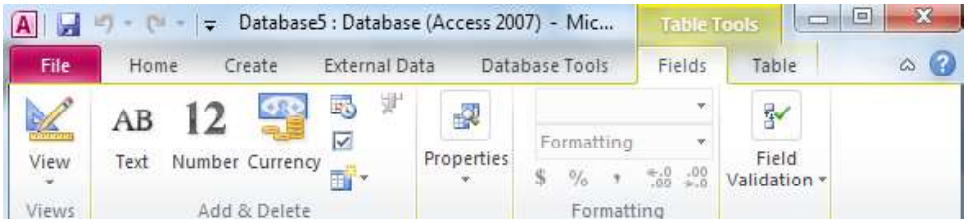
- **Microsoft Access:** Microsoft Access 2010 has all the functionalities of previous editions, such as Protected View, Backstage View, Collaboration and sharing and taking screenshots directly. It is used as a database management system. In this, database can be shared over the web. **Fields** and **Table** tabs are the new tabs which are introduced in Microsoft Office 2010. These are actually the replacement for the **Datasheet** tab of the previous version.

NOTES

Screenshot showing the **Datasheet** tab of the MS Access 2007 is as shown:



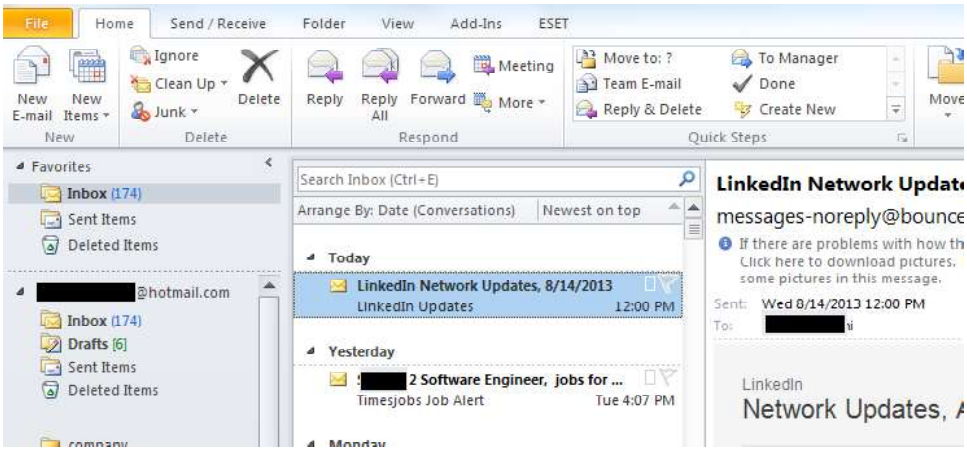
Screenshot showing the **Fields** and **Table** tabs of MS Access 2010 is as shown:



It has improved the features of calculations, formatting, macro builder and design over MS Access 2007.

- **Microsoft Outlook:** Microsoft Outlook is the popular tool for organizing the business. It has all the functionalities of the previous editions, along with the additional feature of storing large amount of e-mails, contacts and tasks on the system. The biggest change in Microsoft Outlook 2010 is the replacement of tabs with the new tabs; user interface has been changed drastically. With the help of Outlook, social connector keeps you updated and your contacts to the social networking site such as LinkedIn, Twitter.

  Screenshot of Microsoft Outlook 2010 is as shown:



