Unit-1:Measuring of Information

1.0 : Objectives

This Unit introduces the concept of measurement and highlights the problems of measurement and lack of devices of measurement of intangible entities like information. This Unit also focuses on problems of definition in regard to information.

After reading this Unit, you will be able to:

- understand the difficulties faced for developing any operational framework for tackling measurement of information and informational objects;
- grasp the feature of measurement and measuring techniques for different types of entities, phenomena, objects or situations; and
- know the notion of Informativeness and how it can be applied to user-text interaction and database evaluation.

1.1 : Introduction

Lord Kelvin (William Thompson), one of the most important physicists of the 19th century, remarked that there should be written on every laboratory door the phrase - "through measurement to knowledge". The result of a measurement is usually expressed by a numerical value. This numerical value is again a real number. The human nature is such and the socio - cultural development of man is such that every person has a better understanding or insight of something if it is expressed as a number or a set of several numbers. Thus, measurements or numerical values play a vital role in human communication and in man's knowledge systems.

The simplest and most straightforward form of measurement is counting. In counting we get values of what we count in positive whole numbers. We can count objects, which exist as discrete entities, usually, the material bodies. We can talk of 543 books, 20 lacs 92 thousand 5 hundred unemployed persons (20,92,500), 25 words per page, 4.2 Megabytes (MB) of memory, 225 fields of a bibliographic record and so on. It shows that if there are objects, which may be physical or even conceptual, which exist or are available in discrete independent forms or units then we can count them and express the result of the counting as a natural number.

But not everything, with which we deal or are concerned with, is available in discrete unitary form. Water or electricity or performance in an examination or records of our senses such as colour of an object or smell or taste of something; or the amount of information in a document or the knowledge that a person possesses is not available in forms, which allow enumeration or counting. Our aspiration is to have all of these expressed as numbers so that we can understand and utilize them better. Water is a tangible material object but cannot be found as discrete or separable units. So water cannot be counted. Therefore, man has found some properties or so-called parameters with which water can be labeled such as volume, weight, density, viscosity, transparency, etc. So we can talk of 2 liters of water (as volume) or 200 grams of water (as mass or weight) or of refractive index for white light through water as 1.2 or of viscosity of water as 0.01 poise at 20°C etc.

Water is a tangible material object. The measuring devices and the scales, with which measurement of properties of water are expressed, are physical devices and physical scales such as a measuring glass, a weighing machine, a refraction apparatus or a viscometer. But when we want to measure the performance of students through an examination, we want to measure knowledge, understanding and capacity of expression of a student on a particular topic or issue. Knowledge, understanding and capacity to express are not tangible. Yet, examiners give marks say, 40%, 81%, 11%, etc. in numerical values. Therefore, to arrive at such results or measures we need scales and devices. Devices and scales for intangible things cannot be usually physical. They are mostly conceptual set up and logical processes.

Till AD 2000 there has not been sufficiently reliable, objective and universally accepted scales and devices for measurement of information but attempts are there for a long time. To understand the problem we have to understand the problem of understanding the nature of information and of the parameters that can be attributed to various aspects of information and the devices that can be invented and developed to measure those parameters. The crucial problems lie in the matter of finding out suitable units of measurement and inventing or devising suitable parameters.

1.2 : Information Revisited

Already in MLIS *01: Information, Communication and Society,* various approaches to information have been considered. It is clear that: information does not mean the same thing to every body. Yet we accept information as a basic entity. Relations of information with knowledge, wisdom, and data were also discussed. Some of the properties were also enumerated. For our purpose we need once again to present a strategic view of information. We can then deal with the issues of measurement of informational properties, of information content of message or text or discourse or document and of information processes including information service and use of databases and texts.

1.2.1 : Status of Information

Information is everywhere. Saying information is everywhere is like saying energy or gravitation is everywhere. Yet we have a feeling that information has something subtler than other fundamental entities or concepts like energy, matter or force.

Everyone talks of information and apparently understands it, but is not clear how information should be meaningfully defined. Indeed there is no workable general definition of information, which can be applied to all types of information processes and things such as multimedia, hypermedia, books and other documents, texts, messages, symbols, art objects, etc.

One can now supply precise well-formulated definitions to a number of fundamentally important concepts such as matter, energy, number, etc. The answers might not be philosophically fully satisfactory, but they are in most cases scientifically sound. By scientifically sound, we mean that the concept has got an unambiguous, literal description (i.e., through a statement or proposition formed in natural language terms) and the concept goes into some mathematical relations as a functional or as a parameter in relation to other concepts. It is neither too general to be obvious, nor too specialized to be applied or useful only in a single context. In essence we search for reality to a concept and try to establish structural solidarity in a conceptual edifice for that reality. Such a system or program is lacking for information.

1.2.2 : Three Attitudes toward Information

One basic difficulty in analyzing information exists because of three widely different attitudes. First, information is existent or meaningful only in `human' context. Only in human cognition and in human communication `information' is relevant. In other words, there cannot be any physics of information.

Second, information is just an expression or measure of organization or orientation of elements in a set. Suppose some (not all identical) items were arranged in a certain way. It depicts a pattern to an observer. Fay alteration of items the pattern is changed. Information in this view gives just an indication of the change, but not of the pattern or the elements (Communication theorists' approach is this).

Third is the action-oriented view. Information can only be considered in terms of a response to a stimulus. Information gives the idea of action or activity of a system being stimulated by input. The input is the information (Idea of Informativeness is just an extension of this view).

1.2.3 : Information in Discourse

Deer gives the requirements of the nature and property of the concept of information in ordinary discourse for defining its nature as:

i) Information be a representation;
ii) The representation be abstract;
iii) It (representation) be meaningful;
iv) It consists of determinations, which have been made;
v) It should have been made of certain objects.

1.2.4 : Defining Information

Belkin (1978) described information as the `structure' of any text, which is capable of changing image structure of some recipient. Mackay (1969) said informatior does logical work on the organizational orientation of a system.

Following and modifying B. C. Brookes' observation we define `information' as al entity manifesting communicable knowledge, capable of reducing or removing logica uncertainty and switching changes of action and/or pattern of organizational stab of a system. The communicable knowledge or information is semantic in essence however; we use the attribute semantic in its broadest sense for want of a bette term. Communication is sharing meaning. Therefore in information exchange thi most important point is that the receiver understands the meaning (semantic import of a message or text that is comnunicated.)

