

Tools and techniques of Scientometric Analysis in the Field of Library and Information Science Research

Dr.M.Manjula

Assistant Professor
Srimad Andavan Arts and Science College
Trichy – 5
e-mail: mnjmeenakshi@gmail.com

R. Lakshmi Narayanan

M.L.I.Sc Student
Srimad Andavan Arts and Science College
Trichy – 5
e-mail: sridhar2194@gmail.com

***Abstract** - This paper focus on the tools of scientometrics in Library and Information science. Scientometric is a very vast area in the field of Library and Information science. Scientometric and Bibliographic is the tool for measuring the impact of publications. This paper focus on the basics of scientometric and impact factor, laws of scientometric and other related citation like impact factor and H-Index.*

Key Words: Scientometric, Citation, H-Index, self -citation, Impact factor.

Introduction:

Statistical tools and techniques are used in all research and development subjects. In recent years statistics has been applied to a number of areas such as planning, industrial and agricultural development etc. Statistics and related techniques when applied in depth to a field may give rise to a new subject formation. Dr. S.R.Ranganathan our father of Library and Information Science clearly explained the progressive decrease of extension and increase of the intention of a basic subject or isolate idea. J.H. Shera explained that denudation is the exposure of a new area of knowledge by the erosion or divestment through research or enquiry. Econometrics is an example. Even in the field of library and information science, the applications of qualitative tools and techniques are becoming increasingly popular.

What is Scientometric:

Scientometrics is concerned with the quantitative features and characteristics of science and scientific research. Emphasis is placed on investigations in which the development and mechanism of science are studied by statistical mathematical methods.

Scientometrics is the science of measuring and analyzing science. In practice, scientometrics is often done using bibliometrics which is a measurement of the impact of (scientific) publications. Modern scientometrics is mostly based on the work of Derek J. de Solla Price and Eugene Garfield. The latter founded the Institute for Scientific Information which is heavily used for scientometric analysis. Methods of research include qualitative, quantitative and computational approaches. One significant finding in the field is a principle of cost escalation to the effect that achieving further findings at a given level of importance grow exponentially more costly in the expenditure of effort and resources. However, new algorithmic methods in search, machine learning and data mining are showing that is not the

case for many information retrieval and extraction based problems. Related fields are the history of science and technology, philosophy of science and sociology of scientific knowledge.

Khosrowjerdi, M., & Bayat, M. K. (2013). The scope of scientometric studies has been increased in recent years. Most of previous studies in this field are just like executive reports, and the international scholars have expressed their concerns about this problem. The nature of scientometric studies has not been studied in previous researches and we do not know that scientometric studies are categorized in multidisciplinary, interdisciplinary or cross-disciplinary domain. This paper explored the use of scientometric approach to study the scientometrics domain. We responded to these questions: 1) Is the scientometric studies classified in interdisciplinary domain? 2) What is the contribution of different scientific fields to scientometric studies? 3) What is the current direction of scientometric studies? The data was extracted from Web of Science (WoS). The results showed that scientometric studies were a part of interdisciplinary studies. Furthermore, the library and information science and computer science had major contribution to this field.

Scientometrics is the science of measuring and analyzing science. In practice, scientometrics is often done using bibliometrics which is a measurement of the impact of (scientific) publications. Modern scientometrics is mostly based on the work of Derek J. de Solla Price and Eugene Garfield. The latter founded the Institute for Scientific Information which is heavily used for scientometric analysis. Methods of research include qualitative, quantitative and computational approaches. One significant finding in the field is a principle of cost escalation to the effect that achieving further findings at a given level of importance grow exponentially more costly in the expenditure of effort and resources. However, new algorithmic methods in search, machine learning and data mining are showing that is not the case for many information retrieval and extraction based problems. Related fields are the history of science and technology, philosophy of science and sociology of scientific knowledge

Bibliometrics is a type of research method used in library and information science. It utilizes quantitative analysis and statistics to describe patterns of publication within a given field or body of literature. Researchers may use bibliometric methods of evaluation to determine the influence of a single writer, for example, or to describe the relationship between two or more writers or works. One common way of conducting bibliometric research is to use the Social Science Citation Index, the Science Citation Index or the Arts and Humanities Citation Index to trace citations.

