CHAPTER 3

Bibliometric:
General Introduction
Citation and Citation Indexes
CHAPTER - 3

BIBLIOMETRIC: GENERAL INTRODUCTION

Bibliometric has become a standard tool of science policy and research management in the last decades. All significant compilations of science indicators heavily rely on publication and citation statistics and other, more sophisticated bibliometric techniques. Examples for such compilations are:

- National Scien29-ce Board
- Observatoire des Sciences et des Techniques
- European Report on S&T Indicators
- Het Nederlands Observatorium van Wetenschap en Technologie: Wetenschaps- en Technologie-Indicatoren
- Vlaams Indicatorenboek

In addition, many extensive bibliometric studies of important science fields appeared during the last two decades. Aim of these studies was to measure national research performance in the international context or to describe the development of a science field with the help of bibliometric means (for instance, Braun et al., 1987).[1]

It is a common misbelieve that bibliometrics is nothing else but publication and citation based gauging of scientific performance or compiling of cleaned-up bibliographies on research domains extended by citation data. In fact, scientometrics is a multifaceted endeavour encompassing subareas such as structural, dynamic, evaluative and predictive scientometrics. Structural scientometrics came up with
results like the re-mapping of the epistemological structure of science based, for instance, on co-citation,”bibliographic coupling” techniques or co-word techniques. Dynamic scientometrics constructed sophisticated models of scientific growth, obsolescence, citation processes, etc. These models are not only of theoretical interest but can also be usefully applied in evaluation and prediction.

Beyond policy relevant applications of bibliometric results, there are recently important applications in the context of studying the linkage between science and technology, or applications to related fields such as library and information science and most recently also Webometrics. Examples for the latter ones are the large ongoing projects EICSTES (European Indicators, Cyberspace and the Science-Technology- Economy System) and WISER (Web Indicators for Scientific, Technology and Innovation Research).

Today, bibliometrics is one of the rare truly interdisciplinary research fields to extend to almost all scientific fields. Bibliometric methodology comprises components from mathematics, social sciences, natural sciences, engineering and even life sciences. The following pages will provide a systematic description of the research structure of the field and a detailed overview of the state-of-the-art in bibliometric methodology.
HISTORICAL REMARKS

The Origin of the Name ‘Bibliometrics’

The terms *bibliometrics* and *scientometrics* were almost simultaneously introduced by Pritchard and by Nalimov and Mulchenko in 1969. While Pritchard explained the term bibliometrics as “*the application of mathematical and statistical methods to books and other media of communication*”, Nalimov and Mulchenko defined scientometrics as “*the application of those quantitative methods which are dealing with the analysis of science viewed as an information process*”. According to these interpretations the speciality scientometrics is restricted to the measurement of science communication, whereas bibliometrics is designed to deal with more general information processes. The anyhow fuzzy borderlines between the two specialities almost vanished during the last three decades, and nowadays both terms are used almost as synonyms. Instead, the field *informetrics* took the place of the originally broader speciality bibliometrics. The term informetrics was adopted by VINITI (Gorkova, 1988) [2] and stands for a more general subfield of *information science* dealing with mathematical, statistical analysis of communication processes in science. In contrast to the original definition of bibliometrics, informetrics also deals with electronic media and thus includes topics such as the statistical analysis of the (scientific) text and hypertext systems, library circulations, information measures in
The Pioneers of Bibliometrics

The statistical analysis of scientific literature began almost 50 years before the term “bibliometrics” was coined. In 1926, Alfred J. Lotka published his pioneering study on the frequency distribution of scientific productivity determined from a decennial index (1907-1916) of Chemical Abstracts. Lotka concluded that

“the number (of authors) making n contributions is about $1/n^2$ of those making one;

and the proportion of all contributors, that makes a single contribution, is about 60 per cent.”

This result can be considered as a rule of thumb even today, 75 years after its publication. At almost the same time, in 1927, Gross and Gross published their citation-based study in order to aid the decision which chemistry periodicals should best purchased by small college libraries. In particular, they examined 3633 citations from the 1926 volume of the Journal of the American Chemical Society. This study is considered the first citation analysis, although it is not a citation analysis in the sense of present-day bibliometrics.

Eight years after Lotka’s article appeared, Bradford (1934) [3] published his study on the frequency distribution of papers over journals. He found that
“if scientific journals are arranged in order of decreasing productivity on a given subject, they may be divided into a nucleus of journals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus when the numbers of periodicals in the nucleus and the succeeding zones will be as $1 : b : b^2 \ldots$”

Zipf (1949) [4] formulated an interesting law in bibliometrics and quantitative linguistics that he derived from the study of word frequency in a text. According to Zipf $rf = C$, where $r$ is the rank of a word, $f$ is the frequency of occurrence of the word and $C$ is a constant that depends on the text being analysed. It can be considered a generalisation of the laws by Lotka and Bradford. He formulated the following underlying principle of his law although he has never shown how this principle applies to his equation.

"The Principle of Least Effort means... that a person...will strive to solve his problems in such a way as to minimize the total work that he must expend in solving both his immediate problems and his probable future problems...." (Zipf, 1949).
STATISTICS.—The frequency distribution of scientific productivity.


It would be of interest to determine, if possible, the part which men of different calibre contribute to the progress of science.

Considering first simple volume of production, a count was made of the number of names, in the decennial index of Chemical Abstracts 1907–1916, against which appeared 1, 2, 3 . . . entries. Names of firms (e.g. Aktiengesellschaft, etc.) were omitted from reckoning, since they represent the output, not of a single individual, but of an unknown number of persons. The letters A and B of the alphabet only were covered. These were treated both separately and in the aggregate, with the results shown in the table and in figures 1 and 2 below.

A similar process was also applied to the name index of Auerbach’s Geschichte der Physik (J. A. Barth, Leipzig, 1910) which cover the entire range of history up to and including the year 1900. In this case we obtain a measure not merely of volume of productivity, but account is taken, in some degree, also of quality, since only the outstanding contributions find a place in this little volume, with its 110 pages of tabular text. The figures and relations thus obtained are shown in the table and in figures 1 and 2.

On plotting the frequencies of persons having made 1, 2, 3 . . . contributions, against these numbers 1, 2, 3 . . . of contributions, both variables on a logarithmic scale, it is found that in each case the points are rather closely scattered about an essentially straight line having a slope of approximately two to one. The approach to this ratio is particularly close in the case of the data taken from Auerbach’s

Figure 1 First page of Alfred Lotka’s famous article on scientific productivity of chemists
Relatively little attention has been paid to these results. The causes for this phenomenon are three fold as:

1. These papers appeared when traditional methods of information retrieval were still sufficient;
2. they applied to different phenomena and the interrelation between these laws which was not completely recognised; and
3. financing systems for scientific research did not yet stand need of quantitative or even sophisticated statistical methods.

This situation dramatically changed when Derek de Solla Price published his fundamental work in bibliometrics.

**Bibliometrics Since de Solla Price**

In his book entitled “Little Science – Big Science” (1963), [5] Derek de Solla Price analysed the recent system of science communication and thus presented the first systematic approach to the structure of modern science applied to the science as a whole.

At the same time, he laid the foundation of modern research evaluation techniques. De Solla Price work was more than pioneering; it was revolutionary. Time was now ripe for the reception of his ideas since globalisation of science communication, the growth of knowledge and published results, increasing
specialisation as well as growing importance of interdisciplinary in scientific research reached a stage where scientific information retrieval began to fail and funding systems based on personal knowledge and evaluations by peer reviews became more and more difficult.

At that time, most basic models for scientific communication were developed. Among these are first models for essential concepts in scientific communication like growth and ageing of information. Literature and information was assumed to grow exponentially, but in individual research disciplines the growth can also be linear or logistic. Finally, the logistic model has been widely accepted since both exponential and linear growth can be considered special phases within the logistic model. The concept of ageing or obsolescence is intimately linked with the growth of science. In information science and bibliometrics, changing frequency of citations given or received over time is assumed to reflect ageing of scientific literature. Some authors have downright considered growth and obsolescence inverse functions, the faster growth of literature in a field, the faster it ages and the literature becomes obsolete in a shorter time (Brookes, 1970, Egghe, 1993, Kärki and Kortelainen, 1998) [6, 7, 8]. Consequently, an exponential model has been proposed for the ageing of literature in the beginning, too (Wallace, 1986, Price, 1963) [9,5]. In particular, the model of radioactive decay has been adopted. Later on, more complex models have been developed (see, Glänzel and Schoepflin, 1995, 1999, Egghe, 1993) [10,11, 12].
Goffman and Nevill have introduced the theory of intellectual epidemics as a model of scientific communication in 1964 [13]. According to this model the diffusion of ideas in a population of scientists could be compared to the spreading of an influenza virus in a population of people, causing an epidemic. This model can be used both, to describe the spread of the disease and to predict the time when the disease reaches its peak, after which it is presumed to decline. The advantage of the model lies in its predictive power. Goffman and Nevill proposed that the same model could also describe the spread of information within the scientific community. According to the model, the population can at any time be subdivided into three groups of infected, resistant and infection sensitive persons. If a published article in a specific topic is considered an infection, it is possible to follow the diffusion of the epidemic by counting the number of publications per author and theoretically make a forecast of its future. Communication between authors builds on attempts to distribute ideas aiming at reception of disseminated information and on providing contact between infection susceptible and already infected persons.

Another general theory characterising processes of scientific communication is the principle of cumulative advantage. Price formulated this in 1976 as follows.

“Success seems to breed success. A paper which has been cited many times is more likely to be cited again than one which has been little cited. An author of many papers is more likely to publish again than one who has been less prolific. A journal which
has been frequently consulted for some purpose is more likely to be turned to again than one of previously infrequent use.”[14].

Bibliometrics/ Scientometrics took a sharp rise since the late sixties. In the seventies, when data collection was often still a matter of manual work, the field bibliometrics was, characterised by the personalities of enthusiastic researchers much in the way of a “hobby” to later integrate interdisciplinary approaches as well as mathematical and physical models on one side, and sociological and psychological methods on the other, not speaking of the long tradition of library science. Later on, since the beginning of the eighties, bibliometrics could evolve into a distinct scientific discipline with a specific research profile, several subfields and the corresponding scientific communication structures (publication of the international journal Scientometrics in 1979 as the first periodical specialised on bibliometric topics, international conferences since 1983, the journal Research Evaluation since 1991).

