Singaporean Journal of Scientific Research(SJSR)
An International Journal (AIJ)
Vol.16.No.1 2024,Pp.56-64
ISSN: 1205-2421
available at :www.sjsronline.com
Paper Received : 04-02-2024 Paper Accepted: 05-05-2024
Paper Reviewed by: 1.Prof. Cheng Yu 2. Dr.Yarab Baig
Editor : Dr. Chen Du Fensidal

Various Tools and Techniques of Scientometrics : A Study

M.Sithi Jagannara
Assistant Librarian
Veterinary College and Research Institute,(TANUVAS)
Namakkal-2

Abstract

This paper deals with the study of various tools and techniques of the Scientometrics. Scientometrics is the Quantitative analysis of publications of scholarly literature. Understanding scientific citations, mapping scientific fields, and the production of indicators for use in policy and management contexts are some of the major research issues. Others are the measurement of impact, reference sets of articles to investigate the impact of journals and institutes, and the production of indicators. Even though scientometrics use very similar methods, their distinct roles are distinguished by their very different contexts. This paper predominantly features the Bibliographic databases, Citation databases, Various tools and techniques of Scientometrics.

Keywords: Scientometrics, Bibliographic databases, Citation databases, Softwares tools and Techniques.

1. Introduction

Scientometrics, as you are aware, is a field of study which deals with methods to quantitatively analyze scholarly literature. As a technique Scientometrics is used basically for studying a) Scholarly communication: tracing the history and evolution of ideas from one scholar to another; and b) Scholarly influence: quantifying impact of articles, journals, scholars, institutions, nations, etc. Both these purposes of Scientometrics have been based on assumptions. They are: • Scholars communicate their findings by publishing articles • Scholars cite earlier related works of others (and sometimes of their own) in their articles to acknowledge intellectual debt and to witness the use of information. There are other reasons for

citation which will be dealt later.

Objectives

The objectives of this paper are to:

- To understand the various parameters used for Scientometric analyses.
- To study various bibliographic and citation databases used as data sources in Scientometrics study.
- Familiarize with some of the software/tools for Scientometric analysis
  - Understand the features of a few software/tools.

Uses of Scientometrics Study:

Scientometric methods have been used to trace relationships amongst academic journal citations. The Scientometric research uses various methods of citation analysis in order to establish relationships between authors or their work. The Scientometric studies are used in

i) Bibliographic control and Comparative assessment of the secondary services
ii) Formulating search strategies in case of automated system
iii) Indexing and Thesaurus
iv) Library Management
v) Measuring the productivity of an author based on the number of published articles.
vi) Measuring the scattering of articles on a subject in various periodicals (Bradford)
vii) Preparation of retrospective bibliographic and Research
viii) Productivity count of literature
ix) Ranking of words in a text based on frequency of occurrence of words
x) To identify the peers, social change and the core journal, etc

2.Data Sources for Scientometrics

Studies Data collection for Scientometrics study has to be done with care and diligence. Data for Scientometrics studies are to be invariably collected from publications. There is a variety of publishing routes these days, and those in different contexts will value different types: journal articles, monographs, blogs & tweets. Traditionally journals are the most valued source and they continue to be so. Hence many of the Scientometrics studies still revolve round the journals. Recent trends in Scientometrics studies show the use of other digital sources as well. But collection data for Scientometrics studies directly from publications is next to impossible task for individual researchers. Hence one has to depend upon some good source from where the raw data could be culled-out. Decision about the data source goes a long way in the output of the study. Going by the literature on the field one could say that data sources for the Scientometrics are: Questionnaires, Bibliographic databases, Citation databases, Journal indices, Library catalogs and Information systems, Institutional information systems, National databases and so on. Normally, results of the Scientometrics analyses are amenable for valid and acceptable generalization only when data collected is considerably large. The method of collecting data through questionnaires and personal inspection of the

original publications are thus considered to be impractical in many situations.

2.1 Databases as Data Sources

The data source for a Scientometric study is mostly a database. Using of multiple databases for a study is also on the rise. Databases developed by commercial establishments or by public or private institutions form the sources of data for Scientometric studies. One may find one or more databases for every established academic discipline. The following are some of the widely used data sources for Scientometrics (the list is just illustrative and not comprehensive):

• **Chemical Abstracts Service (CAS)**: CAS is a division of American Chemical Society. Its objective is to find, collect and organize all publicly disclosed substance information. Arguably, it is the largest database of chemical information. It covers the publications appear in the form of books, journal articles, patents, conference proceedings, and so on. Its coverage is from 1907 onwards.

