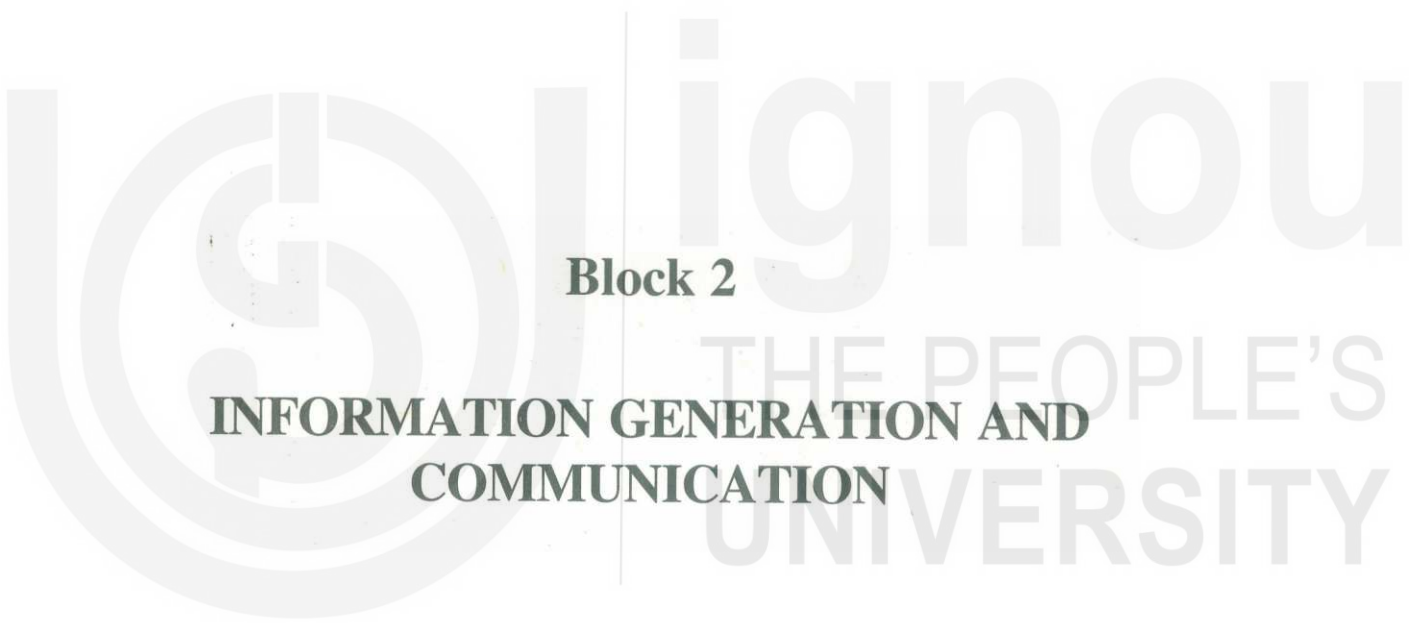


The background of the page features a large, light gray watermark of the Open University logo. The logo consists of a stylized circular emblem on the left, followed by the word "Open" in a large, sans-serif font, and "University" in a smaller font below it. The text "ignou" and "THE PEOPLE'S UNIVERSITY" are also visible in a large, light gray font, partially obscured by the main title.

Block 2

**INFORMATION GENERATION AND
COMMUNICATION**

UNIT 5 INFORMATION COMMUNICATION PROCESS, MEDIA AND DIFFUSION

Structure

- 5.0 Objectives
- 5.1 Introduction
- 5.2 Information
 - 5.2.1 Conceptualisations of Information
 - 5.2.2 Information as Commodity / Resource
 - 5.2.3 Information as Data in the Environment
 - 5.2.4 Information as Representation of Knowledge
 - 5.2.5 Information as Part of Communication Process
- 5.3 Communication: Concept and Genesis
 - 5.3.1 Definition of Communication
 - 5.3.2 Channels of Communication
- 5.4 Types of Communication
 - 5.4.1 Non-vocal and Vocal Communication
 - 5.4.2 Non-verbal and Verbal Communication
 - 5.4.3 Intra-personal, Interpersonal, Group and Mass Communication
- 5.5 Communication Process
 - 5.5.1 Communication Cycle
 - 5.5.2 Elements of Communication Process
 - 5.5.3 Communication Media
 - 5.5.4 Models of Communication
- 5.6 Information Diffusion
 - 5.6.1 Meaning
 - 5.6.2 Diffusion of New Ideas or Innovations
 - 5.6.3 Patterns of Diffusion Process
 - 5.6.4 Factors Affecting Diffusion of Information
- 5.7 Diffusion of Technological Information
- 5.8 Models of Information Diffusion Process
 - 5.8.1 Epidemic Model
 - 5.8.2 Economic Theory Model
 - 5.8.3 Stock Adjustment Model
 - 5.8.4 Vintage Model
 - 5.8.5 Information Diffusion Theory
- 5.9 Information System for Diffusion
 - 5.9.1 R&D Information Flow
 - 5.9.2 Information Services, Information Institutions and Diffusion
 - 5.9.3 Gatekeeping of Technical Information
 - 5.9.4 Information Policy in Support of Information Diffusion
- 5.10 Summary
- 5.11 Answers to Self Check Exercises
- 5.12 Keywords
- 5.13 References and Further Reading

ignou
THE PEOPLE'S
UNIVERSITY

5.0 OBJECTIVES

After reading this Unit, you will be able to:

- understand the meaning of the term information and its different conceptualisations;
- explain the concept of 'communication' and identify different types of 'communication';
- describe communication process and analyse the different elements of "communication process";
- distinguish the different communication media;
- define what "information diffusion" is;
- discuss the patterns of diffusion process;
- explain the various factors affecting information diffusion process;
- elucidate different "models" of diffusion; and
- discuss the role of information systems in information diffusion.

5.1 INTRODUCTION

There is a growing recognition that information and communication are interrelated in very fundamental ways. The disciplines of Communication and Information/ Library Studies have a long tradition of *common interests and concepts*. This tradition has provided increasing interdisciplinary linkages between the two fields. But, "the intersection of Communication and Information Studies cannot be explained only, or even largely, in terms of what might be termed a "Scholarly push". Far more crucial to growing interest in the communication-information relationship, is the momentum and rhetoric of the market place. These forces give urgency to the need for frameworks that clarify the theoretical relationship between communication and information by identifying similarities and differences, exploring ways in which the differences may be complementary and/or supplementary, and facilitating theoretical integration in general". In this Unit, an attempt is made to explain to you the relationship between the concepts "information and communication". In the process, focussed perceptions of 'information' have been discussed. When considering the nature and role of information, from whatever perspective, meaning becomes central. Information is something that one person communicates to another. There is widespread confusion associated with defining of 'communication'. "Not so, however, the attempt to investigate the relationships between information as a phenomenon and communication as process". So far as definitions of the concept communication is concerned, the one employed by Ruben has been commended (*i.e.*)...The process through which individuals in relationships, groups, organisations and societies *create, transmit and use* information to organise with the environment and one another. This Unit discusses the aspects of definition of communication, types of communication, communication process and communication media. Also the Unit describes information diffusion, patterns of diffusion and its implications

to information systems, services and information professionals. It is hoped all these aspects will be useful in providing adequate knowledge to the participants of MLIS programme.

5.2 INFORMATION

The question “What is information”? has eluded answer for a long time. Of course, definitions proliferate and interpretations multiply even though nobody has said the last word. The best minds in information science give considerable importance to the concept and continue to discuss it. As T.D.Wilson observed “it was not so much the definition itself that mattered as the uses to which it was put” [T.D.Wilson, 1981].

The uses to which information is put are countless and therefore, the power of those *metaphors* which link the human and the social organisms in describing information as the *life blood of society* appear to be justified. Without a regular and uninterrupted flow of meaningful information, society would quickly run into difficulties, with business and industry, education, leisure, travel and communications, national and international affairs, etc. In fact, the advanced societies increasingly depend on the enabling powers of information and communication technologies (ICT). It might be stated that information has been a *significant element* in the life of all societies. Therefore, it is common for each society to have its own characteristic information and, knowledge base. However it may be noted that in the case of developing countries it might well be that *indigenous knowledge* rather than that contained in the *World's stores of information* is likely to be more relevant [Menou, Michel, 1994].

5.2.1 Conceptualisations of Information

Information is a concept that is applied in multiple ways in everyday usage as well as in the research literature [Mc Credie, *etal.* 1999]. Here, we identify and illustrate the range of what is meant by information, hoping to shed light on assumptions about information.

5.2.2 Information as Commodity/ Resource

Some researchers emphasise information as a thing or resource [Arrow, 1979, Bates, 1988, Buckland, 1991], a commodity that can be produced, purchased, replicated, distributed, manipulated, passed along, controlled, traded and sold. By focusing upon information-as-thing Michael Buckland [1991] brings into consideration such phenomena as objects and events, seeking in the process to widen our perception of what constitutes data and documents. Buckland's approach is useful both for its breadth of vision and of the way it manages to weave together such concepts as information, knowledge and communication. Also, by demonstrating information as a thing that is embodied in objects or information systems information can be tangible, as opposed to knowledge, which is inherently intangible. By this approach, Buckland sheds more light on the traditional approach of Machlup who saw information as intangible, involving either the telling of something or that which was being told. Buckland points out, in order to communicate knowledge it must be expressed or represented in some physical way as a *signal*, *text* or communication. This

conceptualisation is consistent with a model of sending information as a message from sender to receiver. It may include an assumption that the receiver will interpret and understand the message as intended by the sender.

There are, however, difficulties with the treatment of information as a *resource*. One such difficulty is associated with its intangible nature and also the fact that unlike physical matter, information *per se* is not subject to the kind of laws that would enable it to be treated in economic terms. Some writers have also warned against the dangers of a pre-mature acceptance of the arguments in favour of the *resource characteristics of information*. For instance, Michael Menou is rather sceptical of those claims, which would portray information as a *critical resource in all circumstances- individual, organisational and societal*. [Menou, 1994]. He opines that such claims need to be supported by more than anecdotal evidence and a limited body of empirical research. He emphasises that information is seldom identified to the *level of specificity* required to demonstrate its impact on any given situation or problem. Though, such criticisms are not taken lightly, they have to be balanced against the fact that the notion of *information-as-resource* is by now *well-established* in fact, most evidently in recognition of the related concept of a '*marketplace of ideas*', as reflected in profusion of national and international laws and policies relating to trade in information and its associated goods and services. This aspect serves to highlight another practice of viewing information as a *commodity*.

The notion of *information-as-commodity* has gained considerable currency in recent years, commodity comprising all types of information services, as also transborder data flows. It is stated that the concept of *information-as-commodity* is wider than that of *information-as-resource*, as it incorporates the exchange of information among people and related activities, as well as its use. The concept of information as a commodity is connected closely to the concept of *value chains* with commoditised information gaining in value as it progresses through different steps of creation, processing, storage, distribution and use. It might be stated here that the problems include those of measurement and appropriability, and with treating information as a *public good* as just one more free market commodity.

5.2.3 Information as Data in the Environment

Some writers tend to view information more broadly to include data in the environment available for interaction with human information processing capabilities. This data category includes *objects, artifacts, sounds, smells visual and tactile phenomena, activities, events or the phenomena of nature*. According to Buckland, it is easy to assume that all communication is *intentional*. In practice, however, one is informed also by perception of things that are communicated *unintentionally*. In other words, *neither the individual nor the data* unintentionally engage in communication. "The discoveries one makes in the process of casual (non-goal-directed) browsing [Chang Di Rice, 1993] or the references another makes about an individual's character or performance based on observation of the individual's behaviour when exposed to view, particularly when the individual is unaware of being observed, serve as additional examples of unintentional communication available when one attends to information as data in the environment".

“Taylor (1996) specifically uses the environmental approach in his concept of information use environments, which are intersections of:

- a) sets of people (professors, entrepreneurs, special interest groups, socio-economic groups);
- b) classes of problems (well- versus ill-structured, complex versus simple, assumptions agreed upon versus not agreed upon and familiar versus new);
- c) work settings (involving attitudes towards information, task domain, information access and history/experience, constraints and opportunities) and
- d) what constitutes problem resolution. These environments are contexts for various kinds of information needs.

5.2.4 Information as a Representation of Knowledge

Some of the researchers in information science view information as a representation of, or pointer to, knowledge. The tradition of scientific method and Scholarly publication is a clear example of this conceptualisation [Lievrouw, 1988]. Card catalogues or databases of citations to scientific documents illustrate an abstraction of representation of information, such as in documents, books and periodicals. Traditionally, this view of information has been based on the assumption that printed document is a primary representation of knowledge. Recent years have seen a proliferation of alternatives to print, such as representation of knowledge available on video or audiotape, videodisc, CD-ROM, Internet or other electronic or computer media.

5.2.5 Information as Part of the Communication Process

Some researchers consider information as part of communication process, as part of human behaviour in the process of moving through space and time to make sense of one's world. [Atwood and Dervin, 1982]. From this conceptualisation, meanings are in people rather than words or data and *knowledge is what users do with data rather than what data do to users*. An assumption relating to this conceptualisation is that understanding must be based on observation of human behaviour in information seeking and sense making processes and on the meanings intended and interpreted by participants.

In the foregoing paragraphs, different conceptualisations of the concept information have been discussed and explained to you so that you may be in a position to understand focussed perceptions of the concept 'information'. Let us now attempt to know what communication is, and the different aspects associated with the concept.

Self Check Exercise

- 1) Briefly explain the different conceptualisations of information.

Note: i) Write your answer in the space given below.

- ii) Check your answer with the answers given at the end of the Unit.

5.3 COMMUNICATION: CONCEPT AND GENESIS

The word 'Communication' is derived from the Latin word 'Communis', which means 'sharing'. Communication encompasses all methods used by living beings to express or to convey a kind of thought or feeling. When we communicate with someone, we try to establish a certain degree of commonness with the person we communicate with.

If we look into the evolution of communication, we notice that communication has taken place in many ways and has taken many forms throughout the history of mankind. In the beginning sign languages or expressive body languages were forms of human communication. With the evolution of language, spoken word became the common form of communication. The invention of scripts and writing paved the way for the written form of communication. Written form of communication enabled recording and storing of information for the future generations. Invention of printing provided major technological breakthrough for recording knowledge and information in the form of documents. Other inventions like telegraphy and telephone systems provided mechanisms for people to communicate over long distances. Over the last four decades or so *electronic* and *mass media* became commonplace methods of communication.

5.3.1 Definition of Communication

Communication has been defined in many ways. For example, the Oxford Dictionary defines it as "The imparting conveying or exchanging of *ideas* and *knowledge* whether by *speech*, *writing* or *signs*". On the other hand, Columbia Encyclopaedia defines communication as "The transfer of *thoughts* and *messages* as contrasted with transportation of *goods* and persons. The basic forms of communication are by *signs* (signals) and *sounds* (hearing)".

Also, we come across the following explanations for the concept in literature:

- act or instance of transmitting;
- imparting or interchange of thoughts, opinions or information by speech, writing, etc;
- something imparted, interchanged or transmitted;
- document or message imparting information, opinions, etc;
- process by which information is exchanged between individuals or groups through a common system of symbols, signs or behaviour.

If we analyse the above-mentioned definitions and explanations provided for

the concept, we note that they imply that communication refers to transmission or exchanging of information and that for information to be transmitted, it has to be first mentally conceived, generated or formed by an individual. Only then, can information be communicated to another. The transmission process, could, however, be oral, through visual images or icons, or through written text in any language, or a combination of all these different modes of communication. From what has been stated above, it might be inferred that the media and the method used to communicate ideas through is important. Therefore, let us try to understand the forms and channels of communication.

5.3.2 Channels of Communication

Basically, there are two ways of communication:

- 1) Oral communication,
- 2) Written or Documentary communication.

Oral Communication

Oral communication is a method of communication through speech. It is characteristic of its immediacy and personal touch, which might be difficult to capture in other forms of communication. In this form, there is often an informal or flexible setting (atmosphere), which allows the sender and receiver, interact with questions and answers, comments and response. All this takes place in a given time and results in immediate feedback and permits an instant reaction.

Written Communication

Some basic characteristics of written communication are that it:

- has a more formal perspective, suggesting greater authority and trust worthiness;
- can be a potential record, capable of extended life and of being used again and again;
- can be extracted in parts or portions and can be quoted;
- it is suitable for reading at an individual's convenience, speed and place; and
- is accessible to a large readership which can lead to the benefit of expert criticism and review by peers.

5.4 TYPES OF COMMUNICATION

Communication systems may be classified or grouped into vocal or non-vocal or verbal or non-verbal as per the characteristic of physical phenomenon. They can also be categorised into intrapersonal, interpersonal, group or mass communication. Let us try to understand each one of these categories.

5.4.1 Non-vocal and Vocal Communication

Non-vocal communication refers to the use of *signs, signals and symbols*, gestures, etc in communication process. Signals, signs and symbols are the *three related components* of non-vocal communication process. This type of communication can be found in pre-historic period in all cultures.

There is basic difference between signs and signals. A sign contains meaning of an intrinsic nature while a signal is merely a device by which one is able to formulate extrinsic meanings. In other words, the meaning of a *signal is explicit* while that of a *sign is implicit*. Signals play a prominent role in communication systems. In certain situations signals allow the flow of information and in certain others they prevent. e.g.: use of signals in rail transport system.

Signs usually carry meaning by themselves. The meaning can be understood only by trained staff. Traffic signs are common example of *signs*.

Symbols are generally complex and are not easy to understand. They usually carry seminal meaning. Individual's perception plays a very important role. Example: Symbols used in classification schemes, such as Colon classification. Symbol has been defined as a device with which an abstraction can be made.

These *three* components: signals, signs and symbols have been used in non-vocal communication.

Vocal communication refers to any form of communication that uses the *human vocal chords*. It can range from shouting to spoken languages. Man's vocal apparatus as a device of communication represents an apex of physical and intellectual evolution. It has the potential to express the most basic instructional demands as well as a range of highly intellectual processes including the possible mastery of numerous complex languages, each with a large quantum of vocabulary. The vocal organs permit the production of *sound effects, animal noises, birdcalls* and other *amusing high manipulations of speech*.

5.4.2 Non-verbal and Verbal Communication

Non-verbal communication includes actions such as gestures, vocalisations, facial expressions and other behaviour. Whereas, verbal communication refers to the *use of words*. The era of verbal communication began with the development of language. Communication is generally equated with speaking and writing words (vocabulary) are viewed as the primary means by which information is conveyed from one person to another. Communication by facial expression, by tone of voice, by touch etc. falls under the category of non-verbal communication. Non-verbal communication is of seven types. These are:

- *Para linguistics* is a kind of complementary language which includes meaningful variations in speech corresponding to patterns of loudness, pitch, hesitation;
- *Kinesics* which includes facial expressions, eye movements, gestures, etc;
- *Haptics* which refers to the use of touch to communicate, for example, shaking hands, holding hands, patting, etc;
- *Proxemics* refers to the use of interpersonal space, that is to say how far a person is standing from the other, can communicate different meanings;
- *Dress and appearance*: dress, hair styles, make-up, jewellery, etc, can also constitute non-verbal code of communication;
- *Chronemics* that is to say how time is considered is sometimes a message in itself, a form of non-verbal communication and is called Chronemics; and

- *Iconics, which means interpretation of symbolism, found in objects or designs can be considered a type of communication. The best example of this type of communication is international traffic signs.*

5.4.3 Intra-personal, Interpersonal, Group and Mass Communication

Communication can also be classified in terms of its levels. The different levels of communication are as follows:

- *Intrapersonal communication* means communication within one person, i.e., talking to oneself, reading something, etc. In this type of communication, the source and recipient of information becomes one person. Intrapersonal communication includes a person's thoughts, experiences and perceptions during a communication event. Intrapersonal communication is of different types which include: i) postures, ii) muscle tension, iii) sleep, iv) use of tranquilisers, and v) emotions.
- *Interpersonal communication* means communication between two or more persons. It could be face-to-face, or it could be at a distance with the help of telephone or letters, etc. For effective inter-personal communication we must develop certain skills and tactics, which will help us to establish rapport quickly with new persons. One such skill is to be able to control the communication situation. This does not mean that we must act in a dominant way.
- *Group communication:* In this form of communication many people are involved and the speaker and the listener are shifted among the participants. This type of communication may take place between small groups usually not exceeding 25. It may also take place in large groups where there is communication by one or more persons to an audience of more persons.
- *Public or mass communication:* This generally refers to communication, which takes place between one person or a group of persons through a special media to a large audience. Mostly large-scale communication is disseminated by the media of print, broadcasting, film or the electronic media to large audiences. This type of communication is referred to public or mass communication.

In mass communication the relationship between sources and dissemination is different from that in case of individual or interpersonal communications. Interpersonal communication is one-to-one exchange, whereas mass communication is one-to-many exchange in which a few sources transmit to a great number of destinations.

Self Check Exercise

2) Identify different types of communication with the help of suitable examples.

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

5.5 COMMUNICATION PROCESS

Communication is not something that happens in a disjointed fashion. It is a transfer of a cluster of transactional functions of body and mind of one individual to another. Communication is said to possess the following characteristics: intention of the communicator, nature of communication, emphasis on action, measurability of action, comparability of intention and measured action.

Communication process always requires at least three basic elements: the source, the message and the destination. In fact, source is the point at which messages originate. It is also referred to as *sender* or initiator of information. A source may be an individual (speaking, writing, gesturing, etc.) or a communication organisation like newspaper, publishing house, television station, etc. Message may be in the form of written or printed text, sound or light waves in the air, impulses in the electric current, or in the form of any other signal which is capable of being interpreted meaningfully. Destination is the intended target of the message. Again, destination could be an individual or a group or a mob. Destination is the final link in the communication chain.

It may be stated here that in communication process, we try to establish commonness between the *sender* and the *recipient* of message. For this to happen, two more steps need to be added in practice. The steps are: the encoding of the message to be transmitted and a *mechanism to decode it at the receivers' end*.

The establishment of commonness between the sender and the receiver depends a lot on the field experience of the two. If the field of experience between the two is different, it is difficult to establish commonness between the two. In other words, the source can encode and the destination can decode the message in terms of experience each has had. It might therefore, be said that communication process is one of transmission and reception, the passing of ideas, information and attitudes from person to person. All the steps in communication process must be accomplished with high degree of efficiency if communication is to take place.

5.5.1 Communication Cycle

Hitherto we have been discussing the linear model of communication process in which the source transmits message and the destination receives it. In regular practice, however, the communication process does not end with the receiving of the message by the destination. It must be stated that communication is a continuous process in which each individual or institution functions as the transmitter as well as receiver of messages. In other words, communication process works in a cycle, in which each individual functions as the encoder, interpreter and the decoder of information alternatively. In fact, communication is interactive multidimensional process.

5.5.2 Elements of Communication Process

Most of the communication systems, whether sophisticated or not, are perceived to possess the following basic elements:

- Information source
- Encoder
- Message
- Communication Channel
- Noise
- Decoder and
- Receiver/Destination.

In other words, the act of communication has to originate from a source. The idea/ message of communication has to be encoded using symbols. The process of translating ideas, feelings and information into a code is known as encoding. The device/ mechanism used to carry out the message is called a channel. The success or failures of communication mainly depends on the channel used.

Self Check Exercise

- 3) Explain the basic elements of a communication process.

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

5.5.3 Communication Media

Media is obviously the plural form of 'medium' – which is the conductor, the channel, and the means by and through which something is transmitted. In other words, medium is the channel through which messages or information are transferred to the receiver. A medium could be a spoken word or printed word or it could be an electronic message. The term communication media is often used to refer to both the channel and the source.

Communication media can broadly be classified into conventional and non-conventional media. Conventional media mainly comprises print media, including graphic media while non-conventional media comprises mainly electronic media including optical and hyper media.

Let us discuss, the most important of the media namely print media and electronic media.

Print Media

Print media include: books, learned periodicals, conference proceedings, magazines; newspapers etc. The main function of these media is to provide

information, and in certain cases provide entertainment also. The audience for the print media could both be specialised, diverse and large. As a means of communication the book has monopolised for centuries, followed by the periodical and other forms of communication media, which make available more information for use. Other types of print media like indexing and abstracting periodicals, directories, conference proceedings, etc. provide information about where to find information from primary sources like books, periodicals etc. Print media still remains a powerful communication media.

Electronic Media

Electronic media refers to such media where messages are sent by such mechanisms as telephone, telegraph, facsimile, etc. For example, broadcasting which is carried out by radio and television provides with news and entertainment to general public. These different media are briefly discussed in this section.

- **Telegraph:** This is one of earliest methods of modern telecommunication. It was invented by Samuel Morse in 1830. The invention of telegraph has increased the speed of human communication. It involved the manual keying of messages using a code known as *Morse code*. The code is built up of short and long pulses of current in the telegraph wire, with the transmitter's Morse key. One of the major developments in telegraphy is the invention of Baudot code system. In this system the sender could use a sort of typewriter keyboard; each key press would automatically generate a 5-bit combination (word of current pulses or pulse absence). At the receiving end with a special output unit could reproduce the message, either by punching holes in code patterns in a paper tape or by putting readable characters directly on to a paper tape.
- **Telephone:** The development of the telephone represented a solution to the technological problem of conversion of the sound patterns of human voice into electrical patterns. Telephone happens to be one of the largest established methods of electronic information transfer. It differs from the telegraph in that the electric current carries the more complex patterns of human voice over the wire. Originally telephonic communication between two persons could take place only through a special cable link. The concept of multiplexing is a major development in the direction. This allows a number of different signals to pass through the same links at the same time.
- **Broadcasting Media:** This makes long distance message transmission possible between sources and receivers without the need of transportation or a direct physical (wire) link. In case of broadcasting message is transmitted in the form of energy waves in the electromagnetic spectrum. Broadcasting includes mass communication devices like radio and television, through which messages are sent out regularly by professional communicators through electronic media to large and diverse groups of audience. Broadcasting as a technology implies transmitting electronically generated radio waves to receiving devices. Radio and telecasting stations are the centres for transmission. Recent development in this field includes the use of satellite communication systems. Space satellites now provide

a worldwide television and telephone network that links every country on the globe.

- **Electronic Mail:** It is conceived primarily as an alternative to the conventional postal mail service. It is used for the transmission of messages or documents in an electronic form. In most of e-mail systems, transmission is accomplished via telecommunications network designed for data transmission. The greatest advantage associated with e-mail is that it overcomes the delay connected with traditional postal service. Also, e-mail is economical compared to conventional mail system. The input to and output from an e-mail system can be via a video terminal, or a word processor with a printer, facsimile machine or any data terminal including computer vision and voice communication systems.
- **Videotext and Teletext:** Both these are interactive information services and that these allow individuals to request frames of information. But, they are somewhat different technologies. In other words, Teletext is delivered over the air while videotext is delivered by wire. In videotext, a person can request for information from a central computer for delivery over telephone or cable television lines. Teletext in contrast is a system that enables the users to view on request frames of information already being transmitted but invisible on their television screens.
- Videotext is a generic term that refers to both Videotext and Teletext. Teletext and Videotext are the most radical of the new communication technologies which have brought the powers of computer to the home TV Set and have transformed the entertainment medium into an information appliance.
- It may be mentioned here that electronically the society has progressed from the wired era of telephone and telegraph to wireless era of broadcasting and now stands on the threshold of era of integrated grid. Integrated grid refers to the communication infrastructure that is now taking shape world over, with internet revolutionising the communications network. The developments in ICT are changing the entire nature of telecommunications.

In the foregoing sessions we have learnt about communication media: conventional as well as non-conventional. Since the conventional media are inadequate to cope with the contemporary information environment, the non-conventional media have been developed. These media have not only enhanced the speed of communication, but also increased the chances of information access. Computer mediated communication systems have brought in these advances.

5.5.4 Models of Communication

Models simplify reality, select important elements and indicate the relationships that exist among key elements. Similar is the case with models associated with human communication. Many models of communication include certain basic concepts such as a sender, a process of encoding into signals or symbols, a message, a channel, a receiver, a relationship, a process of decoding, a range of things to which message refers (referents) and an actual or probable effect,

intended or not. Some models also include feedback mechanism between sender and receiver.

Communication models help us to understand how communication systems work. The best-known communication model is that of *Shannon and Weaver*.

Shannon and Weaver’s Model

In this model, a message emanates from an information source. It is converted into a signal or series of signals by a transmitter. Enroute, this signal is mixed with or contaminated by ‘noise’, that is to say, various kinds of unwanted interference coming from ‘noise’ sources. The received signal is decoded by a receiver, being converted back into the original message, more or less, which is what the receiver (or destination) receives.

Shannon and Weaver’s model seems straightforward. It recognises that encoding and decoding occur. It takes into account the problem of interference: ‘noise’ is a term from electrical engineering denoting electro-magnetic interference. Noise arriving with the signal makes it more difficult to decode accurately, as we all know from everyday experience.

This model has been criticised, however, because it is linear. That is to say, it accounts only for the act of sending and receiving. It is a one way model, based on engineering. In fact, this was the dominant model for a quarter century. The thinking behind it was probably reflected in the thinking of many top broadcasters, for example, thought of themselves as primarily responsible for sending out messages, and perhaps for reducing ‘noise’ to a minimum, but not for receiving them back. Broadcasting organisations put most of their resources into production and transmission of programmes, and much less into listening to audience opinion.

To convert Shannon-Weaver model into a two-way model of communication, we must add a feedback channel, complete with transmitter, encoder, noise, decoder and receiver, all to deal with messages running from the receiver to the sender. It could be said that we are simplifying working Shannon and Weaver’s model in reverse, but in fact we are working it almost simultaneously in both directions. In this two-way model, the feedback channel can be vitally important to the original receiver as a means of seeking clarification of the original message. Feedback channels, verbal and non-verbal, continually serve this purpose in human communication.

It may be stated that Shannon-Weaver’s model has been criticised because it did not provide for the complex nature of relationships among humans, some of the scholars like Schramm and Kincaid have advanced models intended to fill this need. Let us briefly learn about some of these communication models.

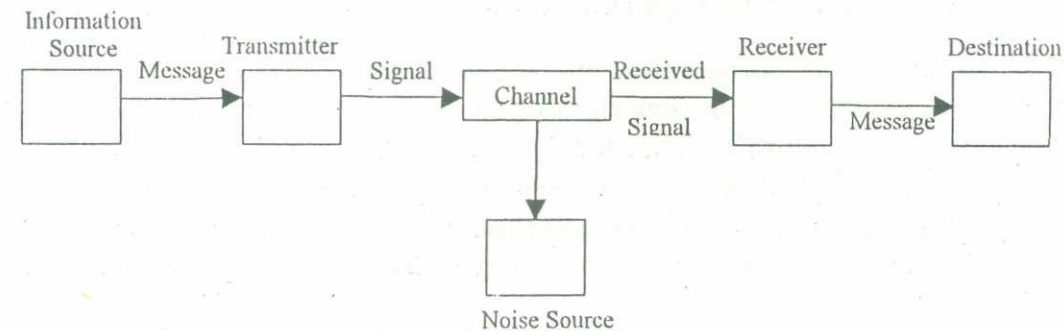


Fig. 5.1: Shannon Weaver Model of Communication Process

The models that need to be considered are:

- Lasswell's model
- George Gerbner's model
- The Schramm model

Lasswell's Model

Lasswell, a U.S. Political Scientist developed this model in 1948. This model is very useful to split the communication process into different components. The value of this model lies in its use as a structuring device and in situations in which the sender has a clear interest to influence the receiver. The model is summarised as "Who says what in which channel to whom with what effect"?