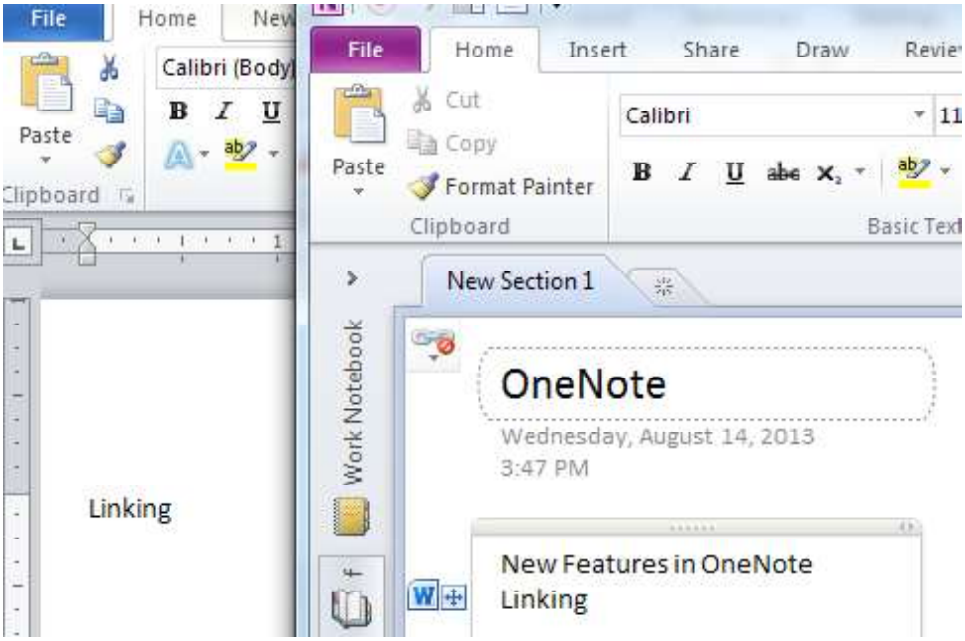
It has improved the features of arranging the mail in the proper manner so that it can be indexed easily at the top. The Organization and storage of the Mail can be done easily with the help of Outlook.

NOTES

Screenshot showing the arrangement of the e-mail is as shown:

- **Microsoft OneNote:** Microsoft OneNote provides the facilities of organizing the notes. It has all the functionalities of previous editions with some additional features which include the replacement of tabs with the new tabs. With the help of OneNote, one can easily establish links with the other websites.

  Screenshot showing the linking is as shown below:



It has improved the features of organization of tabs, sending the content under any section, sharing the notes over the web browser. It is used for accessing files from anywhere: the user's office, home or any other place across the world.

Microsoft office 2010 is better than the previous versions as it is based on the latest technologies. The new features allow us to create more aesthetically appealing documents, presentations, spreadsheets, notes and databases, with added functionalities in comparison to the previous versions.

NOTES

5.4.3 Mathematica

Mathematica is a mathematical symbolic computation program and sometimes referred to as computer algebra system or program which is used in many scientific, engineering, mathematical, and computing fields. It is divided into two parts, the kernel and the front end. The kernel interprets expressions and returns result expressions.

The front end provides a GUI, which allows the creation and editing of Notebook documents containing program code with pretty printing, formatted text together with results including typeset mathematics, graphics, GUI components, tables, and sounds.

5.5 DATA ANALYSIS: SPSS

Several computer packages have been developed to assist researchers in the task of data analysis. They include MS Excel, Mintab and SPSS. These statistical software packages have indeed made the task of researchers easier by facilitating data analysis. MS Excel is a part of MS Office package developed by Microsoft Corporation. There are several user-friendly features in MS Excel. Even if one were not formally trained in MS Excel, one can use interactive 'help' feature to navigate. It uses spreadsheets and help in handling numerical data. There is a tool for data analysis available as a part of MS Excel. It can be used to perform various types of analysis on spreadsheet.

In 1968, Norman H. Nie, C. Hadlai Hull and Dale H. Brent developed SPSS which is now used widely in academic, business, government and other environments. 'SPSS' stands for 'Statistical Product and Service Solutions'. The company owning SPSS seeks to 'drive the widespread use of data in decision making'. SPSS is a statistical software package which has the following functions on its menu bar: Data, Transform, Analyze and Graph. These facilitate data analysis and a variety of numerical operations like tables, graphs, correlation, regression analysis, non-parametric tests, comparing means by one-way ANNOVA test and two-way ANNOVA test, etc.