The main reason for this development can be seen in the availability of large bibliographic databases in machine-readable form and the fast development of computer science and technology. This made it possible that metrics of science could be established also outside the USA. First, license fees and expensive CPU time resulted at least in the 80s in severe limitations but the technology of the 90s brought the breakthrough. “On-line bibliometrics”, however, remained a dream.
The funding of large projects seems to have become the regular way of financing research in scientometrics. From “Little Scientometrics” the field has become “Big Scientometrics”. The publication of several comprehensive books on bibliometrics, among others by Ravichandra Rao (1983), Bujdosó (1986), Van Raan (1988), Egghe and Rousseau (1990), and Courtial (1990), [15, 16, 17, 18, 19] may reflect this process. The fact that bibliometric methods are already applied to the field “bibliometrics” itself also indicates the rapid development of the discipline.

The three “components” of present-day bibliometrics

Present-day bibliometric research is aimed at the following three main target-groups that clearly determine topics and sub-areas of “contemporary bibliometrics”.

➢ **Bibliometrics for Bibliometricians (Methodology)**

This is the domain of basic bibliometric research and is traditionally funded by the usual grants. Methodological research is conducted mainly in this domain.

➢ **Bibliometrics for Scientific Disciplines (Scientific Information)**

The researchers in scientific disciplines form the bigger, but also the most diverse interest-group in bibliometrics. Due to their primary scientific orientation, their interests are strongly related to their speciality. This domain may be considered an extension of *science information* by metric means. Here we also find joint borderland with quantitative research in *information retrieval*.
Bibliometrics for Science Policy and Management (Science Policy)

This is the domain of research evaluation, at present the most important topic in the field. Here the national, regional, and institutional structures of science and their comparative presentation are in the foreground. Finally, we will have a look at how bibliometrics/scientometrics is linked with related fields and application services (see Figure 1).

Figure 1

Bibliometrics/scientometrics is linked with related fields and application services
INDICATORS OF CITATION IMPACT

The Notion of Citations in Information Science and Bibliometrics:

In research evaluation, citations became a widely used measure of the impact of scientific publications. There is controversial discussion about how citations should be interpreted. Susan Cozzens (1989) [20] has argued that citation is only secondarily a reward system. Primarily, it is rhetorical-part of persuasively arguing for the knowledge claims of the citing document. Linda C. Smith, [21] stated that

"citations are signposts left behind after information has been utilized”.

Blaise Cronin [22] defined citations as

"frozen footprints in the landscape of scholarly achievement … which bear witness to the passage of ideas”,

but he also referred to certain problems with regard to reference practices as he concluded,

“If authors can be educated as to the informational role of citations and encouraged to be more restrained and selective in their referencing habits, then it should be possible to arrive at a greater consistency in referencing practice generally.”

Problems with citation analysis as a reliable instrument of measurement and evaluation have been acknowledged throughout the literature. Chapman, for instance, delineated shortcomings, biases, deficiencies, and limitations of citation
analysis. Wouters (1997) [23] has devoted a large monograph on citation culture and in 1998, Leydesdorff has initiated the discussion about reappraisal of existing theories of citation.

According to Westney (1998), [24] citations are nevertheless indicators of scholarly impact:

“Despite its flaws, citation analysis has demonstrated its reliability and usefulness as a tool for ranking and evaluating scholars and their publications. No other methodology permits such precise identification of the individuals who have influenced thought, theory, and practice in world science and technology.”

In recent studies, Glänzel and Schoepflin (1999) [25] have citations more pragmatically interpreted as

“one important form of use of scientific information within the framework of documented science communication,”

however, in general, whether form of nor reason for the concrete information use are not specified. Although citations cannot describe the totality of the reception process, they give, according to Glänzel and Schoepflin,

“a formalised account of the information use and can be taken as a strong indicator of reception at this level.”
This statement, though made from the perspective of information science, is completely in keeping with the above evaluation-related conclusion drawn by Westney.

As mentioned above by Glänzel and Schoepflin, whether neither form of nor are reasons for the concrete information use taken into account in evaluative studies. However, there are many reasons for citing publications. Garfield and Weinstock have listed 15 different reasons for giving citations to others’ work (cf., Weinstock, 1971).

[26]
1. Paying homage to pioneers
2. Giving credit for related work (homage to peer)
3. Identifying methodology, equipment, etc.
4. Providing background reading
5. Correcting one’s own work
6. Correcting the work of others
7. Criticising previous work
8. Substantiating claims
9. Alerting to forthcoming work
10. Providing leads to poorly disseminated, poorly indexed, or uncited work
11. Authenticating data and classes of facts – physical constants, etc.
12. Identifying original publications in which an idea or concept was discussed
13. Identifying original publications or other work describing an eponymic concept or term

14. Disclaiming work or ideas of others (negative claim)

15. Disputing priority claims of others (negative homage)

This list is, of course, not exhausting, but some of the above reasons for being cited, for instance, # 5-7, # 14 and # 15, may illustrate that not all given citations point to quality. But even criticism expresses the reception of documented scientific information. Heavy criticism of a certain scientific work can, in a sense, reflect true impact. By provoking constructive criticism, an erroneous theory may even more contribute to the advancement of a science area than some sound average study. On the other hand, papers of a controversial nature will continue to be cited longer.

These examples may just serve as an illustration of the complexity of citation processes. The general discussion of sociological theories of citations is beyond the scope of this introduction to citation indicators. For further arguments and a detailed presentation of citation contexts we, therefore, refer besides the already cited work to the articles by Small (1978,) and Bonzi and Snyder (1991). [27,28]

These reasons listed by Garfield and Weinstock can be categorised, on one hand, as ‘positive’, ‘neutral’ and ‘negative’ and, on the other hand, as relevant, less relevant and even irrelevant or redundant. Figure 4.1 visualises the ‘weight’ of citations.

*The ‘weight’ of citations from the viewpoint of the use of scientific information*
The question arises in how far citations express the use of information and the reasons according to Garfield and Weinstock in reality. Nevertheless, the refereeing system in documented scientific communication guarantees the observance of relatively strict rules of providing reference citations.

It might also be worthwhile to list reasons for not giving citations to a colleague’s work, that is, for not providing reference citations. The first and most important one is lacking relevance of the topic. Irrelevant topics are obviously not cited. Unawareness is the second reason that is due to insufficient retrieval of published information relevant for an author’s research work. Citations omitted by reason of unawareness are sometimes added by referees reviewing papers prior to acceptance for publication in a journal. According to Garfield (1986) [29] these papers not recognised by unawareness are whether

“victims of cryptomnesia, an unconscious plagiarism in which creative ideas expressed as new are actually unrecalled memories of another’s idea, or were superseded for other reasons remains to be seen.”

Disregard has little do with bibliographic amnesia described by Garfield. Disregard is simply a reason that is already beyond the borderline to unethical communication behaviour. Results by colleagues relevant for the author’s research to be published are this way demonstratively ignored. The fourth reason is a consequence of obsolescence as an expression of ‘natural obliteration. Finally, the fifth reason occurs
rather seldom in scientific literature, as it is an expression of the disappearance of the ‘users’ of information. In order words, literature still being relevant in the context of the research topic is not cited because there are no more authors who could cite it. We can consider such topics extinct.

As already mentioned that, the ageing and obsolescence is expressed by changing frequency of citations (received or given) over time. So, what is actually measured by decreasing citation impact? First and above all, decreasing citation impact reflects, of course, obsolescence. This might be an evolutionary process, that is, cited work gradually vanishes from reference lists, and is replaced by more recent and more relevant literature. However, it might be a revolutionary process if a breakthrough has at once made everything obsolete that was so far relevant in this subject. A third form of obliteration occurs if literature is subject to obliteration by incorporation. This means relevant literature is no longer cited because its substance has been absorbed by current knowledge; its content has thus become "common knowledge" (Garfield, 1977, 1986). Garfield (1997) [30] concluded that

"obliteration … is one of the highest compliments the community of scientists can pay to their author”.

The author or inventor has become eponymous. Examples for such eponymic concept or term in bibliometrics are Bradford’s Law and Lotka distribution.
In this context it should be mentioned that all these arguments and reasons refer to *individual* citations. Bibliometrics, however, deals in most cases with citations to larger sets of publications. As in case of other social processes, a difference in the behaviour of individuals and the behaviour of large groups or masses can be detected in citation processes, too. Critics of bibliometric methods argue that citations might be governed by the will and the (sometimes tendentious) intentions of the authors and may thus result in a deliberate filtering of information sources (e.g., “citation cliques”). Individual citing may be influenced hereby, but this phenomenon is certainly not characteristic for the citations to paper sets published by a larger number of authors. Thus, especially reasons # 10, # 14 and # 15 of Weinstock’s list are less characteristic in citations to large publication sets. Citation analyses are based largely upon citation frequency. Several authors in bibliometrics are downright regarding citation frequencies as a “quality measure”. This interpretation is not only narrowing down the possibilities of application of citation-based methods, it may also have undesired consequences at the micro-level, if publications of individuals are studied.

For instance, the fact that a paper is less frequently cited or (still) uncited several years after publication gives information about its reception by colleagues but does not reveal anything about its quality or the standing of its author(s). Uncited papers by Nobel Prize winners may just serve as an example. However,
“if a paper receives 5 or 10 citations a year throughout several years after its publication, it is very likely that its content will become integrated into the body of knowledge of the respective subject field; if, on the other hand, no reference is made at all to the paper during 5 to 10 years after publication, it is likely that the results involved do not contribute essentially to the contemporary scientific paradigm system of the subject field in question” (Braun et al., 1985).
REFERENCES


Citation and Citation Indexes

Introduction

The practice of giving citations or foot-notes is not new, and also has been well established in scientific writings even when the early periodicals started about three centuries ago. Derek de Solla Price [1] has found out that the earliest name of a foot-note was “Scholia”, which means “relating to scholarship”, foot-noting was considered to be a scholarly practice.

In the sphere of science this elevation is reflected in the sentiment of Newton when he candidly expressed “If I have seen anything further that is standing on the shoulders of giants”. This indebtedness of researchers, whether in sciences or social sciences, has found expressions in citations or bibliographic references, accompanying researcher communication [2].