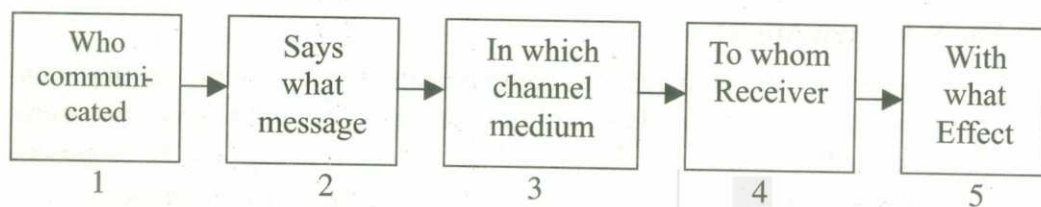


Fig. 5.2: Lasswell's Model of Communication Process

Lasswell's model can be interpreted in the context of Libraries in the following manner:

Who?	Authors, publishers, research bodies, translators, professional bodies, etc.
What?	Symbolic contents of knowledge, use of language notations, symbols etc.
Which channel?	Books, journals, mss, reports, conference proceedings, a.v. materials, computer media etc.
Whom?	Library users or members of society
What effect?	Knowledge addition to individuals of society or users of library.

George Gerbner's Model

This model is conceptually different from the earlier two models. Gerbner developed this model in 1956. the essence of this model is to "connect the communication situation and the participant's perception of and response to the situation and the communication process. This model can incorporate machine as well as human process in different combinations and at various stages.

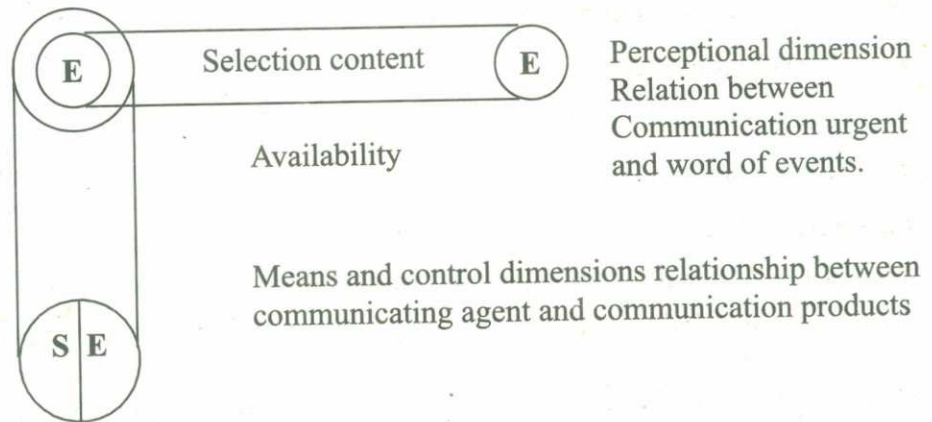


Fig. 5.3: George Gerbner's General Model of Communication

The Schramm Models

Schramm tried to develop a human communication model giving emphasis on experiences of two individuals and the interaction between those two were later designed as models. He developed three models to explain the proposed concept or theory. The three models are diagrammatically represented.

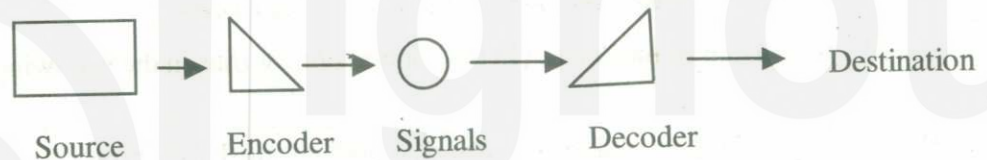


Fig. 5.4: Schramm's Model I

According to this model, communication starts from source (i.e. 1st individual) and it is encoded transmitted through signals. These signals are decoded either manually or mechanically and reach destination (i.e. 2nd individual). This model is more or less similar to Shannon-Weaver Model.

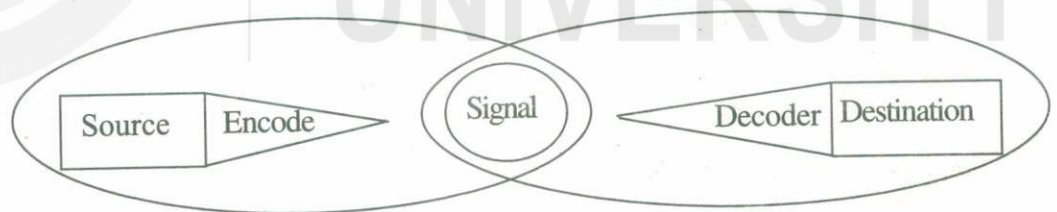


Fig. 5.5: Schramm's Model II

In this model, he introduces the idea that only that which is shared in the fields of experience of both the sources and destination is actually communication because only that portion of the signal is held in common by source and destination.

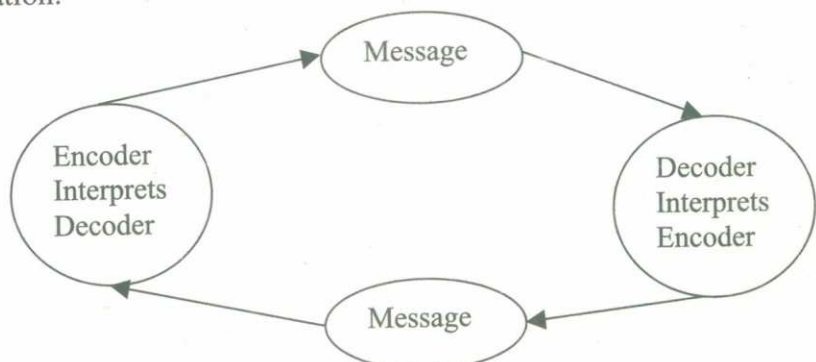


Fig. 5.6: Schramm's Model III

This model deals with communication as an interaction with both the parties encoding, interpreting, decoding, transmitting and receiving signals. In this model feedback and the continuous loop of shared information can be noticed.

Each model is intended to explain certain points, which its creator feels are relevant in communication process or structure. No model can accomplish everything that is desired. In other words, it cannot be “do it all” model. Therefore, it is essential that we select models that would best suit our purpose to solve the problem at hand.

In the above paragraphs, an attempt has been made to explain to you what a communication model is all about and discuss some of the important models available at present. In what follows let us try to learn about Information Diffusion.

Self Check Exercise

- 4) Discuss briefly the Shannon and Weaver Model of communication. List other prominent models of communication.

Note : i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

5.6 INFORMATION DIFFUSION

5.6.1 Meaning

We generally observe that information spreads from person to person in the society. The phenomenon of spread or dispersion through time and space is known as diffusion. Ideas and innovations are dispersed from producer to the user. Innovations are translation of ideas into new practices, processes or products. The diffusion of ideas and innovations is an essential component of social and economic change for better or worse. In other words, the diffusion of ideas and innovations is a source of advancement for the society as a whole. Because of its wide spread influence, the diffusion process has been extensively researched and documented. In this entire process, communication has been found to be a key element.

Research on the diffusion of new ideas has its origins primarily in the field of *rural sociology* with studies focussing on the adoption of new agricultural techniques by farmers in the United States. Subsequently, attention was drawn to other types of innovations, such as new medical products and new industrial processes, as a result, a more international scope emerged. Though, earlier studies were mostly based on desirable changes in society, contemporary studies on diffusion deal with both harmful and beneficial effects of innovations.

Development communication has emerged as a specialised field dealing with factors affecting the diffusion of innovations and attitudes of populations, especially in developing countries. Research on diffusion research also focuses on role of social networks, information flow, demographics and psychological variables “innovativeness” and “resistance to adoption”.

5.6.2 Diffusion of New Ideas or Innovations

The traditional approach to diffusion research was concerned more with social processes and other *demand-related factors* affecting the diffusion of innovations, and not with commercial or supply view point. To fill this gap Lawrence Brown developed a “*market and infrastructure*” model of information diffusion that focussed on *supply-oriented* rather than “*demand-oriented*” factors affecting diffusion. There are three stages in this proposition:

- First Stage: establishment of public or private agency through which an innovation is distributed or made available to the population at large (this provides a special pattern of diffusion)
- Second Stage: conceive and implement strategies to promote adoption among the populations in their service or market areas. This leads to establishing infrastructures such as service and delivery systems and also designs of information dissemination programmes.
- Third Stage: create different levels of access to an innovation, depending on individuals’ economic, locational social characteristics. This ultimately influences the decision of adoption or rejection.

Thousands of new products are introduced each year.. However, a large majority of them fail: Failure to gain acceptance may frustrate both private and public sectors. As public sectors use *public funds* on R and D, they are more prone to criticism. The ‘limited effectiveness’ of technological and social change programmes is attributed to the failure of public agencies to employ sophisticated marketing techniques and strategies.

5.6.3 Patterns of Diffusion Process

A variety of temporal patterns have been associated with diffusion processes. Let us try to understand basic patterns namely: i) temporal pattern of diffusion, and ii) spatial pattern.

Temporal Pattern of Diffusion

While studying the temporal pattern of diffusion, we come across certain questions like:

- When does an idea or innovation first begin to spread within a country, region or organisation?
- How quickly does it spread once introduced?
- What types of people, organisations or firms tend to be early adopters?
- What types of ideas and innovations spread most quickly?

The spread of ideas and innovations generally begins slowly, accelerates as

the diffusion process begins to unfold more fully, slows after some turning point is reached, and levels off as saturation is approached. A graph of the cumulative level of adoption over time follows an *S-shaped (sigmoid) curve*. Figure 5.7, taken from E. M. Rogers (1983) shows the basic sigmoid curve and its relationship to the rate of spread of information.

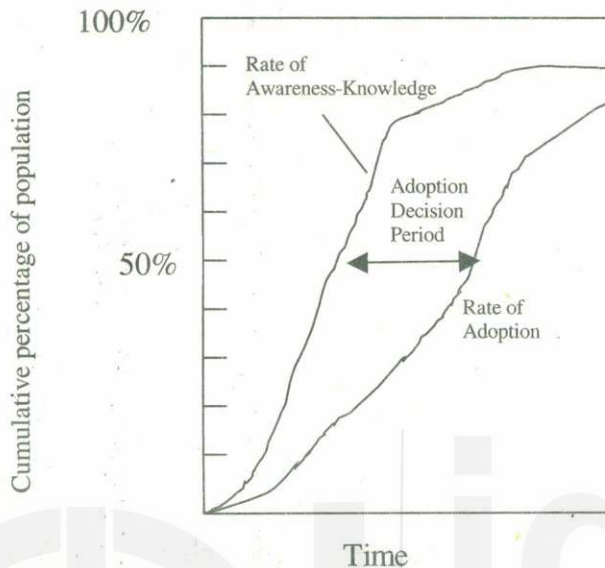


Fig. 5.7: Diffusion over Space and Time

The sigmoid (S-shaped) curve that describes most diffusion processes: rate of awareness-knowledge (uppermost curve), rate of adoption (lower curve), and the adoption decision period (gap between awareness-knowledge and adoption curves). Adapted from Everett M. Rogers, *Diffusion of Innovations* (1962), 3d ed., New York: Free Press, 1983.

The graph indicates the fact that an innovation can be accepted by a person only if the person is aware of the existence of the innovation. In other words, the behaviour of the population toward an innovation is closely related to the distribution of information regarding new phenomenon, even though the availability of information may not, by itself, ensure the acceptance. The graph also depicts the *variable adoption period*, which is stated to be the *shortest* in case of *early adopters*. Another concept (i.e. *the diffusion effect*) is also built into the curve. It may be observed that when 5% of the individuals in a social system are aware of innovation, there is little normative pressure to adopt it. On the other hand, as the rate of awareness of an innovation increases among the population to 20 to 30 per cent, peer pressure begins to develop, and the rate of adoption increases. Once this threshold is passed, a diffusion effect begins to trigger substantial influence from peer networks to accept the innovation. This diffusion or *snowball effect* is partially important in social systems that are highly connected.

Spatial Pattern of Diffusion

In spatial pattern of diffusion, two distinct types been identified. They are: the neighbourhood effect, and hierarchical effect. Let us try to learn about them.

- **Neighbourhood effect:** It has been observed distance affects the rate of distribution. In other words, the neighbourhood effect refers to the time elapsed between the introduction and adoption of an idea or innovation tends to increase with distance from nearest source of information or

access. The source may be a near by adopter, the distributor of an innovation, or some other propagator of information or resource. Distance to a source affects rates of diffusion. It may be stated that personal communications occur more frequently across shorter distances when compared with longer distances. Also, costs of adoption increase with the distance from the source of an innovation, owing to the expenses incurred due to transport, delivery, service and repair. In this case, diffusion occurs in a *wavelike fashion* as depicted in the figure 5.7. In the first instance, adoption is restricted to a compact area. As the diffusion process continues, the circle of acceptance and use expands and distance increases between the points of origin and the area of most rapid adoption. As distance and time lapse increase the innovation waves weaken gradually (see the figure 5.7 Diffusion over Space and Time).

- **Hierarchy or Cascade Effect:** This effect refers to the tendency of large urban places adopting an innovation before small urban places do so. The *per capita rate* of adoption is also greater as the larger cities have the capital to introduce innovations. It may also be mentioned that as the density of population increases, the amount of information also increases. This phenomenon is known as '*Hierarchy*' or '*Cascade*' effect. It has been observed that the *Neighbourhood* and *Hierarchical* Effects often occur in concert. This process is depicted by a *schema* developed by Torsten Hagerstrand in 1967 [Figure 5.7]. This schema portrays diffusion as occurring through a hierarchy of social communication networks. The national, regional, and local levels of spatial aggregation each contain a network within which neighbourhood effect principles of contact operate. Further, a node on one level will have contact with nodes on another level, providing mechanisms for the filtering of contacts down the urban hierarchy. It may be mentioned here that if only the national scale is considered, a *neighbourhood effect* would be observed, but if nodes on all three levels were considered together, the pattern could be seen *hierarchical*.

5.6.4 Factors Affecting Diffusion of Information

Different ideas and innovations diffuse at different speeds. According to Marilyn Brown the following factors affect the rate of diffusion:

- **Relative Advantage:** If the new innovation is perceived to be the superior to the product or practice already in existence, it will replace the existing one. Relative advantage may be considered in terms of cost-effectiveness, a lessening of social conflict, greater comfort or security.
- **Compatibility:** If a new **idea** or technology *requires less change* in the prevalent sociocultural values, behavioural patterns, facilities, existing equipment, and procedures, it has *greater chances of its rapid diffusion*.
- **Impact on social relations:** Many ideas and innovations may lead to alterations in social relationships. The new ideas may inhibit or promote adoption and thus help further diffusion process. For instance, the new agricultural technologies have accelerated urbanisation in some developing countries, and this in turn affected family and community relationships.

- **Complexity:** The use and understanding of an idea or innovation depends on how complex the idea or innovation is. In other words, complexity associated with new ideas or innovations inhibits its adoption and use.
- **Communicability:** If an idea or information about an innovation can be obtained or communicated to the potential adopter easily, it will be quickly disseminated.
- **Observability:** If use of an innovation is observable, knowledge of its existence is facilitated. For example, the diffusion of many agricultural innovations such as new equipment and crops is facilitated by the fact that they are visible and therefore, easily made known to the uninformed.

Divisibility and Reversibility: If an innovation is divisible into smaller parts, then trials are possible and diffusion is promoted. Also, diffusion is more rapid, if it is possible to return to the original condition or situation after its initial trials, in case of failures.

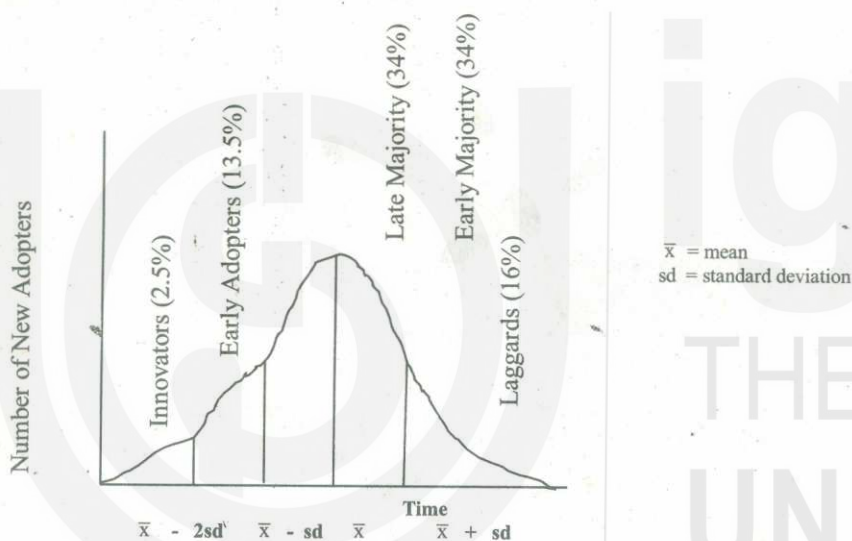


Fig. 5.8(a): Adopter Categories

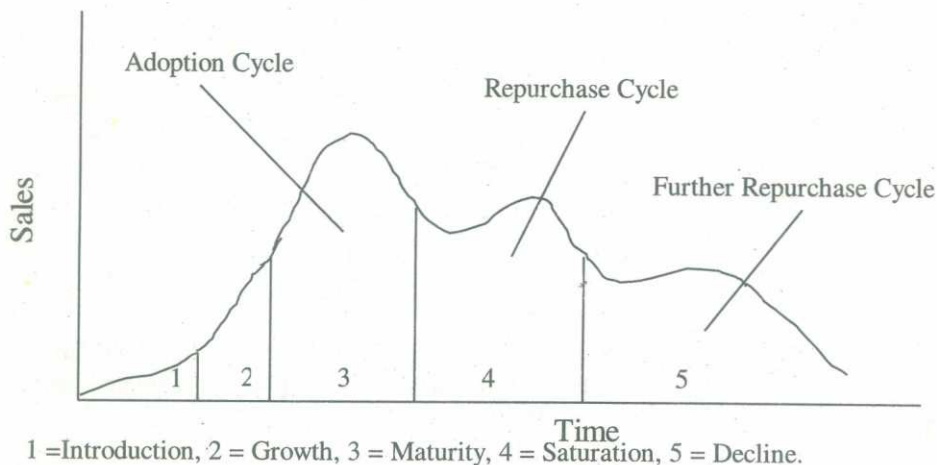


Fig. 5.8(b): Innovation Life-cycle stages

Fig. 5.8: (Diffusion) Adoption and Innovation life-cycle curves. a) Adopter Categories. Adapted from Everett M. Rogers, *Diffusion of Innovations*, New York: Free Press, 1962. b) Innovation life-cycle stages. Adapted from David F. Midgley, *Innovation and New Product Marketing*, London: Croom Helm, 1977.

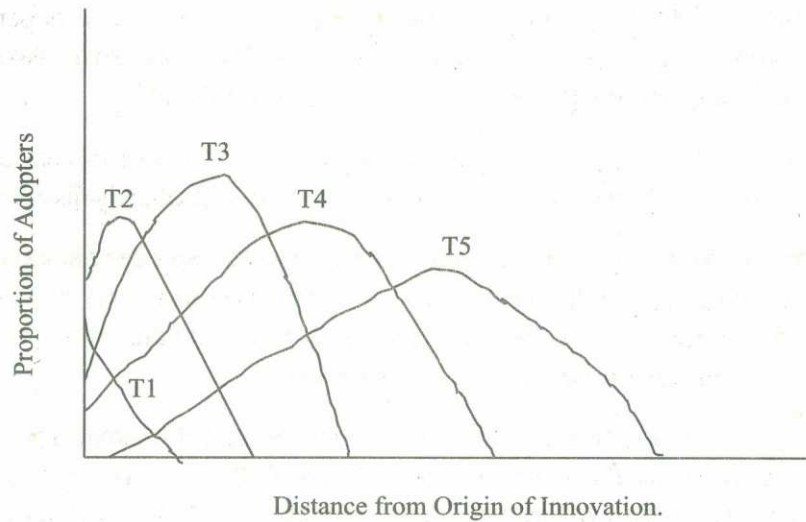


Fig. 5.9: (Diffusion). The shape of diffusion in space and time: proportion of adopters at successive time periods as a function of distance from origin of the innovation. Adopted from Richard L. Morrill, "The Shape of Diffusion in Space and Time," *Economic Geography* 46 (1970): 259-268.

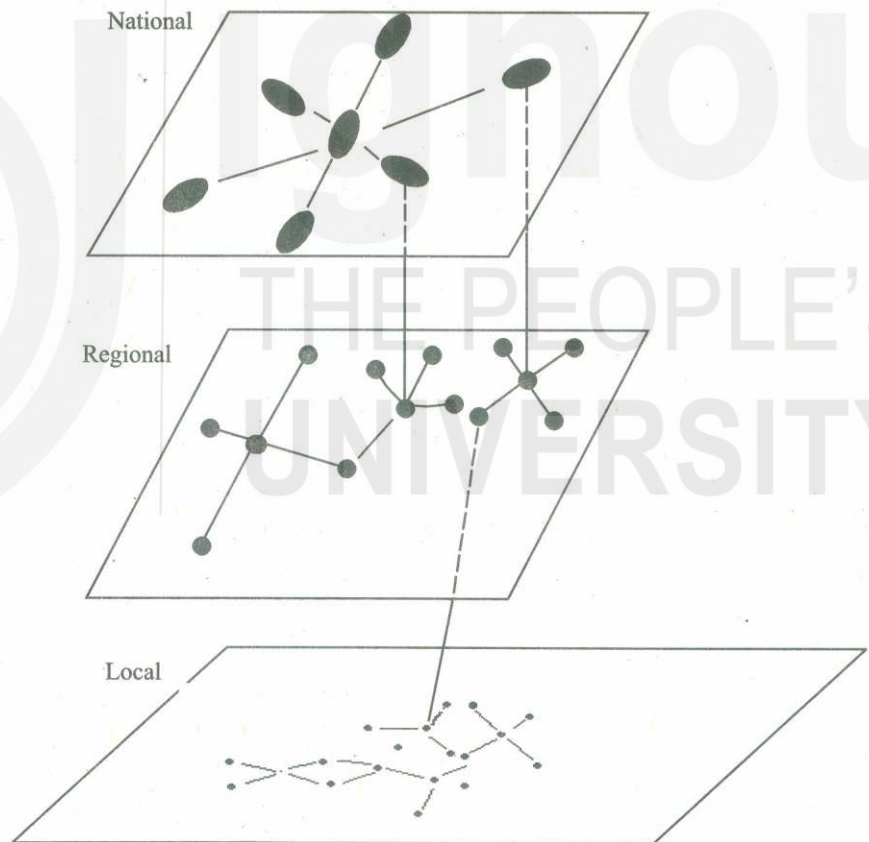


Fig. 5.10: (Diffusion) Schematic portrayal of diffusion networks viewed at three different spatial scales. Adapted from Torsten Hagerstrand, *Innovation Diffusion as a Spatial Process*, Chicago: University of Chicago Press, 1967.

Categorisation of Adopters of New Ideas and Innovations

It may be noted when the absolute number of adopters are plotted into a graph for a distinct time period, a *bell shaped* approximately normal frequency curve is obtained. Using this bell or normal curve, it is possible to talk about the type of people who tend to adopt an idea or innovation during different stages of the diffusion process (Figure 5.8a, 5.8b). It may be stated that people who

adopt an idea or innovation at similar times tend to have similar characteristics. Particularly, innovators and early adopters generally are better educated, have higher social status, are more risk-taking and cosmopolitan, and have greater social participation and opinion leadership than later adopters. They also tend to gain more information from impersonal sources and are less reliant on Interpersonal Communication as an influential source of information. Laggards the last to adopt tend to have traditional values and low social status, education, and income. Their adoption occurs only after information has “tickled down”, when pressures to adopt are strong, and when the prices for the new product have dropped.

Adoption of New Ideas

It may be mentioned here that the adoption of new ideas by organisations and individuals reveals certain characteristics.

Organisations which adopt innovations more readily reflect certain traits such as: openness to communication with their surrounding environment, functionally different subunits, expertise of members, decentralised power, flexible rules and regularisations, and availability of resources, etc.

Apart from organisational aspects, individuals also play important role in the adoption of new ideas. The individual's decision making process can be divided into time periods. E. M. Rogers and F. F. Shoemaker developed a model describing the stages through which an individual passes from first knowledge of an innovation to a decision to adopt or to reject it. This model is known as A-I-E-T-A model of decision-making. As per this model, the decision-making process comprises the following time periods:

- Awareness: The individual learns about the existence of a new idea or innovation,
- Interest: The individual recognises that he/she may have a need for this new idea or innovation.
- Evaluation: The individual assesses the idea or innovation in terms of its benefits and costs and decides whether or not to try it.
- Trial: It is an evaluation of an idea or innovation through a small scale or a limited use. The trial reduces the uncertainties associated with the innovation.
- Adoption: It refers to the adoption or use of new idea or innovation on a continuous, full-scale basis. Problems associated with the use of the innovation may emerge at this stage, particularly when the individuals involved in the *adoption decision* are different from those involved in or affected by the actual implementation.

The A-I-E-T-A model of decision-making can be used to describe the role of different sources and channels of information. In general, mass media and other impersonal sources of information create awareness or interest, but later stages in the process require more personal, specific, and local sources of information.

5.7 DIFFUSION OF TECHNOLOGICAL INFORMATION

In the literature, we come across several studies on the diffusion of technological information in the society (Chatman). Models of information diffusion are constructed. These models help us to estimate the number of people with the information in question at any given point of time. There are different parameters, which may be estimated if the exact values are known. Such parameters include the maximum number of information recipients (N_m) and the number of recipients to whom a transmitter sends information during a time interval (n_i). The diffusion process-taking place is represented in the diagram (see the Figure 5.11).

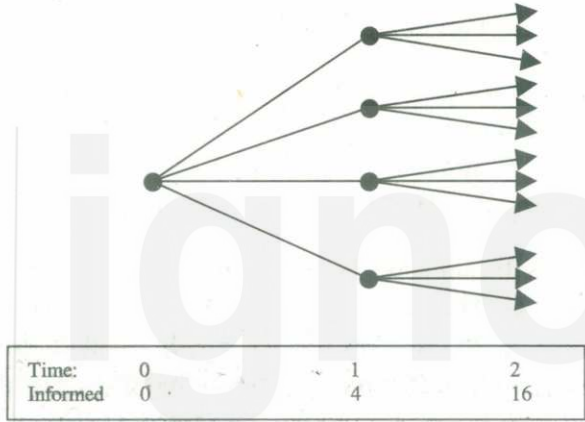


Fig. 5.11: Person-to-Person Diffusion Process (Source: Losee, R. M.)

When information diffuses, it may be seen as having an origin where only one person has information as indicated on the left side of the diagram. It may be noted that in this example population, each information recipient transmits to three additional people.

At the time $t = 2$, the person who originally had the information transmits it to three new people, as indicated by the arrows on the right half of the diagram. At the end of the transmission cycle the total number of people who have received the information will be the number of people who had received the information at the end of the previous transmission cycle plus three times this number. The total number of recipients, who have received information at time ' t ' is $N(t) = 4N(t - 1)$. The total number of recipients, $N(t)$, grows at an exponential rate. If the population were infinite in size and infinitely dense, so that one can never transmitted to some one who already had the information, $N(t)$ could be graphed as a curve 'A' shown in the following Figure 5.12.

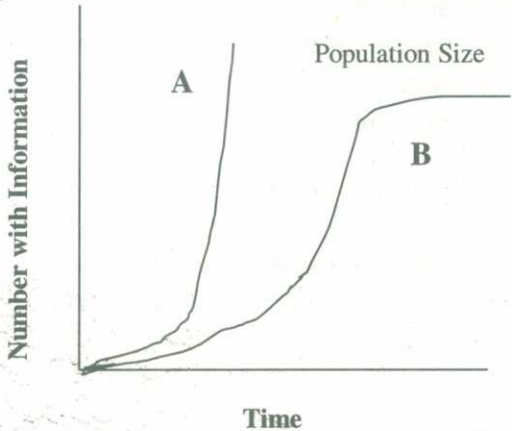


Fig. 5.12: Diffusion Curve (Source: Losee, R. M.)

Neither of these assumptions is ever fully met. As time progresses, more and more people will already have received the information and fewer additional people will be added to the list of informed. The graph of the number of informed people now looks more like curve 'B'. The rate of difference 'r' for curve 'B' is of the form $r = f(t) (N_m - N(t))$. The component $(N_m - N(t))$ will be larger at $t = 0$ and will approach 0 when the number of people with information approaches N_m . If $f(t)$, is $C_i N(t)$, it may be understood as a coefficient of internal influence on diffusion process. Diffusion is seen as occurring only through interpersonal contacts, and the $N(t) [N_m - N(t)]$ may be interpreted as the product of number of people with the information and the number of people without information.

If $f(t)$ is a constant C_e , the number of people already having information not being a factor, the diffusion is controlled by factors external to diffusion environment. When control of diffusion process is mixed between external and internal factors, the rate of diffusion may be modelled as:

$$r = [C_e + C_i N(t)] [N_m - N(t)]$$

Where C_e and C_i are factors representing external and internal coefficients of diffusion, as above.

This entire explanation of Information Diffusion provided above has been taken from Robert M. Losee. 'The Science of Information', 1990, pp.250-253.

So far, we have used the term 'diffusion' rather loosely without precise definition. Mansfield conceives of a three-way definition for diffusion:

- 1) Imitation or Interinstitutional diffusion refers to spread of new process from firm to firm within an industry.
- 2) Intrainstitutional diffusion refers to spread of the process within individual firms.
- 3) Overall diffusion refers to the spread throughout the society as a whole.

Thus, interorganisational diffusion might be measured by the proportion of institutions that have adopted a new process, interorganisational diffusion by the proportion of any one firm's output produced using the new process, and overall diffusion by proportion of the total industry output that is produced using new process. In general, diffusion means *Interinstitutional diffusion*, because it is interesting to study adoption and usage process between institutions.

5.8 MODELS OF INFORMATION DIFFUSION PROCESS

Interorganisational diffusion surveys, study the adoption and adaptation process of an innovation, particularly the analysis of variance between firms. It concentrates on individual innovations diffusing through single type of industries and attempts to explain differences between institutions *in the time taken* to adopt an innovation. This is known as the *institutional characteristic of diffusion*. In the international approach, an explanation of international differences in the speed of diffusion of innovation in terms of the characteristic of the countries and industries is presented.

Rogers argued that the adoption of an innovation mainly depends on communication and often follows an S shaped curve when plotted over time. *Rogers' model* is the most well-known and wide spread diffusion model, but far from the only one. Mahajan, *et al* (1990) presented an overview of diffusion models, arguing that the Bass model (1968) had given the most important impulses to the research on diffusion innovations. Bass identified two forces behind the diffusion process, *mass media communication* and *word of mouth communication*. Mahajan, *et al* mentioned nine assumptions that warrant attention. Four of those considered are of special importance when estimating the diffusion of computers and Internet to households.

These are:

- i) Diffusion of an innovation is independent of all other innovations;
- ii) Nature of an innovation does not change over time;
- iii) Diffusion of an innovation is not influenced by marketing strategies;
- iv) There is only one single adoption by each adoption unit.

Mahajan, *et al* presented several other diffusion models, where attempts have been made to maintain better predictions by handling facts, not fitting the mentioned basic assumptions. Let us try to understand the significance of some diffusion models we come across in the research literature.