• **CiteseerX:** It is an evolving scientific literature digital library and search engine that has focused primarily on the literature in computer and information science. CiteSeerx aims to improve the dissemination of scientific literature and to provide improvements in functionality, usability, availability, cost, comprehensiveness, efficiency, and timeliness in the access of scientific and scholarly knowledge.

• **Compendex:** It is a product of Elsevier. It is the most comprehensive bibliographic database covering all engineering disciplines. It covers peer-reviewed journals, conference proceedings and trade publications. The coverage period starts from 1870 onwards. With 15 million records across 190 engineering disciplines, Compendex delivers the comprehensive, precise information and insights that researchers need. It is available on Engineering Village platform. It covers more than 1000 journals.

ERIC: The Education Resources Information Center (ERIC) - is an online digital library of education research and information. ERIC is sponsored by the Institute of Education Sciences (IES) of the U.S. Department of Education. ERIC provides ready access to education literature to support the use of educational research and information to improve practice in learning, teaching, educational decision-making, and research. ERIC provides unlimited access to more than 1.4 million bibliographic records of journal articles and other education- related materials, with hundreds of new records added multiple times per week. If possible, links to full text in Adobe PDF format are included. Within the ERIC Collection, you will find records for: journal articles, books, research syntheses, conference papers, technical reports, policy papers, and other education-related materials

Google Scholar: In 2004 Google Inc. introduced Google Scholar a citation database for searching scholarly literature. Google Scholar is a freely available citation database. Because of free availability and indexing different forms of scholarly information (book chapters, conference proceedings, books, pre-print servers and other forms) other than

journals has made Google Scholar a major data source for citation analysis and scholarly information for researchers, librarians and other stakeholders.

Inspec: The Inspec database contains 13 million abstracts and specialized indexing to the world's quality research literature in the fields of electronics, computer science, physics, electrical, control, production and mechanical engineering since late 1960s. It contains index and abstracts of articles selected from nearly 5000 scientific and technical journals (1600 of which are indexed from cover to cover), some 2500 conference proceedings, as well as numerous books, reports, dissertations and scientific videos. It is published by The Institution of Engineering and Technology, Stevenage, Herts., U.K.

Library and Information Science Abstracts (LISA): LISA (maintained by ProQuest) is an international abstracting and indexing tool designed for library professionals and other information specialists. LISA currently abstracts over 440 periodicals from more than 68 countries and in more than 20 different languages, selected conference proceedings, book reviews and research report series. The temporal coverage is from 1969 onward. It indexes approximately around 7000 publications annually. In considering candidate journals at the scholarly end, the editor takes account of a range of standard criteria, e.g., publishing standards, timeliness, editorial content, peer review, international diversity of authorship and citation data.

MathSciNet: It is an electronic database of reviews, abstracts and bibliographic information for much of the mathematical sciences literature. Over 100,000 new items are added each year, most of them classified according to the Mathematics Subject Classification. MathSciNet® contains over 2.8 million items and over 1.6 million direct links to original articles. Bibliographic data from retro digitized articles dates back to the early 1800s. Reference lists are collected and matched internally from approximately 500 journals, and citation data for journals, authors, articles and reviews is provided. This web of citations allows users to track the history and influence of research publications in the mathematical sciences.

PubMed: National Library of Medicne (NLM), United States has been indexing the biomedical literature since 1879, to help provide health professionals access to information necessary for research, health care, and education. What was once a printed index to articles, the Index Medicus, became a database now known as MEDLINE. MEDLINE contains journal citations and abstracts for biomedical literature in many languages from around the world. Since 1996, free access to MEDLINE has been available to the public online via PubMed. It comprises more than 22 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites. About 5 lakh records are added every year. Over 5400 biomedical journals published in the United States and 70 other countries and dating back to the 1940s and updated 5 times/week.

Scopus: It is an abstract and citation database of peer-reviewed literature with smart tools that track, analyze and visualize research. The features of Scopus are as follows: Over

20,500 titles from 5,000 publishers worldwide; Contains 49 million records, 78% with abstracts; Includes over 5.3 million conference papers;

Provides 100% Medline coverage; and Interoperability with Science Direct, Engineering Village and Reaxys, a unique chemistry workflow solution. Its covers only English language items since 1995.