Citation indexing is a method of organizing the contents of a collection of documents in a way that overcomes many of the shortcomings of the more traditional indexing methods. The primary advantage of citation indexing is that it identifies relationships between documents that are often overlooked in a subject index. An important secondary advantage is that the compilation of citation indexes is especially well suited in the use of man-machine indexing method that do not require indexers who are subject specialists. This helps to make citation more current than most subject indexes. Furthermore citations,
which are bibliographic descriptions of documents, are not vulnerable to scientific and technological obsolescence as are the terms used in subject indexes. Citation indexing is based on the simple concept that author’s references to previously recorded information identify much of the earlier work that is pertinent to the subject of his present document. These references are commonly called citations in a given collection of documents. Such lists are usually arranged so that the cited document is followed by the citing document [3].

Origin of Citation Analysis

The first practical application of this concept was Shepard’s Citation, [4] a legal reference tool that has been in use since 1873. Shepard’s Citation owes its existence to the fact that American law, like English law, operates under the doctrine of Stare Decisis. Stare Decisis means all courts must follow their own precedents as well as those established by higher courts. The precedents are the decisions handed down in previous cases.

To try a case under Stare Decisis, a lawyer must base argument on previous decision; however, the lawyers must make sure that the decision has not been overruled, reversed or limited in some way. Shepard’s Citation enables the lawyer to do this with a minimum of trouble.
A legal case is always referred to by a code, which consists of the volume and page number of the documents in which the case is reported. Once a case is permanently reported, its reference code becomes fixed for the time. Thus 301 US-356 is references to the case reported in volume 301 of the United States Supreme Court Report on page 356. Statutes are also referred to in a similar manner. Thus, Ch-16 Sec. 24NJRS refers to the chapter 16, section 24 of the New Jersey Revised Statutes.

Taking advantages of coding system, Frank Shepard devised a listing which contains every instance in which a reported decision is cited in a subsequent case. The listing also shows what statutes and journal cite the original decision.

Using Shepard’s citation, a lawyer must first locate a previous decision relating to his current case. He will do this by consulting a digest, index or encyclopedia which will provide him with the case number of any given decision. The lawyer then looks up the case number in Shepard Citations and finds all subsequent citing cases. From the information, he can determine whether the original decision was affirmed or modified in any way.

Citation – Reason D’ Etre

As the doctrine of Stare Decisis provides the logic for Shepard’s Citation, so did the “reference tradition” provide the rationale for citation indexes for science?
A citation represents relationship between cited and citing document. Although there may be a number of reasons as to why a citing author has not cited certain other works. It may be due to the fact that the author was not aware of the document or he could not obtain it or could not read the language in which it was published. The nature of relationship between cited and citing author is difficult to characterize, however, Moravicsik [5], for the first time, developed into the nature of citations process in the literature of physics and classified citations into following four categories:

i. Conceptual/ operational

ii. Organic or perfunctory

iii. Evolutionary or Juxtaposition, and

iv. Conformal or negational.

**Welch Medical Library Indexing Project**

In 1952, Dr. Chauncey Leake was chairman of committee of consultants for the study of indexes to medical literature. This committee was supervising the John Hopkins Welch Medical Library Indexing Project, which was sponsored by the Armed Forces Medical Library. Dr. Leakey suggested that project workers should examine review articles in connection with their investigation of the problems with subject indexes to medical literature.
Garfield

This statement had considerable impact on Eugene Garfield, one of the Welch Project investors. Garfield [6] realized that nearly every sentence in a review article is supported by a citation to a previous work. Thus, a review article could really be considered as a series of indexing statements. The problem then became one of transforming these statements into a consistent format that would be useful as an index.

Adair

In 1953, the Welch project conducted a symposium, news of which was reported in a Colorado newspaper. This article read by William C Adair, whose was a former Vice-President of the firm that produced Shepard’s Citation. Adair wrote to the Welch project and suggested that they consider the methods employed by Shepard’s as possible indexing techniques.

After examining Shepard’s citations, Garfield realized that the “citations” principle could provide a means of indexing review papers, which could be extended to the scientific literature in general.

After Welch Project ended, Garfield began graduate work in Library Science at Columbia University. During this period he continued correspondence with Adair and began writing a detailed article on citation indexing for scientific literature. The
article was completed in 1954 and was edited and referred by Professor Bentley Glass, who was then chairman of the John Hopkins Department of Genetics and on editorial board of Science.

While his own article was awaiting publication; Garfield, who by then was an associate editor of *American Documentation*, suggested that Adair write a shorter article which would explain his concepts. Adair’s article appeared in *American Documentation* in June 1995; Garfield’s article appeared in *Science* in July 1995.

**Laderberg**

In 1958, Prof. Joshua Laderberg of Stanford University wrote to Garfield to inquire if any further work had been done on citation indexing. When informed of the financial problems involved in starting such a project. Laderberg suggested that Garfield should apply for a grant from the government.

**Genetic Citation Indexing**

In 1961, the National Institute of Health initiated a cooperative programme with Garfield’s Institute for Scientific Information (ISI) to prepare a citation index for the field of Genetics.

Garfield soon recognized, however that defining the genetics literature to be covered by a citation index would be quite difficult. Fine judgments would be required as to what was or was not genetics literature. At Garfield’s suggestion, it
was decided to undertake a comprehensive interdisciplinary approach to preparing a citation index and then extract a genetics citations index from that base of information.

**Science Citation Index (SCI)**

The interdisciplinary database was eventually used to produce the first *Science Citation Index* used, which was published in 1963. The first SCI covered the literature of the calendar year of 1961. It covered 613 journals, contained 1.4 million citations, and required five volumes. Nineteen per cent of the citations in the 1961 SCI data were selected by special computerized procedures as ‘having to do with genetics’ and were published separately as the *Genetic Citations Index*. The Genetic Index was also published in 1963 and was complete in one volume.

**Other Citation Indexes**

In addition to SCI and Genetics Citation Indexes, there have been several other efforts at compiling citation indexes or at using the principle of citation indexing information retrieval systems. Most of these have been experimental in nature, extremely narrow in their converge or published on a one–time basis.

Some citation indexes provide coverage of the material published in just one journal one of the earliest examples of this is the cumulative index to volumes 35 through 50 of the Journal of the *American Statistical Association*. Another example of a citation
index with single journal coverage is the one that appears in a cumulative index to volume 1 through 31 of the Annals of Mathematical Statistics.

An example of citation index that covers more than one journal but is limited to a single field is the Citation Index for Statistics and Probability which is being currently produced by Dr. J.W. Tukey at Priceton University. This project was initiated in 1961 and is being conducted in cooperation with the National Science Foundation. In 1968, the Shepard organization itself introduced Shepard’s Law Review Citations. This new publication indexes 117 law reviews and periodicals and shows where any legal article written since 1947 has been cited in the covered journals from 1957 on.

As a final example special indexes can be prepared on demand through system that directly connects the users with the database, as is done with the Technical Information project (TIP) at the Massachusetts Institute of Technology. TIP uses a time-sharing computer connected to remote consoles by telephone cable. The database consists of the full bibliographies of articles from 25 recent physics journals. Thus, a user can obtain a citation index to all the articles or articles form only one of the covered journals or to the articles in a single volume of a covered journal [7].
Social Science Citation Index

The Institute of Scientific Information Social Science Citation Index provide access to current and retrospective bibliographic information, author abstract and cited references found in over 1700 of the world’s leading scholarly Social Science journals covering more than 50 disciplines.

Arts & Humanities Citation Index

The ISI Arts & Humanities citation Index provide access to current and retrospective bibliographic information and cited references found in over 1,120 of the world’s leading Arts & Humanities journals.

Derwent Innovations Index

The Derwent Innovation Index, available through ISI web of science interface, is a web accessible product that merges the Derwent World patent Index with the Derwent Patent Citation Index. Updated weekly it covers over 10 million basic inventions and 18 million patents in all from over 40 patent issuing authorities. As a part of ISI web of knowledge, Derwent Innovation Index will soon be cross searchable with other scholarly content – ISI Web of Science, BIOSIS receiving ISI proceedings and coming soon CAB abstracts and INSPEC.
Definitions

Citation analysis is a technique of bibliometric study of literature based upon some degree of relationship between citing and cited articles or documents. When one author cites another, a relationship is established. Citation analysis uses citation in scholarly work to establish links. Many different links can be ascertained, such as links between scholarly works, between journals, between fields, or even between countries. Citations both from and to a certain document may be studied. One very common use of citation analysis is to determine the impact of a single author on a given field by counting the number of times the author has been cited by others.

According to Martyn “citation analysis has been defined as an analysis of the citations or references which form part of scholarly apparatus of primary communication”.

According to FW Lancaster “One very important group of bibliometric studies relates to what sources authors cite. Citation analysis is concerned with such phenomena as which authors are most cited, which journal are most cited, and what linkages exist through citation (who cites whom, which journals cite which journals, what subject areas are cited in the literature of a particular discipline, and so on )”. 
According to Garfield “Citation frequency is measure of research activity, or of communication about research activity”.

According to E.C. White “Citation analysis plays a prominent role for easy identification of earlier research”.

According to Broads “True Citation analysis is one which deals with works cited as having actually been used in preparation of, or having otherwise contributed to the source paper.”

According to Baughman “It is a systematic enquiry into the structural properties of the subject”.

Dhalig used citations to trace the diffusion of an idea and pointed out that some papers became socio-metric stars while other is isolated. The frequency of the cited documents appearing in a number of citing articles is, in some measure, and indication of its influence or impact on the subject. The technique of citation is largely statistical and is used for arranging those cited material in some kind of rank to study their relative importance.

A citation, that is to say, is not a unit but an event, and only immediately quantifiable in terms of its frequency of occurrence. For some purposes, the simple frequency may be an adequate measure. Bearing in mind that citation is a practice only engaged in by authors, and that authors of published papers area comparatively small proportions of the total population of scientists and
technologists, then raw citation counting can be used to divide the universe of potentially citable documents into groups, those which have demonstrably been of some undermined relevance to the work of authors, and those which have not demonstrable been of use. Common sense suggests that the most heavily cited material are those which have a higher probability of being cited in future, than those which have not been cited, and consequently journal acquisition policies can be based on raw citations counts as a rule-of-thumb measure, although such counts require considerable modification if they are to be anything other rule-of-thumb.

