1

CONCEPT OF INFORMATION

Information is a cover term which signifies the strategic resource determining an organization's competitiveness. It is an imprecise term used to represent data that has been presented in order to alter the understanding of the receiver. It helps in strategic planning and managerial control by adding to a representation, by correcting or confirming some previous information, or by acting as a surprise value to the receiver, who could not predict it. The relation of data to information is that of raw material to finished product. In other words, data which has been processed into useful and meaningful form is called information. More precisely, when the information system processes data obtained in unusable form into a usable form, it becomes information to the intended recipient.

The basic purposes of information are to inform, evaluate, or persuade, and it is needed for identifying problems, solving them, making decisions and plans, and undertaking controlling activities. It has value in specific decision-making contexts in that it changes the probabilities attached to expected outcomes in a decision situation. Generally speaking, information is a fact, datum, observation, perception or another thing that adds to knowledge, and its purpose is to reduce uncertainty about decision situation and its consequences.

Definition of Information

In common parlance, the term 'information' refers to organized data that has been converted into a meaningful and useful context for specific end-users. The systematic sequence of operations performed on data, to transform it into information, is called data processing. In the words of Gordon B. Davis, "information is data that has been processed into a form that is meaningful to the recipient and is of real or perceived value in current or prospective actions or decisions". Information for one person may be raw data for another. The value of information arises from the fact that it affects the decisions or actions to be taken. Transformation of data into information can be represented as:

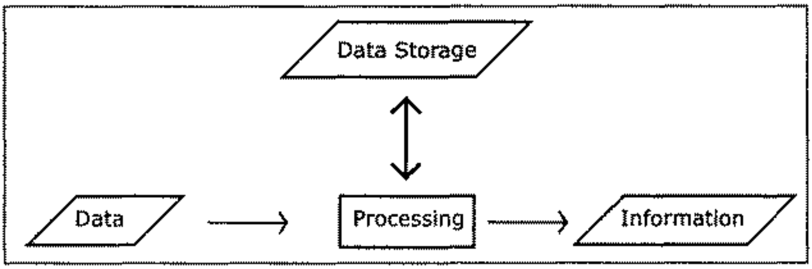

Fig. 1.1: Data Processing

Information in the Mathematical Theory of Communication

Information in the mathematical theory of communication is the average number of binary digits, which must be transmitted to identify a given message from a set of all possible messages to which it belongs. The mathematical theory explains a coding scheme of information. It tries to represent a message with the help of binary digits. The binary digits 0 and 1 represents the absence or presence of a signal, and are called bits. Different combinations of binary digits make it possible to identify a message from a group of messages. If there are two messages, a two-bit code is used to identify them. The theory was developed by Norbert Weiner, a famous mathematician.

Mathematical Definition of Information

Consider a communication system in which the allowable messages are: M_1, M_2....M_k, with probabilities of occurrence : P_1, p_2....P_k, subject to the condition that $P_1 + P_2 + ...P_k = 1$. If the sender selects a message of probability P_k and if the receiver has correctly identified the message, then the mathematical definition of information says that "the system has conveyed an amount of information represented as: $I_k = \log_2 {}^1/P_k$,

Where I_k = Information conveyed.

The definition makes it clear that a greater amount of information has been conveyed only when the receiver correctly identifies a less likely message (message having less probability of occurrence). Thus, we can prove that if two independent messages, M_1 and M_2, are correctly identified, then the amount of information conveyed is the sum of the information associated with each of the messages individually. It can be mathematically expressed as follows.

$$I = \log_2 \frac{1}{P_1} + \log_2 \frac{1}{P_2} + \cdots$$

$$I = \sum_{i=1}^{n} Pi \log_2 \frac{1}{Pi}$$

The information content can be generalized as:

$I = \log_2 n$

Where, n = the total number of equally likely messages.