5.8.1 Epidemic Model

The epidemic model, also known as the learning model, is more popular in sociological and marketing literatures, but has also been used by economists. A common feature of this model is the analogy drawn from the spread of diseases. Consequently, a theoretical tool often used is one of the mathematical theories of epidemics. As reference to these theories is also often made by economists working in this area, an exposition of simplest model of epidemics will provide a useful introduction to the study of this model. The basic hypothesis of a simple epidemic model is represented by the following equation:

$$m_{t+1} - m_t = \hat{a} (n - m_t) m_t / n$$

Where m is the number individuals in a fixed population of n having contacted the disease, between *elapse time epochs* t and $t + 1$. This is proportionate to the product of the number of uninfected, both at time t . The value of $\hat{a}$ (epidemic indicator) will depend on a number of factors such as the *infectiousness of the disease*, and the *frequency of social interaction*.

5.8.2 Economic Theory Model

As a result of the nascent nature of process innovations and of the imperfect information, there may be major differences between firms in the expectations of how profitable the adoption of any innovation will be. This, coupled with the likelihood that firms differ in their attitude to risk general and new innovations in particular, leads to the conclusion that different firms react differently to the same innovation. Specially, the time lag before adopting the innovations will probably differ significantly between firms within the same industry. For economic reasons, it was suggested that an important contributor

to interfirm differences would be the difference in their size (i.e., the size of the firm). This logic has been formalised into a model called “Industrial Engel Curve” which relates probability of a firm having adopted the new innovation to the firm’s size. When the Industrial Engel Curve was fitted with data collection, results were encouraging. The sample innovations, on which the data was based, indicated positive relationships between the speed of adoption of innovation and the firm’s size. Many other correlations could be brought in, such as nature of innovation *vis-à-vis* industrial growth, cyclical factors and growth factors, the role of the firm’s size and industry concentration, etc.

5.8.3 Stock Adjustment Model

This model is used in the study and analysis of investment behaviour. In fact, the Stock Adjustment Model was used to study the stock levels of computers in U.S. and U. K. in the 1960s and 1970s. It helps to indicate the stock level, in particular, equilibrium stocks in any time period. Although the Stock Adjustment Model has limited applications, it is of great help to study interorganisational diffusion process of innovations.

5.8.4 Vintage Model

The vintage model was postulated by Salter and others in 1960 (Productivity and Technical change, 1960) brings in an economic theory of diffusion. The basic structure of a ‘Vintage Model’ may be summarised as follows: “New knowledge assumed to be embodied in new product and so, gross investment in the vehicle of diffusion. Old product is only replaced or scrapped when its operating costs exceed the returns it earns. Similarly new equipment is only installed if its total costs are covered by the revenue it earns. It is usually assumed, for analytical convenience, that plant is invisible and that new knowledge cognate be introduced on old equipment. Thus, each vintage of product embodies the best practice *know-how of its date* of production and is then committed to this nascent knowledge until its expiry. The appearance of a cost saving process innovation within their frame work will have the following consequences:

It will become profitable to replace immediately some proportion of existing product because the total costs of the new products are lower than the operating costs of some old product. The existing capital stock comprises a number of vintages embodying the old know-how the newer vintages being slightly more efficient than the old, due to minor improvements in specification over the years.

However, the simplicity of the model depends heavily on the assumptions of cost minimisation, on the one hand, and the invisibility of plant and the embodiment hypothesis on the other. In, principle, each of these assumptions can be relaxed without destroying the essence of the model. In practice, however it is doubtful whether the model would retain its productive power if all the four assumptions were relaxed simultaneously.

5.8.5 Information Diffusion Theory

Elfreda A. Chatman identifies three diffusion models, namely:

- Elements of Diffusion Model

- Attributes of Innovation Model
- Stages of Adoption Model

She applies these models to study the awareness, use, and diffusion of an innovation in a job environment among working poor.

The diffusion of information is studied with four elements (1) the innovation, (2) communication, (3) social structure, (4) time. Attributes of innovation explain the characteristics that may improve acceptance of innovation or rejection. Finally, stages of adoption are used to explain the decision making process of adoption or rejection of an innovation.

Elfreda A. Chatman presents her analysis of the information diffusion from person, to person, the social structure that helps movement of various models of communication, the impact of variance of these modes. The study also analyses the attributes, such as (1) relative advantage, (2) compatibility, (3) complexity, (4) trialability of innovations. The stages of adoption model, designed for the diffusion of technological innovation appears to have limited applicability to the diffusion of job information. According to the study of information about a job has a finite usefulness when disseminated. Although, the “job adopter” experiences awareness, trial and acceptability of information about a job, the diffusion of the information as useful is limited to the number persons who can be hired for a position. Information diffusion models are some times found overlapping in actual practice and they can be theoretically separated to draw influences.

5.9 INFORMATION SYSTEM FOR DIFFUSION

A nation's competitiveness depends on the capacity of industry to innovate and upgrade. Innovation is based on new ideas, including new combinations of old ideas, the utility of which is tested with applications of many kinds of information. Empirical studies have repeatedly found that in more prosperous enterprises there is greater awareness of information and greater ability to manage it. Sweeny (1989) reports “empirical evidence strongly indicates that *it is the quality of information activities and of the information process by a firm* which makes its innovation commercially successful and gives the competitive edge”. “Information plays a large role in the process of innovation — information that is not sought or available to competitors, or information available to others that is interpreted in new ways. Sometimes it results from sheer investment in market research or R & D. It is striking, though, how often innovators of those firms that are simply looking in the right place, unencumbered by or unconcerned with conventional wisdom” [Michael E Porter, 1990]. It has been observed that innovative ideas are initiated either by the perception of a need or recognition of technical feasibility. Several studies suggest that most major technological breakthroughs are initiated by the discovery of a technological possibility.

Diffusion of innovation happens through the appropriate flow of information. It begins from the R&D stage itself.

5.9.1 R&D Information Flow

Maddock observes that the main function of an R&D establishment as being a

link in the innovation chain to enable the innovation-flow to occur more efficiently. He defines the role of an R&D in the following terms. Firstly, it should be in close contact with relevant advances in science and technology. Secondly, it should provide an opportunity to bring relevant but difficult disciplines together. Thirdly, it should apply science and technology at the lowest cost in terms of time and effort. Fourthly, and most importantly, it should be active in technology transfer and should, therefore, preferably be located close to the users of R&D work. Fifthly, it should be able to assess objectively the feasibility of an innovation project. From what has been stated above it is clear that an R&D establishment has to rely on a large input of information from external sources as shown in the Table 5.1.

Table 5.1: Information in Diffusion of Innovation

INFORMATION IN DIFFUSION OF INNOVATION
The Listening Post
The Interdisciplinary Role
Science / Technology Application at Lowest Cost
In Time and Effort
Technology Transfer
The Feasibility Appraisal Role

In this connection it is important to note difference the between company and corporate research and development establishments, on the one hand, cooperative or national research institutes on the other hand. Whereas the former have a major commitment to industrial innovation in general, and new product development in particular, the later tends to pay more attention to process improvement and to research aimed at improving the technological infrastructure. The national research institutes, therefore, often act as important sources of scientific and technical information to research establishments in industry, although they must themselves rely on other sources of information (Table 5.2).

Table 5.2: Diffusion and Information Orientation

DIFFUSION AND INFORMATION ORIENTATION
CORPORATE RESEARCH AND DEVELOPMENT ESTABLISHMENTS
● Innovation Oriented ● Cost Conscious ● User of Information ● Information of Proprietary Nature: Know-how "Commodity"
NATIONAL RESEARCH INSTITUTES
● Infrastructure Oriented ● Providers of Advice and Information ● Information More of an Open Nature: "Resource"

5.9.2 Information Services, Information Institutions and Diffusion

It may be mentioned here that concepts such as library, information service,

extension services, etc. are manifestations within a continuous spectrum with a shift in the major activity and in the need to adapt to changing requirements. The physical form of the services is determined by the measure of specialisation, which is necessary to satisfy the need for which the service is created or designed (Table 5.3).

Table 5.3: Spectrum of Services

SPECTRUM OF SERVICES		
TYPES OF SERVICES	ACTIVITY	RATE OF CHANGE
Archive Library Special Library Documentation Service Information Service Information Analysis Information Synthesis Extension Service Advisory Service	Storing Interpretation Advising	Low High

Broadly speaking, an information service, as activity, comprises the interpretation of information in terms of user needs. In other words, it is a “repackaging” function. In accomplishing this function of information service, as an organisation can make use of all possible resources or concentrate more on documentary information (such as journals, indexing and abstracting services, monographs, etc.). A typical R&D information service would provide conventional or computerised current awareness and retrospective search services to supply the user with necessary documents. (Table 5.4).

Table 5.4: Information Services in Diffusion

INFORMATION SERVICES IN DIFFUSION
Alerting ● Broad Current Awareness (To Group of Firms) ● Highly Specific SDI (To Individual Scientists, Engineers) ● Retrospective Searching (In-Depth Searches. Quick Reference Enquiries)
Document Delivery ● Acquisition ● Cataloguing. Classification ● Lending. Copies
Interpretation of Information (Analysis of Information in Documents) ● Specialised (In-House) Databases ● Extension (Field Liaison)

Depending on the organisation and the size of R&D establishments, as also the mandate of an information service within such establishment, more sophisticated information services like information analysis and advisory services as well as specialised databases would be provided.

Where R&D information services also provide information services directly to industry, it is not uncommon to find field liaison services, especially in the

case of information services forming part of national R&D establishments. Information services in industrial R&D establishments will probably pay more attention to patent literature and trade intelligence when compared to information services forming part of those national R&D establishments, which are more infrastructure-oriented.

5.9.3 Gatekeeping of Technical Information

Many papers, articles and handbooks have been published on this topic. As such, only some broad guidelines on what industrial firms can do to use external information sources to their best advantage are indicated here. These guidelines concern the establishment of in-house libraries, correspondence information, monitoring and business intelligence systems, as well as the use of external information systems (Table 5.5).

Table 5.5: Information Sources and Diffusion

INFORMATION SOURCES AND DIFFUSION	
•	Library: Books, Journals, Reports Trade Literature, Newspapers, Cuttings, Patents, Standards, Manuals
•	In and Outgoing Correspondence Information System: Technical Subjects
•	The Use of External Systems for Literature Current Awareness or Selective Dissemination of Information
•	Sharing Facilities for Scientific, Technical and Business Databases: Bibliography and Numerical
•	Information Search Unit: Study of Literature to Trigger New Ideas
•	Monitoring or Early Alert System: Identification of Threats and Opportunities
•	Business Intelligence System: Alliances of Information for Decision Support

In this context a distinction can be made between the information needs of entrepreneurs establishing new industries and those of existing industries. It may be mentioned here that the small firms may need advice on how to set up in-house libraries and / or information services mentioned earlier. Many international organisations as well as national documentation centres are able to provide such advice.

No mention has been made, till now, of management information systems, although many firms (especially smaller ones) are in need of unbiased advice on the creation of such systems. There are clear indications, however, that the fields of documentary information, management information, records management, and office automation are converging in a manner similar to the converging of computation and telecommunication.

It must be stressed at this stage that no amount of information can encourage innovation if the management of the organisation is not receptive to risk taking for the sake of innovation, especially in these days of “professional management”.

5.9.4 Information Policy in Support of Information Diffusion

Information flow is a necessary condition for diffusion of information related to innovations. In many countries evidence is available in favour of governmental obligation to provide innovation support to industry. It has been

mentioned that government policies are rather biased to support large industrial firms. But, recent studies indicate that small firms can also contribute extensively to innovations. Chico writing on Technonet Project in South East Asia, lists technical information services, industrial extension services and entrepreneurial development programs as important ingredients for promoting innovation, especially in small firms. However, most of the small industrial firms of developing countries lack these facilities. The developed countries have well designed policies in this regard. In the USA, the policy is to stimulate innovation through removal of impediments, while in other countries direct governmental involvement in the form of subsidies for industrial R&D, based upon selective industrial strategies, as well as the introduction of other innovation support programmes is advocated. Several Western Countries, as well as Japan, have established, or are in the process of establishing formalised information policies for this purpose.

Self Check Exercise

- 5) What is meant by diffusion? Explain the varieties of diffusion.
- 6) Discuss the economic theory of diffusion.
- 7) Briefly discuss some of the important models of diffusion.

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

5.10 SUMMARY

In this Unit, we have attempted to explain to you different conceptualisations of the term 'information'. The main aspects discussed are: Information as commodity / resource; information as data in the environment; information as representation of knowledge and information as part of Communication process.

The second topic covered in this Unit relates to the concept of 'communication'. In this regard, the idea and meaning of communication, the definition of communication is explained to you. The form and channels of communication are described and are represented diagrammatically to have a clear understanding. The essentials or ingredients of communication are dealt with briefly. The different functions of communication along with its types are discussed in detail. Attempt has been made to highlight the significance of personal communication bringing out clearly the importance of models and media of communication.

The third aspect discussed in this Unit happens to be "Information Diffusion". In this regard, the main aspects discussed are: the meaning of diffusion, which is a process of transfer of knowledge, or the movement of technical know-how within a group of users such as farmers, doctors, engineers, etc.; the pattern of

diffusion process occurs in two ways namely temporal and spatial; diffusion could be interfirm, intrafirm, or spread through out the industry; several factors such as relative advantage, compatibility, complexity, communicability, observability, and diversability; affect the information diffusion of innovations; the process of information diffusion has been studied through four models: Epidemic Model, Economic Theory Model, Stock Adjustment Model and Vintage Model, each having its own characteristics. Also, the information system for diffusion is affected through information units operated close to R&D establishments. A spectrum of services is offered by information institutions for the support of innovation, for example Gatekeeping or Technical Information is a unique service adopted to provide information support to innovation. Mention has also been made in the Unit about national efforts providing information support services through a variety of institutions that activate spreading of information.

5.11 ANSWERS TO SELF CHECK EXERCISES

1) Different Conceptualisations of Information:

Information is not just one thing. It means different things to those who expound its characteristics, properties, elements, techniques, functions, dimensions and connections.

In every day life, the word 'information' is closely related to the concept of communication, more specially with the aspect of communication of ideas, thoughts, and knowledge bringing forth an understand of information that it has properties to convey ideas, thoughts, concepts and knowledge.

The table below furnishes different conceptualisations of information.

Conceptualisation	Description	Assumptions
Resource / Commodity	A message, commodity, something that can be produced, purchased, distributed, sold, traded, manipulated, passed along, controlled.	Assumes sender → receiver, assumes receiver makes of message what sender intends.
Data in environment	Objects, artefacts, sounds, smells, events, visual and tactile phenomena, activities, phenomena of nature.	Accounts for unintentional communication
Representation of Knowledge	Documents, books, periodicals, some visual and auditory representations, abstractions of information (e.g. citations)	Assumes printed documents are primary representation of knowledge; assumes primacy of scientific technical knowledge.
Part of process of communication	Part of human behaviour in process of moving through time / space to make sense of world	Assumes meanings are in people, not in words; assumes human behaviour is basis of the process.

Buckland identifies three principal uses of the word information:

- information as process (i.e. the ability to inform),

- ii) information - as knowledge (i.e. the knowledge that is imparted in the process of being informed, and
- iii) information - as thing, concentrating on the various properties of information and its different manifestations and understandings.

By focussing upon information - as - thing, Buckland brings into consideration such phenomena as objects and events, seeking in the process to widen our perception of what constitutes data and documents. Buckland's approach is useful both for its breadth of vision and for the way it manages to weave together such key concepts as information, knowledge and communication. Moreover, by demonstrating that as thing, which is embodied in objects or information systems, information can be tangible, as opposed to knowledge which is inherently intangible, he sheds light on the traditional perceptions of information, most notably that of Fritz Machlup who considered information as an intangible.

2) Communication systems can be distinguished in the following manner:

- i) **Vocal and Non-vocal Communication:** Vocal communication refers to any type of communication where human voice is used; it could be spoken language, shouting or laughing. Non-vocal communication, on the other hand, refers to use of signs, symbols, signals, gestures etc. to communicate something.
- ii) **Verbal and Non-verbal Communication:** Verbal communication refers to the use of words to communicate something; different languages come under the category of verbal communication. Non-verbal communication, on the other hand, refers to those different modes of vocal and non-vocal communication where words are not used. The examples of non-verbal communication are gestures, facial expression, vocalisation etc.
- iii) **Intrapersonal Communication:** Communication within one person such as reading a book.
- iv) **Interpersonal Communication:** Communication between two persons such as writing letters, face-to-face conversation.
- v) **Group Communication:** It could be either a small group ranging from three to about 25 persons or large group where one or several persons communicate to an audience of 25 or more persons. Social gatherings and conferences are examples of group communication.
- vi) **Public or Mass Communication:** Any large scale communication that disseminates through media such as print, broad casting etc. comes under this category. Broadcasting of news through radio and television are examples of public mass communication.

3) The basic elements of communication process are:

- i) **Source:** Which generates information.
- ii) **Encoder:** Translates the information or message into words, signals, signs, etc.

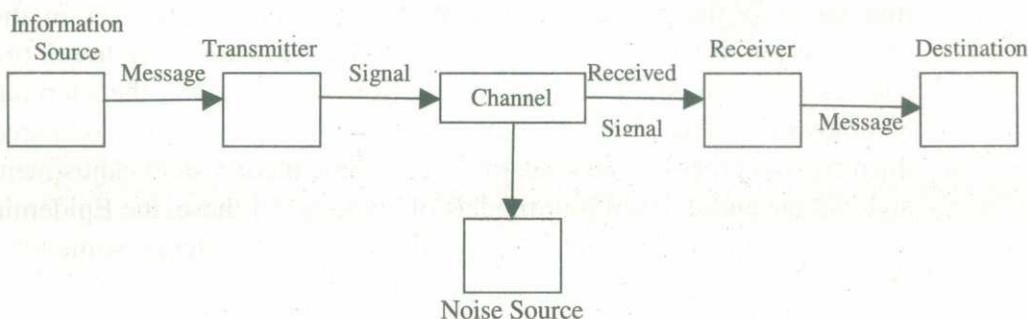
- iii) **Message:** It is the meaningful representation of the thought or ideas of the information source.
- iv) **Channel:** The medium through which the message is transmitted.
- v) **Noise:** Disturbance that interferes with the message in the transmission process and degrades or distorts the information or message.
- vi) **Decoder:** Translates back the message from words, symbols, etc. into thoughts and ideas of the information source.
- vii) **Receiver / Destination:** The person or equipment, which receives the information. In oral communication the listener is the receiver as destination of information. In the case of electromagnetic communication the equipment e.g., telephone, televisions are the receivers and the person or group of persons for whom the information is meant is the receiver.

4) Shannon and Weaver Model of Communication

Shannon and Weaver Model of Communication is one of the best-known models of communication. In this model (depicted in the figure), a message emanates from an information source. It is converted into signal or series of signals by a transmitter. Enroute, this signal is mixed with or contaminated by 'noise', that is to say, various kinds of unwanted interference coming from noise sources. The received signal is decoded by a receiver, being converted back into the original message, more or less, which is what the receiver (or destination) receives.

This model has been criticised, however, because it is linear. That is to say, it accounts only for the act of sending and receiving. It is a one-way model, based on engineering. In fact, this was the dominant model for a quarter of a century.

To convert Shannon and Weaver's Model of Communication into a two-way model of communication, we must add feed back channel, complete with transmitter, encoding, noise, decoding and receiver, all to deal with messages returning from the receiver to the sender. It could be said that we are simply working Shannon Weaver model in reverse, but in fact we are working it almost simultaneously in both directions. In this two-way model, the feedback channel can be vitally important to the original receiver as a means of seeking clarification of the original message. Feedback channels, verbal and non-verbal, continually serve this purpose in human communication. New information technology enables us to engage in far more two-way communication than was possible earlier.



Shannon Weaver Model of Communication Process

Some of the other prominent models of communication are:

- i) Lasswell's Model
 - ii) Gerbner's Model
 - iii) Schramm Model.
- 5) Diffusion is an information transfer process of knowledge. It is the spread or activation of the usage of technology within a population of users, usually within a group characterised by some common element of productive activity viz. farming, mining, etc. Thus diffusion is the movement of technical know-how within the group of users such as farmers, doctors, engineers, etc. This is clearly distinguished from the lateral shift i.e. from one group of users to another e.g. from farming to mining.

Mansfield has conceived three varieties of diffusion. They are:

- a) Interfirm diffusion: It refers to spread of new process from firm to firm within an industry. It is also known as imitation diffusion.
- b) Intrafirm diffusion: It refers to the spread of the process within individual firms.
- c) Overall diffusion: it refers to the spread of transfer through out the industry as a whole.

However, in general diffusion means interfirm diffusion which reflects the adoption and usage process between firms within an industry.

- 6) According to Economic Theory of diffusion there can be major differences between firms within the same industry of how profitable the adoption of any innovation will be. This will depend on the nature of innovations, information about it, and attitude of the firms taking risk. Apart from the size of the firms may also contribute to the interfirm differences. This argument has been formalised into a model called 'Industrial Engel Curve' which relates probability of a firm having adapted to the new innovations to the firm's size. Data collected to Industrial Engel Curve indicate positive relationship between the speeds of adoption of innovations to the firm's size. Many other correlations could be brought in such as nature of innovation vis-à-vis industrial growth, cyclical factors and growth factors, the role of firm size, industry concentration, etc.
- 7) There are considerable number of intra and interorganisational diffusion surveys focussing on the adoption and usage process of innovations. The time taken for the spread of diffusion of innovation is influenced by the institutional level characteristics. At the international level the characteristics of countries and organisations therein affect the adoption of innovation. The researchers of innovation diffusion have compared the diffusion process with epidemics, economic theory, stock adjustment and vintage and devised four models of analogy. Of these, the Epidemic Model and the Economic Theory Model are considered somewhat important. Let us try to know what these models are.

Epidemic Model: This model is known as the 'learning model'. It is more popular in the sociological and marketing literatures, but has also been used by economists.

A common feature of this model is the analogy drawn from the spread of diseases. Consequently, a theoretical tool often used is one of the mathematical theories of epidemics. As reference to these theories is also often made by economists working in this area, an exposition of the simplest model of epidemics will provide useful introduction to the study of this model. The basic hypothesis of a simple epidemic model is represented by the following equation:

$$m_{t+1} - m_t = \beta (n - m_t) m_t / n$$

Where m is the number individuals in a fixed population of n having contacted the disease, between elapse time epoch's t and $t + 1$. This is proportionate to the product of the number of uninfected, both at time t . The value of β (epidemic indicator) will depend on a number of factors such as the infectiousness of the disease, and the frequency of social interaction.

The second important model is that of

Economic Theory Model: As a result of the nascent nature of process innovations and of the imperfect information, there may be major differences between firms in the expectations of how profitable the adoption of any innovation will be. This, coupled with the likelihood that firms differ in their attitude to risk general and new innovations in particular, leads to the conclusion that different firms react differently to the same innovation. Specially, the time lag before adopting the innovations will probably differ significantly between firms within the same industry. For economic reasons, it was suggested that an important contributor to interfirm differences would be the difference in their size (i.e. the size of the firm). This logic has been formalised into a model called "Industrial Engel Curve" which relates probability of a firm having adopted the new innovation to the firm's size. When the Industrial Engel Curve was fitted with data collection, results were encouraging. The sample innovations, on which the data was based, indicated positive relationships between the speed of adoption of innovation and the firm's size. Many other correlations could be brought in, such as nature of innovation vis-à-vis industrial growth, cyclical factors and growth factors, the role of the firm's size and industry concentration, etc.

5.12 KEYWORDS

- | | |
|------------------------------------|--|
| Communication | : Convergence of communication and computer technology. |
| Diffusion | : It is a transfer process in which there is movement of technical know-how within a group of users. |
| Epidemic Model of Diffusion | : Diffusion process examined on the analogy drawn from the spread of diseases. |

Facsimile Transmission	: Transmission of an image over communication line in the form of electric signals in such a way that the image is reproduced at the destination using special equipment and paper.
Industrial Engel Curve	: Relates to probability of a firm having adopted the new innovation to the firms' size.
Interfirm Diffusion	: Spread of new process from firm to firm within an industry.
Intrafirm Diffusion	: Spread of new process within a firm.
Multiplexing	: It is a system of device which takes a number of data communication channels and combines the signals into one common channel transmission.
Optical Fibres	: A thin glass or plastic fibre used to carry data in the form of light waves.
Teletext	: A generic term for information services provided by television companies in which data is super imposed on the television signal in such a way that a suitably adapted television can display the data as full screens of alphanumeric and block graphic characters.
Videotext	: A generic term for system whereby computer based information is made available on an adapted television monitor.

5.13 REFERENCES AND FURTHER READING

- Arrow, K.J. (1979). The Economics of Information. *In: Dertouzos, M.L. and Moses, J (eds.) The computer Age: A Twenty-Year View*. Cambridge: M I T Press. Pp. 306-317.
- Atwood and Dervin (1982). *Challenges to Sociocultural Predictors of Information Seeking*. New Brunswick, NJ: Transaction.
- Baptista, R. (1989). The Diffusion Process Innovations: A Selective Review. *International Journal of Economics of Business*. 61, 107.
- Bates, B.J. (1988). Information as an Economic Good: Sources of Individual and Social Value. *In: Mosco, V. and Wasco, J (eds.) The Political Economy of Information*. Madison; WI: University of Wisconsin Press.
- Brown, L. A. (1981). *Innovation Diffusion: A New Perspective*. New York: Methuen & Co.
- Brown, Marilyn A. (1989). Diffusion. *In: Encyclopaedia of Communication*. New York: Oxford University Press. Vol. 1. Pp. 31 - 36.

Buckland, Michael (1991). Information as Thing. *Journal of the American Society for Information Science*. 42(5), 351-60.

Charpie, R. A. (1970). *Technological Innovation and the International Economy*. Fifth Annual Lecture of the Science Foundation. London: Wiley Interscience.

Chatman, Elfredn A. (1986). Diffusion Theory: A Review and Text of Conceptual Model in Information Diffusion. *JASIS*, 36, 377 - 386.

Davies, S. (1979). *The Diffusion Process Innovation*. Cambridge: Cambridge University Press.

Dholakia, N., and Bakke, J. W. (1996). Institutional Patterns of Information Technology Diffusion. In Belk, R [et al.] (eds.). *Consumption and Marketing Macro Dimensions*. London: International Thomson Printing. Pp. 324 - 356.

Dutton, William, (ed.) (1996). *Information and Communication Technologies: Visions and Realities*. Oxford: Oxford University Press.

Geroski, P. A. (2000). Models of Technology Diffusion. *Research Policy*, 29 (4/5), 603 - 625.

Hall, B. H. (2003). *Innovation and Diffusion*. Preprint. 46p.

Lennstrand, Bo (1998). Diffusion of Information and Communication Technology to Households. Paper presented at the *Twelfth Biennial ITS Conference at Stockholm*.

Lievrouw, L. A. (1988). Four Programs of Research in Scientific Communication. *Knowledge in Society*. 1(2), 6-22.

Losse, Robert M. Jr. (1990). *The Science of Information*. San Diego: Academic Press. Pp. 250 - 253.

Mc Creadie, Maureen and Rice, Ronald E. (1999). Trends in Analyzing Access to Information pt. 1: Cross-disciplinary Conceptualisations of Access. *Information Processing and Management*. 35 (19), 45 - 76.

Maddock, Sir Ieuan (1981). Research Organisations and Their Contributions. *In Conference on Innovation in Industry*. Pretoria: S233, CSIR.

Mahajan, V., [et al.]. (1990). New Product Diffusion Models in Marketing: a Review and Direction for Research. *Journal of Marketing*. 54, 1-26.

Menou, Michael (1994). *The Impact of Information: Definitions and Measurements*. (Preprint, 1994).

Rogers, E. M. (1995). *Diffusion of Innovation*. 4th ed. New York: Free Press.

Shannon, C.E. and Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.

Stoneman, P. (2001). *The Economics of Technological Diffusion*. Oxford: Blackwell's.

Taylor, R. (1996). Information Use Environments. *In*. Auter, E. and Choo, C., (eds.) *Managing Information for the Competitive Edge*. New York: Neal-Schuman.

Torsten, Hagerstrand (1967). *Innovation Diffusion as a Spatial Process*, Translated by Allen Pred. Chicago: University of Chicago Press.

Valente, Tom (1995). *Network Models of the Diffusion of innovations*. New Jersey: Hampton Press.

Witson, T. D. (1981). On User Studies and Information Needs. *J. Doc.* 37 (1), 3-15.

Zettlemeyer, F. and Stoneman, P. L. (1993). Testing Alternative Models of New Product Diffusion. *Economics of Innovation and New Technology*, 2, 283 - 308.



UNIT 6 GENERATION OF INFORMATION: MODES AND FORMS

Structure

- 6.0 Objectives
- 6.1 Introduction
- 6.2 Information
- 6.3 Generation of Information
- 6.4 Modes of Information Generation
 - 6.4.1 Observation
 - 6.4.2 Thought Process, Deliberation, Imagination
 - 6.4.3 Experimentation
 - 6.4.4 Processing of Data
 - 6.4.5 Events
 - 6.4.6 Evolution
 - 6.4.7 Dream
 - 6.4.8 Information Generation in some Selected Subjects
- 6.5 Forms of Information
 - 6.5.1 Oral Form
 - 6.5.2 Information in Sign Language
 - 6.5.3 Hand-written Form
 - 6.5.4 Pictorial Form
 - 6.5.5 Printed Form
 - 6.5.6 Digitised Form
 - 6.5.7 Condensed Form
 - 6.5.8 Coded Form
 - 6.5.9 Simplified Form
 - 6.5.10 Translated Form
 - 6.5.11 Disguised Form
 - 6.5.12 Secondary Form
 - 6.5.13 Tertiary Form
- 6.6 Impact of Information Technology on Information Generation
- 6.7 Summary
- 6.8 Answers to Self Check Exercises
- 6.9 Keywords
- 6.10 References

6.0 OBJECTIVES

After reading this Unit, you will be able to:

- grasp the concept of information in a better way, realise how the concept is moulded in different disciplines to suit the respective needs of the disciplines;

- know the various modes of generation of information such as observation, thinking, experimentation, and so on;
- get an interesting overview of the variegated forms of information such as oral form, written form, printed form, pictorial form, etc.; and
- perceive the impact of information technology on the generation of information.

6.1 INTRODUCTION

The human civilisation has passed through different ages such as stone age, iron age, industrial age, and so on. Now it has entered the information age. The countries rich in information are today in a much more advantageous position compared to those poor in information. In many cases the poor countries are obliged to purchase information from the vendor sometimes at a very high cost. The advent of Internet has proved to be a great boon for accessing any information from any part of the world practically in no time. Even here in many cases, we are to pay for obtaining information. In other words information has become a commodity. The generators of information are churning out various information products, marketing them, and earning profit. The information age has given birth to information industry. The society we are living in has already been termed as information society.

6.2 INFORMATION

In your BLIS programme you have already studied about information and gained some knowledge about it. You have also learnt the meaning of data, information and knowledge and the subtle differences existing among them. In Block 1 of this course, several units have dwelt on the concepts – data, information and knowledge providing definition, outlining types and discussing nature, properties and scope. One peculiarity of information, which you might not have noticed, is that the meaning of information changes from subject to subject. Now, let us see the definition of information in some subjects.

Law – ‘An accusation or complaint made without the intervention of a grand jury’ [The New International Webster’s Comprehensive Dictionary of the English language, 1996]. This definition is clear and does not require any discussion.