SPSS is a comprehensive, integrated software package for statistical data analysis. SPSS for Windows allows one to store data, perform transformations and analyses, and produce charts and graphs of results. Data are entered using a spreadsheet and results are displayed in a separate output window. The data and the output can be saved independently for the next work session. The output tables can be copied to a word processing application for inclusion in papers. There are useful tips on how to create a data set, entering data into a data set, working with a data set, data directory, analysis of data set, computing mean, standard deviation, performing a number of statistical tests like 't' test and chi-square test, etc.

Terminology Used in SPSS

The University of Lincoln has developed a guide which serves as a brief introduction to statistical language as required by SPSS. It requires data in a format where cases

NOTES

are represented by rows, and variables by columns. A collection of data to be analysed is called a *dataset*.

A *case* is the basic unit of analysis; for example, this might be an animal being used in a medical experiment; a single person filling in a questionnaire or a plant used in a genetic study. The items measured and recorded in each case are the variables. These could be a reply to a question of a survey, like the body mass of an insect or length of leaves of a plant, etc. Variables may either be *categorical* or *continuous*.

When within a variable, values are measured on a continuous scale between appropriate limits, for example, 1 to 10, the variable is termed as continuous. In practice, values are measured till a certain degree of accuracy, may be till two decimal places or till the nearest integer. Mean, standard deviation and other such descriptive statistics are calculated often on *continuous variable*.

A variable with values that may only come from a fixed set of choices is named a *categorical variable*. The categories are known as levels and the variables known as factors. There might be no relationship between the categories; gender and occupation belong to such type of category. Bar charts and frequency tables are examples of such type of such categories. Tables may be created for analysis by cross-tabulating categorical variables. An ordered categorical variable may have values divided into low, medium and high ordered sequences. The median for the variable may be calculated in case the values are numeric. Continuous variables may form ordered categorical variable by grouping the values, example, age ranges such as 13–16; 17–20, etc.

Physical measurement has a level of precision. We know that the space or interval between 2 and 3 inches is the same as that between 3 and 4 inches on a ruler. This measurement scale possesses the *interval property*. Most measurements also possess *ratio property*. This means that when the measurement scales states that you now have twice as many units of the variable as before, you really do; for example, when it takes someone 20 minutes to do an exercise, it has taken him twice as long to do that exercise as someone who took 10 minutes. Scales possessing interval property and ratio property are known as *ratio scales*. *Interval scales* do not have ratio property but only interval property. These occur very rarely, e.g., measurement of temperature.

Physical measurement has a level of precision. The inch marks on a ruler are equally spaced; we know that the space or interval between 2 and 3 inches is the same as that between 3 and 4 inches. It can be stated that this measurement scale possesses the *interval property*.

Most measurements also possess *ratio property*. This means that when the measurement scales states that you now have twice as many units of the variable as before, you really do; for example when it takes someone 20 minutes to do an exercise, it has taken that subject twice as long to do that exercise as someone who took 10 minutes. Scales which have the ratio property in addition to the interval property are known as *ratio scales*.

NOTES

Some scales have the interval property and do not have the ratio property. These are called *interval scales*. These occur very rarely, e.g., measuring temperature.

Statistics program

Statistics included in the base software:

- Descriptive statistics: Cross tabulation, frequencies, descriptives, explore, descriptive ratio statistics
- Bivariate statistics: Means, t-test, ANOVA, Correlation (bivariate, partial, distances), Nonparametric tests
- Prediction for numerical outcomes: Linear regression
- Prediction for identifying groups: Factor analysis, cluster analysis (two-step, K-means, hierarchical)

The many features of SPSS are accessible via pull-down menus or can be programmed with a proprietary 4GL *command syntax language*. Command syntax programming has the benefits of reproducibility, simplifying repetitive tasks, and handling complex data manipulations and analyses. Additionally, some complex applications can only be programmed in syntax and are not accessible through the menu structure. The pull-down menu interface also generates command syntax. This can be displayed in the output though the default settings have to be changed to make the syntax visible to the user; or can be pasted into a syntax file using the 'paste' button present in each menu. Programs can be run interactively, or unattended using the supplied Production Job Facility. Additionally, a 'macro' language can be used to write command language subroutines and a Python programmability extension can access the information in the data dictionary and data and dynamically build command syntax programs. The Python programmability extension, introduced in SPSS 14, replaced the less functional SAX Basic 'scripts' for most purposes, although SaxBasic remains available. In addition, the Python extension allows SPSS to run any of the statistics in the free software package R. From version 14 onwards SPSS can be driven externally by a Python or a VB.NET program using supplied 'plug-ins'.