In this way, only three bits would be needed to represent eight messages. It can be represented as:

$I = \log_2 8 = 3$

The codes for eight messages are:

Message Number	*Bits*
1	000
2	001
3	010
4	011
5	100
6	101
7	110
8	111

If the message contains an alphabetical symbol plus a space symbol, the number of bits required will be an average of 4.75 per letter, since letters do not occur with equal frequency. The use of short codes for common letters and space, and longer codes for letters occurring less frequently, reduces the average code size (average information) required to transmit alphabetic text. Similarly, the information required to identify a message will also be different for each one - short codes for messages with many common characters, and long codes for messages with many uncommon characters. Suppose the probabilities of occurrence of messages are 0.50, 0.25, 0.15 and 0.10, then the average information content can be calculated as follows:

Message Number	*Probability of the Message*	$log_2 P_i$	$P_i \log_2 P_i$
1	0.50	-1.00	-0.50
2	0.25	-2.00	-0.50
3	0.15	-2.74	-0.41
4	0.10	-3.32	-0.33
		$\sum P_i \log_2 Pi =$	-1.74

$$I = \sum_{i=1}^{n} Pi \log_2 Pi = -1.74$$

Entropy

Entropy is a quantitative measure of uncertainty. It is the average information per message interval. The term is used in information theory, to describe the uncertainty surrounding an event. It means that entropy is minimum when complete certainty exists, regarding the actual outcome of an event. Thus, entropy for extremely likely and extremely unlikely messages is zero. Entropy can be represented as follows:

$$H = {}^{I}/_{L}$$

Where, H = entropy average information per message

I = Total information content

$$= \left[\sum_{i=1}^{n} Pi \log_2 Pi\right]$$

L = Number of messages.

Redundancy

Redundancy represents duplication of information, due to the communication or storage of the same information using different messages. Redundancy reduces the efficiency of a particular transmission because more codes are transmitted than required to encode the message; and causes wastage of storage space. It also creates problems in the updating of files. But it is very much useful in error detection and control. Redundancy, in the strict sense, means that the listener need not read and decode every letter in order to understand the message. It is built into data communication systems through parity bits. It can be calculated as a percentage of the information coding capability not being used.

R = 1 - In, Im

Where R = Redundancy

In = Information capacity required

Im = Information capacity used

1 = Constant

For instance, a code using six bits, which can code 64 possibilities, can be used in situation that has only 16 possibilities to differentiate. The redundancy in this case is :

In $= \log_2 16 = 4$

Im $= \log_2 64 = 6$

R $= 1 - 4/6 = 2/6 = 1/3 = 0.33$

Data Reduction, Classification and Compression

Data can be defined as the group of non-random symbols, which represent quantities, actions, objects, etc. Data may be in the form of alphabetic, numeric or special symbols, or in the form of text, image or voice. In order to store and use it efficiently, it is essential to sort the data, and only useful data, that too non-redundant, is entered into the system. This is known as data reduction. It simply means reducing the quantum of data to be stored.

Classification of data involves the division of data into various categories like cost data, sales data, production data, etc. The classification is made on the basis of the needs and requirements of the organisation. The data once classified must be organised and stored in data structures like files, tables, database, etc. When information is required, it should be generated by the processing, summarizing, and filtering of available data.

The information system should be designed to generate only relevant data and to screen out the irrelevant ones. In this way substantial economy can be achieved. Data reduction can be effected in two ways:

(1) Transformation: It ensures data reduction by changing the physical form of data, thereby saving storage space, or by making effective data transmission.

(2) Compression: Compression refers to reduction of physical size of data. It can be effected in two ways. One is to physically compress the data, e.g., a cartridge tape can hold upto 240 MB of data in compressed form. Another method is sampling. Random sample may contain all the characteristics of the population and hence can help in reducing the size of data.

Organisational Summarising and Filtering

Summarising is a process of aggregation, a method for providing information more efficiently. In organisations, message summarisation is used to reduce the amount of data transmission required without affecting the purport of the original message. Formal organisational summarising is illustrated by accounting classifications. For example, in accounting, all sales can be summarised into 'total sales for the period'. Thus, it provides more meaningful information for decision purposes. Aggregation is the most popular form of information summarisation, by combining information originating at higher levels of organisations. Another form is compaction, through which irrelevant data is detected.

Filtering is a process by which irrelevant data is segregated and removed. It is a common method for exercising discretion over message transmission, and is called message modification. If too much information is presented to a person within a short time, he may experience information overload and may not be able to properly perceive the data and information. Information filtering is a technique which trims irrelevant information before it reaches the managers. In organizations, information is filtered at each higher level. As such, each level in the hierarchy functions as a filter station, and exercises discretion as to what information should go up to the superior, and when and how. It permits the managers in exercising control over the information flows in the organisation. Sometimes, it may cause certain problems like some crucial information not reaching the superior if it badly affects the manager.