Historical Developments of Citation Analysis

The development of citation analysis has been marked by the invention of new techniques and measures, the exploitation of new tools and the study of different units of analysis. These trends have led to a rapid growth in both the number and types of studies using citation analysis. Since 1920 the US indexing system of the library literature has recorded a number of citation studies. But the term ‘Citation Analysis’ did not appear till 1956. Many of them attempted to rank the importance of journals on the bases of an average number of references to journals cited in research papers. In 1927 Gross and Gross published an article on the college libraries and chemical education. This was first user study, of any significance, based on a more systematic citation count. It just
attempted to rank the journals in chemistry. The method employed was to make numerical counting of all the journals in chemistry cited in a small representative sample of research papers. The study remains important for its historical significance. Because Stevens formed the Bradford’s Law of scattering by using the findings of this study.

The characteristic of research literature used by chemists and physicists in the US was the next significant study by Fusslar. It was the first study on the basis of ‘Citation Analysis’ as understood. Its main objectives are two fold: First to find characteristics of the literature used by the chemists of American origin, and secondly to rank the journals in order of their use. The methodology adopted was by the counting of bibliographic references cited by the American scientists in their research papers. While counting the citations, the study took note of the form of literature shown in each citation. This helps to rank the journals as one form of literature shown in each citation. This helps to rank the journals as one form of literature, and to study the structural characteristics of the literature of chemistry.

During 1950’s and 1960’s a good number of studies smaller in scope but with greater refinement were reported. The methodology was perfected to a degree of precision. But all of them had one uniform conclusion. These are all supported a general hypothesis that in the field of science only few frequently used journals supplied more than half of the citations while the remaining citations were from the less
frequently used journals. Again, these conclusions are in support of findings of Bradford in an indirect way.

Stevens was the first to undertake a comparative study of the materials cited in Ph.D dissertations. A total of 6993 citations retrieved from a sample set of 90 theses. 50 on science and 40 Humanities formed the total data. The analysis revealed heavy use of journal articles in scientific writings and of books in Humanities; a large number of citations are used in Humanities then in Science.

Stevens did not stop this; he undertook a very laborious task of checking the title against the holdings of the libraries of the representative universities that awarded the degrees. He found that the dissertation on History cited a large number of titles which were not in the library. From these findings he derived the conclusions that the scholars in Humanities have to search and make use of materials available in libraries other than their own.

Linda Smith opines that “the easiest techniques to use in a citation count, determining how many citations have been received by a given document or a set of documents over a period of time from a particular set of citing documents. When this count is applied to articles appearing in particular journals it can be refined by calculating the impact factor, the average number of citations received by article, published in journal during a specified time period. This measure allows one to compare the ‘impact’ of journals which publish different number of articles. Pinski and Narin have developed further refinement of citation count which takes into
account the length of papers, the prestige of the citing journal, and the different referencing characteristics of different segments of the literature.

By reviewing few more articles in the development of citation analysis Linda Smith had devised two techniques, to identify documents likely to be closely related, bibliographic coupling and co-citation analysis. Two documents are co-cited when they are jointly cited in one or more subsequently published documents. Thus is co-citation in earlier documents become linked because they are later cited together. In bibliographic coupling later document become linked because they cite the same earlier documents. The difference is that bibliographic coupling is an association intrinsic to the documents, while citation is a linkage extrinsic to the documents, and one that is valid only so long as they continue to be co-cited (dynamic). The theory and practical applications of bibliographic coupling and citation analysis have been reviewed by Weinberg Bellardo. Citation count and bibliographic coupling were the characteristics citation analysis technique in the 1960’s, but in 1970’s co-citation analysis become the focus of much research activity. Co-citation analysis is of particular interest as a means for mapping scientific specialties’. Use of new techniques in citation analysis had been possible by the availability of new tools. Early citation studies frequently were based on lists of reference found in articles appearing in a small number of journals. Citation has been transcribed and manipulated by hand, because of tediousness of this process most studies were necessarily quite limited in scope. The availability of computer has significantly
improved this situation in two ways; through the production of printed indexes which contain citation data from thousands of documents, and through the analysis of citation data available in machine readable form. Product of the Institute for scientific Information (ISI) now provided a wealth of data for citation analysis. Subject coverage has been expanded form the initial Science Citation Index (SCI) to include the Social Science Citation Index (SSCI) and the Arts and Humanity Citation Index (A & HCI) as well.

Although the discussion so far has suggested counting citations only for individual articles of journals, in fact, various levels of aggregation are possible. If one assumes that citations are indicators of importance, then one can use such analysis to determine, the most important scholars, publications departments and so on in a particular discipline. This assumption is just one of the several which deserves close scrutiny if the results of citation analysis one to be understood.

Observations of Citation Analysis

Many critics and reviewers have questioned both the assumption and methods of many studies found in the citation analysis literature. Even all user studies are also brought to criticism. The user study by survey method is criticized as a very imperfect indication of demand. Citation analysis on the other hand is criticized as an indirect measure of the use of library materials. The validity of citation analysis, as a user study is also challenged. The first reason pointed out by
the critics is the difficulty of defining the term ‘use’. In library terminology the term ‘use’ means readership or borrowing frequency. But the citation frequency has no relation as any one of these two measures. It can be argued that what is borrowed need not have been read, or what is read need not have been cited. Jones and others summaries the drawbacks and short comings of this method, for instance “an author need read what he cites, or cites, what he reads”. Dunlop is another writer who is highly critical of the citation analysis and opinion that the studies based on citation count ‘are only suggestive’. Because they examined the product of research (like journal articles or theses) rather than the process of research. Hence their finding cannot be conclusive.

Martyn had made another criticism of citation analysis. The amount of information cited may vary with the section of the paper in which the citation is made, that is, each reference appended to research writing may not be relevant to the whole topic. A typical paper will have an introduction, methodology, and analysis. For example, quantitatively, a larger number of citations will be found under the analysis part. Again the citations referred under the methodology may have no veering to the analysis section. The citations under the introductory part are likely to refer to the whole of the cited articles, whereas the experimental section makes use of bits of materials from the cited articles. This kind of disproportionate use and scattering is a limitation imposed upon the citation analysis.
Edge D states that at best “the citation analysis plays a peripheral role”. In his opinion, citation analysis using ‘formal communication channels’ and deriving ‘empirical data’ from the relatively trivial behaviour of adding citation to papers make efforts to accumulate and average the behaviour of a scientist, but actually destroys the evidence we need of individual variations.

The major objections raised by Edge are:

1. Citation and co-citation counts are just part of the range of the available experimental data, consisting of the wide range in the totality of information, it is difficult to design relative weight age to citation measures;

2. Co-citations and the maps based on it only the perception of authors and fail to establish communication links;

3. The behaviour of scientists in their scientific results is a complex sociological phenomenon. It provides many problems of greater importance than any correlations derived from a computer pointed out listing citation figures. He says, “it is often because individual scientists, and groups do not share the consensus view, defined by co-citation maps that crucial innovative decisions are made”.

4. The concept of a specialty of discipline is a ‘social construct’ and requires a wide measure of agreement but no detailed consensus on its boundaries. The
composition of specialty or discipline through citation analysis is not at all necessary;

5. Citation analysis are based on formal scientific publication such as journal articles, books etc., but not on the informal publications and communication channels which are more significant. It is normally through those informal channels that precise direction of the development of scientific culture can be determined, and

6. Co-citation studies to date have not generated any striking insight, even heuristically.

May observed many abuses of citation indexing. He does not agree that citations give a fair picture of any intellectual links between publications. According to him they are able, at the most of depicting the pictures that authors record. He attributes the unreliability in results of citation studies to the following factors.

   a. Memory failure on the past of the citing author;

   b. Lack of self awareness in the citing author, the editor of the referee;

   c. Plagiarism by other authors in their citations without he having actually used or seen them;

   d. The prevailing widespread custom of not citing obvious sources. For example, if a physicist mention the word ‘relatively’ in his paper, it is reasonable
assumed that the reader of the paper knows that the term is associated with Einstein, and hence find explicit reference to Einstein’s paper unnecessary. The same in the case with C.V. Raman’s paper on ‘Raman Effect’.

e. Carelessness which may result in giving wrong or incomplete citations and

f. Lack of proper selection. Sometimes the author selects citations to serve his scientific, political or personal goals.

Eugene Garfield expressed his views that “unfortunately, some editors, writers and referees in all fields of sciences are not as concerned as they should be about the proper use of cited references. Some authors fail to cite pertinent papers, or cite for reasons which are frivolous or dishonest. Citations may represent on authors attempt to enhance his own reputation by associating his work with greater works or to avoid responsibility by learning heavily on the work of others. Citation can also be intended to carry favour with influential colleagues or references to honour a mentor or friend (a friend in one who cites in return) or to convey the impression of exhaustive knowledge.”

By reviewing the available observations and limitations on citation analysis Lindasmith states as under:

Citation of document implies use of that document by the citing author. This assumption actually has two parts (1) The author refers to all or at least to the most important documents used in the preparation of his work and (2) all documents
listed were needed, used i.e., the author refers to a document only if that document has contributed to his work. Failure to meet these two conditions leads to Sins of omission and commission’, certain documents are under-rated because not all time used were cited, and other documents are over-rated because not all times cited were used’. With respect to understanding it should be evident to any one who has written a paper that citation does not necessarily fully and faithfully reflect usage. Often what is cited is only a small percentage of what is read not all that is read and found useful is cited. Although the author usually does not provide any evidence of omissions, there are exceptions. Consider a paper by Bottle which has as its references. ‘Reference omitted to avoid embarrassing its authors’. With respect to over-rating, Davies offers a ‘fundamental law of reference giving’, it is quite unnecessary to have read or even seen the reference yourself before quoting it. Without looking at the text of both the citing and cited documents, it may be possible to make a judgment as to whether a particular citation does indeed represent use of material in the cited document.

Citation of a document, author, journal etc., reflects the merit (quality, significance, impact) of that document (author, journal etc.).