Communication theory – ‘Any distinct signal element forming part of a message or communication, especially one assembled and made available for use by automatic machines, such as a digital computer: usually measured in bits’ [The New International Webster’s Comprehensive Dictionary of the English Language, 1996]. In this case, information is communicated through signals. In the receiving end the signals are assembled by automatic machines and turned into a language comprehensible to the receiver. What the receiver gets is nothing but information.

Computer technology – ‘Data that have been processed into an organised, usable form and are meaningful to the recipient for the task at hand. Any data that can be coded for processing by a computer or similar device’. [Szymanski, R.A. 1994], A set of data is fed into the computer to generate different types of information to fulfil the needs of the user. Again, any data can be considered

as information in case it can be coded for computer processing. Thus, we find, in computer technology, information has two distinct meanings.

6.3 GENERATION OF INFORMATION

All the while, information is being generated in the world, nay in the universe. The bursting of a supernova in a particular constellation, discovery of a planetary system around a star like the Sun, spotting of a river on the Mars, inundation of areas by a swelling river, eruption of a volcano, invention of a machine, successful testing of a drug, conquering of a deadly disease, birth of a child, and millions of other events are generating information every moment. You may be interested to know whether the generation of information follows any well-defined rule or it generates at random without any regard to any rule.

If you just take a newspaper and try to find out how the news have generated you will notice that they have generated following certain modes. *The Hindustan Times* of 21 July 2004 contained the following headlines in its first page: i) HIV vaccine could come from AIIMS; ii) 'Soften Hurriat with foreign trips' iii) It's almost clear, monsoon's a failure, iv) NCERT's recipe for confusion. On going through the news it will be clear that the first news has resulted due to experimentation, the second news due to deliberation, the third news due to observation, and the fourth news again due to deliberation. In many cases, generation of information involves more than one mode. For example, Newton saw the falling of an apple from a tree. This observation immediately switched his thought process on which ultimately resulted in his propounding the theory of gravitation. In this case, the combination of observation and thought process gave rise to information.

In the next section we intend to discuss various modes of the generation of information.

6.4 MODES OF INFORMATION GENERATION

Information usually generates following such modes as Observation, Thought process, Deliberation or Imagination, Experimentation, Processing of data, Happening of various events, and so on. In certain cases like language, information generates following the path of evolution. Now, we shall discuss all these modes one by one.

6.4.1 Observation

By the word 'observation' here we mean not only seeing with eyes, but also hearing, smelling, tasting and feeling with skin. We can get information about the sky whether it is sunny, cloudy or hazy by looking at it. Many a time, an ornithologist can identify a bird just by hearing its call. Often chemists can recognise a chemical substance, e.g. phenol, by smelling it. Our tongue gives us information about the taste of a substance. A simple touch by the hand is enough to know whether a substance is hot, cold or warm.

Observation may be termed as the most potent mode of generation of information. Charles Darwin observed nature for years to gather information for establishing the theory of evolution. Astronomers all over the world gathered information by observing the celestial bodies for centuries initially with naked

eyes and subsequently with telescopes. Similarly microbiologists gathered information on all microbes observing them with microscopes. A police officer has to observe minutely every detail of the venue while investigating an accident, theft, etc. A scientist conducting an experiment has to observe very carefully the changes taking place in temperature, pressure, colour, etc. and faithfully record the changes. A doctor has to observe the condition of a patient at regular intervals to see whether his condition is improving or deteriorating. Thus, we find, in every walk of life observation is a prerequisite for the generation of information.

Self Check Exercise

1) Explain how Observation helps in the generation of information.

- Note:** i) Write your answer in the space given below.
ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

6.4.2 Thought Process, Deliberation and Imagination

Thought process is the mother of generation of information. Be it observation, experimentation or data collection, thought process is involved in every case to generate information. Man has seen the solar and lunar eclipses for thousands of years and have tried to find out the underlying causes with his limited knowledge and generated information. The ancients observed that during eclipse the sun or the moon is gradually swallowed by something and again it comes out. Hence, the ancient Hindus reasoned that during an eclipse the sun or the moon is gradually gobbled by the beheaded *Râhu*. As it gobbles the celestial body through the mouth it comes out through the cutout throat. Considering the level of knowledge human beings possessed at that time, the reasoning was quite logical. After centuries of observation and reasoning, now we know the real cause of eclipse. The information we generate through our observation, experimentation, reasoning, etc may not always be absolutely true. In many cases, it is subject to correction at a later date. In Arthur Conan Doyle’s novels we have seen both Dr. John Watson and Mr. Sherlock Holmes have visited together the site of the crime. It was always the superior thought process of Sherlock Holmes that was able to pinpoint the culprit.

Be it a household, an office, an organisation or institution, the process of deliberation is encountered everywhere. While studying in class XII, many students are to appear in a number of entrance tests. When a student qualifies in more than one test, the student and the parents are to deliberate a lot to arrive at a decision as to the course the student is going to pursue. The moment the final decision is taken and is made known to others, information is generated.

Before declaring the dates of a general election, the Election Commission (EC) has to consider a number of factors like weather (usually rainy season is avoided), school and college examinations (because schools and colleges are used for setting up polling booths), availability of security personnel, dates suggested by various political parties, and so on. After deliberating on all the factors, when the EC announces its decision, information is generated. In a parliament or a legislative assembly, after a lot of debates, decisions are taken giving birth to plenty of information.

An artist has to imagine about the art piece he is going to create, a chemical engineer has to visualise in his mind the chemical plant he is going to install, an architect is also to picture the building in the canvass of his mind. Once the art piece is complete, the blue prints of the chemical plant or the building are ready and handed over to the persons concerned, information is generated.

Self Check Exercise

2) Explain how imagination helps in the generation of information.

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

6.4.3 Experimentation

If we go through abstracting and indexing services devoted to physics, chemistry, biology, medicine, engineering, agriculture, and other scientific disciplines, we shall find that about two million articles are being included in these databases every year. Most of these articles are based on experimentation. Just from this, one can make out how much information is being generated per year through experimentation. It is however to be noted that experimentation is always attended with observation and thinking process. The results of experimentation usually appear in the form of research papers, short communications, patents and so on.

6.4.4 Processing of Data

You have already learnt that data collected through questionnaire and other methods gives rise to information when processed. Now, we shall see how the processing of data gives rise to information with a concrete example. Let us take the students of a BLIS class of an Indian university as a sample. The data collected about the students is given in Table 6.1.

Table 6.1: Particulars of the Students of a BLIS Class

Sr. No	Name	Age	Sex	H.Q.	Class in H.Q.	Nationality	Religion
1	Nirmalendu	41	M	BA	2nd	Indian	Hindu
2	Gita	33	F	B Sc	1st	Indian	Jain
3	Diana	24	F	BCA	1st	Indian	Christian
4	Aurobinda	27	M	MA	2nd	Indian	Hindu
5	Asit	22	M	BA	2nd	Indian	Hindu
6	Yogendra	23	M	B Com	1 st	Indian	Hindu
7	Laxman	24	M	M Sc	2nd	Sri Lankan	Buddhist
8	Rabi	25	M	MA	2nd	Indian	Hindu
9	Rahul	23	M	B Sc	2nd	Indian	Hindu
10	Rajib	26	M	M Com	2nd	Indian	Hindu
11	Ali	24	M	BA	2nd	Indian	Muslim
12	Monica	23	F	B Sc	1st	Indian	Hindu
13	Rabeya	22	F	M Sc	1st	Bangladeshi	Muslim
14	Ram Bahadur	29	M	BA	1st	Nepali	Hindu
15	Dan Singh	23	M	B Com	2nd	Indian	Sikh
16	Farooq	24	M	BA	1st	Indian	Parsee
17	Latif	22	M	BA	2nd	Indian	Muslim
18	Aparna	23	F	BA	2nd	Indian	Hindu
19	Renu	24	F	B Sc	2nd	Indian	Hindu
20	Prashant	34	M	M Com	2nd	Indian	Hindu

Sr. No = Serial No; H.Q. – Highest qualification; M = Male; F= Female

Now, one by one we shall arrange the columns of the Table in some order and see how information generates from each of the operation.

Information Relating to Names

In Table 6.2, we take column 2 and arrange the names in alphabetical order.

Table 6.2: BLIS Students (arranged alphabetically according to names)

Sr. No	Name	Age	Sex	H.Q.	Class in H.Q.	Nationality	Religion
11	Ali	24	M	BA	2nd	Indian	Muslim
18	Aparna	23	F	BA	2nd	Indian	Hindu
5	Asit	22	M	BA	2nd	Indian	Hindu
4	Aurobinda	27	M	MA	2nd	Indian	Hindu
15	Dan Singh	23	M	B Com	2nd	Indian	Sikh
3	Diana	24	F	BCA	1st	Indian	Christian
16	Farooq	24	M	BA	1st	Indian	Parsee
2	Gita	33	F	B Sc	1st	Indian	Jain
17	Latif	22	M	BA	2nd	Indian	Muslim
7	Laxman	24	M	M Sc	2nd	Sri Lankan	Buddhist
12	Monica	23	F	B Sc	1st	Indian	Hindu
1	Nirmalendu	41	M	BA	2nd	Indian	Hindu
20	Prashant	34	M	M Com	2nd	Indian	Hindu
13	Rabeya	22	F	M Sc	1st	Bangladeshi	Muslim
8	Rabi	25	M	MA	2nd	Indian	Hindu
9	Rahul	23	M	B Sc	2nd	Indian	Hindu
10	Rajib	26	M	M Com	2nd	Indian	Hindu
14	Ram Bahadur	29	M	BA	1st	Nepali	Hindu
19	Renu	24	F	B Sc	2nd	Indian	Hindu
6	Yogendra	23	M	B Com	1 st	Indian	Hindu

Now, from column 1 of Table 2, we find there are four names starting with A, two names starting with D as well as L, one name starting with each of the letters F, G, M, N, P and Y, and six names starting with R. It is observed that the largest number of names is starting with R and no names are starting with the letters B, C, E, and H to K, O, S to X and Z.

Information Relating to Age

Now, we take column 3, and arrange the students in the ascending order of their ages (Table 6.3)

Table 6.3: BLIS Students (arranged according to age)

Sr. No	Name	Age	Sex	H.Q.	Class in H.Q.	Nationality	Religion
5	Asit	22	M	BA	2nd	Indian	Hindu
17	Latif	22	M	BA	2nd	Indian	Muslim
13	Rabeya	22	F	M Sc	1st	Bangladeshi	Muslim
18	Aparna	23	F	BA	2nd	Indian	Hindu
15	Dan Singh	23	M	B Com	2nd	Indian	Sikh
12	Monica	23	F	B Sc	1st	Indian	Hindu
9	Rahul	23	M	B Sc	2nd	Indian	Hindu
6	Yogendra	23	M	B Com	1 st	Indian	Hindu
11	Ali	24	M	BA	2nd	Indian	Muslim
3	Diana	24	F	BCA	1st	Indian	Christian
16	Farooq	24	M	BA	1st	Indian	Parsee
7	Laxman	24	M	M Sc	2nd	Sri Lankan	Buddhist
19	Renu	24	F	B Sc	2nd	Indian	Hindu
8	Rabi	25	M	MA	2nd	Indian	Hindu
10	Rajib	26	M	M Com	2nd	Indian	Hindu
4	Aurobinda	27	M	MA	2nd	Indian	Hindu
14	Ram Bahadur	29	M	BA	1st	Nepali	Hindu
2	Gita	33	F	B Sc	1st	Indian	Jain
20	Prashant	34	M	M Com	2nd	Indian	Hindu
1	Nirmalendu	41	M	BA	2nd	Indian	Hindu

Table 6.3 shows that the minimum age of a BLIS student is 22 and maximum 41. There are three students of age 22, five students of age 23 and another five of age 24. That means the age of 13 (65%) students range between 22 to 24. The age of rest of the students range from 26 to 41. The Table indicates that even slightly elderly students also join this course.

Information Relating to Sex

Now, we move on to the fourth column, and arrange Table 1 according to sex and get Table 6.4.

Table 6.4: BLIS Students (arranged according to sex)

Sr. No	Name	Age	Sex	H.Q.	Class in H.Q.	Nationality	Religion
12	Monica	23	F	B Sc	1st	Indian	Hindu
19	Renu	24	F	B Sc	2nd	Indian	Hindu
2	Gita	33	F	B Sc	1st	Indian	Jain
18	Aparna	23	F	BA	2nd	Indian	Hindu
3	Diana	24	F	BCA	1st	Indian	Christian
13	Rabeya	22	F	M Sc	1st	Bangladeshi	Muslim
15	Dan Singh	23	M	B Com	2nd	Indian	Sikh

6	Yogendra	23	M	B Com	1 st	Indian	Hindu
9	Rahul	23	M	B Sc	2nd	Indian	Hindu
5	Asit	22	M	BA	2nd	Indian	Hindu
17	Latif	22	M	BA	2nd	Indian	Muslim
11	Ali	24	M	BA	2nd	Indian	Muslim
16	Farooq	24	M	BA	1st	Indian	Parsee
14	Ram Bahadur	29	M	BA	1st	Nepali	Hindu
1	Nirmalendu	41	M	BA	2nd	Indian	Hindu
10	Rajib	26	M	M Com	2nd	Indian	Hindu
20	Prashant	34	M	M Com	2nd	Indian	Hindu
7	Laxman	24	M	M Sc	2nd	Sri Lankan	Buddhist
8	Rabi	25	M	MA	2nd	Indian	Hindu
4	Aurobinda	27	M	MA	2nd	Indian	Hindu

Table 6.4 shows that the number of female students is 6(30%), and male students 14(70%). Obviously, there is a preponderance of male students in the class.

Information Relating to Highest Qualification

To generate this information, data of the fifth column has been organised in alphabetical order (Table 6.5). From the Table we find that as many as seven students have BA as their highest qualification, followed by B Sc (4); B Com, M Com, M A, and M Sc (2 each); and one BCA student. It may be noticed that students belonging to Arts are joining the profession more than the students of any other discipline.

Table 6.5: BLIS Students (arranged according to highest qualification)

Sr. No	Name	Age	Sex	H.Q.	Class in H.Q.	Nationality	Religion
6	Yogendra	23	M	B Com	1 st	Indian	Hindu
15	Dan Singh	23	M	B Com	2nd	Indian	Sikh
12	Monica	23	F	B Sc	1st	Indian	Hindu
19	Renu	24	F	B Sc	2nd	Indian	Hindu
9	Rahul	23	M	B Sc	2nd	Indian	Hindu
2	Gita	33	F	B Sc	1st	Indian	Jain
18	Aparna	23	F	BA	2nd	Indian	Hindu
5	Asit	22	M	BA	2nd	Indian	Hindu
1	Nirmalendu	41	M	BA	2nd	Indian	Hindu
14	Ram Bahadur	29	M	BA	1st	Nepali	Hindu
17	Latif	22	M	BA	2nd	Indian	Muslim
11	Ali	24	M	BA	2nd	Indian	Muslim
16	Farooq	24	M	BA	1st	Indian	Parsee
3	Diana	24	F	BCA	1st	Indian	Christian
10	Rajib	26	M	M Com	2nd	Indian	Hindu
20	Prashant	34	M	M Com	2nd	Indian	Hindu
13	Rabeya	22	F	M Sc	1st	Bangladeshi	Muslim
7	Laxman	24	M	M Sc	2nd	Sri Lankan	Buddhist
8	Rabi	25	M	MA	2nd	Indian	Hindu
4	Aurobinda	27	M	MA	2nd	Indian	Hindu

Information Relating to Class in Highest Qualification, Nationality, and Religion

Arranging the data of the columns six, seven and eight in desired manner, we find that while obtaining the highest qualification seven students secured first class and thirteen students second class. Also we find that in the class there are seventeen Indian students, one Bangladeshi, one Nepali, and one Sri Lankan

student. Of the students, as many as twelve are Hindu, three are Muslim, and the rest five are Buddhist, Christian, Jain, Parsee, and Sikh, obviously only one student belongs to each of these religions.

Now, you can see how from a single set of data a substantial amount of information can be generated just by processing it in different ways.

6.4.5 Events

The Concise Oxford Dictionary defines an event as ‘a thing that happens or takes place [Pearsall, Judy (ed), 1999]. A scholar releasing a book, an artist inaugurating an exhibition, a philosopher explaining the concept of time, a saint giving a discourse on a religious matter, a political leader campaigning for vote, legislators debating in a parliament, a lawyer passing a judgement, the prime minister of a country taking oath of office, a war breaking out at a particular region of the world, a patient dying in a hospital due to neglect, a new train being flagged off by a minister, two buses colliding causing death and injuries to a number of passengers, a terrorist hijacking a plane, a comet appearing in the sky, a physicist bombarding an atom with alpha particles, a chemist conducting an experiment to create a new material, a geologist drilling a borehole to prospect petroleum deposit, a paleontologist spotting the skeleton of a dinosaur, a geneticist giving birth to a high-yielding variety of rice, an inventor filing a patent application, a surgeon performing an open heart surgery, a director shooting a new film, umpteen number of sports and games being held every day all over the world, adventurers venturing to conquer a mountain peak, etc. are all examples of events.

Let us go through some of the headlines figured in *The Hindustan Times* of 30th July 2004.

- 1) IA diverts flight, saves Pakistani baby. (p 1)
- 2) Bofor's Ardbo dead (p1)
- 3) Aravali bio-diversity park takes off (p2)
- 4) PM flies, spares traffic (p3) [PM's journey to the airport by a helicopter while proceeding to Thailand]
- 5) 16 lockers cleaned out at Safdarjung Enclave Bank (3)
- 6) 1 killed as bus overturns in NOIDA (p5)
- 7) As expected Pak win easily [in Asia Cup ODI in Sri Lanka](p9)
- 8) CA examination results on August 4 (p19)
- 9) Pakistani hostages killed (p21)
- 10) Madonna makes peace with Jesus in Portugal (p22).

It can be seen that all these news items have generated out of events. A newspaper not only provides information about past events but also of future events (vide headlines against serial no. 8). It is also to be noted that a newspaper also contains items of thought process, (e.g. editorials); data processing (prediction about election results); observation (reporting of an accident); and so on.

It is to be noted that just the happening of an event does not generate information. The information relating to the event generates only when it is seen by someone and reported.

6.4.6 Evolution

Man started communicating by speech some 100,000 years ago [Odham's Colour Library of Knowledge: Language and Communication, 1968]. In those dizzy old days of human civilisation, the vocabulary of human beings of a particular race was only limited. They had only that many words which were required to express their ideas. As they invented newer and newer devices, encountered objects not known before, they started naming them for the purpose of easy identification. This led to the enrichment of vocabulary. When they moved from an old area to a new area they encountered numerous new things such as trees, animals, fruits, tubers, and so on. They also named them. In the course of their endless journey sometimes they encountered an alien race, which resulted either in fighting or friendship. For the exchange of ideas between two different races, need arose for interpretation. How and when the art of interpretation came into being is shrouded in mystery. However, this much is sure that the need for interpretation occasioned the need for linguistic information. That is, how a particular object is called by the alien group. In plain words, they needed information as to the equivalent terms in the language of the alien race. The process of building up of the vocabulary of languages is as mind-boggling as interesting. Sometimes a word has travelled over centuries from one language to the other retaining its original form. For example, the word 'ginger' belonging to a now extinct Middle Indian language travelled from India to Europe through Middle East. In twenty-four languages of the world the word has retained its original form. In many cases words have undergone small changes while travelling from one language to the other. For example, the word 'nine' has taken the following forms in a number of Indo-European languages of the world. It seems as if the word first originated in a particular language and then moved from language to language with small changes in its sound and form. It is depicted below:

Dutch	English	French	German	Italian	Spanish	Swedish	Welsh	Hindi
Negen	Nine	Neuf	Neun	Nove	Nueve	Nio	Naw	Nau

In some cases, of course, we find that a word having the same or similar meaning differs from language to language. Let us take the example of the word 'science'. In Sanskrit it is called 'vijñan', in Russian 'nauka', in German 'Wissenschaften' and so on. In the development of vocabulary, conquest of a country by another, international trade, etc. has also helped a lot. The development of languages, which originated 2,000 or more years ago, is still continuing. Take for example, the English language. Even today, hundreds of new words are being added to this language every now and then. Just with e, we have e-mail, e-journal, e-zine, e-library, e-librarian, and so on. Similar is the case with cyber. We have hundreds of words starting with cyber. Even a decade ago most of these words were unknown. In this way a language has developed over thousands of years following the path of evolution and still the process is on.

Several thousand years ago, man invented the art of writing. Starting from stones down to leaves were used as writing media. Products of human thought thus started getting recorded. At one time the art of translation also began. In the famous 'Rosetta Stone' of 2nd century BC, the decree has been engraved in three different languages, ancient Egyptian hieroglyphics, demotic (the popular language of Egypt at that time), and Greek [Odham's Colour Library of Knowledge: Language and Communication, 1968]. This record bears infallible testimony to the fact that the art of translation was pretty developed in Egypt at that time.

The requirement of translation necessitated reference tools to provide information about the equivalent words in the target language in which the matter is to be translated. In response to the need started appearing bilingual and multilingual dictionaries. For translating, sometimes we need information even for a letter. For example, we come across letters like ॐ in Indus scripts. Scholars have been trying for about a century to extract information from these scripts.

6.4.7 Dream

It is common with every human being to dream. Some of the dreams we remember, other we do not. The dream that we remember and convey about the same to others or record it in our diary, information is generated. Psychologists extract a lot of information about the subconscious mind of a patient through the interpretation of dreams. At times dreams provide the necessary information or clue for solving a problem. Kekule, an organic chemist, was trying to find out the structural formula of benzene for quite sometime, but failing again and again. One night, he dreamt that six snakes had created a ring by biting the tail of each other. This dream immediately gave him the information or clue that the structural formula of benzene would be ring-shaped.

6.4.8 Information Generation in some Selected Subjects

Following different modes, information is generated in various subjects. Taking a few selected subjects, we shall see how information generates in those subjects.

1) Classification

Going through any classification scheme like Dewey Decimal Classification or Colon Classification deeply, we can easily realise the tremendous amount of mental exercise that has undergone for building up the scheme. Sometimes, the mental exercise is preceded by observation or an event. For example, Ranganathan's chance encounter with a meccano set sparked in his mind the idea of a faceted scheme of classification. Moreover, every designer of a scheme of classification had to think of sound canons, principles, devices and so on to maintain consistency, provide logical sequence of subjects, meet the present need and face with confidence the demand of the future. The amount of mental exercise required to build up a classification scheme can be well realised if we ourselves undertake building up a small classification scheme. In this particular subject, information generates mainly through mental exercise.

2) **Philosophy**

Ever since man learned to reason, he started thinking about creation, the creator, environment, life, and all other phenomena he came across. The Greeks, centuries before the birth of Christ tried to understand the nature of the world for the sake of knowledge. In their reasoning, they introduced logical element thus occasioning the beginning of philosophic and scientific thought. Thales (c640 – 546 BC), the Ionian Greek, as well as the first philosopher recorded in history, speculated about the origin and nature of the universe [Odham's Colour Library of Knowledge: Religion and Philosophy, 1968]. From then on till date, the world has seen thousands of philosophers putting forward their ideas on matter, energy, space, time, cosmos, God, knowledge, morality, beauty, and so on. All these are the fruits of their fertile imagination and logical thinking. It will not be very wrong to say that the subject philosophy has borne out of human thinking.

3) **Religion**

In the dawn of civilisation, man with utter helplessness observed numerous natural phenomena like devastating earthquakes, ravaging floods, deadly forest fires, thundering storms, awesome volcanic eruptions, and so on. Even with their limited knowledge, the early men could realise that behind each of these phenomena, there is someone who is enormously powerful and humans are not mighty enough to control and overpower him. Thus dawned in them the idea of supernatural beings basing their imagination and thought process. They reasoned that when these mighty beings get enraged, all natural calamities occur. To appease those mighty beings, developed various forms of worship. Phenomena like diseases, deaths, crop failures, etc. also made them realise that there are many things, which are beyond their control. These also strengthened their beliefs in supernatural entities, which at one time took the form of Gods and Goddesses. In the long journey of human civilisation took birth great religious leaders like Gautam Buddha, Jesus Christ, and others. All of them founded religions basing what they realised out of their meditation or deep thinking about God and related matters. Hence, we find religion is also the outcome of the brainwork of human beings.

4) **Social Sciences**

In social sciences information is generated following various modes. Here, we shall demonstrate this taking the example of statistics, politics and government, and law.

a) **Statistics**

In this field, information is mostly generated through the processing of data. In any statistical survey, e.g., census, data is collected usually using a questionnaire. Once the collection is over, data is cleaned, fed into the computer according to the predetermined format, and then processed to generate the desired information. Of course, the statistical theories, formulas, etc. come into being through process.

Self Check Exercise

- 3) Explain how is statistical information generated.

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

Generation of Information:
Modes and Forms

b) **Politics and Government**

In these two fields, information mostly is generated through events. A political leader inaugurates a conference and delivers a speech that becomes information. The election process generates a good deal of information encompassing such events as nomination of candidates by various political parties, filing of nomination papers, issuance of the manifesto, campaigning, exercising of franchise by the electorates, and formation of the ministry. However, political theories, ideologies, etc are the brainchild of political thinkers.

5) **Pure Science**

In pure science, information is generated through thinking, observation, experimentation and so on. Here, we shall discuss the generation of information in mathematics, astronomy, and physics.

a) **Mathematics**

Whatever we find today in mathematics starting from Hindu numerals (1, 2, 3, etc.) up to any sophisticated calculation has been generated from human thinking. Of course, in applied mathematics, there might be some exceptions. By and large, mathematics is a thought-based subject.

b) **Astronomy**

Unlike mathematics, information has been generated in astronomy mainly through observation. Early men observed the sky years after years. The change in the phases of the moon possibly delighted them. The eclipses obviously generated in them a sense of fear. They noticed some heavenly bodies move faster than others. Gradually and gradually they identified the constellations, zodiac, and the planets visible with the naked eye. At certain point of time, they realised the return of seasons at regular intervals and the concept of year dawned in them. Hence, we find that the entire edifice of astronomy was built on the observation power and the thought process of astronomers. The discovery of telescope enhanced the observation capacity of human beings manifold as far as distant objects are concerned. Using telescopes men could discover many new galaxies, nebulae, stars, and other celestial bodies. When the life of a star comes to an end, it explodes violently throwing the debris with unimaginable force into the space. With naked eyes such phenomenon is rarely seen from the earth. However, in the early morning of 24 February 1987, the phenomenon of an exploding star, named Supernova 1987A, was observed simultaneously

by astronomers Ian Shelton of the University of Toronto, Southern Station, Chile; Oscar Duhalde of Las Campanas Observatory, Chile; and amateur sky watcher, Albert Jones of Nelson, New Zealand. The explosion took place in the galaxy called Large Magellanic Cloud and was visible with naked eyes from the Southern Hemisphere [Britannica Book of the Year, 1988]. The astronomers informed the media and the entire world came to know about the astronomical phenomenon. This is how an important piece of astronomical information generated.

c) **Physics**

Generation of information in physics involves events, observation, and experimentation and thought process. It is, however, not necessary that all the four factors will be required for the generation of information in all cases. Newton saw the falling of an apple from a tree. The event led Newton to think deeply about the phenomenon and his thinking finally gave birth to the famous theory of gravitation. Here, event, observation and thought process were responsible for the generation of information. Michael Faraday conducted a number of experiments by passing electricity through various substances. Once he passed electricity through a vacuum tube and was astonished to see a beam of light travelling from the positive electrode to negative electrode. He further experimented by subjecting the beam of light to strong magnetic field and amused to see that the beam was attracted towards the positive pole of the magnet. From the experiment he could make out that the beam was composed of negatively charged particles. Later on, these particles were named as 'electrons'. Here also we find that experimentation, observation and thinking process – all these three were involved in the generation of information. In the propounding of many theories thinking process becomes paramount compared to observation and experimentation. For example, in propounding the theory of relativity Einstein did not conduct any experiment. His thinking process coupled with mathematical calculations gave rise to the theory. Experiments were conducted afterwards to test the validity of the theory.

6) **Applied Science**

Here we take the examples from medicine. Medicine comprises anatomy, physiology, health, pharmacology, pathology and surgery. Information in these subjects generates following different modes, which will be clear from the following discussion.

a) **Anatomy**

Information on anatomy has been generating since ancient times. In Ayurveda, there is clear-cut instruction as to how information is to be gathered about human anatomy using the dead body of a healthy person. In modern times, doctors dissected body of a dead person to observe bones, muscles, arteries, nerves, veins, nerves and so on. They observed and recorded accurately the exact location, size, shape etc of every part of the body to generate detailed information about the human body. Veterinary doctors did the same thing to reveal the anatomy of animals. Thus, it is clear that information on anatomy has generated mainly based on observation.

b) **Physiology**

Generation of information of physiological processes, in many cases requires a great deal of thinking apart from observation. In some cases experimentation is also necessary. For example, the discovery of circulation of blood by Galen required a great deal of thinking besides observation.

c) **Health**

World Health Organisation brings out publications providing the status of public health in the world. For generating the information, data is collected from various countries of the world through questionnaire and other methods, and then the data is processed to generate information.

d) **Pharmacology**

Every year hundreds of drugs undergo trials throughout the world. The drug, under trial, is given to a set of patients, and another set of patients is given placebo [a harmless substance that exactly looks like the drug, but not the drug in reality]. The result of the drug as well as the placebo is observed taking into account the process of healing, side effects and so on. Interestingly, many patients get cured with the placebo! In case, more patients are healed with the drug than with the placebo, and the side effects are not serious, then the drug is allowed for use and the information generated is made public. Usually, the trial goes on for years throughout the world. Generation of information in this case involves experimentation, observation, and processing of data.

e) **Surgery**

In the case of surgery also, in most cases, prior diagnosis is necessary using x-ray, ultrasound, and pathological and other methods. During surgery, information is generated about the patient's blood pressure, heart condition and so on to help the surgeon to proceed with the surgery. After surgery, the condition of patient is monitored and information generated about the progress of the patient to help the doctor take next step. Here also, several processes are involved for the generation of information.

Summing up, we find that in the field of medicine also experimentation, observation, processing of data and thinking is necessary to generate information.

7) **Arts**

An artist using his thinking process gives birth to a painting, an idol, an object of sculpture, a song or music. Information about many of these is generated through events such as an exhibition. When the painting called *Mona Lisa* by Leonardo da Vinci was stolen from Louvre Museum in Paris, a great deal of information was generated. When an artist organises an exhibition of his paintings information is generated.

8) **Sports**

Generation of information in this case is mostly event-based. A football match, a cricket match, or any other sports or games generate information. Needless to say, international game events like Olympic, Commonwealth Games, Asiad,

etc. generate a huge amount of information.

9) Literature

Authors compose poems, write fictions or dramas, script erudite essays, and so on basing their imagination or thinking capabilities. Information about these writings are generated usually through some events like the release of a book at a function, publication of its reviews in a journal, inclusion of its bibliographical details in a bibliography and so on.

10) Geography

People since time immemorial have gone from one place to another for various purposes. In the process of their journey they observed new places, came across new people and their culture. The information they gathered was passed on to others. Thus, the subject of geography was born. At a later stage, the entire world was mapped applying scientific method of observation. Hence, geography is predominantly an observation-based subject.

11) History

Information on history is predominantly event-based. Thinking also plays a major part. The events like the birth of a king's son, his ascendancy to the throne, his meetings with dignitaries and VIPs, his decisions, and all other activities generate information. In a democratic set up, the election of a president, prime minister and other ministers and their respective activities generate huge amount of information. Recording of all these activities including wars create history.