We say that:

i) For information to exist, communication and so, existence of material system including human beings as generator, receiver, preserver through space and time is necessary;

ii) Any information process i.e., generation, transmission, documentation, receiving etc. is embedded to some form of energetic process, that is, change of energy takes place whenever an information process occurs. But there exists no direct linear relationship between the amount of energy and the amount of information.

1.2.5 : Informetrics and Information Measurement

Informetrics includes bibliometrics as a narrower term and is best manifested by it. It is the tool of quantitative (and to some extent qualitative) accounting of information processes underlying the socio-cultural evolution of man. In bibliometrics we attempt at understanding diffusion of knowledge, growth and decay of invisible college growth of subject fields, productivity of authors, etc. We attempt these through word frequency counts, ranking of journals by counting articles on a topic, counting bibliographic references or counting the number of theorems in mathematics, etc. In doing all these, we are guided by an implicit hypothesis:

`All such items which we try to count are somehow (expectantly linearly related to cognitive knowledge vis-a-vis information per se: '

In effect we use the bibliometric parameters or variables as `information un (IU)'. Choosing a proper information unit is the most important aspect of informetrics. Otherwise we can never be sure whether our bibliometric analysis; can have any relation with cognitive flow of information, which is the basis of sociocultural evolution.

In the recent past there had been some debates about naming the subject, which had been variously called as bibliometrics, librametrics, informetrics, scientometrics, etc. It is now generally accepted that if we use the compound terms ` or the composite term `*sciento-informetrics'* then we can cover all the aspects of the subject. Sciento-informetrics is now applied to documents on internet. This is called *cybermetrics* by some authors. Consequently, we can consider scientometrics, informetrics, bibliometrics, librametrics, cybermetrics, etc. as components of sciento-informetrics.

The essence of sciento-informetrics or so to say of any of its components i.e., librainetrics, bibliometrics, informetrics, scientometrics and cybermetrics is precise measurement. Yet measurements in this area are steeped in ambiguity. Abraham Bookstein has shown the situation prevailing in sciento-informetrics as that of a developing science where measurement is not foolproof or straightforward. He points out that we require theoretical models, which should be able to connect observable parameters to devise methods of measurement to give meaningful results. Such models depend on, at their developmental stage, on intuition and on empirical or theoretical regularities of different variables. Empirical understanding is based on measurements. This understanding helps in model building. Then we rely on the model to guide our investigations towards measurement. That again helps us to develop insight and understanding. This is nothing but a manifestation of spiral of cumulative knowledge (as propounded by Ranganathan and others) and the triad formulated by Karl Popper, the philosopher of science, in his book *Objective Knowledge* as $PP_1 \rightarrow TT_1 \rightarrow EE_1 \rightarrow PP_2 \rightarrow ...$, which means the same spiral of proposition, theoretical modelling, empirical investigation and re-statement of proposition or problems and further development of theoretical model and so on.

So-called informetric laws or sciento-informetric laws of scattering show a set of well-established regularities.

In Lotka's law the number of scientists (N), publishing n papers is proportional to the inverse of

n^a: $N(n) \propto n^{-a}$ *where a* is a constant.

Similarly in Zipf's law the regularity is shown by the rank, r of occurrence of a word in descending order of frequency of use in a text as proportional to the inverse of some power of the frequency, *f* such that constant

Similar also is the case with Bradford's law, which is a relation between the rank (again in descending order) of journals with cumulative number of articles on a subject or topic produced by those journals.

All these three cases can be generalized in some types of regularities of distributions called Zipfian distribution. Unfortunately these regularities cannot be shown as statistical regularities. Bookstein also questioned the concepts that guide data collection in informetrics. What are actually the core journals on which Bradford's law is defined and tested? What constitutes a word? How to define uniqueness of a word? Are specialisation and specialization are two words or one word? Are bay and bays are two words or one

word? Similarly in Lotka's law, how a person is defined as an author or as a specialist in some field? Is a person Chemist because his paper is abstracted in the Chemical Abstracts?

Yet with all these ambiguities and imprecision the regularities are observed with all sorts of bibliographic data sets. It is therefore still an open situation for the question of measurement of sciento-informetric parameters and of the relationship between a group of sources and the items they produce.

1.3 : Framework for Information Exchange

For our purpose we consider that every information system (from human being to a computer to even an atom) has an inherent *K -structure.* At any space - time position, the system is in *a K-state.* Any output of information from the system doe: not necessarily change its K-state, only changes the *Energy State.* But any input of information may or may not change the K-state. K-structural activities within the system may also change the K-state. The 'K-amount' is an abstract subjective .counterpart of `information amount' of a system. K may decrease or increase. The decrease is most easily related to memory loss or forgetting and increase to discover: or gaining new experience in a human system. Example of K-state change b information input is learning.

It is obvious that we have taken `K' as representative of `subjective knowledge or `knowledge within' in an abstract generalized manner. We know that what w know, are not always possible to express or communicate i.e., not always amenable to be transformed in `objective information'. `Objective information' is the `knowledge without' or `communicated knowledge'. `K' is an inherent property of an information system.

Communication is sharing meaning. In other words it is incorporation of information in different K-structures to give rise to change in K-state in an equivalent way. Sharing meaning means the amount of K- state change of different systems ca give rise to the same amount of information and vice versa.

Information to be communicated requires to be coded in symbols transferred through channels or preserved in some material body in the form of messages. Message are embedded in language (we again use such a term for lack of better one). The information can be shared only if the sharers can produce and receive sernantic elements in *atomic information units (AIU or aiu).* AIU's are information-carrying concepts. There may be information units (IU or iu) or semantic elements which may carry meaning of a large string of aiu's. By analogy we may say that IU's at molecular concepts. For the purpose of measurement IU s may be taken as equivalent to the units of information. A unit of information is an item (informational) c convenience. It can be a byte, a word, a document, a part of a document and anything that can have an 1U counterpart.