SPSS places constraints on internal file structure, data types, data processing and matching files, which together considerably simplify programming. SPSS datasets have a two-dimensional table structure where the rows typically represent cases (such as individuals or households) and the columns represent measurements (such as age, sex or household income). Only two data types are defined: numeric and text (or "string"). All data processing occurs sequentially case-by-case through the file. Files can be matched one-to-one and one-to-many, but not many-to-many.

The graphical user interface has two views which can be toggled by clicking on one of the two tabs in the bottom left of the SPSS window. The 'Data View' shows a spreadsheet view of the cases (rows) and variables (columns). Unlike

spreadsheets, the data cells can only contain numbers or text and formulas cannot be stored in these cells. The 'Variable View' displays the metadata dictionary where each row represents a variable and shows the variable name, variable label, value label(s), print width, measurement type and a variety of other characteristics. Cells in both views can be manually edited, defining the file structure and allowing data entry without using command syntax. This may be sufficient for small datasets. Larger datasets such as statistical surveys are more often created in data entry software, or entered during computer-assisted personal interviewing, by scanning and using optical character recognition and optical mark recognition software, or by direct capture from online questionnaires. These datasets are then read into SPSS.

SPSS can read and write data from ASCII text files (including hierarchical files), other statistics packages, spreadsheets and databases. SPSS can read and write to external relational database tables via ODBC and SQL.

Statistical output is to a proprietary file format (*.spv file, supporting pivot tables) for which, in addition to the in-package viewer, a stand-alone reader can be downloaded. The proprietary output can be exported to text or Microsoft Word. Alternatively, output can be captured as data (using the OMS command), as text, tab-delimited text, PDF, XLS, HTML, XML, SPSS dataset or a variety of graphic image formats (JPEG, PNG, BMP and EMF).

SPSS server is a version of SPSS with a client/server architecture. It had some features not available in the desktop version, such as scoring functions (Scoring functions are included in the desktop version from version 19).

Versions

Early versions of SPSS were designed for batch processing on mainframes, including; for example, IBM and ICL versions, originally using punched cards for input. A processing run read a command file of SPSS commands and either a raw input file of fixed format data with a single record type, or a 'getfile' of data saved by a previous run. To save precious computer time an 'edit' run could be done to check command syntax without analysing the data. From version 10 (SPSS-X) in 1983, data files could contain multiple record types.

SPSS version 16.0 runs under Windows, Mac OS 10.5 and earlier, and Linux. The graphical user interface is written in Java. The Mac OS version is provided as a universal binary, making it fully compatible with both PowerPC and Intel-based Mac hardware.

Prior to SPSS 16.0, different versions of SPSS were available for Windows, Mac OS X and Unix. The Windows version was updated more frequently, and had more features than the versions for other operating systems.

SPSS version 13.0 for Mac OS X was not compatible with Intel-based Macintosh computers, due to the Rosetta emulation software causing errors in calculations. SPSS 15.0 for Windows needed a downloadable hotfix to be installed in order to be compatible with Windows Vista. The latest version of SPSS is 19.0.