6.5 FORMS OF INFORMATION

We are already familiar with the forms of documents such as books, periodicals, patents, theses, standards, catalogues, dictionaries, encyclopaedias, bibliographies, and so on. The form of information is different from the form of documents. In a document, the information may be in hand-written form, printed form, coded form, simplified form, disguised form, and so on. Here, we are going to deal with oral form, hand-written form, printed form, digitised form, condensed form, coded form, simplified form, translated form, and disguised form. This is not an exhaustive list.

6.5.1 Oral Form

When we talk to a person sitting or standing in front of him or through telephone, the information exchanged between the two is oral form of information. This form of information is extensively used in our day-to-day conversation, in the lectures delivered by teachers, speeches delivered by ministers, notes and letters dictated by officers, cross examination undertaken by lawyers, interviews conducted by interviewers, commands given by military officers, and so on. In our Vedic period also this was the most predominant form of information. People used to listen (*shruti*) and remember (*smriti*). They also did not have any system of writing. Illiterates all over the world use this form of information to express themselves. Using this form of information does not require the knowledge of the scripts and the spelling of words. There are tribes in the

world who speak languages that do not have any script. Obviously, those tribes use oral form of information for communication.

6.5.2 Information in Sign Language

Deaf and dumb people cannot use oral form of information. Hence, sign languages have been developed for them whereby they communicate. In this case, information is generated using hands, fingers, and other parts of body. For example, using our forefinger and middle finger we produce the shape of V to indicate victory. Many a time we indicate our consent with a nod. People all over the world use sign languages to communicate. Some primitive tribes use sophisticated sign languages for communication.

6.5.3 Hand-written Form

People of ancient Egypt, Indus Valley, China developed systems of writing much before the advent of Christian era. With this development the written form of information came into being. Different languages of the world developed varied scripts and alphabets. Sometimes a group of languages adopted the same script and alphabet with slight change here and there. For example, languages like English, German, Spanish, Italian, Portuguese, and Rumanian use the Roman scripts. Even in our country, Roman scripts are used for writing in languages like Mizo. Some ethnic groups like the Romans preferred writing from left to right; the Arabs preferred writing from right to left, and the Chinese from top to down. For recording hand-written information numerous recording materials like papyrus, parchment, vellum, *bhurjapatra*, palm leaves, terracotta tablets, bamboo strips, etc. were used. The advent of paper extensively reduced the use of other writing materials. Lakhs of manuscripts preserved in thousands of libraries of the world are all hand-written. Scribes thrived in those centuries when printing was still a very distant dream. Copying manuscripts was the order of the day and a respectable means of earning livelihood.

6.5.4 Pictorial Form

In newspapers, everyday, we see cartoons conveying some message in a humourous or satirical way. General maps and atlases present information in pictorial form about places, rivers, lakes, mountains, universe, and so on. Anatomical atlases present information about various parts of the body. There are other types of atlases as well. A photograph of Taj Mahal, a portrait of Shakespeare, a statue of Gandhiji, etc. tell us about their look. Children like to read comics such as *Amar Chitra Katha*. In many cases, say in comics and cartoons, words and/or sentences are also added to make it more comprehensible and interesting.

6.5.5 Printed Form

The Chinese developed the technique of block printing by 8th century AD or before since the oldest known example of block printing recording Buddhist charms dates to around 765 AD. The Chinese and Koreans also developed the technique of printing from movable types. However, their technique remained confined within the region. The technique of printing from movable types that spread the world over was developed by the German printer Johannes Gutenberg

in mid-1450s [9]. The impact of Gutenberg's technique was so great that by the end of 15th century some 9,000,000 books comprising mostly religious writings, and classical works of Greek and Roman authors were in circulation all over Europe. The world has never seen the generation of so many titles in just less than 50 years! [Odham's Colour-Library of Knowledge: Languages and Communication, 1968]. Even today, printed information reigns supreme as maximum amount of recorded information, is in printed form.

6.5.6 Digitised Form

With the advent of computers, digitised form of information came into being. Here information is recorded using only two digits 0 and 1. For digitisation, there are codes like ASCII and EBCDIC. ASCII stands for American Standard Code for Information Interchange and EBCDIC for Extended Binary Coded Decimal Interchange Code. ASCII is 7-bit code, and EBCDIC 8-bit code. However, extended ASCII is 8-bit code. In extended ASCII, the word INFORMATION will be recorded as 1100100111001110110001101100111111010010110011011100000111010100 110010011100111111001110. This is digitised form of information. Presented in this form, a computer will be able to easily handle the data.

6.5.7 Condensed Form

We are all aware of abstracts and summaries. These are nothing but condensed form of information. An article of twenty pages may have an abstract of half a page only. Abbreviations like ILA, BLA, etc. are also condensed form of information. In fact, condensed form is secondary form of information as it is always derived from the primary form of information.

6.5.8 Coded Form

In the coded form of information, usually numbers, letters and symbols are used. Sometimes, they are used singly, and sometimes in combination with one another.

We find coded form of information in a number of subjects including our own. When we represent the subject of a book on science with a class number like 500 or A, we put the information in a coded form. A class number like 954 or V2 immediately tells the classifier that the number signifies 'history of India'. The above examples show us that the information can be codified simply with numbers, letters, or a combination of both.

The information exchanged in espionage is normally in codified form. Each country has its own codes for the purpose. For deciphering the coded information there are keys to the codes. Once keys are available, coded information can be deciphered without much of a problem. There are also experts who can break the codes even without the keys.

In trade and commerce we use \$ to mean dollar, £ to mean pound, ¥ to mean yen, @ to mean at the rate of, and so on.

We use PIN code in postal communication. For example, 110019 indicates Kalkaji area of New Delhi encompassing Kalkaji, Nehru Place, Chittaranjan

Park, Alaknanda and few other places. From the term 'PIN code' itself, it is clear that the information is in coded form.

The number written on the number plate of a car or scooter is also coded information. From the number it is possible to make out whether the vehicle belongs to the government, an embassy, or an individual. The state wherefrom the license for the car has been issued can also be identified. For example, the number TN01S9899 indicates that the vehicle belongs to Tamil Nadu. The number DL1T 2345 informs that the vehicle is a Delhi taxi.

In mathematics, we encounter symbols like +, −, ×, ÷, <, =, > etc. meaning respectively plus, minus, into, divide, less than, equal to, and greater than. In geometry ∠ indicates an angle, Δ a triangle, O a circle, □ a square, and so on.

Astronomers also use various symbols like O, ☉, ♀, ☿, to mean respectively the sun, mars, full moon, new moon, and star.

Abbreviations and equations in physics are also coded form of information. The formula that provided the base for the manufacture of atom bomb is as simple as $E = mc^2$ where 'E' means energy, 'm' means mass and 'c' velocity of light. It indicates that the annihilation of one kilogram of mass can generate $1 \times 300,000,000 \times 300,000,000 = 9 \times 10^{16}$ joules of energy.

Chemists all over the world use coded form of information for representing names of elements as well as compounds. For example, to them Fe means iron, Pb lead, NaCl sodium chloride and so on.

In telegraphy, Morse code is being used to send telegrams. As in digitisation, in Morse code also, two symbols are used to represent all the 26 letters of Roman alphabet and the numbers. The two symbols are • (dot) and - (dash). For example in Morse code A is represented by • -, B by - •••, and so on. When a ship or an aircraft in danger sends the distress signal SOS, it is sent with the following Morse code ••• - - - •••.

6.5.9 Simplified Form

Writings in many subjects are pretty difficult to understand for a common man. One requires special knowledge to grasp. Moreover, an expert in a subject may be a total novice in another. Hence, simplification of information becomes essential for laymen, school children, and in certain cases even for scholars.

Now let us see how difficult-to-understand sentences can be made simple and easily understandable. John Keog of Australia patented a 'circular transportation facilitation device' in the year 2001 and was awarded the Ig (Ignoble!) Nobel Prize [11] for his novel invention [Australian Patent No. 2001100012]. The sentence may not be comprehensible to every one because of the jargon used. If we rewrite the sentence as – John Keog of Australia patented a 'wheel' in the year 2001 and was awarded the Ig (Ignoble!) Nobel Prize [Abrahams, M, 2004] for his novel invention, then everybody will understand the sentence without any difficulty. Let us take another sentence – Mr. Salim Ali was numero uno in ornithology in India. The sentence may not be comprehensible to a layman or school-going children because of the two terms numero uno and ornithology. If we rewrite the sentence as 'Mr. Salim Ali was the number one bird specialist in India', everybody will get the meaning of the sentence quite

easily. Thus we find difficult writings can be simplified to make them understandable to a layman.

Information is generally presented in simplified form in encyclopaedias meant for children and laymen. In these encyclopaedias, usually jargons; long, difficult and highly complex sentences are avoided and whenever a technical term occurs, its meaning is explained in simple terms.

6.5.10 Translated Form

There are numerous languages in the world, which are extant. There are also languages, which are extinct. Some of the extinct languages of the world had even written form, e.g. hieroglyphics of Egypt. The people of Indus Valley civilisation had also a language in written form.

The need for translation of information from one language to the other was felt by men even in ancient times. We find in Rosetta stone, the content in hieroglyphic and its translation in demotic and Greek. It may be noted here that the translation of the hieroglyphic content into Greek was of great help to the linguists like Champollion in deciphering writings in Egyptian hieroglyphics. The deciphering of the Egyptian hieroglyphics opened up before the world the entire history of Egypt.

We have already discussed about the oral and written form of information. The translation can be from oral to oral form, oral to written form, written to oral form, and written to written form. In the written form we include the printed form as well.

i) Oral to Oral Form

In the UN Assembly, international conferences, and in many other occasions the speaker delivers his speech in a particular language and immediately it is translated into several other languages by interpreters. When two persons do not know the language of each other, they take the help of an interpreter to convey their messages. In these cases the translated form of information is oral.

ii) Oral to Written Form

In Purnea Zila School where I studied in early 1950s, the teachers in subjects like social studies used to deliver the lecture in Hindi. The Bengali students in the class immediately translated the same mentally and took down the notes in Bengali. This was a case of oral to written form of translation.

iii) Written to Oral Form

In India and many other countries of the world an English teacher in a non-English medium school orally translates the poems, short stories, essays, etc from English language to local language (Bengali, Hindi, etc.) to make the student understand the thought content of the piece of writing. In this process, the student understands the thought content of the piece of writing much better. In bilingual mode of teaching, the practice is quite common all over the world.

iv) Written to Written Form

Every year thousands of books are translated from one language to another. For example, Nehru's *Glimpses of World History* has been translated into Arabic, Assamese, Bengali, Croatian, German, Gujarati, Hindi, Japanese,

Kanarese, Malayalam, Marathi, Mongolian, Oriya, Persian, Russian, Swedish, Urdu, and so on. Apart from books journals, articles are also translated. Once upon a time, INSDOC used to translate articles and books from around twenty non-English languages of the world to English language. Many Russian journals are translated from cover-to-cover.

There are various other modes of translation such as natural language to digitised language and vice-versa; sign language to natural language and vice versa; and so on.

6. 5.11 Disguised Form

In a disguised form of information, sentences appear pretty innocuous to a common man. Only the person, for whom the information is meant, can get the real meaning of the sentence. One of our drivers at INSDOC used to stop the car at a particular point on the road and say, "Sir, I am going to fetch my medicine". Only those close to him used to know that he was going to fetch a bottle of wine! In 1942 Arthur H Compton, head of the nuclear fission project at the University of Chicago telephonically told James B Conant, Director of the National Defense Research Commission of the USA, "The Italian navigator arrived at the shores of the new world and found the natives were quite friendly. It is a smaller world than he believed" [Cane, P, 1961]. The sentences looked as if somebody like Columbus had found out a new continent. It was not so. In reality, the information was about the successful achievement of the first nuclear chain reaction by Enrico Fermi, the very first step towards the manufacture of atom bomb. In the aforesaid quotation the 'Italian navigator' was Enrico Fermi, 'the new world' meant the world of atomic energy, 'natives were friendly' connoted that the reaction could be controlled, and 'smaller world' gave an idea of the amount of uranium needed [Cane, P, 1961]. I think the example makes it clear what the disguised form of information is.

Self Check Exercise

- 4) How do you differentiate form of documents from the form of information? Explain briefly.
- 5) Why sometimes information is passed in disguised form?

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

6.5.12 Secondary Form

So far we have talked about the generation of primary information. Secondary information is generated basing the primary information. When somebody

prepares an abstract of a research article, the abstract becomes the secondary information. Most of the editorials in newspapers are based on primary information. However, the editorials themselves are secondary information. Suppose, the railway budget has been presented in a parliament. Some newspapers have written editorials on the budget. Here, the railway budget provides the primary information and the editorials on it are secondary information. Encyclopaedias, dictionaries, indexing periodicals, abstracting periodicals, etc. usually contain secondary information.

6.5.13 Tertiary Form

In many cases tertiary information is generated basing secondary information. For example, if a scientist writes a review article basing the abstracts of the original articles, then the review article provides tertiary information. Sometimes review articles are based on both primary and secondary information.

6.6 IMPACT OF INFORMATION TECHNOLOGY ON INFORMATION GENERATION

Information technology has impacted information generation in a number of ways. Let us take the case of a fledgling author. The author writes an article and tries to get it published. In many cases, he has to move from pillar to post to see his writing in printed form. Sometimes, he is successful, and sometimes not. Where he is successful, information is generated. Where he is not the article with the passage of time vanishes from the world remaining totally unknown to anybody. Now, the author has options. He can place the article in his own website, if he has one. Otherwise, there are websites where he can place his articles. The article will come to the notice of thousands of people. Thus, information will be generated. To what extent information contained in the article is authentic is a difficult question. It may be authentic, partially authentic and may be devoid of authenticity. The user will have to judge it. In Internet, there are many articles written by school children, where you will find spelling mistakes, grammatical errors, factual inaccuracies, and so on.

The IT era has given birth to a number of publications in electronic form. These publications among others comprise books, periodicals, and conference proceedings. *Cybermetrics*, an electronic-only journal as well as a virtual forum is devoted to the study of the quantitative analysis of scholarly and scientific communications in the Internet. The URL of the journal is: <http://www.cindoc.csic.es/cybermetrics/cybermetrics.html>. Nowadays, like books and periodicals, there are many articles which are available on Internet only. For example: the article by Kalyane and Sen on Tibor Braun, is available only on Internet. The number of electronic-only publication is continuously increasing opening up a new gateway for the generation of information as well as access to information.

IT has also helped in the generation of quicker as well as better information. Before the advent of IT era, information search used to take a huge amount of time. Moreover, information that could be searched was not very up-to-date. For example, a journal from US or European countries used to take a month or more to reach India by sea mail. Hence, in a place like India we used to get at least one-month old information, sometimes more. Today, the time for locating

and accessing information in many cases has been reduced considerably. In all probability, this has increased the productivity of researchers and many other writers. Moreover, they are now in a position to provide more updated information in their publications.

Self Check Exercise

- 6) Describe briefly the impact of information technology of the generation of information .

Notes: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

6.7 SUMMARY

In the beginning, we have learnt about various objectives of the Unit. Subsequently, we have seen how the concept of 'information' connotes different meanings in different disciplines. We, LIS professionals generally deal with the information that is recorded in some media. Information generates through various modes such as observation, thought process including deliberation and imagination, experimentation, processing of data, and occurrence of various events. The generation of information in different areas follows different modes. To illustrate this point, the generation of information has been discussed in areas like classification, philosophy, religion, statistics, politics and government, mathematics, astronomy, physics, anatomy, physiology, health, pharmacology, surgery, sports, literature, geography and history. The form of information is different from the form of documents. In the form of documents, the presence of a document is a must. In the form of information, document may or may not be present. In a document, the information may be in hand-written form, printed form, coded form, simplified form, disguised form and other forms. In this Unit we have discussed oral form, hand-written form, printed form, digitised form, condensed form, coded form, simplified form, translated form, and disguised form of information. Generation of secondary information and tertiary information has also been mentioned. At the end we have discussed the impact of information technology on information generation.

6.8 ANSWERS TO SELF CHECK EXERCISES

- 1) Observation is one of the most potent modes of generation of information. We observe an object, an event, an experimentation, and so on and record that in our brain. When we communicate this to other, it becomes information. We can communicate by speaking, sign language, writing

or recording the information in some other media. Suppose, a shepherd has seen a wolf in the jungle bordering his village. The moment he conveys this message to a villager, information is generated. Monica has visited Lakshmi Narayan temple in Delhi. She conveys this fact to her friend Sonia through a letter. To Sonia, this is an information. Here, Monica generated information by recording it in a letter. Astronomers all over the world generated information by observing the celestial bodies for centuries initially with naked eyes and subsequently with telescopes. Similarly, microbiologists generated information on all microbes observing them with microscopes. A scientist generates information by conducting experiments. A doctor generates information by examining a patient about the disease the latter is suffering from. Thus we find, in numerous cases we generate information by observing.

- 2) An artist first of all imagines about the art piece he is going to create, a chemical engineer also visualises in his mind the chemical plant he is going to install, an architect also draws a picture of the building he is going to design in canvass of his mind. As the work of all these starts, many a time people come to know about these from the person concerned and information is generated.
- 3) Statistical information generates mostly through the processing of data. In any statistical survey, e.g. census, data is collected usually using a questionnaire. Once the collection is over, data is cleaned, inputted into the computer according to the predetermined format, and then processed to generate the desired information. If the amount of data is small, it can be processed manually. Of course, for the generation of statistical theories, formulas, etc. a great deal of thinking is necessary.
- 4) Documents are found usually in the form of books, periodicals, patents, theses, reports, standards, catalogues, dictionaries, encyclopaedias, bibliography and so on. The form of information is different from the form of documents. In the form of documents, the presence of a document is a must, which is not the case with the form of information. Information in oral form, sign language, etc. do not involve any document. However, documents are involved in hand-written form printed from coded form, simplified form, disguised form, etc. of information. Hence, form of documents are different from the form of information.
- 5) Some information needs are to be kept secret. It is as true in the case of a family as in the case of a nation. Information on defence of a country is usually kept secret. Only a few can have this information. When the secret information is to be passed on to somebody it is passed either in coded form or in disguised form. Even if this information accidentally goes to a third hand, he will not be able to make out what it is, and thus the interest of the country, organisation or a family will be protected.
- 6) Information technology has impacted information generation in a number of ways. Previously, a budding author had to struggle a lot to get his writing published. Now, he can straightway place his writing in Internet to attract the attention of others without wasting any time. The process will generate information and to a certain prevent information loss.

The IT era has given birth to a number of publications in electronic form. these publications among others comprise books, periodicals, and conference proceedings. *Cybermetrics*, an electronic-only journal as well as a virtual forum is devoted to the study of the quantitative analysis of scholarly and scientific communications in the Internet. The URL of this electronic journal available at Internet is: <http://www.cindoc.csic.es/cybermetrics/cybermetrics.html>. The number of electronic-only publication is continuously increasing opening up a new vista for the generation of information as well as access to information.

IT has also helped in the generation of quicker as well as better information. Before the advent of IT era, information search used to take a huge amount of time. Moreover, information that could be searched was not very up-to-date. For example, a journal from US or European countries used to take a month or more to reach India by sea mail. Hence, in a place like India we used to get at least one-month old information, sometimes more. Today, the time for locating and accessing information in many cases has been reduced considerably. This has helped in increasing the productivity of researchers and many other authors. Moreover, authors are now in a position to provide more updated information in their publications.

6.9 KEYWORDS

- Coded Form of Information** : In this form, information is recorded using some codes. For example in the classification of books we provide information in coded form.
- Condensed Form of Information** : Information in brief provided in the form of summary, abstract, etc.
- Consolidated Form of Information** : When scattered information on a topic is gleaned first and then put together in a logical order. This gives rise to consolidated form of information.
- Digitised Form of Information** : Presentation of information using the digits 0 and 1 gives rise to this form of information.
- Disguised Form of Information** : Information is communicated using common language. The hidden meaning becomes apparent only to the person for whom the information is meant.
- Handwritten Form of Information** : Information written on a medium like paper. Before the advent of paper, people used to write on papyrus, parchment, vellum, palm leaves, bhurjapatras, and so on.
- Oral Form of Information** : Information communicated orally.

Pictorial Form of Information	: Information communicated through pictures instead of alphabets and symbols. Examples: maps, atlases, photographs, portraits, etc.
Printed Form of Information	: Information printed on a medium like paper.
Secondary Form of Information	: Information generated from primary information e.g. an abstract.
Simplified Form of Information	: Information presented replacing the jargons, difficult words and long complex sentences with simple words and sentences to make it understandable to children and layman.
Tertiary Form of Information	: Information generated basing secondary information, sometimes primary information as well.
Translated Form of Information	: Original information is presented in a different language called the target language retaining the original meaning.

6.10 REFERENCES

Abrahams M. (2004) Weird Science Discoveries that will make you laugh, then make you Think. *Readers Digest* (June). P129.

Britannica Book of the Year 1988: London: Encyclopaedia Britannica. 136.

Cane P. *Giants of Science*. Pyramid ed. New York: Pyramid Books, 1961::276.

Odhams Colour Library of Knowledge: Language and Communication. 1968. P54

Odhams Colour Library of Knowledge: Language and Communication. London: Odham Books, 1968:: 17.

Odhams Colour Library of Knowledge: Religion and Philosophy. P83.

Pearsall, Judy. (ed.). (1999). *The Concise Oxford Dictionary*. Indian ed. New Delhi: Oxford University Press. P. 493.

Szymanski, R.A. [et al.]. (1994). *Computers and Information Systems*. Upper Saddle River (NJ): Prentice Hall.

The New International Webster's Comprehensive Dictionary of the English Language. (1996). Deluxe Encyclopaedic ed. Naples (Florida): Trident Press International.

UNIT 7 INFORMATION THEORY

Structure

- 7.0 Objectives
- 7.1 Introduction
- 7.2 Approaches to Information Theory
- 7.3 Information Basics
- 7.4 Information Measure
- 7.5 Information Entropy
- 7.6 Information Communication
 - 7.6.1 Efficient Communication
 - 7.6.2 Reliable Communication
- 7.7 Semantic Information Theory
- 7.8 Summary
- 7.9 Answers to Self-Check Exercises
- 7.10 Keywords
- 7.11 References and Further Reading

7.0 OBJECTIVES

After reading this Unit, you will be able to understand and appreciate:

- importance of information in the present-day society
- recognise the need for information theory
- different perspectives of information
- different approaches to information theory
- a scientific definition of information
- information as part of human thought process
- how to measure information
- what is information entropy
- how to calculate information entropy
- how to achieve efficiency in information transfer
- what is source coding and its purpose
- channel coding for reliable information transfer
- basics of semantic information theory
- relative information content measures
- different parameters that define the context

7.1 INTRODUCTION

Quest for knowledge has been a central theme of human evolution. Information is a key component in the growth of knowledge. From early civilisation of mankind, information has played a significant role in societal development and in improving living standards of human beings. Information is closely linked with the growth of economic, political,

health, cultural, educational and other sectors of a nation. It is now well recognised that effective use of information can turn hitherto non-productive resources into value added economic resources. Good examples of this are biogas and fuel pellets made out of human waste. Information contributes to political strength of countries so much, that we talk of *information-rich* nations being more powerful than *information-poor* nations. Successful application of science and technology (S&T) to social and economic developments depends on the effective use of information. Many countries have set up special purpose S&T information centres accessible to common man. India has over a dozen such centres. The increase in life expectancy and the growing population have resulted in large-scale governmental operations that call for extensive use of information. Thus, at present times, information has come to occupy a central role in national development and is reckoned as a driving force for all human activities. Consequently, the present society is termed as **information society**.

Information is useful only when communicated from originators to other potential users of information. Information communication is as ancient as information itself. In the early days, messengers used to carry information from one person to another. Birds were trained to carry messages. There were other techniques that used free space as the communication medium. Beating drums, waving flags and lighting fires are some of these ancient techniques. Inherent in these techniques is the concept of **coding** where a particular action conveyed a certain predetermined message. For example, waving a red flag may be a warning of an impending danger. The next major step in information communication is the postal network that is used quite extensively today. Modern telecommunications started with telegraphy in 1837. These systems transport information via electrical, optical or electromagnetic signals. Until about 1950s, the telecommunication systems were based on analog technology. Principles of digital communication were propounded in the second half of 1930s and first digital computers were built in mid 1940s. Since then, the digital technology has been advancing leaps and bounds both in the fields of communications and computers. **Binary coding** is used extensively in these systems.

Yet another aspect of information that needs attention is its enormous volume. In the early days of human civilisation, information generation was a slow process. The population was small and only a few individuals were involved in the process of creating new knowledge. The advent of industrial age accompanied by an increase in world population has brought about significant growth in information generation and dissemination. By the year 1800, the quantum of information generated was doubling every 50 years and by the year 1950, it was doubling every 10 years. The quantum of information generated by industry, governments and the academic world reached unmanageable proportions by the middle of 20th century that a need was felt to devise new ways for managing information. A search in this direction has given birth to the new **information technology (IT)**.

The amount of information generated by the academic community is gauged by the fact that there are around 150,000 journals and periodicals being published at present in the fields of science, engineering, technology, medicine, social sciences, arts and humanities. This means that on an average 15 million articles are written by the academic community every year. Industry is no different. It is said that the weight of

the drawings of a jet plane is greater than the weight of the jet plane itself. Remote sensing satellites gather terabytes (10^{12} bytes) of information everyday, which is equivalent to a million books of about 300 pages each. The banking and finance industry has a large volume of financial and personnel information stored in its vast data banks. The quantum of Government information is mind-boggling too. Land records, population records, voter lists, police records, licensing records, transaction records, accounting records, policies, rules, regulations, laws, judgements and other innumerable pieces of information are ever growing.

Thus, information has become the central theme of living these days. It is treated as a commodity and traded for a price. *Information economics* has emerged as a subject of recent interest. The world is witnessing a phenomenon of *information explosion*. Consequently, the present period of human civilisation is aptly called the **information age**. Historically, the information age is supposed to have set in since early 1970s and is expected to last for another century or two.

In the context of information society and information age, a number of questions related to information have arisen. What constitutes information? How can we transmit information reliably and efficiently using modern telecommunication systems? How can we store large volumes of information in a compact form? Is there a measure for information? Can we evaluate the information content by attaching value to information? Such questions have led to the development of **information theory** that deals with the following aspects:

- Concept of information
- Information measure
- Information content
- Information communication
- Information storage

This Unit is a study of the various aspects of information theory.

7.2 APPROACHES TO INFORMATION THEORY

Studies in information theory have been pursued using three different perspectives of information:

- Syntactic perspective
- Semantic perspective
- Contextual perspective

Studies using syntactic perspective concentrate on the **source characteristics** and its **symbol set** usage. These studies do not concern themselves with the semantic aspects of information. Their primary focus is on how to represent and communicate information effectively and reliably via the modern communication systems. They view information as something conveyed by **messages** put out by the source. The messages are constructed using the symbol set of the source. They measure the information content of messages by analysing the occurrences of the constituent symbols. Consider the following two sentences:

1. Dr. Jaideep Sharma is co-ordinating the preparation of this Unit.

2. The preparation of this Unit is being co-ordinated by Dr. Jaideep Sharma.

The two sentences are syntactically different but convey the same meaning. A syntactic analysis may yield different values for the information content of the two sentences. The values may, however, differ only marginally. A syntactic technique may code the two sentences in different ways to achieve efficiency in transmission. For example, let us consider some form of binary coding and a binary transmission channel. If it is known that the first syntactic form is used more often, then it can be coded as a smaller binary string and the second one as a larger string. Such coding would, on an average, result in less number of bits being transmitted on the channel and in the consequent efficient utilisation of the channel.

Semantic perspective is concerned with complete and precise meaning of the messages as well as relative information content between messages. Contextual perspective of information derives the meaning of messages from not only what is contained in the message but also from the context in which the message occurs. Contextual perspective is also known as **pragmatic perspective** of information. Consider the following three messages pertaining to the same situation:

1. There is a traffic jam on the National Highway No. 3 (NH 3) between New Delhi and Agra in India. Time: 11:30 Hours. Date: 2 April 2005.
2. There is traffic jam on this highway.
3. There is traffic jam on this highway between Faridabad and Palwal.

The first message is complete and precise and makes sense to anyone in the world. The second is relevant only to those who are on NH 3 at the time of receiving the message. This message is context dependent. The context is place and time. The message does give meaningful information to those at a particular place (NH 3) and at a particular time (11:30 Hours on 2 April 2005). The message if seen in isolation is incomplete and useless. But it is pragmatic information that has value in a given context. The third message is also context dependent. But, the information content of this message is higher than that of the second one. (Note: Faridabad and Palwal are two towns on NH 3). One may speak of relative measure of information content between messages 2 & 3.

In the case of semantic and context oriented approaches, the recipient is an important component of the study. If the second message above were sent to someone on NH 8, it would give wrong information and may be said to have negative value for a person on NH 8. The third message is more valuable to a traveller near Faridabad or Palwal than to one near Agra. (Note: The two towns are far away from Agra on NH 3). Similarly, the first message is more valuable to a non-resident Indian (NRI) in USA whose family is travelling on NH 3 at that time than for a NRI who has none on the highway. Thus, the value of information is closely linked to the recipient. The recipient being a human being, the studies involve the disciplines of psychology, philosophy, behavioural science, biology and logic. Often, recipients are considered as part of the context itself and all recipient-centred studies are placed under pragmatic information studies. In this course material, we treat user or recipient as part of the context itself.

Studies using semantic and pragmatic perspectives have been carried out predominantly by British scientists. Some of the main

contributors include Ackoff, MacKay, Carnap, Bar-Hillel and Hintikka. Due to British-dominant studies, the semantic and pragmatic approaches are often termed the British tradition of information theory.

The genesis of the syntactic approach may be traced to a seminal paper titled '*A Mathematical Theory of Communication*' by U.S. scientist Claude E. Shannon published in 1948. Shannon's primary interest was in digital communication systems. He addressed two principal questions in his seminal paper:

1. How to convert analog signals to digital ones optimally without losing the original information content?
2. How to realise efficient and error-free transmission of digital signals over transmission channels that are affected by noise?

Semantics was not in Shannon's mind at all. Shannon proposed the now famous **sampling theorem** as the solution for the first problem. We study more about the sampling theorem and analog to digital conversion in Unit 8. For the second problem, Shannon suggested an **information measure** and **coding** at two levels as the solution. Shannon proposes **source coding** for efficient representation of information from a given source and **channel coding** for error-free transmission. These aspects are discussed in this Unit. Although Shannon is considered the founder of information theory, his 1948 paper was based on two important earlier theoretical contributions by H. Nyquist and R.V.L. Hartley. Nyquist, in 1924, arrived at the minimal sampling rate required to convert analog signal to digital without loss of information. Shannon built his sampling theorem on this result. Hartley, in 1928, proposed a measure for syntactic information for the first time. Shannon added probability concepts and generalised Hartley's result to arrive at his information measure. Subsequent to Shannon's seminal paper, syntactic perspective has been studied predominantly by U.S. scientists. As a result, syntactic approach is often considered as American tradition of information theory. One of Shannon's close associates, W. Weaver presents a holistic view by adding semantic perspective to Shannon's study.