An iu or aiu may be transmitted through strings of message symbols in some code. I f the symbolic string breaks up or get disorganized the semantic essence of the 1U is lost. And one cannot divide an aiu and preserve the information. Information (and the K-state) is then a particular pattern or organization of information elements based upon organization and pattern (as well as state) of some symbols or representations embedded of material and energetic elements (of molecules, atoms, larger objects, acoustic waves, electro-magnetic (EM) waves, movements of electrons, etc.).

Let us now represent information by I, information units by J, energy by E and K-state by K, measure of any parameter by m, such as m (i), information on or about information by I(I), uncertainty by U, etc. We shall use other symbols in common use with their common meanings.

Information equations and relations may be considered as follows:

1) $$I = |I_1 - I_2| = {}^{*}J_{jq}({}^{*}j_{qr})$$

*representing composition on p and r.

J are the AIU's or core or prominent information elements and j are the aiu's or supporting

information elements. A statement like `want articles on air pollution in Indian cities' would give a composition of core information elements of pollution, air, articles with supporting elements of cities, Indian. Some of the concepts may not be atomic as is cities. Question would be which cities.

2) $m(\Delta I) = in(*J)(*j)$

$= m(a_P\ a_r)$

a_P and a_r representing values of aiu's equivalent to J and j.

The values of a's are to be determined in regard to the pattern of arrangements in some units (by proposition).

3) Output ΔI: $K \rightarrow K$, when ΔK amount is output as ΔI amount of information, such that $\Delta I \geq 0$; there is no change in K even if ΔK is output.

4) Input $\Delta I \rightarrow K \rightarrow K + \Delta K$, $\Delta K \geq 0$, $\Delta I > 0$. This means that when someamount of information (ΔI) is input to a system from outside, its K amount may change if the knowledge is new ($\Delta K > 0$) or does not change if the knowledge is already known ($\Delta K = 0$).

5) $\Delta K = f.(E)$, $\Delta I = g(E)$; $\Delta I > 0$, $\Delta E >> 0$, $\Delta K \geq O$. This means that whenever there is some change in information or knowledge, there is a corresponding change in energy (E) of the system.

6) $m(I(I)) << m(I)$. This means amount or measure of information about information is much less than the amount or measure of information itself. For this reason, indexing or cataloguing or classification is important.

If there is a decisional problem and uncertainty is say, U_1, then

7) $U_1 + \Delta I \rightarrow U2 < U1$

This illustrates that information AI reduces uncertainty from U2 to U1.

We have talked of logical uncertainty because uncertainties inherent in a *system* like that connected with uncertainty principle of quantum physics cannot be *reduced* or removed by any amount of information from or regarding that system.

By logical we have primarily meant `inductive'. Information processing or output is essentially inductive. We should remember statistical accounting is inductive. Accounting of pattern or arrangement is also inductive. Our approach is here only indicative and the research dealing with these issues is at the preliminary stage.

1.4 : Measurement

The subject of measurement is a complex but well-established domain in its own right. Measurement is essential in any science. Without measurement no science can be developed; indeed no useful knowledge is possible without measurement.

Simply defined, measurement is the assignment of numbers to observed phenomena according to some rules excepting any type of randomization rule. In fact measurement enhances information value on any system and reduces uncertainty or randomness.

Defined in a slightly different way, a measurement is a quantified observation o: some attribute or aspect of (parameter related to) an object, process, project o system (and a set of objects, processes, projects or systems). Mathematically speaking, measurement is a correspondence or a one to one mapping from some assignments that may be natural or artificial on an observable through an observation on to the real line (i.e., the real number system). It is important in measurement theory and practice that only those parameters, which are observable, are to b measured. The correspondence or mapping preserves the observed relationship and operations performed on the parameter in the same manner on

the real line. This is called homorphism' or homomorphism' (in mathematical terminology).

1.4.1 : Requisite Conditions of Measurement

Measurement is used differently for different types of phenomena, parameters and contexts. The way we measure time is very different from the way we measuring the action of a drug. Both these are very different from the way we measure the performance of a student in an examination. Again it is different when we try measure (determine) the marketable value or price of a product or of a service.

Any measurement -

i) Is an energetic process; it needs effort and it involves expense of energy. The process of measurement involves a stimulus-response complex; this affects the system under measurement. It generates a fluctuation in energy distribution in the system. In other words, it enhances entropy. This reduces the precision of result of measurement and thus, makes a deviation from expected exact information. Because of involvement of exchange of energy, attempts of increasing sensitivity of the measurement mechanism or of measurement instrument results in increasing noise. The issue of exactitude or precision in measurement is the problem of reducing noise, which is the problem of minimizing the effects of energetic fluctuation, which again means minimizing informative fluctuation.

ii) Interferes with the system on which measurements are made. Therefore the measurement is always imprecise and tells about a condition, which may not permit repetition or second time measurement in any future time. The process of measurement increases entropy.

iii) Is to be made on a well defined parameter with an exact expectation about the outcome of the measurement, not in terms of the `value' or `result' of the measurement but in terms of its `meaning'.

iv) Is to be expressed as a `datum', a value that can be translated or transformed into a numerical expression. Therefore each and every measurement must follow a `code' (encoding process), a `scale', a `unit', a `point of reference' or `origin' or `zero' of the scale; in other words any measurement must have a well-developed `frame of reference'. The scale is usually a linear scale.

v) Requires a device or an `equipment' or a `tool' or a `measuring instrument' and a set of `step by step' measuring process.

However measurements follow some conditions to be performed. Then there are some general facts about measurements. All measurements are: (i) arbitrary to some extent; (ii) approximate; (iii) relative; and (iv) in some sense or other a `validation' and an `evaluation'.

1.5 : Scales of Measurement

A scale is a function or a functional *f* '. Mathematical concept of a scale is a homomorphism between two relational systems. An n-dimensional scale is a homo-morphism in of an irreducible empirical relation system (a,, t_1'..., t_m) into an n' dimensional numerical relation system (R^n, Si, ...S_m) where R^n

is the nth Cartes product of the set of real numbers. The empirical relation system is the assignation, on the parameter, which are to be represented as numerical measurement values.