Web of Knowledge (WoK): Thomson Reuters (formerly ISI) Web of Knowledge is today's premier research platform for information in the sciences, social sciences, arts, and humanities. It is a suite of databases containing about 25 different databases. The most important among them are Science Citation Index, Social Science Citation Index, and Arts & Humanities Citation Index. The database includes the following: 23,000 academic and scientific journals (including Web of Science journal listings); 23,000,000 patents; 110,000 conference proceedings; 9,000 websites; Coverage from the year 1900 to present day (with Web of Science); Over 40 million source items; and Integrated and simultaneous searching across multiple databases. Web of Science has been a subject for criticism also. The most important among them is its bias towards English and US. It is said that it does not cover even 10% of India's scholarly journals. Another limitation of WoS is that does not cover books, dissertations and theses, patents and other kinds of literature.

3. Types of Databases:

Selection from these kinds of database depends upon the kind of Scientometric study to be conducted.

3.1 Bibliographic databases or Indexes are good for finding additional materials written about a particular subject. Just like a bibliography at the end of a paper, a bibliographic database can provide you with citations for further study and documentation on a subject. They contain bibliographic information (title of article, journal name, author, date of publication, volume #, issue, page #, etc.) about various types of publications and formats (print, video, audio, software, etc.). Among the databases listed above - CAS, Compendex, ERIC, LISA, Inspec, MathSciNet, and Pubmed are basically bibliographic databases. One can't use these databases for studies which call for cited data.

Citation Databases: On the other hand, the Citation databases slightly differ in their content with that of bibliographic databases. Citation, as you are aware, is a best practice among scholarly community to acknowledge the ideas taken from earlier works. The acknowledgement will be in the form of references at the end of the article. Citation databases are specific for presenting each article included in the database also by the respective list of references in addition to bibliographic record. These lists of references are called cited references. The search according to cited references is more complete because it enables target follow up of a particular topic through all articles on the topic which are included in the database. Citations are presumed to be related to the topic of the current paper by their contents, irrespective of the reasons for their citing. In addition to allowing for literature searching according to topics, citation databases provide data on

the number of citations received by a particular journal, author, or paper. CiteseerX, Web of Knowledge, Google Scholar and Scopus are examples for Citation databases. The databases, like the ones listed above, contribute to the Scientometric studies in two different ways: a) they act as the reliable data sources for Scientometric studies; and b) databases do provide some analytical tools for Scientometric studies.

As a source of data: Different Scientometric studies can be conducted using the databases as discussed in various studies .The following are different fields/data elements in the databases on which one collect data for Scientometric studies.

- Subject oriented fields (e.g. classification codes, descriptors, identifiers, keywords, words in the title, words in the abstract, words in the full text).

- Type of publication (e.g. journal paper, conference paper, book, patent, report, etc.).
- Source (e.g. journal title, CODEN, ISSN number, ISBN number, patent number, year of publication, volume, number of issue, pages, name of publisher, place of publication).
- Responsibility (e.g. name of authors , editors, translators).
- Geographical and institutional information (e.g., country of its editor, name and corporate affiliation of the authors - name of organization, city, country).
- Language(s) of publication.
- Secondary source (e.g. year, volume and number of the abstract).
- Citations or references (eg. in the three ISI citation databases)
- Analytical tools: Manual Scientometric analysis is often cumbersome and tedious in nature. Thanks to the developments in ICT. The databases provide fast, inexpensive, advanced, domain dependent, reliable and reproducible analytical tools. Article counting on different attributes, removal of duplicate items (when multiple sources are used), frequency analysis, defining of subset, ranking on specific criterion, h-index calculation, link analysis, mapping, visual representation, integration with external programs, etc., are all possible with modern databases.

Software/Tools for Scientometric Analyses:

Quantification is important in all aspects of life. Even in scholarly world, the academic and research activities need to be measured. Scientometrics has become a dominant tool for measuring the value of research activity. Collecting and analyzing huge amount of data for Scientometrics is not always an easy task. Thanks to technology. Now we have very qualitative and reliable databases. More than that there are a number of analyzing tools also. These tools are heavily used by Scientometricians. In this section, a bird's-eye-view is presented about the software/tools available for Scientometric analysis. We do not intend, as we proposed in the previous section also, to provide a comprehensive list of all products for want of space and time.