Inferences

Inferences are drawn from a body of data and are communicated in place of original data. The inferences may be on the basis of quantitative data, such as statistical inference, or may be more subjective, like the anticipated role of trade union in mergers during the coming period. It helps in substantial reduction of data. The problem of inferences is that the user does not know anything about the data on which the inferences are based.

Age and Quality of Information

Organisations collect data and store it in the database on a routine basis. Such data is subject to periodical updating. The periodically updated data has a time factor associated with it. The generation of information from such data takes some time. The length of this time, which ranges from the occurrence of the event and the use of the information about it, determines the age of information. For explaining the concept of age of information, two types of data are defined.

1. Conditional data, which relates to a point in time (e.g., stock at 22/12/1995)

2. Operating data, which reflects changes over a period of time (e.g., stock used during a month)

The causes of delay in generating information from these types of data are information interval and processing delay. Information interval is defined as the interval between reports. It is the time period embedded, like one week, one month etc. Processing delay is defined as the processing time between the end of the information interval and the issuance of the report for use. it is the time taken to convert raw data into information.

Age of information is the sum of the information interval and the processing time. The maximum, average and minimum age of information can be defined with the help of these two factors.

	Conditional information	*Operating information*
Maximum age	d + i	d + 1 ½ i
Average age	d + ½ i	d + i
Minimum age	d	d + ½ i

Where i = information interval

d = processing delay

For conditional information, the minimum age is the period of processing delay. For instance, if the processing delay is 10 days, then the value of stock for March 20 will be at least 10 days old, before it is received on March 30. If reports are prepared weekly (i = 7 days), the maximum age of information is 17 days (10 + 7) . The average age during the period is 13 ½ days (10 + 3 ½).

The average age of operating information is one - half of the interval, the minimum age is d + ½ i, and the maximum age is

d + 1 ½ i.

The quality of information is determined by how it motivates human action and contributes to effective decision-making. Roman R. Andrus suggests that information can be evaluated in terms of the percepections of the decision-maker. He identifies four utilities of information, such as :

1. Form utility: The closer the form matches the decision maker'srequirements, the greater the value.

2. Place utility: Information has greater value if it can be accessed or delivered easily.

3. Time utility: If information is available when needed, it has greater value.

4. Possession utility: The possessor of information greatly influences its value by exercising control over its dissemination.

In other words, the attributes of information which influence its quality can be classified in to there categories, such as:

1. *Time-related attributes including:*

   (a) Timeliness,

   (b) Currency / up to date,

(c) Frequency, and
(d) Time period

2. *Content-related attributes including:*
   (a) Accuracy,
   (b) Relevance,
   (c) Completeness, and
   (d) Brevity.
3. *Form-related attributes including:*
   (a) Clarity,
   (b) Detail,
   (c) Order,
   (d) Presentation, and
   (e) Media

Application of Information Concepts to Information System Design

The mathematical information theory throws light on certain concepts which are valuable for information system design. Such valid concepts include:

1. Value of information: information has value only if it changes decision behavior. It identifies the requirements of the end-user and develops the system design accordingly.

2. Reduction in uncertainty: information influences the selection of alternatives in decision-making situations and thus reduces uncertainty.

3. Redundancy: It is highly useful in detection and control of errors and helps to overcome noise. It ensures correct transmission of messages.

4. Information presentation: information presentation concept like summarisation and message routing are relevant in designing information system.

Conclusion

The concept that information has value only if it alters a decision, provides a useful guideline for information. The existence of noise can be effectively overcome with the help of redundancy. Thus, these concepts provide a useful technique in designing a valid reporting system, in which information is a valuable resource.

Exercise

Short Answer Questions

1. What do you mean by information?
2. Define information.
3. Explain the mathematical theory of communication.
4. Discuss the mathematical definition of information.
5. Define the term entropy.
6. What do you mean by redundancy in information?
7. Explain the concepts of data reduction, classification and compression.
8. What do you mean by filtering of information?
9. Describe the age and quality of information.

Essay Questions

1. Briefly explain the application of information concepts to information system design.
2. Describe the concepts of (a) entropy, (b) redundancy, (c) data reduction, (d) classification, (e) compression, and (f) information filtering.