Keeping aside this abstract (mathematical) approach of an n - dimensional se let us consider a linear or one dimensional scale, that is, when n = 1. A 1 dimensional scale (the special case for n = 1) is a mapping from an empirical to the real number system R. Most of the time we use such a linear scale. Bu times we may use a set of real numbers in order of preference

usually called tuple or in specific cases binary or duplet, triplet, quadruplet or 4-tuples, 5-tup etc. when we need to represent measurements (or vectors) of multi-dimensional representations.

Scales are neither unique nor the same for all purposes, they can be of differ types. The major types are: (1) Category scales which are also called nominal qualitative; (2) Sequence scales - these are usually words or numbers in ordered sequence with uniform spacing between them like as the months of a 3 -January, February, etc.; they are also called ordinal scales; (3) Quantitative sc which can be of many varieties but always give results in numbers; major sub-t} are interval, difference, ratio and absolute scales.

Nominal (or category or qualitative) Scale:

Function `f is a nominal scale if the admissible transformation • from empirica numerical is the set of all one to one mapping from R (real line) to R. Such a s consists of an ordered or unordered series of words and/or numbers, that na identify or describe observable. Each word or number defines a distinct cater that connects one or more entities say for example *good, fair, average, poor* say, *very informative, fairly informative, poorly informative, not informative all.* Such qualitative assignments can easily be changed into numerical nom assignments through some rational weightage e.g., 10, 6, 3, O.

Ordinal Scale:

Function f is an ordinal scale if the admissible transformation is set of monotonically increasing continuous mapping from R to R. The examples are ratings or ranking or numerical rolls such as accession number in an Accession Register or scores which can be expressed first, second, third, fourth, (1st, 2nd, 3rd, 4th), in a competition, say. In many of the sequence scales the numbers are used identification and do not have quantitative significance. Time series and order occurrence scales are the most widely used sequence scales.

Interval Scale:

Function *'f"* is an interval scale if the admissible transformation is the set a positive linear transformation of *'f"* with the functional relation *off(x) = ax* examples are almost all scales of physical measurements like temperature, ti distance, etc.

Numbers are used for denoting the values of the parameter under measurers and not just for identification or tagging.

Difference Scale:

Function *'f"*is a difference scale if the admissible transformation is the se mapping differing from *f* by a constant: *G (f') = f (n) + C.* It is sometimes ca step up scale. A difference scale must have a unique unit of measure but may I an arbitrary origin.

Function *'f"* is a ratio scale if the admissible transformation of the set of mapping differing from *f* by a positive multiple which is called similarity: *f(x)* = ax. Examples may be sizes or weights or such expression as `three times more informative', etc. Ratio scales have a unique origin but an arbitrary unit of measure. Zero volume or zero weight is unique as origin but a glass of water may be x milliliters or y drums or w grams or v ounces. Similarly, zero informativeness or zero relevance (not relevant at all) may be true in a unique sense in context of information retrieval or informational assessment; but on the contrary a document may be considered as of relevance weight zero point five (0.5) or zero point seven (0.7) or zero point three (0.3) depending on the nature of ratio scales used for assessment.

Absolute Scale:

Function f' is an absolute scale if the admissible transformation of the set of mapping is the identity transformation *f(x)* = x. This is the case of all types of counting. There is arbitrariness in case of counting. In the context of information measurement in most cases, the first step is counting such as, number of words, number of sentences, number of articles, number of books, etc

1.6 : Measurement Techniques

Techniques of measurement differ widely for different parameters in different subject areas. In physical science usually a physical scale and a physical device is used. Results of measurements are usually observed optically. In most cases the records are manifested as changes in length or in angle through a graduated instrument. The temperature is recorded from thermometer represented by the length or height of a mercury column. The time is recorded on a circular scale of a watch or clock by recording the angle covered by the hands of the clock. In case of measurement of volume we use sometimes graduated cylinders or pots, the graduation .represents a length. Now-a-days, however, numerical values of measurements or records can be shown directly as numbers on a display device, such as a liquid crystal device (LCD) which are called digital scales or digital devices.

Records of measurement are also obtained from comparison or equivalence or equality by standardizing a unit of the observable parameter. Volume or weight of liquids is measured in this way. In the first case a pot or vessel with a standard measure is used to compare volumes by pouring liquid into that vessel. In a common balance weights are compared to a standard weight say one kilogram or one gram.

1.6.1 : Utility of Measurement

The most important use of measurement is for Identification; but is also important for Comparison, Value judgment, Classification and knowing better. All these are very important for measurement of information because it depends upon the information state and structure of the system and upon the user's capacity to decode the message embedded in the informational item. In sciento-informetrics the utility of measurement can be best understood by consideration of exploiting o bibliographic data.

Bibliographic data represent socio-cultural output. Folk information, which is unrecorded and communicated orally, becomes bibliographic data if recorded or documented Manipulation of bibliographic data can therefore lead to in understanding of sock cultural activities. So far bibliographic data have been applied to the scientific and technological activities through bibliometrics. We have a vast field opening before us for collecting and processing other types of bibliographic data for understanding of different types of socio-cultural activities.

1.6.2 : Counting and Measurement

Counting is also called enumeration. Counting is only possible if items under counting or the things to be counted are available as a set of discrete objects. One may count human beings or a set of grains or stones but can not count a glass of water. For measurement there must be a device of measurement and a scale explicitly defining the unit of measurement. Counting is the most primitive way of measurement. In sciento informetrics all types of bibliographic data are mostly counted.

1.6.3 : Classification and Counting

Unless something is available as a unitary object with a certain property it cannot be counted. Counting is possible only when the observable property is uniquely defined over a set of countable items. When a sub-class of species is formed from a generic class it is formed on the basis of some common property or properties defining the species. In that case by counting the items we get a measure of the speciating parameter. Thus, we can count cows because we can form a sub-class of cows with common, property of cow-ness using it as a speciator from, the class of animals. In a similar manner we can count words

or sentences -or documents. Thus, counting of an intangible qualification or entity can be possible if it ha a manifestation through countable observable. Therefore, many intangible entities in be counted because they manifest through countable symbols or they can be artificially represented through observable and countable manifestation.