The underlying assumption in the use of citation counts as quality indicators is that there is a high positive correlation between the author of citations which a particular document (authors, journal etc.) receives and the quality of that document (author, journal etc.). The use of citation analysis for evaluative purposes is the issue that has
generated the most discussion. While Bayer and Folger note that measures derived from citation counts have high face validity, Thorne argues that citation counts have high spurious validity because documents can be cited for reasons irrelevant to their merit. Nevertheless, this assumption has been tested and has found support in a number of studies, including studies of scientific papers, journals and scholars. In each case some non-bibliometric measure’. Further more, one cannot automatically assume that an infrequently cited document is without merit. In the case of journals, for example the usefulness of citations as a measure of the journal; new journals may be of a high quality but frequently cited. Until more is understood about the reasons for citing citation counts viewed as a rough indicator of quality. Small differences in citation counts should not be interpreted as significant, but large differences may interpreted as reflections of differences in quality and impact, results of citation count should be compared with alternative quality indicators to look for correlations. The validity of the measures is most fragile in citation counts for individual documents and authors. One can have more confidence in comparisons of counts based on larger units such as journals.

Citations are made to the best possible works. One can better understand the nature of citations if one knows the population from which they are selected. If one assumes that authors shift through all of the possible documents that could be cited and carefully selected those judged best. But studies of science information use have suggested that accessibility may be as important factor as quality in the selection of
an information source. Soper conducted a study to investigate the effect of physical accessibility upon the selection and use of references. She found that the largest proportion of documents cited in author’s recent papers was located in personal collections, a smaller proportion was located in libraries in departments and institutions to which respondents belonged, and the smallest proportion was located in libraries of other cities and countries. Thus a paper might well have been cited because it was the ideal paper to cite. Accessibility of a document may be a function of its form, place of origin, age and language of a journal article, its accessibility may be determined by the journals circulation, reprint policies and coverage by indexing and abstracting services. Just as a document may be more or less accessible, a researcher may be more or less visible. Other scientists work may came to the author’s attention as a result of their discoveries, their leadership in scientific community, or their activities in the world of politics and controversy. As with documents, researcher cited therefore, do not necessarily represent the most outstanding in a particular field. It may be that anything which enhances a researcher’s visibility is likely to increase his citation rate, irrespective of the intrinsic quality of his work.

While observing the above pros and cons one can conclude that further judgment is needed to analyze citation context and draw inference. Nevertheless, both mechanical and intellectual refinements after alternative to treating citations as masses of undifferentiated units. Although for some applications it is sufficient to
CHAPTER-3

treat citations equally, for others it is appropriate to investigate the structure of citation practice.

**CONCEPT AND DEVELOPMENT**

The concept of citation analysis had been existing under different names. The word Bibliometric, the other name of citation analysis was first coined by Allen Pritchard in 1969 to mean “an application of mathematical and statistical method to books and other media process of written communication and of the nature and course of a discipline.” Earlier however Cole and Eales through their work, ‘A history of comparative Anatomy’ in 1917 analyzed a comparative anatomy paper through counting a number of publications country-wise from 1543 to 1860 which was regarded as the first counting technique in evaluating the international scientific activities. Owing to its quantitative approach to the bibliography in literature, they called this technique as statistical literature analysis. Gross and Gross in 1927 used citation count to rank the periodicals in chemistry which was regarded as the first user study of any significance based on a more systematic citation count that later became the basis and the methodological direction to the Bradford’s law of Scattering. S. R. Ranganathan, and then (1949) in 1948 coined the term ‘Librametry’ for quantitative studies and analysis of library activities.
Bibliometric Measurements

The bibliometric measurements are derived from the concept of citation indexing. It was based on the English legal system, which operates under the doctrine of Stare Desisi Precedent on the basis of which Garfield developed Science Citation Index, Social Science Citation Index and Art and Humanities Citation Index.

The simplest technique of bibliometrics is counting the total number of publications a scientist or a group of them have published. While the publication count gives a quantitative measure of the total volume of research output, the qualitative aspect of the published work remains to be assessed. The publication counts, however, have its other limitations also. Multiple authors has shown sharp increase during past few decades owing to prevalence of collaborative research, poses problem in the process of publication count. Jones studied the correlation between the publication counts and other measures of scientific merits such as research funding, cited ness, etc. and found a reasonable correlation.

Direct Citation Counting

Citation counting is a technique that determines how many citations in a given document, author, journal etc. has received over a period of time, originally used by Gross and Gross [9]. The rational for this is that citations are objective indicators of use and therefore an article, author, and journal that is frequently cited,
is more useful or productive, as the case may be, than one that is less frequently cited. In order to offset the limitations of citation counting, some modified measures have been suggested. The “Impact factor” and Immediacy Index are two such measures. Garfield [10], who coined the term impact factor defines it as “the ratio of the number of times a journal is cited in a given time period to the total number of times a journal is cited in a given time period to the total number of source items published in a journal, during specific period of time”. The impact factor is a measure of the frequency with which the average cited article in a journal has been cited in a particular year. It offset the effect of age, size, the frequency of a publication of a journal on the frequency of citation. The immediacy index is a method of showing the frequency with which a material received by the article during the year to the number of articles published. Hirst [11] has suggested the discipline impact factor (DIF), a method of determining core journals for a discipline, which is similar to the impact factor, which measure the number of times a paper in journal is cited in the core literature of the given discipline.

**Bibliographic Coupling**

*Fano* [12] first expressed the notion of bibliographic coupling although it was *Kessler* [13] in 1963 who introduced the bibliographic coupling as a measure of the subject similarity between two documents. Two documents are bibliographically coupled when both of them cite one or more reference in common. The numbers of
such common references determine the strength of bibliographic coupling. The strength of bibliographic coupling increases with the increase in number of references cited in common. The concept has since been applied to the problems of information retrieval and for mapping the geographical structure of science.

**Co-citation Coupling**

The concept of Co-citation was for the first time suggested independently by Small [14] and Marshaova [15] almost simultaneously in 1973 and later developed by Small. Co-citation is also defined by many authors such as Cawkell and Bellardo as a subject’s similarity indicator. Small describes it as “the frequency with which two documents are cited together”. He also states that cited documents are linked together through the process of co-citation, and this process is similar to the similarity measures of the co occurrence of words.

Cawkell defines co-citation as “subject similarity indicator, and demonstrates co-citation and bibliographic coupling clearly through a citation matrix.”

Bellardo defines co-citation, as “a process where by an author cites to earlier documents in a new work.”

Garfield explains that the strength of co-citation can be calculated using the following formula:
CHAPTER-3

Co-citation of A+B
\[ S = \frac{(\text{Total citations of A+B})-(\text{Co-citation of A+B})}{(\text{Total citations of A+B})-(\text{Co-citation of A+B})} \]

**Bibliographic Coupling Vs Co-citation Analysis**

Co-citation and bibliographic coupling is related to each other in a kind of mirror symmetry. While the co-citation is a dynamic measure i.e., the strength of co-citation links increases as subsequent scientists cite their predecessor’s work. The bibliographic coupling of two papers does not change over time and cannot be interpreted in a straightforward manner.

Bibliographic coupling and co-citation analysis have their application in clustering or classification of the micro and macro structures of citing or cited documents and their mapping. While the one generate the cluster of cited papers the other generate cluster of citing papers. Both the techniques have their application and neither should be assumed to be superior as a measure of association or relatedness of documents. The choice between the two techniques is determined by the nature of the phenomenon to be investigated and the interpretation sought.

**Co-word Analysis**

The co-word analysis involves identification of keywords and their co-occurrence in an attempt to generate a map index of papers linked by the degree of co-occurrence of the keywords. Methodology was developed by the center de
sociologies De’l innovation (CSI), [16, 17] Paris and is independent to any particular database or a secondary service based on selective coverage. The institute has also develop a computer aided indexing techniques for full text databases called LEXINET. The database is linked with online thesaurus, which is updated interactively each time a new document is analyzed. The CSI has also developed a computerized programmer called “LEXIMAPPE” for co-word mapping. The technique is developed as a science policy tool and has already been used to analyze publications from research on various specialties.

Need of Citation Analysis

A plethora of knowledge is being unfolded everyday. This situation of overwhelming mass of available information is been denoted by ‘information explosion.’ It is a common knowledge that during the present generation more science related information has been produced than in the past. This has happened because of the ever growing specialization and diversification of knowledge, need of distinguishing oneself in academic circles and increasing numbers of printed and recorded information in various forms.

Today the amount of information available in books, articles, periodicals, abstracting and indexing periodicals, conference proceedings, seminar papers, bibliographic databases has grow phenomenally. According to a rough estimate S & T information now increases at 13% per year. Therefore information in these fields
will be doubled in every 12 months. Electronic databases and databanks have proliferated.

According to UNESCO there are over 6,00,000 documents being published every year in more than 80 written languages. Out of these document 3,00,000 alone are traditional books, 150000 are periodicals and rest 1,50,000 other type of documents like reports, patents, govt. publication, papers, etc. The world growth of scientific literature is estimated at 6-7% a year. The rate of growth is so rapid that the scientific literature is getting doubled every year.

Cost is another factor for concern. It is said that price of books has gone up by 40% since January 1991, whereas the cost of periodicals keeps on increasing by 15-20% each year According to a survey of expenditure carried by Roorkee University, the cost of journals doubles every fourth year. This has outstripped the purchasing capacity of libraries all over the world. It is said that even a very rich library like the Library of Congress of the USA, whose budget is in several crores of rupees, will not be able to procure each and every document that is being published from different parts of the world and in different forms and languages. Therefore, to avoid difficulties created by inflation on the one hand and the shrinking fund position on the other, a purposeful study of evaluating the types of literature and of selecting periodicals according to their use value is of immediate interest and need. The demands of the Universe of knowledge are multidimensional.
and ever-changing. Therefore, the evaluation of literature of a sample is essential and has been practice of librarianship. Citation analysis of any journal is helpful to know the origination of knowledge form which disciplines the journal deals.

**Importance of Citation Analysis**

Citation analysis as presently practiced is largely used for putting things in order. The things ordered can be journals, paper in journal, authors of papers in journals or organizations to which authors of journal papers are affiliated. The type of order can be linear, as in ranking, or multidimensional, as in the generation of citation networks. Studies of obsolescence rates of journals or documents may be considered to be special cases of ranking, over time. Whatever the type of analysis preformed, the interpretation of the results hinges on the nature of relationship between the citing and cited documents.