CHECK YOUR PROGRESS

6. What is Microsoft Outlook?
7. What is SPSS?

5.6 SUMMARY

- Progressive educational institutions can conduct classroom sessions using what is known as CBT. In this case, each student then sits at a computer terminal which operates software that presents course material in interactive sessions.
- Many research studies cannot be carried out without the use of computer particularly those involving complex computations, data analysis and modelling. Computer in scientific research is used at all stages of study, i.e., from proposal/ budget stage to submission or presentation of findings.
- In the setting of a distributed client-server environment, typically, a digital library will require both server side and client side software and hardware.
- A telecommunication network is used by the client and the server for communication purpose over a well-defined protocol.
- In digital libraries, the data is saved in digital files, which could be created from the digitization of original record materials or from the most common or standard data formats.
- A video file is a moving image recording, and it has audio synchronized created with original digital video or analog. Some of the characteristics that are important in a video file are pixel array, frame rate per second, aspect ratio, bit rate, field order, color space and the definition being standard or high.
- When the various components of a digital library and the services associated with them can be interchanged both functionally and logically, they are said to be interoperable. These components are structured according to a set of well-defined publicly known interfaces.
- For libraries that work with electronic content, a huge matter of concern is the archiving and preservation of content. This relates to content that might have been converted in-house, acquired via subscription services, purchased in digital media, or any other way.
- Linking is a feature that is well supported by digital libraries. Since the Web is a system based on hypermedia, it enables the linking between the various electronic resources that are stored on servers which might even be located at different geographical locations with massive geographical distances between them.

NOTES

- DOI or Digital Object Identifier is a unique identifier which is used to support links to bibliographic databases or to full text databases. DOI serves as a system that provides interoperability for exchanging and identifying intellectual property that is available in the digital environment.
- The proposal for a digital library must contain goals, benefits, scope, target users, time needed for establishing the library, deliverables, issues associated with the implementation and the costs that will be incurred.
- MS Office 2010 is the popular and extensively used application suite of Microsoft with the number of highly beneficial applications and products such as Word, PowerPoint, Excel, Access, Publisher and Outlook.
- Microsoft Word 2010 is a word processing application that has improved features of searching, editing words and pictures. It includes the Screenshot feature, Photoshop feature, visual designs and formatting tools that allow you to create professional looking documents and files.
- Microsoft Excel is a useful spreadsheet application of MS Office suite 2010 with some new features and innovations such as Slicers, Sparklines, Pivot Table, etc.
- A wide range of publications and marketing material, such as brochures, newsletters, invitations and business cards can be produced efficiently using this desktop publishing platform.
- Microsoft Office 2010 provides a feature of Protected View. Documents downloaded from the web are opened in protected view.
- One of the unique features which were introduced in Microsoft Office 2010 is the feature of Backstage View. Backstage View is a group of commands which are used to perform different types of tasks on documents. It provides you with the features of managing documents, sharing files and providing greater security.
- Sparkline and Slicers are the two important features which are included in MS Excel 2010.
- Fields and Table tabs are the new tabs which are introduced in Microsoft Office 2010. These are actually the replacement for the **Datasheet** tab of the previous version.
- Microsoft Outlook is the popular tool for organizing the business. It has all the functionalities of the previous editions, along with the additional feature of storing large amount of e-mails, contacts and tasks on the system.
- SPSS is a comprehensive, integrated software package for statistical data analysis. SPSS for Windows allows one to store data, perform transformations and analyses, and produce charts and graphs of results. Data are entered using a spreadsheet and results are displayed in a separate output window. The data and the output can be saved independently for the next work session.
- The many features of SPSS are accessible via pull-down menus or can be programmed with a proprietary 4GL command syntax language. Command

NOTES

syntax programming has the benefits of reproducibility, simplifying repetitive tasks, and handling complex data manipulations and analyses.

- SPSS places constraints on internal file structure, data types, data processing and matching files, which together considerably simplify programming.
- SPSS can read and write data from ASCII text files (including hierarchical files), other statistics packages, spreadsheets and databases. SPSS can read and write to external relational database tables via ODBC and SQL.