1.7 : Standardization of Measurement

The usability of any measure or measurement depends upon a uniform practice, which can be established and ensured through standardization at international. and national levels. The standardization covers the point of origin or the zero in the scale, unit and sub-units. The scale of the dimension if it is needed (i.e., if the measurable parameter is not a pure number), the equivalence or exchange values with other scales and other units (if required) and techniques and devices of measurements are to be specified. Thus, for the temperature there is a standard absolute zero, the scales of measurement such as Fahrenheit and Centigrade: have established transformation rule from one to the other. A thermometer is to be tested for its effective working and keeping the standard. In case of information measurement we do not have any standard practice as yet.

1.7.1 : Accuracy and Stability

The first condition of any measurement is that it must be accurate but we have already noted that all measurements are in some way or other approximate. This means accuracy or precision in measurement needs to be considered in the background of purpose, content and the nature of measurable parameter. When one talks of the distance from one place to another, say from railway station `A' to railway station one expresses the distance in kilometers knowing fully well that the more accurate or more exact figure may be few meters more or less, but when one has to talk of the distance between two mirrors in a precision optical measurement the distance may be needed to be expressed in the order of accuracy of $1/10^9$ meters. Even then there will be an inaccuracy beyond 10th place of the decimal in the measurement. We, therefore, talk of the limits of error or so-called deviations almost usually expressed by standard deviation or standard error. So, we express the speed of light in vacuum as 299792 4 kilometers. This means limits of accuracy is 4 kilometers more or 4 kilometers less than the stated value which again means the exact value may lie between 299788 and 299796 kilometers.

Coming to the cases of dealing with social or intangible parameters the situation is much worse and imprecise. Because of this we usually talk of `true' score and `observed' score. It is a fact that nobody can have a direct guess about the value of the true score. From different measurements of the same `state' one can get an idea of a true score around which the observed or measured scores may be available or obtained. Take the case of evaluation of answer scripts. A student S_1 answering a set of questions Q gets 52 in a 101-point scale from an examiner E_1. The same student writing answers to the same set Q at another point of time may get something different from 52 from the same examiner E,. Another examiner E_2 may award still another marks to S_1, against answers of Q. Only with a good number of trials we can arrive at an average score which may be considered as the best approximation to the true score. With such trials we shall also be able to, estimate the limits of error or deviation by calculating the standard deviation of all the scores, which is usually called the standard error.

Another way of approaching a precise or accurate value is to device a method of arriving at results of measurements in an exponential manner either in the descending or ascending order or both. These are done mostly through a technique called iteration.

Repeatability of observation is another pre-condition to precise measurement. This requires a stable environmental condition and a stable response. This is most important when we want to observe or measure something on the basis of human activity or human response. Answers to the same questions put to the same person in different manner or at different times but within the framework of same context should bring forth the same answer. This is the condition of stability. If the answers differ then confusion may arise. The two answers may be independent or interdependent depending on the person's memory, understanding and outlook. This point is important when we want to assess or measure the effectiveness of information services or when we try to assess the information content of a document through responses from different persons.

1.7.2 : Reliability and Validity

Reliability is the condition that the technique of measurement should give the same value, which is dependable in different situation and at different points of time. In many cases we can measure reliability through the judgment of error between 0 and 1. In all cases of statistical estimates reliability can be expressed through the ratio of true score variance to the observed variance i.e. It = st^2/sx^2 for measurement of a variable (or parameter) x.

A measure is valid if it can be assessed in relation to an observed measure. It is intuitively clear that no measure can be valid without also being reliable but a reliable measure is not necessarily a valid one. Measurement validation is a loop consisting of the functions or activities of `define (or specify) - collect - present - use -' and their feedback upon each other.

1.8 : Information Objects and Content

M. Buckland observed that information can exist or rather information can be approached in three ways information as thing; information as process and information as knowledge. When information is recorded, the record or its components exists as information as thing. This may be called informational objects. In most cases such objects can be counted. Typically enough information objects range from a whole document like as a book or CD-ROM to an alphanumeric symbol or bit. In many of the information exchange content amount of message which is considered to represent information is measured through the number of such objects. We talk of 32 megabytes of computer memory, or 3000 words for an article or 800 papers in informetrics, or 125 books issued from a library desk in a day. Such measures are considered related to the amount of information in one context or other.

1.8.1 : Information Content

It is easily understood that the number of informational objects by counting can not provide a true picture of the information content in any such object or record or message. We therefore consider something, which we call semantic value, is the amount of meaning that is carried by an informational object. For a record R, a document Ranganathan used the word `thought content'. Here, lies the difficulty. Till now we don't have any unit for measuring semantic value or thought content. In natural languages the word is usually the lowest unit carrying semantic value but the same concept can be expressed by different words in different languages and even in the same language. Thus, the problems of synonymous or equivalence terms create problem in using words as units of information. More importantly, when words are combined to form sentence or

sentences are combined to form texts or message there is no uniform rule of combination of the meaning or the semantic values. Most importantly in such combination semantic values or information content are never ad S_1ditive. Indeed they do not follow any uniform rule so that any method of measurement scale can be applied taking the word or the sentence or any component of the text as unit. Classificationists attempted a way out by taking the universe of knowledge as a whole and assigning subject components in a hierarchical order to represent thought contents of documents. But the classification number can only represent the nature of the thought content of a document but not the total amount of its information content and the total amount of semantic value contained in a document

1.8.2 : Range of Information Content

There are many debates about what should be considered as information content in case of transmission or exchange of information. Should we take into consideration each and every component of a message or a text including all the details, redundancies and exemplification or should we consider an indicative usable package from which the recipient of the information would be able to work out the details.

Let us take for example the subject of plane (two-dimensional) classical Euclidean geometry. A sender may send statement of all the axioms and theorems working them out with solved problems and examples. On the other hand the sender may send to the recipient only the axioms and statements of some of the theorems and the methods or rules of solving the theorems and developing the subject. Are these two messages informationally equivalent? If so, there must be some way of showing this equivalence. Unfortunately, we don't have any means so far. There is also a similar problem when we use conceptual contraction. In the development of the subjects it happens often that a new term is coined to represent some idea or concept which required a number of terms to represent the same concept or idea earlier. In such cases the single term and the whole chain of description have equivalent information content. To overcome these a number of attempts have been made in terms of response to a stimulus, reduction of uncertainty, change of informational status or state, assessment of informativeness, etc. some of which will be discussed in latter sections of this Unit.