Citation analysis is very often fruitfully applied to derive the following benefits:

**To lead the reader to further studies in the field**

This is perhaps, the primary purpose of citations. Readers can verify the correctness of the information and thereby convince themselves.
For the preparation of bibliographies

The first use of citation indexing was made in Sheperd’s Citations published in 1873. This technique of citation indexing has been perfected by Eugene Graifield and others since the early 1960s. It is a fact that compilation of bibliographies in new fields is really difficult. In such circumstances, analysis of citations of articles may be the only way together information. The very fact that the citations have been verified, evaluated and recommended by authors who are experts in their own fields make them all the more acceptable for inclusion in a bibliography.

To study the use pattern of different types of documents

Citations given may be of books, journal articles, reports, standards, these/dissertations etc. the relative use of each of these types can be ascertained based on the frequency of citations. For example, various citation studies have shown that journal articles are the most preferred source consulted by scientists since they constitute about 70-80% of the total citations. Similarly citation practices among social scientists indicate that they give equal importance to books and journals.

To find out the relative use of different languages

Since English has emerged as a world language, especially in science and technology, there is a predominance of English language publications in all
CHAPTER-3

branches. This can easily be understood from citation analysis. In the mid-sixties, for instance, the share of English language papers in Mathematics and Chemistry was more than 50%. Russian occupied the second position with about 20% followed by German and French.

Citation practices have also shown that the relative amount of literature in different subjects produced by different countries changes with time. It has been observed that German has declined very much in the 20th century, especially in the field of Chemistry where publications in this language reigned supreme.

To study the use of literature from different countries

From the citations, the country of their origin can be identified in all types of materials like journal articles, books, reports etc. In many subject areas, U.S. Publication are found to be used more heavily. In medicine, biochemistry, physiology and pharmacology, Sengupta had identified the leading role played by U.S. journals. Journals of U.K. occupied the second position, but they come nowhere near their American counterparts in the frequency of use. Similarly, Martyn and Gilchrist had found that in sixties one in every eight citations was to British publications. Some of the user studies in India have shown that India publications are also equally cited in certain subjects.
To study the scattering of subjects

Studies about the dispersion or scattering of subjects in different sources as evidenced by citation analysis have brought out interesting results. For example:

- Social science and arts subjects show a wider scatter of publication than the science.
- Research publications in technology show a greater dispersion than those in science.
- A new branch of science, especially an interdisciplinary one, shows a greater dispersion than an older branch of science.
- There can be differences in scatter between sub-fields within a subject as also among major subjects.
- The rate of scatter within the same subject alters with time.

To decide the obsolescence rate of documents in different subjects

Citations in subsequent literature and usage pattern in libraries are considered as two indicators of the obsolescence of literature. Analysis of citations by age of the cited document can show the useful life of a document. In order to measure the decay or obsolescence rate of document, the concept of ‘half life’ has been borrowed from Nuclear Physics. Using this measure Burton and Keblee had suggested a range of half-lives for different subjects. The fast growing subjects
would have lesser half-lives compared to established disciplines. The above study had shown the half-life of Metallurgical Engineering as 3.9 while that of Botany is 10 years. These time scales are highly useful in the planning of library holding.

**To determine the interdependence and lineage of subjects**

The interdependence of basic and applied fields can be understood by citation studies. Establishment of this interdependence can be of use in the acquisition policy of special libraries or information centers. The analysis of citations of the Annual Review of Medicine for the year 1965-69 by I.N. Sengupta has established the contribution made by journals in the fields of biochemistry and physiology to the medical research. Further studies by him have brought to light the mutual contribution of biochemistry, physiology and microbiology.

**To prepare ranked list of periodicals**

Ranked list of periodicals can be prepared by two methods:

- by actual citation counting; and
- by counting the number of entries in indexing and abstracting periodicals.

In the first method, information is collected from the references cited in source articles. By studying the average number of citations, one can develop a list of cited journals in the ranked order.
In the second method, the number of items contributed by different periodicals during a specific period of time is calculated from the secondary source and the ranked list is prepared based on the productivity of journals. Such ranked lists are very often used as guidelines in the acquisition of periodicals as other materials in the library.

**To study the rate of collaborative research**

Collaborative research can be effectively measured from the number of authors in papers. Such studies can be conducted to understand global trends, national trends on trends in different subjects. Studies in this direction have indicated that collaboration varies from discipline to discipline, within the same discipline from time to time, and from country to country. However, the extent of collaboration may not be revealed from the citations.

**For the analysis of scientific journals**

Citation analysis provides a number of interesting and useful insights into the networking of journals. These insights are developed from five different citation measures, which are perfected by *Institute for Scientific Information (ISI).*

**Citation rate of a journal**

This is the number of times a journal has been cited. It can consist of all the references to the cited journal counting even duplicate references from the same
source article as a separate citation. It can also be calculated by counting only the number of source articles that cited the journal. A third method of calculation citation rate that is followed by ISI is by counting the number of references to the cited journal, but treating duplicate references from the same source article as only a single citation link.

**Impact factor**

Impact factor (IF) is the average citation rate of journal’s articles. It is basically a ratio between the citation rate of the journal and its citation potential. Citation rate is defined as the number of time cited, whereas citation potential defined as the number of citable items published. Therefore

\[
\text{Impact factor} = \frac{\text{the number of times a journal was cited}}{\text{the number of citable items the journal published}}
\]

**Self-citing rate**

This is a measurement of the frequency with which journals cite themselves. It shows what percentage of a journal’s references cite articles it published.

**Self-cited rate**

This, again, is a measurement of self-citation. It shows what percentage of citations received by a journal originated in articles published by the journal. These
self-Citation rates serve as indexes to newness, size and isolation of the intellectual universe in which a journal operates.

**Immediacy index**

This is a method of showing how rapidly the materials published by a journal are picked up and used. It is calculated by counting the number of citations received by articles in a journal during the year in which they were published.

**Citation index**

A citation index is an ordered list of cited articles which is accompanied by a list of citing articles. The cited article is identified as a reference and the citing articles as a source. The association of ideas existing between the cited and citing articles is utilized in the preparation of this index. It may perhaps be said that cited articles are, ancestors and the citing articles are descendants and this descending relation of subject is reflected through the index.

Citation indexing is a relatively new method of organizing the contents of a collection of documents is in a way that overcomes many of the shortcomings of the more traditional indexing method. The primary advantage of Citation indexing is that it identifies relationship between documents that are often overlooked in a subject index. An important secondary advantage is that the compilation of citation indexes is especially well studied in the use of man-marching indexing method that
do not require indexers who are subject specialists. This helps to make citation more current than most subject indexes. Furthermore citations, which are bibliographic descriptions of document, are not valuable to scientific and technological obsolescence as are that the terms used in subject indexes.

Citation indexing is based on the simple concept that author’s references to previously recorded information identify much of the earlier work that is pertinent to the subject of this present document. These references are commonly called citations in a given collection of documents.

**Limitations of Citation Analysis**

Generally, all studies using citation and co-citation analyses suffer from two major limitations: the assumptions which underlie citation analyses, and the problem with the sources of citation data. Some problems which need to be considered are.

i. **Self citation**

While papers are written by a single author, the elimination of self citation is easy; but a further checking may be needed for multi-authored papers. The elimination of group self citation is the more difficult problem.
ii. Multiple authorship

The citation indexes include only the first-named authors of cited articles. For finding all citation to publications of a given author, including those in which he/she is not the first author, a bibliography of his or her works is needed so that all articles can be checked in the citation index. An online search can be helpful for publications in one of the source journals of ISI’s database.

iii. Homographs or homonyms

To differentiate among scientists with the same name and initials publishing in the same field, extra information such as institutional affiliation is needed. Otherwise citation could be incorrectly attributed to an author, particularly if he/she has a common name. Even this problem is more difficult with Chinese or Japanese names than with English names.

iv. Synonyms

Until establishing a standard form for the author’s name, citation will be scattered. A famous example of “synonym” includes an author’s name in the field of informatics is given by the different forms of Derek J de Solla Price’s, name, a woman’s maiden and married names, different treatment of foreign names and misspellings.
v. **Types of sources**

Line shows that in citation analyses the type of source can influence the results. He found differences between the analysis of references drawn from journal and monograph in: date distributions, forms of material cited, subject distributions, the proportion of references made beyond the social science, and countries of publication cited.

vi. **Fluctuations in time**

Citation data should not be too restricted in time, while there may be large variations in citation counts from one year to another.

vii. **Field variations**

There are different referencing characteristics among different segments of the literature: example, the number of citations per publication varies greatly from field to field. This may lead to difficulties in cross-discipline comparisons.

viii. **The incompleteness of the ISI-database**

While the ISI-database does not cover all periodicals it may create problems for national studies.
ix. Dominance of English as a scientific language

While either English or French are official languages in Canada, French Authors who publish in English are more often cited than those who publish in French.

x. Errors

Citation analysis, including those based on citation indexes can be more accurate than the raw material used. Although processing of citations for inclusion in citation indexes may introduce some errors while eliminating others, many errors due to citing authors remain. These can include errors in cited authors names, journal title, page, volume and year.

Citation Analysis: Applications

Techniques of citation analysis have been put to a spectrum of applications from studies at the macro level dealing with the large aggregate of publications to the micro level analysis of a single author or a single paper; from the linear ordering of citation data to obtain a ranked list of periodicals to multi-dimensional mapping of key citations. With the assumption that the citations are indicators of scholarly importance, their analysis can thus be used to determine the most important scholars, publications, departments, etc. In other words, unit of analysis in a study can vary from individual articles or journals or authors to industrial organizations,
academic departments, universities, countries or geographic regions. The major applications of bibliometric studies and citations analysis, reported in the literature are summed up under the following categories with a few leading examples under each category:

**Measuring Science: The Size and Growth of Literature**

Application of bibliometrics to evaluate the size and growth of literature had occupied a large number of writers since the early part of the present century. A number of bibliographic parameters like total number of pages, words, letters, articles, chapters, etc. have been used effectively for measuring knowledge and its growth. The phrases like “Exponential Growth” and “Information Explosion” used widely even today are the result of analysis conducted decades ago.