In summary, we may place approaches to information theory under four different categories:

1. Semantic-centred approach
2. Context-dependent approach
3. Recipient-centred approach
4. Semantic-independent approach

The first three approaches are in some way connected with the semantics of information. Hence, some authors place all the three under one heading *semantic information theory*. The fourth approach completely ignores semantic aspects and concentrates on the information communication and storage aspects. To contrast from the other three approaches, the fourth approach is usually called *syntactic information theory*.

Semantic-centred approach is **context independent**. The main emphasis of context-independent approach is on the relative information content of different messages. Relative measures are neither concerned with the source nor with the recipient. They just treat messages as they are, without regard to the context that includes the source and the recipient. They compare different messages and rate their information content in a relative or normalised way.

Context-dependent approach is also known as **pragmatic information** approach. Pragmatic information studies take into account the contextual aspects like place and time before evaluating the information content of the messages. Recipient aspects may or may not be considered in pragmatic information studies.

User-centred approach may be context-dependent or context-independent. User perspective of information is based on his needs. Different users may assign different information values to the same message depending upon the relevance of the message to them and the value assigned may vary from one context to another.

The main focus of semantic-independent approach is efficient and reliable information communication and storage of information. This approach is totally unconcerned with semantic aspects. In a broad sense, it may be said that this approach looks at syntactic aspects of information. In a strict sense, this is not true. It is more concerned with sender's choice of a message and the symbol set that makes up the messages. Hence, it is also called source-centred or sender-centred approach. A symbol set may be as basic as the letters in an alphabet or as complex as sentences that are used to form messages. The term *syntactic studies* is used widely in the literature to denote studies that ignore semantic aspects of information. At present, information theory is dominated by syntactic studies. One of the main reasons for this is the mathematical approach used in syntactic studies. In particular, statistics and probability theory play an important role in these studies. As a result, syntactic studies are sometimes called as *statistical information theory*. On the other hand, the semantic-centred approach is based on abstract disciplines and hence, the studies are more abstract and somewhat subjective.

This Unit deals with both syntactic and semantic information theories. We discuss syntactic information theory in Sections 7.4 through 7.6 and semantic information theory in Section 7.7. In Section 7.3, we discuss certain basic aspects that are applicable to both syntactic and semantic information theories.

Self Check Exercise

1. Present in a tabular form the different approaches of information theory along with the corresponding information perspective(s) that is/are used for studies in each approach.

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

-----	-----
-----	-----
-----	-----
-----	-----
-----	-----
-----	-----
-----	-----
-----	-----

7.3 INFORMATION BASICS

In Section 7.1, we explained the importance of information and the central role it plays in the present day society. But, what precisely is information? In this section, we give a definition of information and place it in proper perspective in the context of human thought process.

Information has been defined variedly by different authors. Not all definitions have received wide acceptance. In this Unit, we present and use a definition that is scientific in nature:

Information is defined as a description of the state of an object.

There are three keywords in this definition: *object*, *state* and *description*. The object may be animate and inanimate. The term *object* here refers to the entire spectrum of things in the creation: the tiniest particles like electrons or molecules, living beings like human beings and animals, inanimate objects like mountains and weather, and large systems like planets, solar system and the milky way. The second keyword in the definition is the *state*. Every object is in one of finite or infinite number of states associated with it. For example, an electric bulb may be *on* or *off*. It may be mounted in an electric circuit or lying on a shelf. It may be in working or fused condition. For a very large number of objects that we encounter in practice, the number of states is finite. In this course material, without loss of generality, we deal with only finite number of states. The third keyword is the *description*. Information describes the state of an object in some way. The description of the state may be verbal or non-verbal, i.e. written, pictorial etc. Interestingly, the above definition of information is equally applicable to both syntactic and semantic studies.

Information is a part of human thought process at a certain level of abstraction. The human thought process is abstracted usually at four levels:

- Data
- Information
- Knowledge
- Wisdom

These levels are depicted in Fig 7.1 in rectangular boxes with the rounded boxes indicating the inputs or the processes that lead to the next higher level of abstraction. There are no clear-cut boundaries among these levels. What is considered as a piece of information in one context may be treated as a piece of knowledge in another context. As a result, these levels are considered a part of a continuum with overlapping areas. This is called *knowledge continuum*. Notwithstanding this perception, the four-level abstraction model of human thought process is widely accepted. At the lowest level of abstraction, we have the raw data that is a collection of facts as observable from nature, or as obtained from experimental outcomes or values of certain quantities that are measured. Examples of raw data include population census, temperature values and the outcome of games played.

When the raw data is processed or value is added to it, data becomes information. The first level of processing is usually statistical in nature involving computations like averages, maxima and minima. Value addition may be qualitative or quantitative in nature. Examples of

information include maximum temperature value in a day, percentage of errors in a text and the number of persons in different age groups. In general, information creates awareness in human beings. When

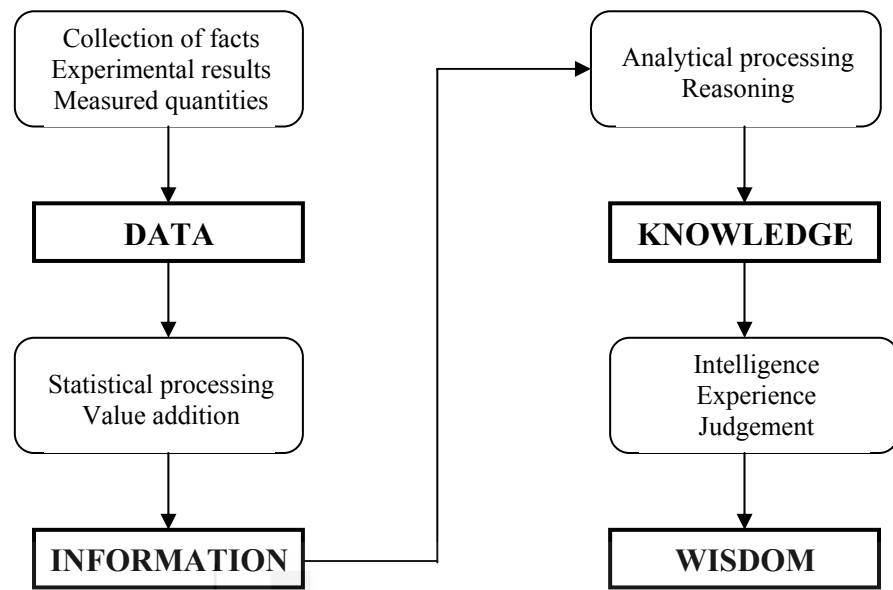


Fig. 7.1 Abstraction levels in human thought process.

information is further processed and reasoning is applied to it, information becomes knowledge. The processing at this stage is usually analytical in nature involving reasoning, inference, extrapolation and other complex mathematical operations. A statement in the knowledge domain may be something like '*Poverty level is decreasing all over the world*'. In general, knowledge implies understanding of a subject. Hence it may be said that a human being moves from an *awareness domain* to an *understanding domain* when he processes information to acquire knowledge. At the highest level of abstraction in human thought process is wisdom. When knowledge is interpreted intelligently taking into account past experience and sagacious judgements are made, wisdom is said to be in display. It may be said that experienced experts in particular fields tend to display wisdom in their respective domains of expertise.

Computers have traditionally acted as tools in performing functions related to human thought process. They used to be termed as *data processing machines* for about 40 years since their inception. Today, they are termed as *information processing machines* and are serving the information society. The information processing machines are more powerful than data processing machines. In particular, software support is superior. It is forecast by a number of world thinkers that today's information society would evolve towards a *knowledge society* in the future. At that time, the computers with additional capabilities may be termed as *knowledge processing machines*. These machines may support knowledge bases and intelligent processing.

To illustrate what document material constitutes different levels of abstraction of human thought process in different contexts, we present in Table 7.1 some details pertaining to the fields of education and industry.

Table 7.1 Examples of Data, Information and Knowledge

Abstraction Level	Education	Industry
Data	Syllabus, References, Indexes	Market, Sales and Financial data

up his friend and says, "I had breakfast today". The information content of this message is rather low because the information conveyed by the message is the most expected status of the object in question. If, on the other hand, the person tells his friend that he has not had his breakfast that day, then the information content is high because the message conveys something that is less probable. Thus, intuitively, we may relate the information content of a message to the probability of occurrence of the state conveyed by the message. If the probability is high, the information content is low and vice versa. Hence, we may propose that the measure of information is inversely proportional to the probability of occurrence of the message, i.e.

$$I \propto (1/p) \quad (7.1)$$

Where

I = Information content of the message

p = Probability of occurrence of the message

Let us suppose that the constant of proportionality is unity in Eq. 7.1. Then, if $p = 0$, the value of I is infinity and if $p = 1$, the value of I is unity. If the probability $p = 1$, as in the case of the message "There is day light here" in the example considered above, we would like the value of I to be zero rather than unity. To achieve this, we adopt a logarithmic expression that also serves as the constant of proportionality as in Equation 7.2 below:

$$I = \log_2 (1/p) \quad (7.2)$$

The value of $\log 1$ is zero. Therefore, when $p = 1$, the value of $I = 0$ and when $p = 0$, $I = \infty$. Shannon chose the base of the logarithm as 2 as he was interested in binary digital systems. Sometimes the definition uses the base 10 or natural logarithm with base e . The quantity I is a dimensionless number, but by convention a unit is assigned to it. When the base is 2, the unit of information is called *bit*, when the base is e , it is called *nat* and when the base is 10, it is called *decit*. The unit *decit* is also known as *Hartley* named after R.V.L. Hartley who first proposed a measure of information. The use of base 2 is especially convenient when binary signals are used to convey a message. Equation 7.2 can be rewritten by using the laws of logarithms. You may recall that

$$\text{Log } (A/B) = \log A - \log B \quad (7.3)$$

$$\text{Similarly, } \log_2 (1/p) = \log_2 1 - \log_2 p = -\log_2 p, \text{ as } \log 1 = 0$$

$$\text{Therefore, } I = -\log_2 p \text{ bits} \quad (7.4)$$

Equation 7.4 is the famous Shannon measure of information content of a message. As an example of application of the information measure that we have developed above, let us consider the following. If the probability of occurrence of a message is $1/4$, then from Equation 7.2 we obtain its information content of the message as $I = \log_2 (1/(1/4)) = \log_2 4 = 2 \text{ bits}$.

Shannon's approach is source-centric. Shannon was concerned with the probability with which a source may put out messages and proposed Eq. 7.4 as a measure of information. But, interestingly, the Equations 7.2 or 7.4 can be applied to recipient-centred approach as well. If we replace the probability p with the recipient's expectation of information, then the same expressions in Eq. 7.2 and 7.4 represent the informational value of a message from the recipient's point of view. In fact, we have used a recipient-centred reasoning in arriving at Eq. 7.2 without explicitly stating so. The fact that Shannon's information measure is applicable to both

source-centric and recipient-centric studies makes it fundamental to information theory.

Equation 7.2 or 7.4 is a measure for the information content of a single message. Now let us consider a set of messages put out by a source that describes the state of an object like, say, weather. As stated earlier, the number of states associated with an object may be finite or infinite. Accordingly, the set of messages put out by the source is finite or infinite. Without loss of generality, we consider a finite set of N messages for our further study. Let the N messages $m_1, m_2, \dots, m_K \dots m_N$ originate from the source with the probabilities $p_1, p_2, \dots, p_K \dots p_N$ respectively. From the theory of probability, the probabilities of all possible outcomes must sum up to one. We then have

$$p_1 + p_2 + \dots + p_N = 1 \quad (7.5)$$

The information content of the k^{th} message m_K is given by

$$I_K = \log_2 (1/p_K) \quad (7.6)$$

If the messages are statistically independent, the amount of information conveyed by two or more messages is the sum of information content of each message. Thus, for two statistically independent messages m_j and m_K , we have

$$I_{j,K} = I_j + I_K = \log_2 (1/p_j) + \log_2 (1/p_K) \quad (7.7)$$

The total information content of the source having a repository of N messages is given by

$$\begin{aligned} I_{\text{source}} &= \log_2 (1/p_1) + \log_2 (1/p_2) + \dots + \log_2 (1/p_N) \\ &= \sum \log_2 (1/p_K) \text{ for } 1 \leq K \leq N \end{aligned} \quad (7.8)$$

If all the N messages are equally likely, then $p_1 = p_2 = p_j = 1/N$, and the information content of any message I_j is given by

$$I_j = \log_2 (1/(1/N)) = \log_2 (N) \quad (7.9)$$

Where, $1 \leq j \leq N$. In this case, the information content of all the messages is the same. Equation 7.9 is the one developed by Hartley and is a special case of Shannon's expression in Eq. 7.2, where $p = 1/N$.

There are other cases that are of interest. If $N = 1$, it means that there is only a single possible message with probability $p = 1$. In this case, no useful information is conveyed by the message and the information content of this message is zero. At the other extreme as $p_j \rightarrow 0$ (p_j tends to 0), $I_j \rightarrow \infty$ (I_j tends to ∞). If $N = 2$, there are only two messages. If the probability of occurrence of one message is p , then by Eq. 7.5, the probability of the other message is $1 - p$. Applying Eq. 7.4, the information content of individual messages are $-\log_2 p$ and $-\log_2 (1 - p)$ bits respectively.

Self Check Exercise

4. A weather bureau puts out one of four different messages m_1, m_2, m_3, m_4 predicting the weather for the next day. The probabilities of the messages m_1, m_2 , and m_3 are $1/2, 1/4$, and $1/8$ respectively. What is the probability of the message m_4 ? Calculate the information content of each message and the total information content of the source.
5. How many *bits* of information constitute one *nat* and how many one *Hartley*?

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

7.5 INFORMATION ENTROPY

As mentioned earlier, Shannon's approach is source-centric. It is based on a simplistic view of the generally observed human communication process. In everyday life, it is the communicator (source) who decides what is to be said from all that could be said. In other words, source is the repository of messages and it is the source that decides what message is to be put out. Shannon's model assumes that the source acts independently without regard to prevailing situation or the condition and nature of the recipient. This is not true in human communication process. A human communicator, subconsciously or consciously takes into account the current situation and the capability and interest of the recipient before saying something.

In section 7.4, we discussed the information content of individual messages. In this section, based on Shannon's source-centric model, we discuss the average information content of a message in a sequence of messages put out by a source over a long period of time. The average information content per message is referred to as **source entropy** or **information entropy** and is represented by the symbol H . As before, we consider a source with a repository of a finite set of N messages $m_1, m_2, \dots, m_N$ with probabilities $p_1, p_2, \dots, p_N$. From this set of N messages, suppose that L messages have been gathered over a long period of time. If L is much larger than N , then the sequence of L messages will contain different messages from the set of N messages in the same proportion as their probability of occurrence, i.e. m_1 occurs $p_1 \times L$ times, m_2 occurs $p_2 \times L$ times and so on. The information content of individual messages is given by Eq. 7.2 or 7.4. Applying Eq. 7.2, the total information content of a sequence of L messages is

$$I_{\text{total}} = p_1 L \log_2 (1/p_1) + p_2 L \log_2 (1/p_2) + \dots + p_N L \log_2 (1/p_N) \quad (7.10)$$

The average information content per message is given by

$$H = (I_{\text{total}}/L) = p_1 \log_2 (1/p_1) + p_2 \log_2 (1/p_2) + \dots + p_N \log_2 (1/p_N) \quad (7.11)$$

$$= \sum p_k \log_2 (1/p_k) \text{ for } 1 \leq k \leq N \quad (7.12)$$

Readers are advised to compare and appreciate the differences amongst Equations 7.8, 7.10 and 7.12.

Applying Eq. 7.4, Eq. 7.12 may be rewritten as

$$H = -\sum p_k \log_2 (p_k) \text{ for } 1 \leq k \leq N \quad (7.13)$$

Equation 7.13 is the famous Shannon's equation for information entropy which is also called as source entropy. **Entropy is the average information content per message of a source.**

We have seen that in the case of individual messages, as $p \rightarrow 0$, $I \rightarrow \infty$ and, as $p \rightarrow 1$, $I \rightarrow 0$. But, in the case of information entropy, the contribution of an individual message to the average information content tends to zero, if its probability of occurrence is very low or very high. In other words, the quantity $p \log (1/p)$ tends to zero when p tends to 0 or 1. Thus, the average information content of both an extremely likely and extremely unlikely message is zero. This is a very interesting result. Extremely likely message has a very low information content and hence its contribution to entropy is very low. Extremely unlikely message hardly occurs to be able to contribute significantly to the entropy. Figure 7.2 shows the variation in the contribution of a message to the information

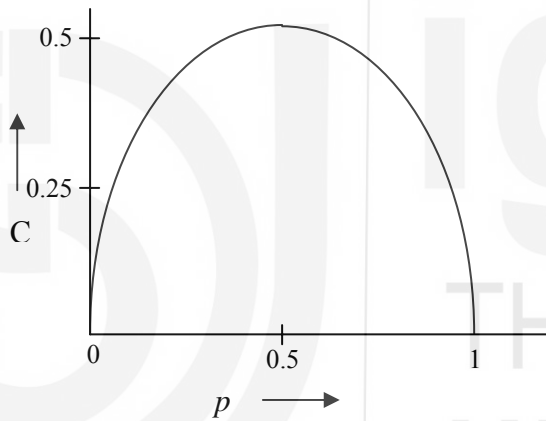


Fig. 7.2 Contribution of a message to the entropy

entropy as its probability of occurrence varies. It may be noted that $C = 0$ at $p = 0$ and at $p = 1$. The maximum contribution occurs when $p = 0.5$. It is important to recognise the difference between the maximum contribution of a message to the entropy and the maximum value of entropy. The contribution of a message is maximised when its probability of occurrence is 0.5. This does not mean that the entropy is maximised. Entropy depends on the entire set of messages and hence on the probability of occurrence of all the messages. In fact, entropy is maximised when all messages occur with equal probability. We illustrate this fact in the following.

Consider the case of only two messages in the message set. With $N = 2$, let $p_1 = p$, then $p_2 = (1 - p)$. The information entropy works out to be

$$H = p \log_2 (1/p) + (1 - p) \log_2 (1/(1 - p)) \quad (7.14)$$

If $p = 0.5$, then we have

$$H = 0.5 \log_2 2 + 0.5 \log_2 2 = \log_2 2 = 1 \text{ bit / message.}$$

For all other values of p , H has a value that is less than 1 bit/message. Similar results apply for cases where $N > 2$. The case of $N = 2$ has special significance in the context of binary digital transmission systems. If there are only two messages, they can be represented by '1' and '0' in the binary

7.6 INFORMATION COMMUNICATION

The main focus of syntactic studies as founded by Shannon is information communication. In information communication, we are concerned with two main goals:

- Information communication is not concerned with the semantics of information at all. In fact, Shannon states so explicitly in his seminal paper. The main concern is how to communicate a given piece of information efficiently in an error-free manner. Shannon developed a theoretical framework to quantify the problem of communicating with high efficiency and accuracy. In Section 7.4, we defined information and developed a measure for information. In Section 7.5, we introduced the concept of information entropy. Information measure and entropy are fundamental concepts developed by Shannon as part of the theoretical framework for information communication. In this section, we apply those concepts in the context of efficient and reliable communication of information. We study the efficiency and reliability aspects one by one. First, we take up efficiency aspect.

7.6.1 Efficient Communication

As we know, Shannon's approach is source-centric. Shannon proposed **source coding** as a method of achieving efficiency in communication. We use the model shown in Fig. 7.3(a) for understanding the concepts involved in achieving efficiency. In this model, the communication channel is assumed to be error free and reliable. An ideal channel can

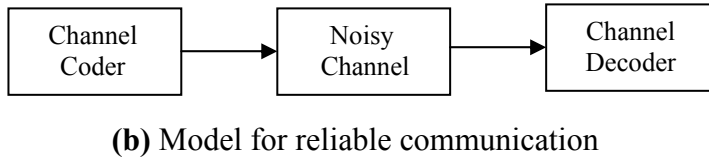
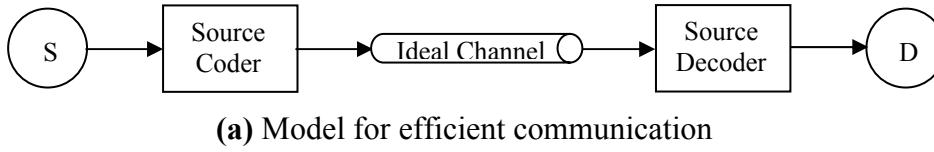


Fig. 7.3 Communication system models

transport information with perfect accuracy. It is not affected by noise and it does not cause impairment to the signal in any way. We now extend the idea of source entropy to what is called **source information rate** (SIR). We know that entropy is the average information content of a message. If a source, on an average, delivers s messages per second, then SIR is given by

$$\text{SIR} = s \times H \text{ information bits/second} \quad (7.15)$$

SIR is the average information rate from a source. SIR is not a bit-rate in the conventional sense. It does not represent binary digits being transmitted through a channel. H represents the average information content per message and sH represents the information rate. To make a distinction between bits of information and binary digits we use the term *information bits* when we talk of information measure. We now turn our attention to a quantity known as **channel capacity** C . The capacity of a channel is defined as the maximum rate at which the channel can transfer information with perfect accuracy. This definition implies that the channel is ideal. We can now relate C and SIR as

$$C \geq \text{SIR or } \text{SIR} \leq C \quad (7.16)$$

Equation 7.16 is the famous Shannon's channel capacity equation and is based on simple common sense reasoning. If a certain quantity of water is to be delivered within a specified time, then we need a pipe of appropriate size. Similarly, the channel must have a minimum capacity that is equal to or greater than the rate at which information is delivered by the source. It is important to recognise that Eq. 7.16 is at an abstract level. It does not concern itself with practical transmission systems. For example, the determination of the channel capacity of a practical channel is still an unresolved problem. Shannon and many others have derived expressions for the capacities of practical channels under different simplifying assumptions of noise pattern etc. Our next step is to see some practical aspects of source coding. We do this by taking an example in the following.

Let us consider a source with four messages with the corresponding probabilities as $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ and $\frac{1}{8}$. The entropy for the source works out to be 1.75 *information bits*/message. Readers are advised to calculate the entropy using Eq. 7.12. Let the source put out messages at the rate of 8

messages per second. Then, the SIR works out to be $8 \times 1.75 = 14$ *information bits* per second.

Let us now assume that the messages are to be communicated using a binary digital transmission system. As you may know, a binary communication system uses '1' and '0' to represent and transmit information. [Binary system fundamentals are dealt with in Unit 8]. Source coder performs the function of representing the messages in binary form. One way to represent four messages in binary system is to use two bits that give us four unique patterns as 00, 01, 10 and 11. Each pattern represents a message. Each message is represented by a code of uniform length of 2 bits. When 8 messages per second are put out by the source, $8 \times 2 = 16$ bits per second (bps) are transmitted. The bit rate out of the source coder (SCR) is 16 bps whereas the SIR is 14 *information bits/sec*.

In the above example, two binary bits carry 1.75 bits of information. Intuitively, one may feel that for maximum efficiency, the two rates SIR & SCR must be the same. As discussed in Section 7.5, each binary bit is capable of carrying a maximum of 1 bit of information provided the binary '1' and '0' occur with equal probability. If the source coding is done in such a way that each binary bit carries maximum information, then the maximum efficiency will be achieved. In other words, the entropy and the average length of the coded patterns must be equal. Two source coding techniques have been proposed to arrive at average code word length that is close to entropy value: **Shannon-Fano coding** and **Huffman coding**. These schemes are named after their inventors. Both schemes are based on variable length coding with shorter code length for messages with higher probability and longer ones for messages with lower probability. A detailed discussion of these techniques is beyond the scope of this course. In general, Huffman coding has been shown to be superior to Shannon-Fano coding.

7.6.2 Reliable Communication

We now turn our attention to reliable communication. In practice, we do not encounter ideal channels and find that the channels are noisy. Hence, our problem is that of ensuring error-free transmission of information using noisy channels. Shannon proposed channel coding as the solution for this problem. We use the model shown in Fig. 7.3(b) to study this problem. The output of the source coder is fed to the channel coder that codes the input bit string for error-free transmission. At the receiving end, the channel decoder recovers the source-coded information and feeds the same to the source decoder.

Shannon stated and proved that *provided Eq. 7.16 is satisfied there exists a channel coding scheme that would ensure arbitrarily small probability of transmission error*. Shannon, however, did not work out as to what that coding scheme was. It was left to his successors to work on coding schemes. Since then, considerable work has been done in the area of coding. Today, we have a fairly matured branch of study known as *Coding Theory*. Work on coding theory has resulted in the evolution of two broad classes of codes:

- Error-detecting codes
- Error-correcting codes

In a binary digital system, error means that when a binary '1' is transmitted, it is interpreted as '0' at the receiving end and a '0' as '1'. The error performance of the system is usually indicated by a parameter called

bit error rate (BER). BER is specified as 1-bit error in a block of n bits, e.g. 1 in 10^4 . Both error detection and error correction are done at the receiving end. Error correction involves the detection of error first and the correcting the error. Error correction implies automatic recovery from error. When error detection codes are used, error recovery is done by retransmission. If an error is detected, the receiver requests the transmitter to retransmit the information. Most widely used error detection codes include *parity check*, *checksum*, and *cyclic redundancy code* (CRC). *Block parity* and *Hamming code* are popular error correcting codes. All these coding schemes take a block of information bits, add some error check bits according to a mathematical formulation and transmit both the information and error check bits. We call the information bits and the error check bits together as *transmission block*. At the receiving end, the same or an inverse mathematical formulation is used to determine whether the information has been received correctly. It may be noted that an error may occur in any of the bits of the transmission block including the error check bits. Of the above mentioned coding schemes, we discuss the parity check and block parity scheme in the following. Other coding schemes are beyond the scope of discussions for this course.

In parity check, one check bit is added to the chosen block of information bits. The mathematical operation is simple and involves counting the number of binary '1s'. The error check bit is set to a binary '1' or '0' such that the total number of '1s' in the transmission block is either even or odd. Whether the number should be even or odd is predetermined and is known both to the sender and the receiver. The sender sets the desired parity (odd or even) and the receiver checks the received parity. If the parity is incorrect, then the transmission is in error. Otherwise, the transmission is assumed to be without error. Parity check failure occurs whenever one or odd number of bits are in error. If two or any other even number of bits go wrong, the parity condition would be satisfied and the errors would go undetected.

In block parity, the information block is arranged in the form of a matrix and a parity bit is assigned to each row and column of the matrix. An illustrative example is shown in Fig. 7.4 with 30 information bits arranged in the form of 5×6 matrix. Both odd and even parities are illustrated. The sixth row and the seventh column are the parity bits. The parity bits in the seventh column are called longitudinal parity check (LPC) bits and the ones in the sixth row as vertical parity check (VPC) bits. These bits are set according to the parity scheme chosen for the row and the column information bits respectively. The diagonal bit of the sixth row and the seventh column (shown in $\hat{}$ box) may be set to '1' (or '0') as shown in Fig. 7.4. Alternatively, it may be set as the parity bit for the VPC row or the LPC column, or as the parity bit for all the other bits in the transmission block.

The error recovery in this case of block parity scheme proceeds as follows. At the receiving end, both LPC and VPC bits are checked first. If an error is noticed in both of them, then the information bit at the

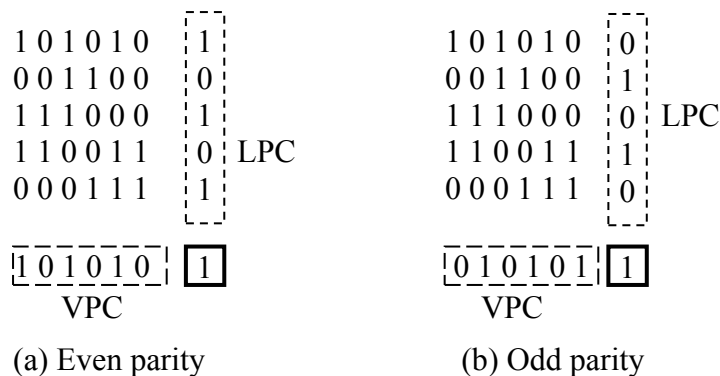


Fig. 7.4 Block parity

intersection of the failing LPC row and the failing VPC column is error. The bit is corrected by inverting it. If an error is noticed only in LPC or VPC, then it implies that the corresponding parity bit is error. No correction is required in this case as all the information bits are in tact. Block parity scheme can detect and correct all single bit errors. It is also capable of detecting many two bit and multiple bit errors. In general, multiple errors result in parity check failures in multiple columns and/or rows.

Since the parity schemes are ideally suited for recovering from single bit failures, the information block size must be chosen such that the transmission block does not experience more than one bit error. This is done by knowing the BER of the channel. For example, for the BER value of 1 in 10^4 , the transmission block size should be less than 10,000 bits.

Self Check Exercise

8. Draw a communication model that depicts both efficiency and reliability aspects of information communication.
9. Given that the BER of a system is 1 in 10^5 , calculate the probability of one bit being in error and the probability of one bit being error free.
10. Work out the parity bit value for even and odd parity schemes for the information block 1 1 0 0 0 1 0 1 1 0. What is the size of the transmission block in this case?

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

7.7 SEMANTIC INFORMATION THEORY

As mentioned in Section 7.2, studies in semantic information theory have proceeded in three different directions:

- Semantic-centred studies
- Context-dependent studies
- User-centred studies

Semantic-centred studies are context independent. They consider both the sender and the receiver as part of the context and hence ignore their role in arriving at information values. They are concerned with relative **information content** of messages. The content measure is based on substantial information that a message carries instead of the surprise element or the unexpectedness of the message as in Shannon's measure. In measuring content, a message is considered to be composed of a number of *atomic statements* or *constituents* from a set. Atomic statements must be such that they convey some information. Statements that express truths or are in the nature of axioms are considered not to have informational value. For example, the information content of the expression $11 \times 13 = 143$ or the statement 'All the sides of an equilateral triangle are equal' is treated as zero.

Many ways or approaches for assessing the information content of messages have been discussed in the literature. None of these ways appears to have an indisputable mathematical basis, with the result that none has come to be accepted widely. We illustrate the basic approach in semantic-centred studies in the following. Consider the following atomic statements:

1. It is raining.
2. The wind is blowing.
3. It is humid.

Let each of these atomic statements have a complement as follows:

4. It is not raining.
5. The wind is not blowing.
6. It is not humid.

Now, a message consists of one or more atomic statements such as:

7. It is not raining.
8. It is raining but the wind is not blowing.
9. It is raining, the wind is blowing but it is not humid.

Clearly, the substantial information content of the statement 9 is the highest among the statements 7, 8 & 9. The number of atomic constituents in a message is called its *width*. The larger the width, the higher is the information content. In this approach, all known conditions are *explicitly* mentioned. Missing conditions constitute lack of information.

Another approach to assessing information content is to accept *implicit* representation of information in messages. Consider only the three atomic statements 1, 2 & 3 without their corresponding negation statements. A message may now contain one, two or three atomic statements. The absence of an atomic statement implies its negation. For example, the message '*It is raining*' implies that the wind is not blowing and it is not humid.

One may consider logical connectives and the probability of occurrence of the constituents to arrive at a more sophisticated measure of the content. The logical connective AND provides higher information

content than the OR connective. For example, the information content of the message 'It is raining AND the wind is blowing' is more than the message 'It is raining OR the wind is blowing'. This approach combined with probabilities of occurrence yields a result similar to that of Shannon. If we assume that the atomic statements are statistically independent, then the AND connective implies joint probability of the atomic statements in the message. The joint probability is a multiplication of the individual probabilities and has a value less than any of the individual probabilities. The larger the width of the message, the smaller is the joint probability and higher is the information content.