1.9 : Informativeness

Tague-Sutcliffe states that information is an, intangible aspect of interaction between a text and a reader. The way a reader benefits from a text can be used for comparing the usefulness, hence information content of a text for a reader or for different readers. One may say that a certain text P is more informative than another text *Q* for a reader R,. Similarly, the text *P* may prove to be less informative to a reader R2. The view of informativeness as a measure of information is dynamic and is similar to the concept or view of relevance presented by L. Schamber, M. B. Eisenberg and M. S. Nellan. According to them relevance is a dynamic concept and depends on the judgments of users on the quality of the relationship between information provided by the information item and information need felt by the user at that particular point of time. Degree of relevance can be used in quantitative terms if it is approved from the user's point of view both conceptually and operationally. Thus relevance can be related to the amount of ^*information wanted'* and ^*informativeness'*. A text, which is more relevant for a particular user, should be more informative to that user. If measurableness of information depends on such a concept of informativeness then it is actually an approach for measurement of information services. An information service and hence measurement of informativeness is also somewhat contextual. Ultimate user interaction with text depends on how the collection in the database has been developed, how an item of collection and the collection as a whole have been organized and described (indexed), how the text or information items are retrieved and how

they are packaged and repackaged and supplied.

From the foregoing discussion it is understood that a text is informative to a user only when it adds information to the knowledge store of the user (K + • $I \rightarrow$ K> K in the framework presented at section 1.3). Therefore, informativeness can never be negative but it can be zero. When it is zero, the user considers the text as non-relevant, useless or uninformative.

So we can say -•

Property 1 - Informativeness is a non-negative number associated with interaction of records with user. Informativeness of a record or a text may vary from user to user.

Property 2 - Informativeness can not be measured directly. However, user's reference for ranking of records relative to the amount of information. preserves any ordering of informativeness values. So a user may be asked to help prepare a rank order list of informativeness for a number of texts, documents or records.

Property 3 - Informativeness is not necessarily commutative or additive under concatenation. Concatenation means different records or texts are read or used in sequence (i.e., one after another). The order in which records are concatenated (sequence for interaction or use) may affect their informativeness ranking.

Assumptions :

Let us consider that

1) there are two texts T1 and T_2 which are related through content or topic;
2) there are two clients C_1 and C_2 with nearly equal background information state in respect of the topics of texts $T_1$and T_2. Let this information background state be j;
3) C_1 is given the text T_1, and T_2 in this sequence i.e., T_2 is given after T_1 is read. But C_2 is given the text in alternative sequence i.e., T_1 after T_2 is read.

If C_1 and C_2,are asked to give judgments about informativeness I of T_1 and T_2. Will their judgments be same or similar ? The answer is no.

There is no cornutativity in concatenation. If the texts are of different subjects, then pure additivity or additiveness may be valid in concatenation.

$I(T_1), 1(T_2) \sim I(T_2), I(T_1)$ [if the two texts are on similar topic].

There is no cumulativeness in concatenation. If the texts are of different subjects, then pure additiveness may be valid in concatenation. $I(T_1) * I(T_2) = I(T_2) * I(T_1) = 1(T_1) + 1(T_2) = I(T_2) + I(T_1)$, if the two texts

are of different topics and both topics are unknown to C_1 or C_2.

Property 4 - The total informativeness of a record or a document is invariant under granularity or partitioning in different ways. Partition means dividing into disjointed segments. This property asserts that sum of the informativeness of the partitions is equal to the informativeness of the whole record (or document). But the sequence of the partitions should be maintained. Any breaking or alteration of the sequence of partitions does not guarantee this property. This condition is called *subsequence.* The partitioning or granularization can be taken down to the level of sentences.

Property 5 - When records are ordered according to non-increasing user reference i.e., the more informative record is given lesser rank than the less informative one; in case of ties the ranks may be arbitrarily chosen, the informativeness of a subsequence is approximately proportional to the logarithm of the number of records in the subsequence.

This indicates that effective informativeness of documents or records goes on decreasing with the use of documents i.e., acquiring of knowledge: There is a fallacy in this proposition. Unless the records or documents have related topical content/thought content, decreasing nature of informativeness can not hold good. However, for most specialist users

probability of acquiring new information decreases or in other sense probability of encountering already acquired or known information increases more or less exponentially. If informativeness decreases exponentially (or equivalently in a power law sequence) then logarithm straightens out the index (power) and gives rise to direct proportionality or linearity. That is when $I(T_a)$, ¥ k^n (or, a.exp bn); then log $I(T)_a$ = c.n where $I(T)$, is the informativeness of nth text and k, a, b, c are arbitrary constants.

Property 6 - The informativeness of the response of an information service to a user query is related to the completeness and ordering of the records perused in the retrieval process with respect to those in an ideal chain of records where this chain represents the records that may completely satisfy user information need.

"Queries can be categorized in terms of the number of records in the ideal chain, the number of alternative ideal chains that could be provided, and the extent in which the records and ordering of the ideal chain will vary from person to person with the same query". Queries requiring short actual answers (fact-finding or short-range questions) will have short ideal chains and will vary very little from one user to another. On the other hand ideal chain for a major research study or a new knowledge area may be very different *for* different users.

Property 7 - The informativeness of an information service organization for a user community (actual or potential) is a measure of the extent to which the service satisfies the information needs of the community. It is a function of the degree to which the community needs are served by the organization and the timeliness and informativeness of each service.

Note : The informativeness measure must be capable of aggregation (accumulation and cumulation) over the records and users if it is to serve *as* a basis for evaluation of information service organization.

1.9.1 : Use of Informativeness Measure

Tague-Sutcliffe mentioned and illustrated uses of informativeness measure in lot areas of :

1) *Collection development* and assessment of information content of a collection or database. This can be done by comparing user assessment of informative values of different databases or for the same database at different stages of development. One may also assess the amount of information that a single document may provide to the user community over a period of time and the information, the document may be expected to provide for a certain length of time in future.

2) *Document description :* The measure may be used for assessment of effectiveness of descriptions of documents for different groans of users. The description includes all varieties of indexing, abstracting, classifying, cataloguing, extracting, etc.