BibExcel:

It is a free-ware for academic and non-profit use. It is developed by Olle Persson of Sweden. Its popularity lies in the fact that it can do most type of commonly done Scientometric analysis. Frequency distribution (Authors, Titles, Citations, or any field specified), and Co-

occurrence analysis (includes Co-citation analysis, Bibliographic coupling, Co-author analysis, Co-word analysis) are the most widely used functionalities in BibExcel. One unique feature of the tool is it uses two counting methods – Whole Counts and Fractional Counts. The distinction between the two is not difficult to understand. For example, for a three-authored article, while the whole count method assigns one count for each author; the fractional count method assigns one third of the count for each author. Both these counting methods are in vogue in Scientometrics.

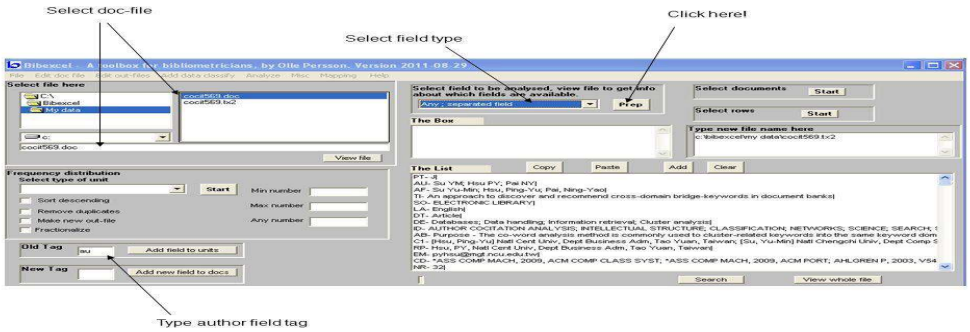
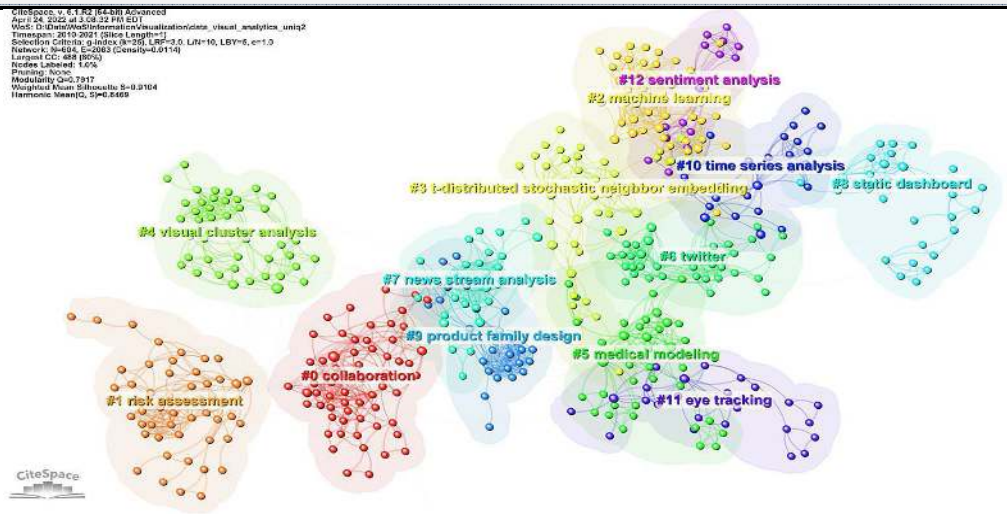


Figure: Bibexcel

A useful feature in Bibexcel is the one that enables us to produce data matrices for export to statistical software. It allows easy interaction with other software, e.g. Pajek, Excel, SPSS, etc. The program offers the user high degree of flexibility in both data management and analysis and this flexibility is one of the program's real strengths. It is, for example, possible to use other data sources than Web of Science, and Bibexcel can in fact deal with data other than bibliographic records.

CiteSpace:

Chaomei Chen created a tool to visualize and analyze trends in scientific literature called CiteSpace. It is a free Java application that can be downloaded by the users. The input data sources for CiteSpace are Web of Knowledge, PubMed, arXiv, ADS, and NSF Award Abstracts. A unique feature of CiteSpace is that records from Derwent World Patents Index can also be visualized.