Let us now turn our attention to contextual information studies. The theory here is based on certain fundamental aspects like laws of nature, context or situation, events and perception ability of individuals. There are certain postulates that govern this theory. They are:

1. Laws of nature, like gravitation, exist always. They existed in the past, they exist in the present and they would exist in the future. Laws are not created but are discovered from time to time. Some laws are discovered and some remain undiscovered. What is discovered may be correct or incorrect. A discovery may be proved wrong in future. The number of laws is constant all the time.
2. Time, place (space) and environment constitute the context or the situation. The set of all objects in the universe including human beings form part of the environment.
3. Information is the measurement and description of the conditions of the context. Information exists always. It is not created.
4. Governing laws and the context determine the future outcomes called *events*.
5. Events can be predicted correctly provided the understanding of the laws is precise and the conditions of the context are measured accurately, i.e. the information gathered is accurate.
6. Perception ability of human beings varies from individual to individual. Every individual is unique. The perception ability of a human being determines the level of his/her understanding of the laws and the conditions of the context and hence his/her ability to predict the future.

The above postulates indicate clearly the philosophical nature of context-centred studies. Human beings play a central role bringing in subjects like psychology, biology, behavioural science and even astrology in an indirect way into the studies. Attempts have been made to characterise individuals in order to determine their perception abilities. Characterisation of individuals considers factors like date, time and place of birth, race and religion etc. apart from factors like educational status and economic status etc.

In context-centred studies, the human beings are characterised to determine their perception abilities. In recipient-centred studies the concentration is on how precisely the recipient gets the meaning that the sender intends to convey.

Self Check Exercise

11. Given that the probabilities of occurrence of the atomic statements 1, 2 & 3 above are $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{4}$ respectively, apply Shannon's measure and determine the information content of the following messages:
 - a) It is raining AND it is humid

b) The wind is blowing AND it is NOT humid.

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

7.8 SUMMARY

This Unit is a study of information theory. The Unit starts emphasising the need for a theory of information by bringing out the importance of information in the present-day society. Different theoretical perspectives of information and different approaches to information theory are then brought out. Syntactic information theory, also known as *statistical information theory* or *mathematical theory of communication*, is then taken up for a detailed study. First, information is defined and placed in perspective in the context of human thought process. Then, an information measure based on an element of unexpectedness is evolved. Shannon's equation for information measure is derived. The concept of information entropy and the method of calculating the same are then presented. The issue of efficient and reliable information communication is then addressed. Source coding and channel coding as techniques for improving communication efficiency and reliability respectively are discussed. Finally, the basic aspects of semantic information theory are presented. Semantic-centric, context-dependent and recipient-centred approaches are discussed. Different approaches to measuring information content are introduced. While Shannon's approach does not concern itself with the semantics of information in messages, semantic information theory places emphasis on substantial information content in messages.

7.9 ANSWERS TO SELF-CHECK EXERCISES

1. There are four approaches to information theory. Some of the approaches are known by more than one name. There are three information perspectives. One or more perspectives are used in each of these studies. The approaches and the corresponding information perspectives are shown in a tabular form below.

Table: Information theory approaches and information perspectives

Approaches	Perspectives
Syntactic or Semantic-independent	Syntactic or Non-semantic

or Mathematical or Statistical	
Semantic-centred or context-independent	Semantic
Context-dependent	Pragmatic or Contextual
Recipient-centred	Semantic or Pragmatic

2. Information is defined as *the description of the state of an object*. There are three keywords in the definition: Object, State and Description. An object is anything in the creation: Living things such as micro-organisms, animal kingdom and human beings; Inanimate things like man-made objects, earth and mountains; and large systems like the planets, galaxies and milky way. Each object is in one of infinite or finite number of states at any point of time. (Illustrate by giving examples). Description of the state in some way constitutes information. The description may be oral, written, pictorial etc.
3. Knowledge continuum is a four-level abstraction of human thought process. (Discuss each level by presenting Fig. 7.1. Illustrate by examples that what is data for someone may be information for another, what is information for one may be knowledge for another etc.)
4. From Eq. 7.5, we know that $p_1 + p_2 + p_3 + p_4 = 1$.
Therefore, $p_4 = 1 - (0.5 + 0.25 + 0.125) = 0.125$.
From Eq. 7.2, we calculate the information content of each message as:

$$I_1 = \log_2 (1/2) = 1 \text{ bit.}$$

$$I_2 = \log_2 (1/4) = 2 \text{ bits.}$$

$$I_3 = \log_2 (1/8) = 3 \text{ bits.}$$

$$I_4 = \log_2 (1/8) = 3 \text{ bits.}$$

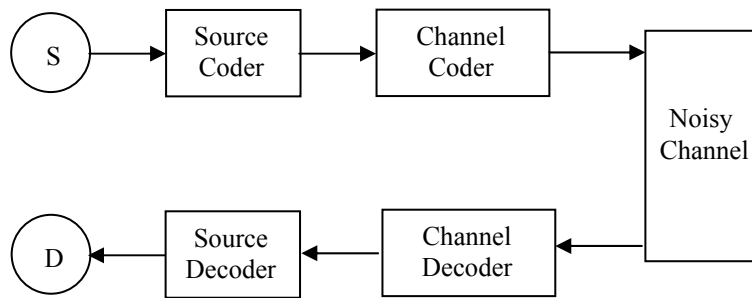
$$I_{\text{Source}} = 1 + 2 + 3 + 3 = 9 \text{ bits.}$$
5. From the laws of logarithms, we know that
If $\log_a x = n_1$ and $\log_b x = n_2$, then n_1 and n_2 are related as

$$n_2 = n_1 \times \log_b a.$$

Now, $1 \text{ Hartley} = \log_{10} 10$. Let $1 \text{ Hartley} = n_2 \text{ bits}$. Then,
 $n_2 = 1 \times \log_2 10 = 3.322 \text{ bits}$. Similarly,
 $1 \text{ nat} = 1 \times \log_2 e = \log_2 2.719 = 1.443 \text{ bits}$.
6. Information entropy H is the average information content of a message of a given source. It is given by Eq. 7.11. In Exercise 4, we have considered a source with four messages. Therefore, the information entropy of this source is given as

$$H = 0.5 \log_2 (1/2) + 0.25 \log_2 (1/4) + 0.125 \log_2 (1/8) + 0.125 \log_2 (1/8)$$

$$= 0.5 + 0.5 + 0.75 = 1.75 \text{ bits.}$$
7. The information entropy of two messages is given by Eq. 7.14. H is calculated for different values of p and plotted with p on the X-axis and H on the Y-axis. The resulting curve looks similar to the one in Fig. 7.2 with the maximum value of H as 1 when $p = 0.5$.
8. A model for both efficiency and reliability incorporates both the source coder and the channel coder at the sending end, and the source decoder and the channel decoder on the receiving side as shown below:



Model for both efficient and reliable communication

9. The probability of one bit being in error P_1 is given by BER as $= 10^{-5}$
The probability of one bit being correct $= 1 - P_1 = 1 - 0.00001 = 0.99999$
10. The information block is 1 1 0 0 1 0 1 1 0 and has 10 bits. One parity bit is added and therefore the size of the transmission block is 11 bits.
Even parity pattern: 1 1 0 0 1 0 1 1 0 1
Odd parity pattern: 1 1 0 0 1 0 1 1 0 0
11. Both the messages contain two atomic statements each connected by logical AND. The probability of occurrence of the messages is given by the joint probability of the constituents.
The probability of message (a) $P_1 = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$.
Therefore, the information content of message (a) $I_a = \log_2 (\frac{1}{8}) = 3$ bits.
The probability of message (b) $P_2 = \frac{1}{3} \times (1 - \frac{1}{4}) = \frac{1}{3} \times \frac{3}{4} = \frac{1}{4}$.
Therefore, the information content of message (b) $I_b = \log_2 (\frac{1}{4}) = 2$ bits.

7.10 KEYWORDS

Bits, nats and hartleys	: Units assigned to dimensionless measure of information depending on the base of logarithm used. <i>Bits</i> when the base is 2, <i>nats</i> when the base is e and <i>hartleys</i> when the base is 10.
Channel capacity	: The amount of information that can be carried by a channel with perfect accuracy in unit time.
Channel coding	: Coding for reliable or error-free communication via a noisy channel.
Contextual information	: Where informational value of a message is assessed based on a given context. Also known as pragmatic information.
Efficient communication	: Maximally utilising the capacity of a channel by the use of coding techniques.
Information entropy	: Average information content of a message in a set of messages.
Information measure	: Assessing the informational value of a message by using one of a variety of techniques.
Message	: An informational statement that describes the state of an object in the universe.

Object	: Anything in the creation: animate, inanimate etc.
Pragmatic information	: Where informational value of a message is assessed based on a given context. Also known as contextual information.
Recipient-centred	: Where informational value of a message is assessed based on the recipient's perception of the meaning. Also called user-centred.
Reliable communication	: Achieving error-free communication by the use of coding techniques.
Source coding	: Coding of source messages to achieve efficiency in information communication.
Source entropy	: Average information content of a message from the source.
Semantic information	: Where informational value of a message is assessed by the amount of substantive information contained in it without regard to the context, sender or recipient.
State of an object	: The condition in which an object exists. All objects in the universe exist in one of finite or infinite number of states.
Syntactic information	: Where informational value of a message is assessed without regard to the meaning of the message.

7.11 REFERENCES AND FURTHER READING

- Brewster, R.L. (1986) Telecommunications Technology. New Delhi: Affiliated East-West Press Pvt. Ltd.
- Hintikka, Jaakko and Suppes, Patrick. (1970) Information and Inference. Dordrecht-Holland: D. Reidel Publishing Company.
- Kasiwagi, Dean. (2003) Information Measurement Theory (IMT). Encyclopaedia of Information Systems, Volume 2. USA: Elsevier Science.
- Lebow, Irwin. (2000) Understanding Digital Transmission and Recording. New Delhi: Prentice Hall Of India.
- Lubbe, Jan C A van der. (1997) Information Theory. Cambridge: Cambridge University Press.
- Verlinde, Patrick. (2003) Information Theory. Encyclopaedia of Information Systems, Volume 2. USA: Elsevier Science.

UNIT 8 DIGITAL INFORMATION

Structure

- 8.0 Objectives
- 8.1 Introduction
- 8.2 Nature of Digital Information
- 8.3 Digital Fundamentals
 - 8.3.1 Binary Coding
 - 8.3.2 Binary Numbers
- 8.4 Digital Text
- 8.5 Digitising Documents
 - 8.5.1 Scanning
 - 8.5.2 Image Compression
 - 8.5.3 Character Recognition
- 8.6 Analog to Digital Conversion
- 8.7 Digital Audio
- 8.8 Digital Video
- 8.9 Digital Formats
 - 8.9.1 Document Formats
 - 8.9.2 Image Formats
 - 8.9.3 Audio Formats
 - 8.9.4 Video Formats
- 8.10 Legality of Digital Documents
- 8.11 Summary
- 8.12 Answers to Self Check Exercises
- 8.13 Keywords
- 8.14 References and Further Reading

8.0 OBJECTIVES

After reading this Unit, you will be able to understand and appreciate:

- how digital information is created;
- the nature of digital information;
- the features of a digital document;
- basics of digital technology;
- how digitised information is stored within the computer;
- the process of digitising documents;
- how to digitise text;
- how to convert analog information to digital form;
- what is digital sound;
- what is digital video;
- different multimedia formats; and
- legal aspects of digital documents.

8.1 INTRODUCTION

Information is central to our daily activities these days. Advances in computer and communication technologies have brought about the representation, recording and communication of information in electronic form. Information may be put in electronic form using analog or digital technology. For example, in a conventional audiocassette, information is recorded using analog technology whereas on a CD-ROM information is recorded using digital technology.

Analog technology has been known for a long time (100 years or more) whereas digital technology is relatively new (40–50 years). Digital technology is preferred over analog technology for reasons of efficiency and reliability. At present, there is a perceptible trend towards the use of digital technology in both communication and computer fields. Everything electronic is moving towards digital technology. One may say that there is a digital revolution that is currently sweeping the world. As a result, electronic information is also going digital. Even sound and video are being recorded using digital technology. Many of you may be aware that many cinema theaters have modernised their projection system and use digital (Dolby) sound systems. Digitally recorded audio and video CDs are available today.

Electronic information in digital form is called **digital information**. In many texts, no distinction is made between electronic information and digital information. The two terms are used synonymously. You must remember that electronic information may be analog or digital whereas digital information is entirely digital. In other words, digital information is electronic but electronic information is not necessarily digital. This Unit is concerned with representation, recording and communication of information in digital form. The Unit also touches upon the legal aspects and the issues of copyright of digital information.

8.2 NATURE OF DIGITAL INFORMATION

Digital information is created and managed by using three digital technologies: digital computer, digital communication and digital storage technology. In addition, there are end devices that acquire information in digital form. Examples of such devices are digital voltmeters, digital telephones and digital facsimile. Digital information is capable of being stored inside a computer, processed by a computer and transmitted over a digital communication system.

A large volume of information in this universe is in non-electronic or analog form. This information needs to be digitised before it can be handled by digital technologies. For example, printed information may be digitised and stored inside a computer using a scanner or a digital camera that is attached to the computer. The computer controls these devices, and acquires and stores the digital images produced by them. Alternatively, digital information may be created directly on the computer by entering information via the keyboard or other input devices like mouse. For example, this Unit is prepared directly on the computer by inputting text using the keyboard and making drawings using the mouse. Thus, we may say that *digital information is created either directly*

on a computer or from other sources with the help of a computer. Examples of digital information include e-mail messages, computerised files, digital books, e-news, textual databases, on-line journals and encyclopaedias on CD-ROMs.

Recorded information constitutes a document. Hence, a document that contains digital information is a **digital document**. Although digital document is the precise nomenclature for computer-generated documents, the terms *e-document* and *e-paper* are used commonly to refer to digital documents. We use these terms interchangeably in this course material. A print document may contain text, figures, tables, graphs and photographs. Unlike print documents, digital documents are **multimedia** in nature. In addition to what is contained in print documents, digital documents may contain voice, music, animation, and motion video. Thus, a digital document may contain information in the following forms:

- | | |
|----------------------|-------------------------------|
| ● Digital text | ● Digitised text |
| ● Digital images | ● Digitised images |
| ● Digital sound | ● Digitised sound |
| ● Digital video | ● Digitised video |
| ● Computer animation | ● Computer generated drawings |

We call each of the above as a multimedia component. It is important to distinguish between digital multimedia component and digitised multimedia component. Digital component refers to items directly created on the computer whereas digitised component refers to items converted to digital form from a original source that is non-electronic or analog. In general, conversion of non-electronic information to digital form involves a scanning or a photographic process. Conversion of analog information to digital form involves the use of analog to digital converters. We shall study these aspects later in this Unit. Digital text consists of alphanumeric and special characters. When text is digitised by a photographic process, an image map of each character is created inside the computer. These image maps are not processable by the computer in the same way as the characters of a digital text are done. The image maps need to be further processed to make them appear as digital text characters for the computer software. These aspects are discussed later in Section 8.5. Most computers have facilities for creating simple line drawings and graphs. These are stored in digital image form. Digitised images and computer-generated drawings represent what are called static images. Moving or dynamic images are produced by digital video and computer animation. The representation and storage formats of different multimedia components of a digital document vary and follow certain standards. We learn about these in later sections in this Unit.

Associated with digital documents are certain features that make them very versatile when compared to paper based documents:

- Digital text documents are computer searchable. A specified string of characters may be searched for and located instantly. The string of characters may be keywords representing subject topics.
- Digital image documents can be processed by image processing software to locate certain features or to compare with an existing image etc.
- Digital documents are computer editable which means that the document

can be easily corrected, reorganised, sequence of presentation changed, etc.

- Digital documents are linkable to one another. A related document can be reached via links thereby providing continuity across documents.
- Digital documents can be annotated with texts that are searchable.
- Digital documents are shareable by many users at the same time.
- Digital documents are capable of being transferred from one location to another by means of communication networks. When transferred, the original document remains intact and only a copy is sent across.

Self Check Exercise

- 1) What is the distinction between digital sound and digitised sound? Give examples for each case.

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

8.3 DIGITAL FUNDAMENTALS

Underlying the representation, storage and communication of information in digital form is the binary system. In binary system, we have only two symbols, 0 and 1. It is interesting to know that we can represent any information by using just two symbols. Any two symbols, say P and Q or α and β , could have been chosen arbitrarily. But the digits 0 and 1 are chosen because they can represent both numeric and non-numeric information and binary arithmetic can be performed on numerical quantities represented by 0s and 1s.

8.3.1 Binary Coding

The underlying idea in information representation is that we form strings of the two symbols to any desired length and pre-assign a meaning to these strings. This is called *coding*, i.e. we are coding known information as binary strings. Then, a sequence of strings may represent a meaningful message. For example, consider the following strings with their preassigned meanings:

0001	A
0011	C
0000	P
0100	T

We have listed four strings above. In each string, there are four binary digits. A binary digit is called a *bit*. Therefore, we say that each string has 4 bits. If a sequence is transmitted comprising bit strings 2, 1 and 4 in that order, i.e. 0011

0001 0100, then the information conveyed is the word CAT. Similarly, the sequence of strings 2,1 and 3 conveys CAP, the sequence 1,2 and 4 ACT and the sequence 3,1,2 and 4 PACT.

With four bits, we can have a maximum of 16 unique combinations. Each bit position has two possible values, either 0 or 1. There are four bit positions giving us a total combination of $2^4 = 16$. Sixteen combinations are not adequate to represent even all the letters in English alphabet. We need to choose strings longer than four bits to represent a bigger character set. Standard coding schemes for character sets are discussed in the next section.

It is not only characters that we can represent using binary symbols. Literally, everything in this universe can be represented using binary strings. To illustrate the ideas involved further, let us consider the representation of days of the week and rainbow colours inside the computer. In each case, there are seven items. Three bits would give us $2^3 = 8$ combinations which are adequate to represent these items. An arbitrary coding scheme for these examples is shown in Table 8.1. A few observations regarding the coding shown in Table 8.1 are in order. Of the 8 combinations, the pattern 111 is unused in coding the days of the week and the pattern 000 is unused in coding rainbow colours. The selection of patterns is arbitrary. The idea is to choose as many patterns as required from an available set of patterns. Some days and colours have identical coded pattern. For example, Thursday and the colour green are coded as 100. How then do we distinguish between the two items inside the computer? The coding is context dependent. A program dealing with colours would interpret 100 as green and another program dealing with days would interpret the same pattern as Thursday. What if a program deals with both colours and days? In this case, the total number of items is 14 and we would need 4 bits to code the items. The situation is analogous to student roll numbers in different classes. Inside a classroom, the roll number uniquely identifies a student. In the context of the whole school, a student is uniquely identified only when he states the class, section and the roll number.

Table 8.1: A Coding Scheme for Days of the Week and Rainbow Colours

Day	Code	Colour	Code
Sunday	000	Violet	001
Monday	001	Indigo	010
Tuesday	010	Blue	011
Wednesday	011	Green	100
Thursday	100	Yellow	101
Friday	101	Orange	110
Saturday	110	Red	111

8.3.2 Binary Numbers

Let us now turn our attention to see how we represent numbers using binary digits (bits). As you are aware, we use a *place value* concept while representing numbers in decimal system. For example, the number 5657 has a value equal to

$$\begin{aligned}
 &5 \times 10^3 + 6 \times 10^2 + 5 \times 10^1 + 7 \times 10^0 \\
 &= 5000 + 600 + 50 + 7 = 5657
 \end{aligned}$$

Each place in a number is assigned a value that is a power of ten and the digit in the place is multiplied by that value. If we identify the place of the digits starting from right, then we have 7 in place 1, 5 in place 2, 6 in place 3, and 5 in place 4. The digit 5 has a value 50 in place 2 and a value 5000 in place 4, i.e. the value assigned to a digit depends on its place in the number. We use a similar place value system to represent numbers in binary system. Since there are only two symbols in the system, the place value is a power of 2. For example, the string 1101 has a value

$$\begin{aligned}
 &1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \\
 &= 8 + 4 + 0 + 1 = 13
 \end{aligned}$$

Much as the way in decimal system, very large numbers can be represented by using a long string of bits. For example, a 32-bit string allows us to represent numbers upto 429, 49, 67,296; i.e. approximately 430 crore or 4.3 billion. There are certain binary string lengths that are used widely in information representation. They are 4 bits, 8 bits, 16 bits and 32 bits. A 4-bit string is called a **nibble** and a 8-bit string a **byte**. The others are referred to by their actual length like 16-bit or 32-bit word.

Binary arithmetic is similar to decimal arithmetic. Addition, subtraction, multiplication and division are performed in the same way as in decimal arithmetic. Since there are only two symbols, addition of two 1s leads to a carry, i.e. $1 + 1 = 10$ much as the way $6 + 4$ leads to a carry in decimal system. When a 0 is added to a 1 or to another 0, there is no carry, i.e. $0 + 1 = 1$ and $0 + 0 = 0$. Similar considerations apply to other arithmetic operations.

Self Check Exercise

- 2) Devise a coding scheme to represent months in a year.
- 3) Add the two binary numbers 1101 and 0101.

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

8.4 DIGITAL TEXT

Text consists of alphanumeric and special characters. In English language, we have 26 upper case and 26 lower case letters in the alphabet. Indo-Arabic numerals have 10 digits, 0 – 9. There are special symbols like +, –, &, * and

%. All these add up to a total of 95 characters. We call this sum total as the character set for text. With a 6-bit string, we can have a maximum of 64 (2^6) combinations which are inadequate to represent all the characters in the text character set. If we choose a string of 7 bits, then the maximum number of combinations we can have is 128 which is adequate to represent all the characters. Hence, a 7-bit string (code) is used to represent characters inside a computer.

The most widely used 7-bit code is the *American Standard Code for Information Interchange* (ASCII). We have an equivalent Indian Standard code called *Indian Standard Code for Information Interchange* (ISCII). Since we have only 95 characters, we are left with 33 combinations in a 7-bit code which can be used for some other purposes. In fact, many of the spare 7-bit combinations in ASCII are used for controlling various aspects of information communication process. These combinations are called *control characters* or *control codes*. Control characters are non-printable and are invisible on a computer monitor. They, however, communicate special signals to devices like printers and other communication devices. The ASCII characters along with their 7-bit codes are presented in Table 8.2. The entry *np* in the Table represents special control characters that are not printable but used to control devices. The 7-bit code is presented with a space after three bits from the left in order to enhance readability. In computer representation, there is no intervening space and the 7 bits are continuous. In practice, ASCII is an 8-bit code (byte) inside the computer with the extra bit being used for error detection. Other than ASCII, there are code sets that are in use. One such well known code set is *Extended Binary Coded Decimal Interchange Code* (**EBCDIC**) used on large IBM computers. EBCDIC is also an 8-bit representation of characters.

As we know, documents are more than plain texts. They contain formatted text; i.e. paragraphs, sections, chapters etc. For example, this course material is a document containing formatted text. Digital text documents are prepared inside the computer using software packages called *word processors*, *text editors* or *text processors*. These documents are stored as files inside the computer. The software packages tend to use their own formatting standards and own file formats for storing these files. When it comes to transporting a document from one system to another to be processed by a different software package, we need standards for conveying formatting information as well as standard formats for files. The most widely used standard for transporting text documents across different computers and different software packages is *Rich Text Format* (RTF). We learn more about RTF in Section 8.9.

Table 8.2: Coding in ASCII

Code	Ch	Code	Ch	Code	Ch	Code	Ch
000 0000	np	010 0000	sp	100 0000	@	110 0000	`
000 0001	np	010 0001	!	100 0001	A	110 0001	a
000 0010	np	010 0010	“	100 0010	B	110 0010	b
000 0011	np	010 0011	#	100 0011	C	110 0011	c
000 0100	np	010 0100	\$	100 0100	D	110 0100	d
000 0101	np	010 0101	%	100 0101	E	110 0100	e
000 0110	np	010 0110	&	100 0110	F	110 0100	f
000 0111	np	010 0111	‘	100 0111	G	110 0100	g
000 1000	np	010 1000	(	100 1000	H	110 1000	h
000 1001	np	010 1001	)	100 1001	I	110 1001	i
000 1010	np	010 1010	*	100 1010	J	110 1010	j
000 1011	np	010 1011	+	100 1011	K	110 1011	k
000 1100	np	010 1100	,	100 1100	L	110 1100	l
000 1101	np	010 1101	-	100 1101	M	110 1101	m
000 1110	np	010 1110	.	100 1110	N	110 1110 110	n
000 1111	np	010 1111	/	100 1111	O	1111	o
001 0000	np	011 0000	0	101 0000	P	111 0000	p
001 0001	np	011 0001	1	101 0001	Q	111 0001	q
001 0010	np	011 0010	2	101 0010	R	111 0010	r
001 0011	np	011 0011	3	101 0011	S	111 0011	s
001 0100	np	011 0100	4	101 0100	T	111 0100	t
001 0101	np	011 0101	5	1010101	U	111 0101	u
001 0110	np	011 0110	6	101 0110	V	111 0110	v
001 0111	np	011 0111	7	101 0111	W	111 0111	w
001 1000	np	011 1000	8	101 1000	X	111 1000	x
001 1001	np	011 1001	9	101 1001	Y	111 1001	y
001 1010	np	011 1010	:	101 1010	Z	111 1010	z
001 1011	np	011 1011	;	101 1011	[	111 1011	{
001 1100	np	011 1100	<	101 1100	\	111 1100	
001 1101	np	011 1101	=	101 1101	]	111 1101	}
001 1110	np	011 1110	>	101 1110	^	111 1110	~
001 1111	np	011 1111	?	101 1111	_	111 1111	np

sp = space np = non-printable control characters

Self Check Exercise

4) Write the ASCII code for the character string ‘IGNOU’

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

8.5 DIGITISING DOCUMENTS

Non-electronic documents and analog electronic documents need to be converted to digital form in order to make them digital documents. In this section, we learn the process of digitising a non-electronic document. In the next section, we learn about converting analog documents into digital

documents. A non-electronic document may be a paper document or use any other medium like palm leaves and may contain hand- or typewritten text, photographs, illustrations, drawings, artwork, graphs etc. The steps involved in converting a non-electronic document to a digital document are shown in Fig.8.1. There are two major steps: *scanning* or *imaging* and *compression*.

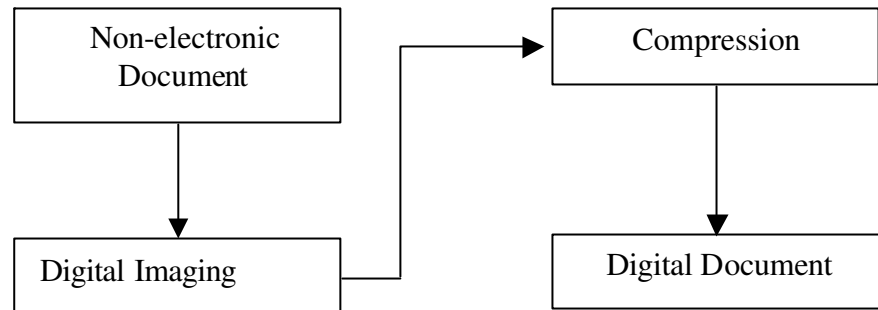


Fig. 8.1: Digitising Documents

8.5.1 Scanning

The first step in digitising a document is to image the document. This may be done by means of a scanning or a photographic imaging process. The scanning process uses a scanner and the photographic imaging process uses a camera. The scanner and the camera may be analog devices or digital devices. An analog device produces wave like electrical signals as output whereas a digital device produces voltage levels representing binary digits as output. Both the outputs represent the information contained in the non-electronic document. If the devices are analog, an additional step of converting analog information to digital, as discussed in Section 8.6, is required. For the present, we assume that these devices are digital. A scanner resembles a photocopier and the process of scanning is similar to that of photocopying or xeroxing. The non-electronic document is placed on a flat bed transparent surface that is then scanned by focussing a light source over the document, measuring the reflected light and presenting the value of the reflected light by means of binary strings. In the case of photographic imaging, the camera is focussed on the document and it produces digital output.

The scanning or photographic imaging is a microscopic process. The surface is scanned from the top left corner to the bottom right corner in a sequential order. The surface is divided into a collection of horizontal lines. Each horizontal line is conceived to be made up of a large number of dots called *pixels* or *pels*. The word pixel or pel is a short form for picture element. The density of dots could vary from 75 dots per inch (dpi) to 2400 dpi. The horizontal line density is also specified in terms of dpi and is usually the same as the density of dots in a line. The dot density and the line density together are called the *scanning resolution*. The commonly used scanning resolutions in the present day scanners are 600×600 dpi, 1200×1200 dpi and 2400×2400 dpi. For a surface of given size, the number of dots or pixels on the surface increases as the scanning density increases. When light is shined on the surface to be scanned, each pixel reflects light according to its contents. The contents may be in colour or in black and white (B&W). We will consider colour scanning later in this section. First, we consider scanning of B&W documents.

In a B&W surface, the content is either white or black of varying shades like dark, light, etc. The varying shades including white are called grey (gray) levels. The quantum of light reflected by each pixel depends on the grey level of the pixel. Each pixel value, i.e. the amount of light reflected by it, is represented by a binary string. Once the scanning of the surface is complete, there are as many binary strings in the output as there are pixels on the surface. While scanning a B&W surface, 16 or 256 levels of shades including white are recognised. Sixteen levels can be distinguished by using 4-bit string (nibble) and 256 levels call for 8-bit string (byte). Commercial facsimile (fax) machines, which also use a scanning technique, recognise only two grey levels, i.e. black and white, requiring only one bit for representing the value of each pixel. The number of bits used to represent the grey levels or colours is called *bit depth*.

The speed of scanning is usually specified in terms of number of pages per minute and is dependent on the scanning resolution. Higher the resolution of the scanner, the longer is the time taken for scanning. Some scanners take as much as a few minutes to scan one sheet of paper.

8.5.2 Image Compression

Scanned image files are generally very large in size. Consider scanning a post card size (6" × 4") photograph using 600-dpi scanner. In six inches there are 3600 pixels and in four inches 2400 lines. If we assume a bit depth of one byte, the size of the scanned image file works out to be $3600 \times 2400 \times 1 = 86400$ bytes. Sixteen such pictures would occupy an entire floppy. Storing scanned images as they are would need a large amount of storage space. In order to reduce the storage requirements for an image, scanning is always followed by compression before storing the image. There are two broad classes of image compression techniques:

- Information preserving techniques
- Approximation techniques

Information preserving techniques ensure that the integrity of the contents of the scanned surface is fully maintained. Approximation techniques tend to approximate the scanned image and in the process may lose some information aspect of the document. Approximation techniques are able to achieve much better compression than the information preserving techniques. In general, information-preserving techniques are used for documents containing factual data and approximation techniques for documents containing photographs, pictures etc. The compression efficiency of a technique is measured by a parameter called *compression ratio CR*, which is defined as:

$$CR = \frac{\text{Size of the uncompressed image}}{\text{Size of the compressed image}} \quad (8.1)$$

Most of the compression software packages produce a compression ratio in the range of 10 – 20. Some very sophisticated packages produce a compression ratio in the range of 40 – 60.

We now discuss some aspects of colour scanning. Human vision recognises radiation in the frequency range of $4 \times 10^{14} - 8 \times 10^{14}$ Hz. It is in this range that the frequencies of the main colours of the rainbow (VIBGYOR) lie. White

light comprises the wavelength of all visible colours. We see an object and perceive a colour when a specific frequency component (colour) of the white light falling on the object is reflected and detected by the human eye. Radiation of different frequencies produces the sensation of different colours in the eye.