3) *Retrieval processes to* estimate the information of the sequence of documents or records retrieved using a particular search strategy for a particular user for a particular need at a particular point of time. Simultaneously one can use the measure to assess the amount of useful or potentially pertinent information, which remain non-retrievable or missed. Retrieval efficiencies of different search strategies may be assessed.

4) *Repackaging and consolidation* of information can be evaluated against usefulness and against the original items of information. This may be an important application in the new development of knowledge management or content analysis and synthesis in cyberspace

1.9.2 : Observations

1) Informativeness of a text is a relative measure (not a universal or an absolute measure).

2) Informativeness values of texts depend upon:

i) topical content or thought content of the texts;

ii) informational state or knowledge state or K-state of the clients interacting the texts;

iii) concatenation or sequence of interaction of the texts by the clients;

iv) the amount I(T) (which can be hypothesized but cannot be measured or evaluated mechanically or directly) that is information content or thought content or semantic content of the texts;

v) the commonality or disjointedness of I(T)'s of texts;

vi) the informativeness I of texts cannot be additive unless I(T)'s are disjointed.

C_1: $K_{.1}$-- $I(T_1) \leftarrow K_1 + [\Delta K]$

ΔK represents the absolute information amount of Text T_i.

1.10 : Appendix

Algorithm to Determine the Informative Chains (This algorithm is taken from Tague-Sutcliffe, 1995).

**Input N and array PREFER

*N is the number of records in the informative set.

* The input array PREFER is defined as follows :

PREFER (I, J) = 1 if the Ith record is more "
informative than the Jth record,

= 0 if the Ith and Jth records are equally informative,
= -1 if the Jth record is more
informative than the I th record,
= 9 if the Ith and Jth records are imcomparable.

INPUT N
INPUT ARRAY PREFER

```
**Determine array NUM

*NUM (I) is the number of records I such that the Ith record T(I) is more informative than T.

*Q1 is an index that determines if NUM has to be recalculated because of use] revisions to
preferences

Ql=I
REPEAT TO LABEL (2) UNTIL Q1 = 0 Q1 =0
ARRAY NUM = ZERO
FOR I = 1 TO N
        FOR J = I TO N
        FOR J=I T O N
IF PREFER (I, J) = 1 THEN NUM (I) = NUM (I) + 1
        NEXT J
NEXT I

**Determine pairs for which NUM is not an ordinal function
*NON contains the non-transitive triples if NUM is not ordinal
*M is the number of such non-transitive triples


M=0
FOR J= l TO N
        FOR K=1 TO N
                IF {PREFER (J,K) = 1 AND N(J) > N(K)}
                OR {PREFER (J,K) = -1 AND N(K) > N(J)}
                OR {PREFER (J,K) = 0 AND N(J) = N(K)}
                OR {PREFER (J,K) = 9
                THEN
                        GO TO LABEL (1)
                ELSE
                        FOR H= 1 TO N
                        IF {PREFER (J, K) = 11 AND J.< > H AND K < > H
                        AND {PREFER (J, H) = -1 OR PREFER (J, H) = 0}
                        AND {PREFER (H, K) = -1 OR PREFER (H, K) = 0} THEN
                        M = M + 1
                        NON (M, 1) = J : NON (M, 2) = K : NON (M, 3) = H
                        END IF
                        NEXT H
                END IF
LABEL (1) NEXT K
NEXT J
**Revise preferences to satisfy negative transitivity
IF M > 0
THEN
        Q 1 = 1
        PRINT "PREFERENCES FOR THE FOLLOWING RECORDS ARE NOT
        CONSISTENT. PLEASE REVISE"
        FOR J = I TO M
                FOR K = 1 TO 3
```

```
            A = NON (J, K)
            S = K M O D ( 3 ) +  1
            B NON (J, S)
            PRINT A, 13, PREFER (A, B)
      NEXT K
            FOR    K= I TO 3
                  A = NON (J, K)
                  S = K MOD (3) + 1
                  B = NON (J, S)
                  PRINT ‘ ‘ PREFERENCE FOR : ` ` ; A; ` ` VS. ‘ ‘; B
                  INPUT PREFER (A, B)
                  IF PREFER (A, B) = 1 THEN PREFER (B, A) = -1
                  IF PREFER (A, B) -1 THEN PREFER (B, A) = 1
                  IF PREFER (A, B) = 0 THEN PREFER (B, A) = 0
                  IF PREFER (A, B) 9 THEN PREFER (B, A) = 9
            NEXT K
      NEXT J
END IF
LABEL (2) CONTINUE
**p Calculate the array OUTT
*The array OUTT (I, J) is calculated as follows
```