5.7 KEY TERMS

- **Microsoft Access:** A relational database application meant for allowing users or programmers to manage, sort and filter large amounts of data. It is a powerful tool for creating and running database solutions and producing meaningful reports.
- **Microsoft Word:** A word processing application that has improved features of searching, finding, editing words, pictures and also making additions in a proper and unambiguous way over its previous editions.
- **Microsoft Excel:** A useful spreadsheet application of MS Office Suite 2010 with some new features and innovations such as Slicers, Sparklines, and Pivot Table, etc.
- **Microsoft OneNote:** A note-taking and information gathering program. With this, files or documents can be accessed from anywhere, that is, from home, workplace or on-the-go.
- **Microsoft Outlook:** A web application and a personal information manager for e-mail communication using which you can store e-mails on your system. This tool is also well-known for sending, receiving e-mails, managing messages, accounts and contacts.
- **Microsoft PowerPoint:** A presentation program which is built for the Windows operating system and also made available for use in other operating systems. It is used for creating and organizing complex, animated and transitive slide shows.

5.8 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The components needed to build a digital library fall into five categories:
   a) Collection Infrastructure
   b) Access Infrastructure
   c) Computer and Network Infrastructure
   d) Digital Resource Organization
   e) Manpower Training

NOTES

2. The digital constituents of a digital library are as follows:
    a) Collections available with the library in digital formats
    b) External digital collections to which the library gains access
    c) In-house content that was created in digital format (born digital) for printing
    d) Collection of existing print media that is converted to digital format
    e) Scanned documents converted to machine readable format via OCR programs
    f) Organized and structured guide to electronic resources available on the Internet via portal sites or gateways to the electronic collections on theWeb
3. For a digital resource, the access infrastructure would comprise the following:
    a) For library Catalogs, WebPACs and/or multi-WebPACs would be used.
    b) For specialized image-based local collection, specialized collection Websites would be used.
    c) For Web resources, portals or subject gateways would be used.
    d) For local collections, search and browse interface would be used.
4. Some of the software that are generally needed in a digital library are as follows:
    a) Software for scanning an image capturing
    b) Software for image manipulation and enhancement
    c) Software to automate the integrated library systems
    d) Software to run Web servers
    e) Software to perform information retrieval
    f) Software for OCR
    g) Software for database management
    h) Software for right management
5. Digital Object Identifier (DOI) System has four components:
    a) Enumeration: Assigning an alphanumeric string to the identified digital object
    b) Description: Forming a description of the identified entity
    c) Resolution: Making the identifier actionable
    d) Policies: Rules governing the system's operation
6. Microsoft Outlook 2010 is a web application and a personal information manager for e-mail communication using which you can store e-mails on your system. This tool is also well-known for sending, receiving e-mails, managing messages, accounts and contacts.

7. SPSS is a comprehensive, integrated software package for statistical data analysis. SPSS for Windows allows one to store data, perform transformations and analyses, and produce charts and graphs of results.

NOTES

5.9 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define CBT.
2. List the components of digital library.
3. What are the software that are generally needed in a digital library?
4. Write a short note on digital libraries.
5. Outline the concept of IPR and copyright in digital libraries.
6. Differentiate between linking, cross-linking and interlinking in digital libraries.
7. Define SPSS.
8. What are the different versions of SPSS?

Long-Answer Questions

1. Explain the components needed to build a digital library.
2. Discuss the various formats and media types used in a digital library.
3. Describe the process of creating and developing a digital library.
4. What are the different applications of MS Office 2010?
5. What are the new features and tools incorporated in office 2010 in comparison to earlier versions?
6. Explain the terminalogy used in SPSS.

5.10 FURTHER READING

William, Brin K. Stacey C and Sawyer. 2007. *Using Information Technology: A Practical Introduction to Computers and Communications.* Ohio, US: McGraw-Hill Irwin.

Long, Larry and Nancy Long. 2004. *Computers: Information Technology in Perspective*. Upper Saddle River, New Jersey: Prentice Hall.

Sinha, P. K. 2004. *Computer Fundamentals: Concepts Systems and Applications*. New Delhi: BPB.

Norton, Peter. 2008. *Introduction to Computers*. Ohio, US: Tata McGraw.

Basandra, Suresh K. 2007. *Computers Today*. New Delhi: Galgotia.