The term bibliometrics was first used by Alan Pritchard his article “Statistical Bibliography” in 1969. The bibliometrics helpful to the library to identify obsolence of documents to facilitate to formulate weeding out policy. More over it is helpful to identify the updatness of a subject so as to have a efficient and effective acquisition policy.

Purpose

The bibliometric studies are useful for the following reasons in the librarianship.

Laws of Bibliometrics

One of the main areas in bibliometric research concerns the application of bibliometric laws. The three most commonly used laws in bibliometrics are: Lotka's law of scientific productivity, Bradford's law of scatter, and Zipf's law of word occurrence.

Lotka's Law

Lotka's Law describes the frequency of publication by authors in a given field. It states that ". . . the number (of authors) making n contributions is about $1/n^2$ of those making one; and the proportion of all contributors, that make a single contribution, is about 60 percent" (Lotka 1926, cited in Potter 1988). This means that out of all the authors in a given field, 60 percent will have just one publication, and 15 percent will have two publications ($1/2^2$ times .60). 7 percent of authors will have three publications ($1/3^2$ times .60), and so on. According to Lotka's Law of scientific productivity, only six percent of the authors in a field will produce more than 10 articles. Lotka's Law, when applied to large bodies of literature over a fairly long period of time, can be accurate in general, but not statistically exact. It is often used to estimate the frequency with which authors will appear in an online catalog (Potter 1988).

Bradford's Law

Bradford's Law serves as a general guideline to librarians in determining the number of core journals in any given field. It states that journals in a single field can be divided into three parts, each containing the same number of articles: 1) a core of journals on the subject, relatively few in number, that produces approximately one-third of all the articles, 2) a second zone, containing the same number of articles as the first, but a greater number of journals, and 3) a third zone, containing the same number of articles as the second, but a still greater number of journals. The mathematical relationship of the number of journals in the core to the first zone is a constant n and to the second zone the relationship is n^2 . Bradford expressed this relationship as $1:n:n^2$. Bradford formulated his law after studying a bibliography of geophysics, covering 326 journals in the field. He discovered that 9 journals contained 429 articles, 59 contained 499 articles, and 258 contained 404 articles. So it took 9 journals to contribute one-third of the articles, 5 times 9, or 45, to produce the next third, and 5 times 5 times 9, or 225, to produce the last third. As may be seen, Bradford's Law is not statistically accurate, strictly speaking. But it is still commonly used as a general rule of thumb (Potter 1988).

Zipf's Law

Zipf's Law is often used to predict the frequency of words within a text. The Law states that in a relatively lengthy text, if you "list the words occurring within that text in order of decreasing frequency, the rank of a word on that list multiplied by its frequency will equal a constant. The equation for this relationship is: $r \times f = k$ where r is the rank of the word, f is the frequency, and k is the constant (Potter 1988). Zipf illustrated his law with an analysis of James Joyce's *Ulysses*. "He showed that the tenth most frequent word occurred 2,653 times, the hundredth most frequent word occurred 265 times, the two hundredth word occurred 133 times, and so on. Zipf found, then that the rank of the word multiplied by the frequency of the word equals a constant that is approximately 26,500" (Potter 1988). Zipf's Law, again, is not statistically perfect, but it is very useful for indexers.

Impact Factor (If)

The impact factor (IF) of an academic journal is a measure reflecting the average number of citations to recent articles published in the journal. It is frequently used as a proxy for the relative importance of a journal within its field, with journals with higher impact factors deemed to be more important than those with lower ones. The impact factor was devised by

Eugene Garfield, the founder of the Institute for Scientific Information. Impact factors are calculated yearly for those journals that are indexed in the Journal Citation Reports.