OUTT (I, J) = 1 if the I-th and J-th records are comparable and have yet to be assigned to a comparable set, 0 otherwise.

```
FORI= 1 TON
      FOR    J = 1 TO N
            IF PREFER (I, J) = I THEN OUTT (I, J) = 1 : NUM (I) = NUM(I)+1 IF
            PREFER (I, J) = 0 THEN OUTT (I, J) = 1
            IF PREFER (I, J) = -1 THEN OUTT (I, J) = 1
            IF PREFER (I, J) = 9 THEN OUTT (I, J) = 0
      NEXT J
NEXT I
**Sort NUM in descending order.
ID(I) is the original number of the Ith sorted record
FORK = 1 TO N
      ID(K) = K
NEXT K
FORK = N TO 2 STEP -1
      F O R J = K - 1  TO 1 STEP -1
      IF NUM (1+1) > NUM (J)
      THEN
      TEMP 1 = NUM (J) : TEMP 2 = ID(J)
      NUM (I) = NUM (J+1) : ID (J) = ID (J + 1)
       NUM (J+1) = TEMP I ID (J+1) = TEMP 2

            AND IF
      NEXT J
NEXT K
**Determine comparable sets
*The matrix COMPARE will be calculated as follows :
```

```
        COMPARE (I, J) = 1 if the Jth record is in the Ith comparable set,
                       = 0 if the Jth record is not in the Ith comparable set.
*P is the number of the comparable set currently being created.
*Determine the first comparable set, containing the record ID(1).
P = 1
K1 = ID(l)
COMPARE (P, K) = 1
OUTT (K1, K1) = 0
F O R J = 2 T O N
                K2 = ID(J)
                IF PREFER (K1, K2) = 9
                THEN
                        TEST = 0
                ELSE
                        TEST = 1
                        FOR K = 1 TO J - 1
                        K3 = ID(K)
                        IF COMPARE (P, K3) = 1 AND PREFER (K3, K2) =9 THEN
                        TEST = 0
                NEXT K
        END IF
IF TEST = 1
THEN
                COMPARE (P, K2) = 1 FORK= 1 T O J
                K3 = ID(K)
                IF COMPARE (P, K3) = 1 THEN OUTT (K3, K2) = 0 :
                OUTT (K2, K3) 0
        NEXT K
ELSE
        COMPARE (P, K2) = 0
END IF
NEXT J
**Iteratively determine if all pairs have been assigned to a comparable set. If not, create
another comparable set.

F O R I = 1 T O N
        Kl = ID(I)
        FOR J= 1 TO I
                K2 = ID(J)
        IF OUTT (K1, K2) = 1
        THEN
                P = P + 1
        COMPARE (P, ICI) 1 : COMPARE (P, K2) = 1
        OUTT (K1, K2) = 0 : OUTT (K2, K1) = 0
        FOR H = 1 TO N
                K3 = ID(H)
                IF PREFER (K1, K3) = 9 OR PREFER (K2, K3) = 9 THEN
                        TEST = 0
                ELSE
        TEST = I
        F O R K = I T O H - 1
```

```
        IF COMPARE (P, K) = 1 AND PREFER (K3, H) = 9
        THEN TEST = 0
        NEXT K
END IF
IF TEST = 1
THEN
        COMPARE (P, K3) = 1
        F O R K = 1 T O J
                K4 = ID (K)
                IF COMPARE (P, K4) = 1 THEN OUTT (K3, K4) =0
OUTT (K4, K3) = 0
                        NEXT K
                ELSE
                                COMPARE (P, K3) = 0
                        END IF
                NEXT H
                END IF
        NEXT J
NEXT I
**Determine chains
*The output matrix EDGE is defined as follows:

EDGE (I, J) = K, K>0, if the Jth record is in the Kth edge of the Ith chain,
            = 0 if the Jth record is not in the Ith chain.
FOR I = I T O P
*Redetermine array NUM for the Ith chain
        F O R J = 1 TON
        NUM(J) = 0
        IF COMPARE (I, J) = 0
        THEN
                NUM (J) = -1
ELSE
        FORK= 1 TON
                IF COMPARE (I, K) = 1 AND PREFER (*J, K)=1
                THEN
                        NUM(J) = NUM(J) + 1
                END IF
        NEXT K
        END IF
        NEXT J
*Sort array NUM in descending order
*ID(I) is the original number of the Ith sorted record
        FOR K = N TO 2 STEP -1
        FOR J = K -1 STEP -1
        IF NUM (J+1) > NUM(J)
        THEN
                TEMP I = NUM(J) : TEMP2 = ID (J)
                NUM(J) = NUM(J+1) : ID(J) = ID(J+1)
                NUM(J+l) = TEMPI : ID(J+l) = TEMP2
END IF
NEXT J
```

```
NEXT K
*Assign records to edges in Ith chain based on value of NUM
NUM(N+1) = -1
R = 0
FOR J = N TO 1 STEP -1
        IF NUM(J) NUM(J+l) THEN R = R+l
        EDGE {I, ID(J)} = R
NEXT J
NEXT I
END
```

1.11 : Summary

Measurement of information is not yet directly possible. This is because there is no suitable working definition of information and there is no effective device for measurement. Informetrics is the subject of studying quantitative aspects of informational processes. Therefore, it is essential to have clear understanding of the measurement aspect of information and informational processes for studying informetrics.

Keeping this in view we have in this Unit re-introduced the idea of information from a number of points of views. We have shown that every information system can be considered to have a K-structure (abstract knowledge or information structure - which may be considered as a sort of hardware and software capable to receive, understand, output information in an information exchange process). Based on the status of information residing on K-structure, the system is in a K-state at a particular time (analogous to active `memory'). K-structure is fixed; K-state may go on changing.

If the information is `input' from outside to the K-structure, K-state. may or may not change depending on the `informativeness' of the `input' and the K-state. In case of output from K-structure, K-state

remains the same (i.e., no loss of memory or knowledge). Through this model we can consider measure of information in a message through atomic information units which are semantic i.e., have unique elementary quantum of meaning (they may be key concepts in a discourse).

We have also made it explicit that informetrics, bibliometrics, scientometrics and® librametrics may be considered as different components (they are not independent of each other) of a compound term `scientometrics and informetrics' or a composite term used by some authors as `sciento-informetrics'. We have also noted that the most crucial part in sciento-informetric studies is the determination of `information units'. Bookstein has shown that precise definitions of concepts relating to such units are lacking as yet. The nature of informetrics is that of an early stage of a developing science.

We have then presented definitions of `measurement' in the context of general theory of measurement. Any measurement is assigning a `number' or `quantified' tag to a parameter based on a unit of

measurement. Any measurement is an energetic process, it interferes (through the measuring device) with the system on which measurement is done, and is to be expressed as a `datum'. It is made with appropriate expectation of the outcome in terms of its `meaning' and just not `value'. A scale for measurement is essential. There are various types of scales. They are nominal scales, ordinal scales, interval scales, difference scales, ratio scales, and absolute scales. There are measurement techniques. Much of measurement depends on the technique used. Classification and counting are two simple universal means of measurement. Measurements need to be standardized in terms of accuracy, stability, reliability and validity.

In the last section we have discussed nature of information objects and their informational contents, and then defined the idea of `informativeness'. Informativeness is very useful for measuring effectiveness of information services. Following Jean Tague-Sutcliffe we have

enumerated seven properties of informativeness. In short they constitute the following: Informativeness is a non-negative number that is not usually additive and cannot be measured directly. As informativeness relates to an information object or document or record the total informativeness remains the same even if informativeness of parts are taken separately. Informativeness can be service satisfies the information needs of its user community (clientele). Informativeness measures can be used for collection development, document description (evaluative indexing), retrieval process and repackaging or consolidation of information (indeed for content management or knowledge management).