*Hulme* [18] who coined the term “statistical bibliography”, the predecessor of today’s term “bibliometrics” concentrated his studies on the quantification of growth of scientific knowledge. He analyzed author and journal entries in the “International Catalogue of Scientific Literature” and measured the contribution of various countries. In 1961, Derek J. De Sola Price first published his book “Science since Babylon” [19] followed by another related work “Little Science, Big Science”[20]. The two books neatly outline the boundaries of the scientific enterprises. Price in these treatises published some observations on the exponential rate of growth in the number of scientific journals along with the corresponding exponential rate of
growth of abstract journals. Bourne [21] in 1962, used list of serial titles to estimate the quantity of the world scientific and technical literature and its linguistic, national and subject origin for abstracting and indexing purposes. Barr in 1967 published the first comprehensive census of the then currently existing periodicals based on 26,000 scientific and technical periodicals actually being received by the National Lending Library in 1965. In 1969, Price used bibliometric techniques to show correlation between the scientific productivity of a country and its gross national products (GNP).

Garfield [22] evolved the concept of journal impact (impact factor) computed by obtaining the ratio of citation to particular years of a journal and the number of articles published in these years. Impact factor, thus, is a measure of the frequency with which the “average article” in a journal has been cited in a particular year. The Institute of Scientific Information (ISI) first published a list of top fifty high impact journals of science in 1972. In the same year United States National Science Foundation, National Science Board, published its report “Science Indicator’s [23] 1972, which has a section of use of scientific publications as an output measure in science. The report notes “there are certain relatively direct results of research and development which provide indicator’s for comparing the scientific and technical performance of nations. Primary among these are report of research published in scientific and technical journals and patents covering new products and processes. Narin [24] measured frequency of citations from one country to another in 1975.
Sengupta [25] proposed a new formula for ranking of journals not only by the number of citations from periodicals but also by reference to the number of years of publications of each periodical.

**Scientography: Mapping the Geography of Science**

The literature is comprised of documents forming a continuously changing network of relationships, groupings and patterns. It is characterize by uneven growth and varying levels of associations as some areas of research spring to sudden spurt of activities while other lie dormant. An understanding of the nature and direction of these relationships is required by all groups working in the information field. In information science, the structure of science has a direct impact on the design of classification systems and its ability to cope with the changes required in terms of maintaining effective library stocks responsive to the changes in the literature. Scholars and researchers require this information to identify peripheral subjects and relevant developments in them. Sociologists are concerned with the structure of science in terms of patterns of informal communications, group structure, invisible colleges, etc. In science policy studies, there is a need to know more about where the action, is, who is involved and where. Historian and philosopher of science are interested in such relationship in terms of paradigms or clusters and communities, i.e., social and collective aspects of science, the publishers of the secondary services a need the data to enable them to obtain the most effective
coverage pattern. This data may also help editors of primary journals in determining present and future policies for the acceptance of submitted papers.

*Cason and Lubotsky, [26]* in 1936 advocated use of scientific journal to study the functioning of scientific community. Their paper discussed the interdisciplinary influence and dependency relationship between psychological journals. The paper, for the first time, put forth the idea of a cross-citing network in the literature through a 28 x 28 element matrix for most significant journals. Daniel and Louttit in 1953, analyzed the structure of the psychological literature by formulating a journal –to-journal cross-citation matrix. Daniel and Loutitt were the first to use the cluster analysis and mapping technique for related journals.

*Xhignesse and Osgood [27]* in a 1967 paper, analyzed the citation characteristics of psychological journals and introduced the concepts and parameters of networking theory to the field of bibliometrics. The concept of interpoint variable distances was introduced for the first tie to match the actual rank order of drawing a graphic portrayal. *Van Cott and Zavala* [28] applied the technique of factor analysis to the subject classifications of articles in the physics and concluded that it should be possible to extract from the literature of a given discipline some relatively stable structure that reflect the true structure of that discipline.

Citation maps into a language structure by substituting frequently occurring terms or phrases from the citation using context analysis for each document on the map. *Small* [29] in 1981 used co-citation analysis to explore the relationship of natural
and quantitative way to group or cluster the cited documents since the patterns of co-citation changes with time as new discoveries are made and introduced through the literature resulting into generation of new linkages. Several studies have since been made to locate networks of co-cited papers and using the resulting networks to define scientific specialties and the nature of activities at the research front.

**Journal-to-Journal Influence Mapping**

While Small [30] and his colleagues concentrated on defining the micro-scale structure of the literature, Narin and his team began to classify journals on macro-level by fields and subfield. In their 1972 paper they mapped the interrelationship between individual journals and fields. They developed a method called the influence measures for mapping the hierarchies of influence among journals. In an attempt at further refinement in journal classification, Carpenter and Narin developed a procedure for clustering scientific journals based on their crossed citation patterns. Pinski and Narin [31, 32] devised a method to evaluate impact of journal considering length of papers, the prestige of the citing journal, and referencing to different segment of literature.

**Information Storage and Retrieval**

A cited document is related in content to the citing documents. Balrup [33] in 1969 in an experiment in which authors were asked to assess the degree of
relatedness of citations to their own work, found that 72% of articles to be definitely related and only 5% to be definitely not related.

Citation links have been used in developing document representation in automatic classification and various retrieval algorithms, which make use of the ability to find related documents in the file irrespective of words and languages. Citation as retrieval tool have the advantage that they are unaffected by changing terminology and they reveal relevant papers which can not be searched using conventional indexing. Yermish described and interactive information retrieval system, which he developed to manipulate citation relations existing among bibliographic records efficiently. Chapman and Subramanyam [34] described the use of co-citation in a search strategy to retrieve relevant document on given topics using commercially available search systems and the citation index databases. Schiminovich [35] supporting the earlier findings that the search conducted on retrospective literature on the basis of citations resulted in remarkably high recall and relevance ratios close to 100%. Gerard Salton,[36] based on the results of several studies conducted by him, concluded that the bibliographic citations are generally useful for document analysis and information retrieval systems should assign a bibliographic citation code the documents in addition to standard content indicators. Bichteler and Eaton [37] developed a linkage similarity measure for combined use of bibliographic coupling of documents and their co-citations for document retrieval. They concluded that the combined use of the two techniques does offer improvement in document
retrieval performance over what is obtainable by the use of bibliographic coupling alone.

**Bibliometric Analysis of Various Disciplines**

This subset of bibliometrics is very large consisting of thousands of papers with their titles typically having wordings such as “The language distribution...” or “Bibliometric analysis of literature of ...” Most of these publications look at the citations in a particular subject area to describe patterns of citations, identify the core journals, and conduct bibliometric distribution of publication by country, language, subject and form-wise. Source of citations data may be limited to a single journal or a group of journals or they may include other type of material in addition to journals. Mostly Bradford’s law of Scattering is tested on the collected citations and the growth and decay rates are calculated. Other characteristics of citations examined frequently include types, age, highly cited authors, etc. and changed in these characteristics over a period of time.

**Type of Literature Studies**

Citation analysis has been used to measure use of publications in certain type of literature such as government publications and patents. Chambers and Heraley [38] analyzed the citation given in these to evaluate the local collections. Kubata analyzed citations of graduate theses submitted to the school of Library and Information Science, University of Keio, Japan. Goehlert [39] studied the use of
government publications by international organizations by analyzing the references cited by their publications.

**Citation Count as Measure of the Scholarly Eminence**

The citation analysis can provide objective information about the usefulness of a scientist’s work to the scientific community. Garfield in 1955, while proposing construction of a citation index suggested use of citations to measure the impact number of publications of a scientist. Price [40] in a study published in 1965 on networks of scientific papers also discussed scientific productivity as measured by citation rather than total number of publications. He concluded that the concentration of productivity in talented individuals as shown of citations to a relatively small number of papers. Clark [41] compared various measures of contributions such as number of publications, offices hide, listings in biographical directories, etc. highest correlation however existed between the significant contributions and the volume of citations received by them. On the basis of Clark’s findings, a faculty selection committee at the University of Toronto undertook to evaluate citation as a valid measure of scientific eminence (Myers). The committee found high degree of correlation and concluded that such an analysis can be objectively applied for the professional appointments. It is, however, clear that citations do not reflect teaching ability, administrative talents and other departmental contributions neither it guarantee continuing higher quality of
scientific research. Thus citation performance should be treated as only one more, though distinctive, indicator of candidate’s influence and can be compared with others used in traditional evaluation procedure. Cole [42] found that citation counts to be highly correlated with various other measures of eminence. He found that the quality of work of 120 physicists, as indicated by citation counts, to be correlated with each scientist’s visibility to his colleagues and with the number of awards he had received. Ziman [43] opined that “success in science is not merely a matter of winning prizes; it is rather having other scientist cite your work”. He observed that empirical evidence suggesting positive correlation between numbers of citations of person’s published work and other available measures of prominence. Narin and Carpenter [44] used citation counts as well as the volume of publications as measures of the scientific activity of nations, with the USA ranking first the USSR second.

Countless studies have attempted to evaluate the role publications have played in gaining scientific recognition in terms of awards, academic appointments, salaries etc. Hamermesh [45] compared citation frequencies of scholars with their salaries. He found that the number of citations is more strongly correlated with increase in salaries than is either professional experience or number of publications. In a recent study Sauer [46] found that the numbers of citations and published journal articles rather than books are the most important productivity indicators in determining salaries.
User’s Study

The user’s study has implications for collection development and design of services for a particular library. Reference list of articles written by a user or a group of users of a specific collection/ library were analyzed to determine the types of materials, age, subject, language and whether the referred material is locally owned or borrowed. The hypothesis like “the scientists want to write but not read and the technologists want to read but not write” can be tested by the user’s study. The citation analysis can be used to compare behavior of users in different fields in different time span. Skelton [47] used citation data to compare the information requirements of scientists of scientists and social scientists.