CiteSpace

A user guide describes the following steps for visualizing information on CiteSpace:

• Collect Data - The primary source for data is Web of Science, and default input data format is ISI Export Format.
• Create a Project - Consists of two directories: input data files and files generated by CiteSpace for analysis and visualization.
• Adjust Parameters - Change time slicing, node types, term sources, term selection, links, pruning, and visualization options.
• Generate Visualizations - Available visualizations include Cluster View, Time-Zone View, Show Networks by Time Slices, and Show Merged Networks.
• Explore Visualizations
• Generate Clusters - CiteSpace uses a spectral clustering algorithm to decompose a network, and the resultant clusters are mutually exclusive (one item to one cluster).
Generate Cluster Labels - Labels can come from three sources: title terms, abstract terms, or index terms.

HistCite:

Eugene Garfield, popularly known as the father of Citation Analysis, developed a new software tool called HistCite for individuals to make it easier for individuals to perform Scientometric analysis and visualization tasks. HistCite is a system designed to help selectively identify the significant (most cited) papers retrieved in topical searches of the Web of Science (SCI, SSCI and/or AHCI). Once a marked list of papers has been created, the resulting Export file is processed by HistCite to create tables ordered by author, year, or citation frequency as well as historiography which include a small percentage of the most-cited papers and their citation links. Scientometric analysis uses the bibliographic information such as authors, titles, dates, author affiliations, references, etc., to measure and/or study various aspects.

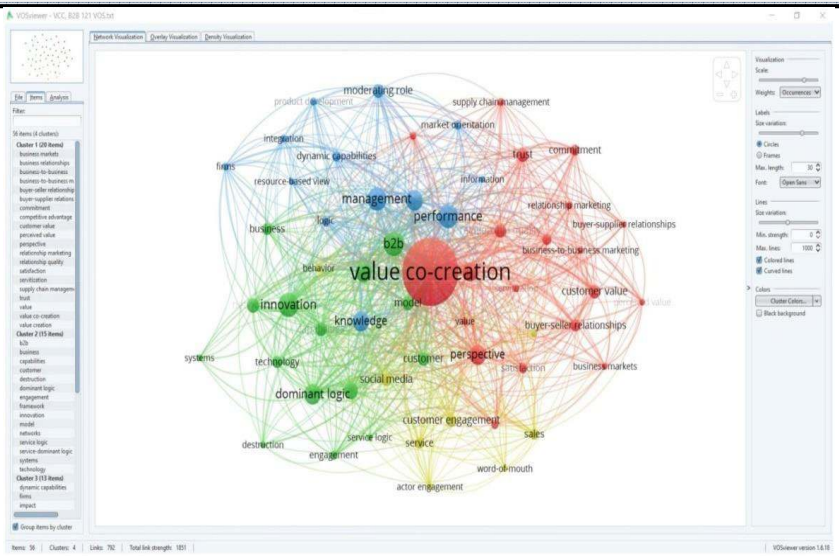


Figure : Histcite

Conclusion

Scientometrics is a quantification tool which uses scientific communication between scholars as the basis for analysis. Journal articles, monographs, blogs & tweets are the different media of communication. The bibliographic and citation data for Scientometric analysis are collected through Questionnaires, Bibliographic databases, Citation databases, Journal indices, Library catalogs and Information systems, Institutional information systems, National databases and so on. There is a subtle difference between Bibliographic databases and Citation databases. Bibliographic databases contain only the bibliographic details whereas the citation databases contain in addition to bibliographic details contain citation data as well. CAS, Compendex, ERIC, LISA, Inspec, MathScinet, and Pubmed are exemplars for bibliographic databases; and CiteseerX, Web of Knowledge, Google Scholar and Scopus are examples for Citation databases. Selection of data sources for Scientometrics is always a tricky question as each source has its own merits and demerits. A few studies have been conducted to compare the relative merits among these databases. Scientometrics software and tools are used for bibliographic analyses. The popular Scientometric software/tools are: BibExcel, CiteSpace, HistCite, Scholarometer, Scholar h-index Calculator and so on.

References

- Aguillo, I. F. (2012). Is Google Scholar useful for bibliometrics? A webometric analysis. Scientometrics, 91(2), 343–351.
- Elaine Bergman. (n.d.). Bibliometrics: From Garfield to Google Scholar. Technology. Retrieved from http://www.slideshare.net/librarian68/upstate-ny-special-librariesassociation-bibliometrics-presentation
- Hood, W. W., & Wilson, C. S. (2003). Informetric studies using databases: Opportunities and challenges. Scientometrics, 58(3), 587–608.
- Hsieh-Yee, I., & Coogan, J. (2010). Google Scholar vs. Academic Search Premier: What libraries and searchers need to know. Presented at the Bridging the Spectrum Symposium, Washington, DC.