How Impact Factor is Calculated

"The annual JCR impact factor is a ratio between citations and recent citable items published. Thus, the impact factor of a journal is calculated by dividing the number of current year citations to the source items published in that journal during the previous two years."

Journal Impact Factor:

The journal impact factor measures the importance of a journal and "is a measure of the frequency with which the 'average article' in a journal has been cited in a particular year or period"

H-Index

The h-index is an index that attempts to measure both the productivity and impact of the published work of a scientist or scholar. The index is based on the set of the scientist's most cited papers and the number of citations that they have received in other publications. The index can also be applied to the productivity and impact of a group of scientists, such as a department or university or country, as well as a scholarly journal. The index was suggested by Jorge E. Hirsch, a physicist at UCSD, as a tool for determining theoretical physicists' relative quality and is sometimes called the Hirsch index or Hirsch number.

Citation Index

A citation index is a kind of bibliographic database, an index of citations between publications, allowing the user to easily establish which later documents cite which earlier documents. The first citation indices were legal citations such as Shepard's Citations (1873). In 1960, Eugene Garfield's Institute for Scientific Information (ISI) introduced the first citation index for papers published in academic journals, first the Science Citation Index (SCI), and later the Social Sciences Citation Index (SSCI) and the Arts and Humanities Citation Index (AHCI). The first automated citation indexing was done by CiteSeer in 1997. Other sources for such data include Google Scholar.

Citation is the process of acknowledging or citing the author, year, title, and locut of publication (journal, book, or other) of a source used in a published work. Such citations can be counted as measures of the usage and impact of the cited work. This is called citation analysis or bibliometrics. Among the measures that have emerged from citation analysis are the citation counts for:

- an individual article (how often it was cited);
- an author (total citations, or average citation count per article);
- a journal (average citation count for the articles in the journal).

Citation Count

Citation counts measure the impact of a publication or an author by counting the number of times either is cited by other works. Although this sounds simple, it is complicated by the fact that there is no single citation analysis source that covers all publications and their cited

references. This section introduces the sources available to the ASU community for creating citation counts and explains their coverage and method of searching.

Self-Citation:

"The practice of self-citation can be considered at many levels, including author self-citation, journal self-citation, and subject category self-citation.

Limitations of bibliometrics and scientometrics

Some of the limitations of bibliometric and scientometrics studies are summarized below:

- It describes text, not for the meaning and context
- The peculiar type of research publication monograph are not covered
- The citation index produce selective coverage in English language only.
- Only the first author is cited in the citation indexes

Conclusion

Bibliometric and Scientometrics studies are the most effective studies in the field of Library and Information Science research. These elements are calculated by the extensive use of qualitative and quantitative techniques. These techniques is very useful in to identify the research trends in a subject. Trends in authorship and collaboration in research In the field of author productivity have been helpful to mapping of disciplines. In the above fact there are so many tools and techniques and web analyzing software were developed in this purpose. It is also applied in the various areas including library and information science research.

References

1. Bradford, S C (1934). Sources of Information of specific subjects, Engineering, 137,p85-86.
2. Garfield, E. (1996). Fortnightly review: How can Impact factor be improved? British Medical Journal, 313,pp 411-413.
3. Kessler.M.M (1983). Bibliographic coupling between scientific papers. American documentation, 14. Pp.10-15.
4. Lotka, AJ (1926). The frequency distribution of scientific productivity, Journal of Washington Academy of Sciences 16, p.317.
5. Moya-Anegón, F., Vargas-Quesada, B., Herrero-Solana, V., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., & Munoz-Fernández, F. (2004). A new technique for building maps of large scientific domains based on the cocitation of classes and categories. *Scientometrics*, 61(1), 129-145.
6. Rousseau, S., & Rousseau, R. (1997). Data envelopment analysis as a tool for constructing scientometric indicators. *Scientometrics*, 40(1), . Towards a novel unified framework for developing formal, network and validated agent-based simulation models of complex adaptive systems. *arXiv preprint arXiv:1708.02357*.