Historical Studies

The citation counts have also been used to identify important contributions to a discipline. Garfield suggested the use of citation indexing for historical research in 1955 them. Methodology for use of citation indexing in sociological and historical research was, however, standardized after publication of Science Citation Index. The published papers represent the scientific work and relationship between discrete pieces of work is represented by references in papers. While in conventional bibliography, a simple chronological listing of publications may not give even the faintest hint of historical development of a particular subject, but citation indexes,
and networks derived from it, display the interrelationship among important landmarks and events. A new tool termed as “histograph” devised by the Institute for Scientific Information for graphical display of citation data that shows key scientific events, their chronology, inter-relationships and relative importance. The histograph would display key papers and break through events identified by citation analysis, reduce large, number of seemingly unrelated events to a coherent pattern and would keep track of influence of one event on another. Such histograms provide a legitimate starting point for the historian of science, the technique can eliminate the drudgery associated with scholarly scientific historical writings.

Citations have been sued to trace the chronology of events, relationships amongst them, and their relative importance. In 1964, Garfield [48] compared the history of DNA research according to Asimov’s “Genetic Code” with the citation network of key citations derived for the Science Citation Index. He concluded that citations studies are valid for creating accurate historical descriptions of scientific fields. Ellis [49] using citations in patents generated the citation networks to display the history of technical developments. The patterns observed through such studies can be validated through interactions with specialist and questionnaire surveys.

More recently McCain [50] used co-citation cluster analysis of forty-nine authors to investigate changes in the structure of Drosophila genetics literature over the year 1979-1983. He observed that the literature on history of Drosophila genetics consist of a set of author cluster representing classical genetics, surrounded by more
specialized research clusters namely: development genetics, molecular genetics, neurogenetics and population genetics.

**Communication Patterns**

Contemporary science depends almost entirely on the journal literature, which in turn, provides an effective channel for scientific communication. Citation may be used to trace patterns in scientific communication. Citation linkages do not necessarily reflect social contracts. The assumption can be helpful to identify problem areas in communication, including linguistic isolation, limited dissemination of new ideas and barriers between basic and applied science or between specialists and the public at large. Typical studies that have been carried out include formal and informal communication, publication vs. communications, co-authorship and inter-organizational mobility. Brillouin Information that can be communicated through an information system. Informal collectives of closely interacting scientists, generally limited to a size that can be handled by interpersonal relationships is termed as “Invisible Colleges” which are considered significant social and cognitive formations that advance the research fronts of science, a concept largely confirmed by the early studies of Crane [51] Mullins [52] and Chubin[53]

**Collection Development**

Citation analysis has been used as a tool for collection development although it has attracted several criticisms. Journal ranking by citation counts may bear little
relation to the frequency of their use in a particular library as citation analysis and use analysis measure different activities. Besides, non-cited journals like trade and technical journals and professional magazines are not considered in the analysis.

In spite of these criticisms, the ranked derived from citation counts can still be used to:

i) Identify highly ranked journals not available locally and within subject scope as worth examining in more details;

ii) Low-ranked journals available at the library may be examined; and

iii) Peripheral journals which may not be available in the collection may be of use and can be acquired for local use on inter-library loan.

Broundus [54] opined that in the absence of subject specialist on library staff, citation studies can be of considerable value in choosing serials and monographs. Gross and Gross [55] were, however, the first to use citation to evaluate the key scientific journals in a subject or a discipline. The paper described and interpreted a count of literature (including books and journals) of comparative anatomy for determining contributions and performance of various countries for over three centuries. Triggered by this seminal paper, several thousands of subsequent papers have used this basis counting techniques to analyze the literature referred to by a given journal or a group of journals. In 1949, Fusser’s [56] paper signaled the resurgence of post war interest in the bibliometrics. Fusser’s paper addressed to the problem of optimal collection size,
book selection policy and important serial titles for chemistry and physics.

*Gottschalk and Desmond* [57] in a related work produced a census of the journal literature to aid libraries in housing comprehensive collections in 1963.

*Garfield* [58] in an extensive paper in 1972 discussed citation analysis as a tool for journal evaluation. Using a citation base of 1,000,000 from 2,200 journals covered in SCI till 1969, Garfield proposed ranking of journals by their impact factor. He listed 152 most frequently cited journals along with their impact factor obtained by dividing total number of references to 1967 and 1968 articles by number of articles published in 1967 and 1968. Garfield observed that a good multidisciplinary journal collection need not contain more than a few hundred titles. He also noted that average cited paper is cited only 1.7 times a year.

**Obsolescence**

Obsolescence has been defined as the decline in the validity of the information. The obsolescence studies are of interest to the collection managers to help them decide as the item to keep and which to discard. The obsolescence study culminates in simple mathematical formulae, which could be applied with equal success to any libraries. Line and *Sandison* [59] had in a comprehensive review of 200 papers summarized their data and made studies and their results. *Menard* [60] discusses at length the relationship between the growth rate of field and bibliometric measures. *Clark* [61] studied the obsolescence of patients.
Burton and Kebler [62] put forth the concept of half-life, a term borrowed from studies of radioactive isotopes, interpreted as the mean age of citation in a given study. The half-life can also be treated as an indicator of the rate at which a journal article becomes obsolete. This, in turn, may reflect the rate of obsolescence of information in the subject area covered by that journal.

**Automatic Classification**

A document classification aims at grouping the similar documents together. Citation indexing is a natural or automatic system of classification; here the material to be classified orders itself through its conceptual links. Citations have been used to make meaningful connection between the atomized fragments of information. Garfield [63] developed a computer-based system for automatic classification of scientific articles determined by citation and cogitation patterns from the Science Citation Index database. Classification of clusters of the citing papers emerged on the basis of their citation patterns are determined by examination of high frequency word pairs from the title of citing papers. Classification of new items is performed automatically based on clusters they cite. Garfield described the techniques of co-citation clustering as an automatic self-generating system considering that each citation is a facet of document citing it. Small used two-step bibliographic coupling linkages for journal clustering. The results obtained were evaluated by comparing them with an independently
CHAPTER-3

derived manual classification of the same set, which was found to be in good agreement. Small also observed that nearly all highly cited papers are linked together deduced that the distinct specialties of science are not totally isolated from one another but somehow connected by weak, although perhaps important links.

Model of Scientific Changes

Garfield [64] proposed use of citation data to investigate the patterns of evolution in the growth of science and scientific specialties by examining shifts in clusters over a specified time. The manner in which the cluster of cited documents change over a specified time, the manner in which the cluster of cited documents change over time, would reflect the change in scientific specialties. The methodology involves identifying corresponding clusters in successive years through highly cited documents in the cluster; the changes observed may be depicted and analyzed by various graphing techniques.

A four years study conducted at ISI on 31 specialty clusters reveals that the changes from one year to the next in the span of four years, about one third of the 31 specialties experienced major shifts in the set if cited documents- i.e, all but one or two of the cited documents in the cluster dropped out, and an almost entirely set appeared. On the basis of this study, Garfield predicted that on an average, a specialty would undergo one revolution every 12 years. The inwards
and outwards movement of documents from a cluster in –group is interpreted as shifts in the leadership of the specialty from one school to another. Garfield also suggested that the existing models of scientific change could be put to test using the technique described above.

**Generation of Computerized Synopsis**

Small [65] used co-citation cluster to develop synopsis of scientific specialties. Small referred to these reviews as “Specialty Narratives”. The process – involved analysis of documents within the cluster to identify passages that cite the cluster’s combine and sequence such passages to form speciality narratives. Small defined such narratives as “a combination of statements by several individuals from several sources brought together by common usages, and selected to typify that usage”.

**Citation Indexes**

The first practical application of the concept of citation indexing was Shepard’s Citations lists all precedents cases followed by subsequent citing cases. Science Citation Index was conceived on the similar lines as a structured list of all citations in a given collection of documents arranged so that the cited documents are followed by the citing documents.

Garfield [66] and Adair,[67] in two separate articles in 1955, first floated the idea of citation indexes for science as a new and valuable way of indexing
scientific literature. In 1961, Garfield as the Founder President of the Institute of Scientific Information received a grant from the National Institutes of Health (later transferred to the US National Science Foundation) for the production of a citation index to the feasibility, production methodology and disciplinary scope of citation indexes to the scientific literature. Garfield soon realized, however, that defining the genetics literature to be covered by the citation index would be quite difficult. It was thus decided to undertake a comprehensive, interdisciplinary approach to prepare a citation index and then extract a Genetics Citation Index from that base of information. This interdisciplinary database was eventually used to produce the first Science Citation Index, which was published in 1963. The first Science Citation Index covered the literature of the calendar year 1961. It covered 613 journals, containing 1.4 million citations from 102,000 source articles and required five volumes. Nineteen % of the citations on genetics were extracted from the Science Citation Index database using a special computerized procedure and was published separately as the Genetics Citation Index in one volume. The US National Science Foundation did not accept the recommendation for publication of a multidisciplinary science citation index, but the Institute for Scientific Information decided to continue the publication of Science Citation Index on their own subsequent to its first appearance in 1963.

In 1973, Institute for Scientific Information introduced the Social Science Citation Index and the journal Citation Reports, an annual companion volume to
CHAPTER-3

Science Citation Index. The report presents a bibliometric analysis of references processed in the precious year. Arts and Humanities Citation Index was launched in 1978 along with a volume covering 1977. The three citation indexes and their three annual companion volumes (JCRs) offers host of data for bibliometric analysis, interpretation and manipulation.

Limitations in Application

We have seen that many researchers have questioned the very logic of depending on citations, which are given according to the whims and fancies of authors. They believe that any result that is obtained through such a study of citations leads to erroneous conclusions. Maurice B. Line is a strong votary of this point of view.

The common arguments leveled against citations analysis are the following:

- Negative citation i.e., citing a paper just to repudiate it. The result is that controversial papers will get more citations than really worthwhile papers.
- Too mush of self-citations and in – house citations.
- Practice or citing only to get the favors of the powerful or appears to others.
- Citation given just to dress up the proper.
- Variation of citation rate during lifetime of paper.
- Variation of citation rate with type of paper and specialty.
It is a fact that there are extraneous considerations in giving citations. But they do not totally undermine the value of them. As Sengupta argues, “the numbers of scientist who are capable of doing such mal-practices are not significantly high to make citations studies unworkable or misleading”. Further, it is too much to question the honesty and integrity of the majority of scientists.
REFERENCES


63) Garfield, E., and Malin, M.V. (1968). *Can noble prize winners by predicted?* (Paper Presented at 135th Annual meeting of the AAS, Dallas, USA)