In television and colour digital images, different colours are formed by mixing three primary colours red, green and blue (RGB). Colour scanners measure the intensity of reflected signals from each pixel at the frequencies corresponding to these three primary colours. The intensity of each colour is measured with 256 levels, calling for one byte to represent the intensity value. Thus, colour scanning produces 3 bytes of digital data for each pixel and a colour image size is three times larger than that of a B&W image for a given scanning resolution. High-resolution colour scanners use 16 bits to represent intensity values and produce six bytes for each pixel.

8.5.3 Character Recognition

Consider scanning one line of text containing 60 characters at 600-dpi resolution. Let the height of the characters be 0.25" and the line width 8". The size of the scanned image file works out to be $600 \times 8 \times 600 \times 0.25 = 7200$ bytes. In Section 8.5, we learnt that characters can be stored inside a computer using ASCII bytes. If this line of text were to be stored as ASCII characters, then we would require a space of only 60 bytes, i.e. 120 times less space than the scanned image. Clearly, the images take up a lot of computer storage when compared to text. It makes good sense to store the information in ASCII form if the document being digitised contains text predominantly. This is achieved by further processing the scanned image by *character recognition software* that has the capability to recognise character patterns and reconstruct a text file from an image file. The process is shown in Fig.8.2. Since the process involves optical scanning before character recognition, it is called **Optical Character Recognition (OCR)**.

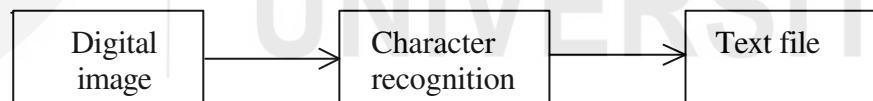


Fig.8.2: Obtaining Text Files from Image Files

The character recognition is not always hundred per cent correct. If the original document is typewritten or printed, character recognition is likely to be highly successful. If, on the other hand, the document is hand-written, character recognition may only be partially correct. In general, the output of character recognition software needs to be manually edited to ensure fully correct recognition. To aid the editing process, software packages that check for spelling, sentence construction etc. may be used.

In Section 8.2, it was brought out that digital documents are far more versatile than paper documents because of the associated computer processing and communication possibilities. The idea that paper based information can be very effectively managed once it is converted to digital image or text has led to the emergence of what are known as **document management systems**. These systems are useable effectively in office environment. Every paper document is converted to a digital document, which is then used to take follow

up actions. Such documents are easily retrieved, distributed with annotated instructions and managed in an automated mode. A typical document management system consists of a scanner, character recognition software, office management software, a personal computer, a printer and optical storage devices supporting writable optical disks.

Self Check Exercise

- 5) Consider a B&W document containing 20 pages of dimension $10'' \times 10''$ being scanned at a resolution of 1200×1200 dpi. The bit depth is 4 bits. The compression ratio of the software is 20. Determine the:
 - i) Size of the scanned image file before compression
 - ii) Size of the image file after compression.
- 6) If the document in Q.5 above contains only text and is processed using OCR software, estimate the file size required and the saving in storage. Assume that each line of the text contains 100 characters and each page contains 40 lines of text. What do the results indicate?

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

8.6 ANALOG TO DIGITAL CONVERSION

A large number of physical quantities measurable or observable in this world are **analog** in nature. By analog we mean that these quantities take on values that vary continuously with time. For example, the day temperature values vary continuously over a period of time. When the values of such quantities are plotted as a graph with time as the X -axis, the curves that represent these values are continuous. In nature, a large number of information signals are also analog. For example, human speech and music produce analog signals. Information may be recorded, processed and communicated using analog technology. In fact, this was the case entirely in the past. Even today, analog technology is in wide use. Analog technology suffers from certain disadvantages. First, analog signals are susceptible to external noise and their reception becomes unreliable in the presence of noise. Second, analog devices are temperature sensitive and their performance is affected by variations in the ambient temperature. Last, analog signals of different quantities such as voice and video are very different in their electrical characteristics like voltage, current, frequency and power. This necessitates the design of new systems whenever new quantities represented by analog signals are to be stored, processed or communicated.

Search for an alternative to analog technology has given birth to the **digital technology**. Digital technology is more rugged and reliable when compared to analog technology. Digital signals have better noise immunity, quality, consistency of reproduction and ease of processing. Early digital computers were built in mid 1940s and the first digital communication system became operational in 1962. Since then, the digital technology has been advancing leaps and bounds both in the fields of communications and computers. As we know, today's computers are hundred per cent digital. The telecommunication networks world over are fast evolving towards digital networks. Information representation is also fast becoming digital. This is the reason why we are studying about digital information in this Unit.

As mentioned above, a large amount of information produced in nature happens to be in analog form. For example, sound is in the form of air pressure waves that are analog. Our ears are tuned to hear analog signals rather than digital ones. In order to be able to use digital technology, analog quantities need to be converted to digital form. Digitising analog information is done by means of **Analog to Digital Converters** (ADC). While digital technology is used to store, process and communicate information, for actual consumption by human ears and eyes the information needs to be presented in analog form. Therefore there is the need to convert digital information back to analog form. This is done by means of **Digital to Analog Converters** (DAC). The principles underlying ADC and DAC are similar to converting graphical representation to numerical form and vice versa respectively. When a point value is read from a curve, it is a number or a numerical value. The graphical representation corresponds to analog form and the numerical to digital. If we read off points closely from a curve, we can form a table of values that represents the curve. We can reconstruct the original curve from the table by interpolating between the successive points. Thus, both analog and digital forms represent the same information and one form can be derived from the other.

How closely do we need to read off points from a curve in order to preserve the information content and to reconstruct the original curve with full information content? If we take too few points, we are bound to lose information content. If we take too many points, the size of the table becomes unnecessarily large. In other words, we will be overloading the digital system unnecessarily. Therefore the question is what is the optimum number of points that would preserve information content, and at the same time reduce load on the digital system? The answer to the question lies in **sampling theorem** stated and proved in 1933 by Shannon and Nyquist. According to the sampling theorem, in order not to lose the informational content, the analog signal must be sampled at a rate f_s which is equal to or greater than twice highest frequency component f_m of the analog signal as defined in Eq.8.2. The process of sampling is equivalent to reading off points from a curve.

$$f_s \geq 2 f_m \text{ samples/sec.} \quad (8.2)$$

Sampling time interval, which is the inverse of the sampling rate is given as

$$T_s \leq 1/(2f_m) \text{ seconds} \quad (8.3)$$

The analog signal is sampled at regular time interval of T_s . The minimum sampling rate of $2f_m$ is called the **Nyquist rate**. Usually, the sampling is done

at a rate slightly higher than the Nyquist rate. Figure 8.3 shows sampling of an analog signal. The X-axis represents time and the Y-axis the amplitude of the analog signal. The sampled values appear as vertical arrows. They look like pulses or spikes.

The next aspect of digitising analog signals is *quantisation*. The sampled values of the analog signals may have any value in a continuous spectrum of values varying between the minimum and the maximum amplitude of the analog signals. Digital presentation of continuous values calls for very long binary

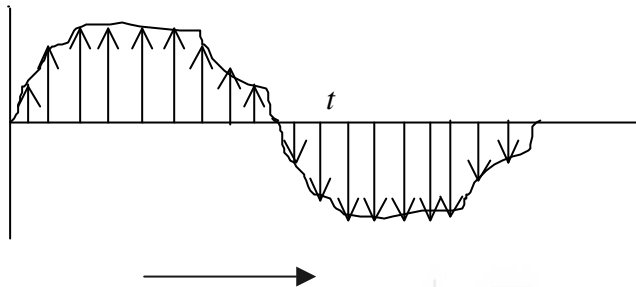


Fig. 8.3: Sampling Analog Waveforms

strings of ones and zero. But practical considerations limit the bit string length to 4, 6 or 8 bits. The number of bits determines the number of discrete values that can be represented between the minimum and maximum values of the analog signals. With 4 bits we can represent 16 (2^4) values, with 6 bits 64 (2^6) and with 8 bits 256 (2^8) values. The values vary in steps and are fixed. It now becomes necessary to approximate the sampled signal values to the nearest fixed value in the range of specified values. This process of fixing a set of specific values and approximating the sampled value to the nearest fixed value is known as **quantisation**. Obviously, quantisation introduces error in sampled values. But the design of the system is usually such that the error levels do not affect the quality of signals in any significant manner.

The next step in digitisation is the coding process, i.e. representing the quantised values by means of a binary string. Since the analog signal may have both positive and negative amplitudes, one bit in the binary string is used to denote the sign and the remaining bits represent amplitude values. The number of bits used to represent a quantised sample value is called *sample resolution*.

In the above described A-D conversion process, since we generate pulses by sampling, approximate their values to previously fixed amplitude levels (quantisation) and then code them into binary strings, the process is called **pulse code modulation (PCM)**. When telephone speech is digitised using standard PCM, quantised sample values are represented by 8-bit strings, i.e. sample resolution is 8 bits. The most significant bit represents the sign of the analog signal and the remaining 7 bits the magnitude. There are other techniques of ADC such as *differential pulse code modulation* and *delta modulation*. A discussion on these techniques is beyond the scope of coverage for MLIS course.

As we have seen above, each sample value is represented by a byte when the sample resolution is 8 bits. Then a sequence of bytes represents the original analog signal. This sequence can be stored in a computer or transmitted over digital communication systems to other destinations. To reconstruct the original

signal, we need to feed the sequence of bytes to a discretiser and a signal smoothening filter. The discretiser takes each byte of digital information and produces the corresponding quantised voltage level. The sequence of bytes processed by the discretiser produces a sequence of quantised voltage levels as pulses. These pulses are then passed through a smoothening filter that interpolates the values to produce analog waveform. This entire process of PCM ADC and DAC is depicted in Fig 8.4.

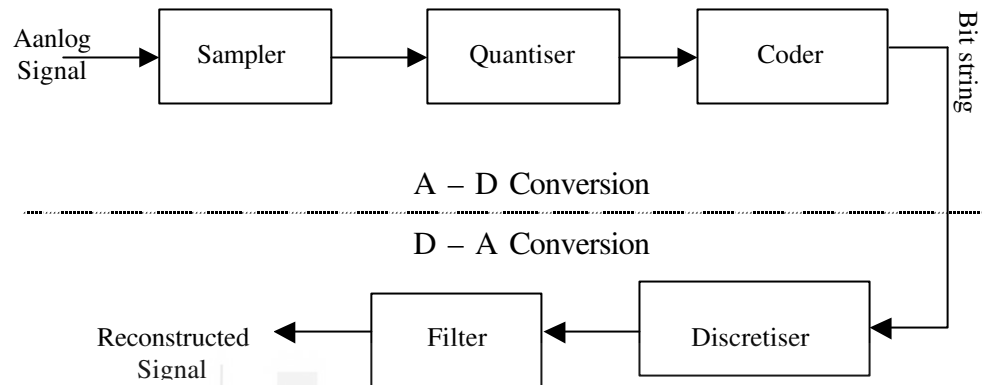


Fig.8.4: ADC and DAC of Information Signals

Self Check Exercise

- 7) A file containing digitised information corresponding to 5 seconds of analog signal has a size of 400 kB. The sampling resolution is 8 bits. If the sampling has been done at Nyquist rate, what is the maximum frequency content of the analog signal?

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

8.7 DIGITAL AUDIO

Human ear is sensitive to frequencies in the range of 20 Hz to 20 kHz. This frequency range is called the *audio spectrum*. Audio information like speech and music occupy different portions of the audio spectrum. Frequency ranges for different types of audio information are shown in Table 8.3. Human speech lies predominantly in the range 300 Hz to 7 kHz. A portion within this range, 300 Hz to 3.4 kHz, is called intelligible speech. By intelligible speech, we mean that it is both recognisable and understandable. We can recognise the person speaking and understand what is being said. All telephone networks are designed only to carry intelligible speech. Hence, intelligible speech is referred to as *telephone quality* speech. FM radio broadcasting stations broadcast music up to 10 kHz or 15 kHz depending on the technology used. Audio CDs record music up to 20 kHz.

Table 8.3: Audio Frequency Spectrum

Information Type	Frequency Range
Full Audio Range	20 Hz – 20 kHz
Speech Spectrum	300 Hz – 7 kHz
Intelligible Speech	300 Hz – 3.4 kHz
Low Fidelity Music	100 Hz – 10 kHz
High Fidelity Music	50 Hz – 15 kHz
Very High Fidelity Music	20 Hz – 20 kHz

In the digital domain, telephone quality speech, often called *toll speech*, is digitised using 8 k samples per second, a little more than Nyquist rate. The sample resolution used is 8 bits. When we transmit digital speech over telephone channels, we are actually transmitting 8 kilo samples per second with each sample represented by 8 bits. On a serial communication link, this amounts to a bit rate of $8\text{ k} \times 8\text{ bits} = 64\text{ kbps}$. In the context of Internet connections, we often hear line speeds of 64, 128 and 256 kbps. These speeds come about from the fact that one or more digital speech channels are assigned for Internet connectivity. In the United States, 7-bit sample resolution is used for telephone speech, giving a data rate of 56 kbps.

For different quality of sound signals, digitisation calls for different sampling speeds and sample resolutions. Table 8.4 gives the sampling rates and sample resolutions used in different audio products. One major disadvantage of digital audio is that it produces a large volume of data. For example, a floppy can hardly hold 10 seconds of digital audio. It, therefore, becomes necessary to compress digital audio before storage, much as the way it is done with digitised documents. A number of compression standards are used for this purpose. One

Table 8.4: Digital Audio Parameters

Media	Channel	Sampling rate	Resolution	Bit rate
Telephone	Mono	8000	8 bits	64 kbps
Audio CD	Stereo	44,100	16 bits	1.41 Mbps
Digital Audio				
Tape (DAT)	Stereo	48,000	16 bits	1.536 Mbps
Digital Radio	Stereo	32,000	16 bits	1.024 Mbps

among them is *audio compression-3* (AC-3). When PCM is used for digitisation and AC-3 is used for compression, the digital sound is known as **digital dolby** sound, the name that we come across in cinema theatres these days.

8.8 DIGITAL VIDEO

Digital video and computer animation fall in the motion video class unlike digitised images which are static and single frame images. The underlying principle of motion video is that a moving image can be represented by a sequence of still images that are projected one after another at a certain rate.

This principle was first applied in motion pictures using photographic film technology. The current trend is to produce high quality moving images using computers applying the same principle. The principle works because of the *persistence of vision* property of the human eye. Any image projected on the human eye persists for about 40 – 50 ms. If a sequence of still images depicting progressive stages of motion is projected on the human eye every 30 – 40 ms, the eye perceives the sum total of projections as a continuously moving picture. Each still image is called a *frame* and we need a frame rate of 25 – 30 frames per second (fps) to produce the effect of continuous picture. A higher rate, say 30 fps, produces a smooth picture and a lower rate, say 15 fps, produces a jerky picture that strains the eyes.

An image captured by a video camera is analog in nature. This needs to be digitised to form digital images. Analog to digital converters (ADC) are used for this purpose. The ADC may be placed externally between the camera and the computer. Alternatively, it may be in-built within the camera in which case the camera is called a **digital video camera**.

Compression techniques applied to individual frames are similar to the ones used in digitised images. In addition, redundancy in neighbouring frames is used to obtain further compression. Unless there is a scene change, the adjacent frames differ very little in contents. In principle, a few frames can be recorded in full and then only the differing aspects of subsequent frames are recorded. A widely used file standard for storing digital video is Moving Picture Experts Group (MPEG) format. Stored images are retrieved and decompressed to form full frames of still pictures that can be projected on a TV screen or a computer monitor at the required frame rate. These devices require analog signals and hence digital to analog conversion is carried out before the images are sent to the monitor. Much as the persistence of vision, computer monitor and TV screen have the property of persistence of display which gives the impression of continuous picture on the screen.

Animation is the process of creating a moving image by playing still frame drawings at 15 – 20 fps rate. Traditionally, an artist hand draws animation frames that are then imaged by photographic film process. Presently, the trend is for the artist to create animation frames using computers. An example of simple animation is the hourglass displayed on the computer screen when the processor is busy on a particular activity.

8.9 DIGITAL FORMATS

The prolific use of digital information, witnessed in the last 10 years, has led to the emergence of a number of formats and standards for storing and delivering digital information. Awareness of different digital information standards has become important for the library staff, particularly in the context of converting conventional libraries to digital ones. This section presents a brief overview of the standard digital formats that are widely in use for documents, audio, still images and motion video. In general, the standard formats deal with one or more of the following aspects:

- Storage and/or transfer
- Information structuring
- Information presentation.

8.9.1 Document Formats

Digital document formats fall under three classes: *basic text formats*, *presentation formats* and *structured formats*. We briefly discuss the formats under each of these classes.

1) Basic text formats

Text formats are the simplest form of digital formats and are largely used for documents containing predominantly textual information. There are three text formats used for text representation: **ASCII**, **Unicode** and **RTF**. Of these, the first two are used for encoding characters. We have discussed ASCII in Section 8.4. ASCII is used to represent Western language characters, i.e. Latin characters. Unicode is proposed as a multi-lingual extension of ASCII to represent characters in major written languages of America, Europe, the Middle East, Africa, India and the Asia Pacific region. Unicode is a 16-bit code that has the capacity to represent 64k characters. At present, 38,885 characters have been defined. Both ASCII and Unicode are pure character codes and do not support formatting or page layout features other than those created by the user using the character set.

Rich Text Format (RTF) is an enhanced text format that supports some minimal formatting features like font types and sizes, margins, paragraphs, bold, italic and underlined characters and justification. RTF is widely used for transporting text documents across different computers and different software packages. RTF is not a multimedia format. Being pure text format, multimedia contents and hyperlinks are not supported in RTF. All text processing software packages accept and deliver RTF files. They have a mechanism to convert own file formats into RTF and vice versa. While RTF provides a standard file format, its ability to support formatting features are limited. Advanced features like columnar text, tables and drawings may not be successfully transported by RTF. In general, there is this caution that some formatting information may be lost when converting a word processor file to a RTF file.

2) Presentation formats

Presentation formats are meant for on-screen display or printing. They are based on page description languages that preserve the look and feel of the original layout with precise location of graphical elements. Two well-known presentation formats are *Postscript* and *Portable Document Format* (PDF). Both the formats are developed by Adobe Corporation and need the special software package distributed free by the corporation under the trade name *Adobe Acrobat Reader* for browsing. PDF is an improved version of Postscript that supports features like table of contents, internal hyperlinks and thumbnail views.

3) Structured formats

Structured formats are somewhat like presentation formats but are more flexible. They do not retain the original look and feel of the documents but are used for on-screen display and printing. They are based on *mark-up* principles that are practised by the publishing industry. The mark-up, however, takes place in the electronic domain instead of the conventional markings on paper documents. There are three structured formats that are in use:

- Standard Generalised Mark-up Language (SGML)
- Hypertext Mark-up Language (HTML)
- Extensible Mark-up Language (XML)

SGML was first developed by International Standards Organisation (ISO) for use among typesetting machines used in the publishing industry. The language definition is very comprehensive and therefore complex. A simplified version of SGML is HTML for use by non-experts. HTML is used extensively on Internet. XML is an enhanced version of HTML. It retains the simplicity of HTML but offers more features.

8.9.2 Image Formats

There are three commonly used formats for storing and transferring digital images obtained from a scanning or a photographic process:

- Tagged Image File Format (TIFF)
- Graphics Image Format (GIF)
- Joint Picture Expert Group (JPEG) format.

The first two of these formats use information-preserving compression techniques and the last one uses a lossy compression technique. TIFF has been developed as the common format for image scanners and DTP software. Since TIFF uses loss-less compression, it preserves the original exactly, retaining layout features, graphics and any character form. Being the bitmap of the original, it has to be passed through OCR software before the text, if any, in the original can be made editable. GIF has been developed for use on the Internet. GIF uses 8-bit representation for the pixels and hence can represent only 256 colours or grey levels. In this sense, it has limited resolution but the file sizes are small and can be transported easily across Internet. JPEG format is an image coding standard that has been optimised for continuous tone products such as photographs. It supports 16 million colours. It performs lossy compression by discarding information that is considered non-essential to the image. Hence, it achieves very high compression ratios but the quality of the image suffers. Options are available (typically three) to choose between picture quality and the compression ratios to be achieved. There are software packages that convert images from one format to another.

8.9.3 Audio Formats

There are a number of digital audio formats proposed and used by different manufacturers and expert groups. Important ones among them are WAV by Microsoft, AIFF by Macintosh, AU by Sun Micro Systems and MP3 by Motion Picture Expert Group (MPEG). All these formats use a standard file structure as shown in Fig. 8.5. In Fig.8.5, *Wrapper* contains management information such as licensing conditions from the copyright owner of the product. Header contains information about sampling rate, sample resolution and the type of compression used. Certain audio formats support *streaming* facility. Streaming enables a user to listen to the early part of the file while the rest of the file is being downloaded. Playback begins as soon as several seconds of audio data has been downloaded and stored. Downloading continues while the playback is on.

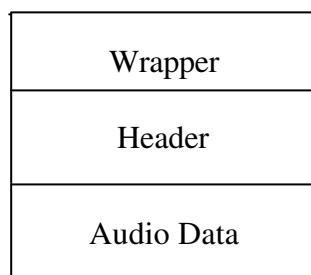


Fig. 8.5: Digital Audio File Structure

8.9.4 Video Formats

Digital motion video formats are standardised by Motion Pictures Experts Group (MPEG) set up by ISO. These standards are used for recording video on CDs and digital videotapes in compressed form. Standards for transferring real time video on telecommunication networks are evolved by International Telecommunication Union (ITU). At present, there are three MPEG standards in vogue: MPEG-2, MPEG-4 and MPEG-7. Two observations are in order with regard to motion video. First, motion video is always accompanied by audio these days. Second, motion video may be visualised as a sequence of still frames played out at certain rate. As a result, MPEG standards draw upon digital audio and digital image standards to a large extent. Audio CD standard at 44.1 kbps and DAT standard at 48 kbps are used by MPEG standards for recording audio.

8.10 LEGALITY OF DIGITAL DOCUMENTS

Much as the paper based documents, electronic documents must provide for information integrity, authentication, accessibility, and confidentiality in order to qualify as a legally acceptable document.

Integrity of an electronic document means that the document is so preserved as to represent accurately the information originally generated, transmitted or received without loss, damage or manipulation. The format of preservation may be the same as in the original document or may be different. If the preservation format is different, then there must exist a means by which it can be demonstrated that the integrity of the original information is unaffected.

Accessibility means the ability to gain access to the original document for subsequent references in future. The requirement in conventional law that information shall be in writing or in the typewritten or printed form is actually met by the accessibility criterion of electronic documents. Conventional written documents ensure non-repudiation by contracting parties at a future date. Similarly, electronic contracts must also provide for binding the parties concerned to the document in such a manner that none of the parties would be able to deny the content of the document.

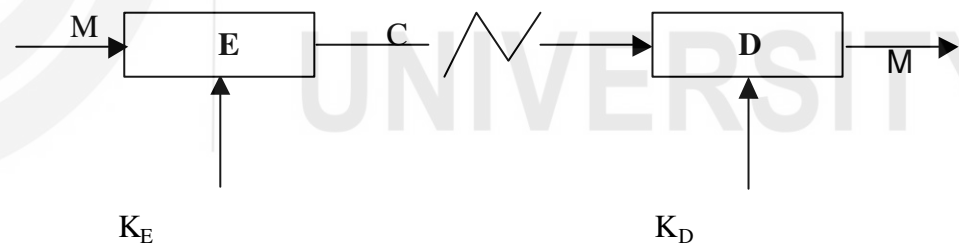
The requirement of any conventional law that affixing the signature of the person(s) concerned shall authenticate a document is met by **digital authentication procedures** in electronic documents. A digital document is authenticated by *digitally signing* or by affixing a *digital signature*. Much as the paper based signature, digital signature also identifies the originator of the electronic document, and conveys the express agreement to the contents of the

document. Digital signatures must be reliable enough for a third party to verify and confirm that the document is actually created by the originator and has not been tampered with by anyone else.

Confidentiality implies a provision to be able to send documents to selected persons only. A confidential document can be opened (accessed) and read by only those who are authorised to deal with such documents. Confidentiality provision also covers privacy aspects and private communication. It should be possible to define different levels of confidentiality such as:

- Confidential
- Strictly confidential
- Secret and
- Top secret

Secrecy in communication systems and digital storage has been almost always achieved by the use of cryptographic techniques. Cryptography may be defined as the art of hiding the significance of information while communicating or in storage. Applying an encryption method and an encryption key to the plain text produces cryptographed text, known as cipher text in technical terms and as coded message in popular parlance. The cipher text is decoded by applying a decryption method and a decryption key. Figure 8.6 depicts a general scheme of cryptography. In Figure 8.6, if the encryption and decryption keys, K_E and K_D are identical, the cryptographic system is known as **private key cryptography** or **symmetric crypto system**. If the two keys are different, but form a unique pair with certain properties, the cryptographic system is known as **public key cryptography** or **asymmetric crypto system**. The public key cryptography system is the one used for authentication of digital documents.



C = cipher text D = decryption method E = encryption method
 K_D = decryption key K_E = encryption key M = message

Fig. 8.6: General Scheme of Cryptography

The main purpose of cryptography is to protect user data from intruders or attackers. In the recent times, the term *hacker* has come to signify intruders of information bases. Although the terms *attacker* and *hacker* are used interchangeably, an attacker is one who attempts to break a security system whereas a hacker is one who attempts to expose or exploit a loophole in the security system. Hackers are further classified as *white hat hackers* and *black hat hackers*. White hat hackers are usually interested in exposing loopholes whereas black hat hackers exploit the loophole for personal gains or to harm unsuspecting victims.

Digital information, like print information, needs to be copyright protected. A digital product is protected from copying by incorporating anti-piracy measures using techniques like encryption. Copyright provisions of digital documents treat an attempt to break or circumvent the anti-piracy measures as a crime. However, anti-circumvention provisions are exempted for non-profit libraries, archives, educational institutions, academicians and graduate students to varying degrees and limited extents. Digital products are ideally suited for distance education. How to promote distance education while ensuring copyright protection is a subject of study at present.

8.11 SUMMARY

This Unit deals with representation of different kinds of information in digital form. Information is multimedia in nature comprising text, pictures, drawings, audio, video, animation and computer graphics. When represented in digital form, information of any kind appears as a string of ones and zeros. This helps in building systems that are capable of handling ones and zeros only and such systems can be made very robust. This is the underlying consideration for adopting digital technology. After having discussed the nature of digital information, the Unit places the digital fundamentals in perspective. The two distinct aspects of digital fundamentals; i.e. digital coding and binary number system are discussed. Representation of text in digital form is then discussed. Conversion of textual information in print form to digital text is then presented. This conversion process involves scanning, compression and optical character recognition. A large volume of information in nature appears in analog form that requires to be converted to digital form. The Unit then discusses analog-to-digital and digital-to-analog conversion processes. Representation of audio and video information in digital form is then discussed. The different standards that are currently used for representing multimedia information components are then presented. Finally, the Unit touches upon the legal and copyright aspects of digital information.

8.12 ANSWERS TO SELF CHECK EXERCISES

- 1) When sound is recorded directly on the computer using a microphone that is attached to the computer, it constitutes digital sound. Example is your voice recorded on the computer.

When previously recorded sound in analog form is passed through an analog to digital converter, it constitutes digitised sound. Example is a digital audio CD of an old song.

- 2) Since there are twelve months in a year, we need 12 binary combinations to represent them. With three bits we have $2^3 = 8$ combinations which are not adequate. Four-bit strings give us $2^4 = 16$ combinations and we may choose any 12 of them to represent the months in a year. The 16 combinations and a coding scheme are given below:

Code	Month	Code	Month
0000	Unassigned	1000	August
0001	January	1001	September
0010	February	1010	October
0011	March	1011	November
0100	April	1100	December
0101	May	1101	Unassigned
0110	June	1110	Unassigned
0111	July	1111	Unassigned

- 3) We perform binary addition much as the way we do in decimal arithmetic as illustrated below:

Carry	1101
Number 1	1101
Number 2	0101
Result	10010

- 4) From Table 8.1, it is seen that the ASCII code for the letter 'I' is 1001001. Similarly, by looking up the Table the ASCII code string for the character string 'IGNOU' is obtained as 1001001 1000111 1001110 1001111 1010101.
- 5) Size of the uncompressed file = $20 \times 10 \times 1200 \times 10 \times 1200 \times 0.5 = 1.44$ GB.
Size of the compressed file = $1440/20 = 72$ MB.
- 6) Total no. of characters in the document = $20 \times 40 \times 100 = 80000$. Therefore the size of text file obtained after OCR = 80 kB. Factor of saving in storage from uncompressed file = $1440000/80 = 18000$. Factor of saving in storage from compressed file = $72000/80 = 900$. Clearly, OCR process results in significant savings in storage.
- 7) Digitised data in one second is $400/5 = 80$ kB. This corresponds to 80 kilo samples per second, as the sampling resolution is 8 bits, i.e. each sample is represented by a byte. This rate is given as Nyquist rate. Therefore, the maximum frequency component in the signal is $80/2 = 40$ kHz.

8.13 KEYWORDS

Accessibility	: Ability to gain access to the original document
Analog Information	: Information represented by continuous signals like curves in a graph
ASCII	: American Standard Code for Information Interchange

Audio Spectrum	: The frequency range that is audible to the human ear
Binary Coding	: A system of coding information in binary form
Binary Number System	: A system of representing numerical quantities using only two symbols '1' and '0'
Compression Ratio	: Ratio of uncompressed image file size to the compressed image file size
Cryptography	: The art of hiding the significance of information
Digital Document	: A document that contains digital information
Digital Information	: Information in digital form represented by ones and zeros
Digital Signature	: The process of digitally signing a digital document
Digital Text	: Text represented in digital form
Digitised Text	: Text originally in print or other form converted to digital form
Grey (gray) Levels	: Different black & white shades in a picture
Integrity	: Preservation of the original contents
Multimedia	: Comprising text, picture, diagram, image, sound, video, and computer graphics and animation
OCR	: Optical character recognition
PCM	: Pulse Code Modulation
Primary Colours	: Red, green and blue colours. A mix of these colours is used for representing different colours in a colour image
Quantisation	: The process of approximating a sampled value to the nearest standard value
Sample Resolution	: Number of bits used to represent a quantised sample value
Sampling Theorem	: A theorem that specifies the minimum sampling rate for digitising analog signals
Scanning Resolution	: A specification of how closely the dots and lines are chosen for scanning a document

8.14 REFERENCES AND FURTHER READING

Cleveland, Gray (1999). *Selecting Electronic Document Formats*. Ottawa: IFLA UDT Core Programme.

Sharda K. Nalin (1999). *Multimedia Information Networking*. New Delhi: Prentice Hall of India.

Steinmetz, R. and Nahrstedt, K. (2002). *Multimedia Fundamentals*. Vol. 1: *Media Coding and Content Processing*. New Delhi: Prentice Hall of India.

Vijayashankar, N. (1999). *Cyber Laws*. Bangalore: Ujvala Consultants Pvt. Ltd.

Viswanathan, Thiagarajan (2002). *Telecommunications Switching Systems and Networks*. New Delhi: Prentice Hall of India.



ignou
THE PEOPLE'S
UNIVERSITY

SOSS-IGNOU/P.O. 2.5T/May, 2006



ignou
THE PEOPLE'S
UNIVERSITY

ISBN-81-266-2045-5