


Subject: Library and Information Science

Production of Courseware

 -Content for Post Graduate Courses



Paper No : 10 Informetrics and Scientometrics

Module : 02 Data Sources and Software Tools for Bibliometric Studies



Development Team

Principal Investigator
&
Subject Coordinator

Dr. Jagdish Arora, Director
INFLIBNET Centre, Gandhinagar

Paper Coordinator

Dr I K Ravichandra Rao
Retd Professor, Documentation Research and
Training Centre

Content Writer

Dr N S Harinarayana
Associate Professor, University of Mysore

Content Reviewer

Dr I K Ravichandra Rao
Retd Professor, Documentation Research and
Training Centre

Module 2

Data Sources and Software Tools for Bibliometric Studies

I. Objectives

The objectives of this module are to:

- Understand the parameters used for bibliometric analyses.
- Know various bibliographic and citation databases used as data sources in bibliometric studies
- Appreciate the relative merits and limitations of these databases
- Familiarize with some of the software/tools for bibliometric analysis
- Understand the features of a few software/tools.

II. Learning Outcome

You will gain knowledge about different data sources and the software tools. You learnt only a few databases and a few software tools.

III. Module Structure

1. Introduction
2. Why Scholars' publish?
3. Data Sources for Bibliometrics Studies
 - 3.1 Databases as data sources
 - 3.2 Kinds of Data Sources
4. Case Studies on Comparison of Citation Databases
 - 4.1 Case Study 1
 - 4.2 Case Study 2
 - 4.3 Case Study 3
5. Conclusion on data sources for bibliometric studies
6. Software/Tools for Bibliometrics Analyses
7. Summary
8. References

1. Introduction

Bibliometrics, as you are aware, is a field of study which deals with methods to quantitatively analyze scholarly literature. As a technique bibliometrics 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 bibliometrics 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.

2. Why Scholars' publish?

Scholars consider publishing their works or ideas as a paramount activity. The catchphrase 'publish or perish' is quite popular among scholars and indicates the kind of importance given for publishing activity. In fact as Merton (1957) says, publishing research of their work is an obligation on the part of the scholars. Reward system for the scholars like promotion, recognition, awards and so on is normally based on their publication activity. The three purposes served by the scholarly publication are: spreading scientific findings, protecting intellectual property and gaining popularity among peers.

The publications of scholars are the basis for studies adopting bibliometric techniques. The common metrics used in bibliometric studies include but not limited to the following:

- Article counts with attribution by country (see example 1), by institution and by author (see example 2 and example 4)
- Impact factors (see example 3)
- H-index and other indices (see example 4)
- Citation scores at article level (See example 4)
- Co-citation scores ((the number of times that two papers are cited together in a single paper) (see example 5)
- Visitor numbers (or other info) for online articles
- And many others... e.g., blog entries, tag, etc.

All of these techniques combine to give more detailed and more effective measurements. Results are presented in various forms, such as mapping, in order to depict the relationships between participants and expand the means for analysis.

In this chapter, you will learn about a few data sources and software tools commonly used in bibliometric studies. Please note that we do not intend to provide a comprehensive list of all possible sources/software. The lists, given in the subsequent sections, are only illustrative.

3. Data Sources for Bibliometrics Studies

Data collection for bibliometric study has to be done with care and diligence. The question is where one would get the details about publications? How to collect publication details - say,

by authors, by institutions, by nations and so on? How to get citation data for bibliometric analyses?

Data for Bibliometrics 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 bibliometric studies still revolve round the journals. Recent trends in Bibliometric studies show the use of other digital sources as well. But collection data for bibliometric 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 bibliometrics 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 bibliometric 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.

3.1 Databases as data sources

The data source for a bibliometric 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 bibliometric studies. One may find one or more databases for every established academic discipline. The following are some of the widely used data sources for bibliometrics (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 Medicine (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.

3.2 Kinds of Data Sources

At this juncture, as a student of bibliometrics, it may be useful for you to know a subtle difference between **Bibliographic databases** and **Citation databases**. Selection from these kinds of database depends upon the kind of bibliometric study to be conducted.

- **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. Namely, citations are presumed to be related to the topic of the current paper by their contents, irrespective of the reasons for their citing (i.e. favorable, such as paying credit to, or for criticism and correction). 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 bibliometric studies in two different ways: a) they act as the reliable data sources for bibliometric studies; and b) databases do provide some analytical tools for bibliometric studies. It will be examined below.

- **As a source of data:** Different bibliometric studies can be conducted using the databases as discussed in various studies (Stefaniak, 1987; Deogan, 1987; and Hood & Wilson, 2003). The following are different fields/data elements in the databases on which one collect data for bibliometric 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 (see example 4), editors, translators).
 - Geographical and institutional information (e.g., country of its editor, name and corporate affiliation of the authors - name of organization, city, country (See example 1)).
 - 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) (see example 4).

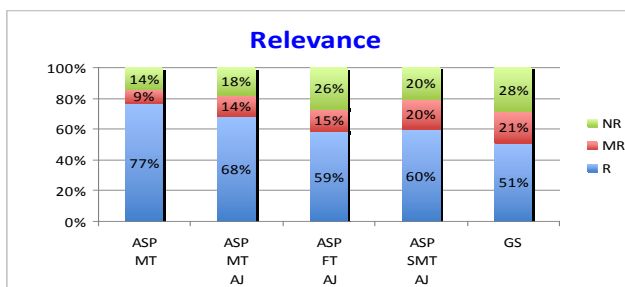
- Analytical tools:** Manual bibliometric analysis is often cumbersome and tedious in nature. Thanks to the developments in ICT. The databases provide fast, inexpensive, advanced, domain dependent, reliable and reproducible analytical tools. Article counting on different attributes, removal of duplicate 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.

4. Case Studies on Comparison of Citation Databases

We have seen in the previous sections that there are a number of databases – bibliographic and citation. The number is growing. The bibliometric researchers always confront with the question which one is the better among their rivals. A few studies have been conducted to compare the databases. The gists of those research works are presented here as case studies.

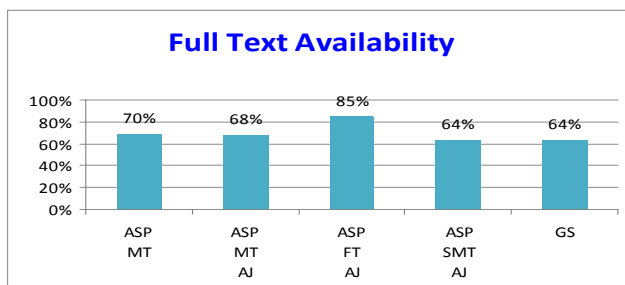
4.1 Case Study 1

Hsieh-Yee and Coogan compared two databases in their work ‘**Google Scholar vs. Academic Search Premier: What Libraries and Searchers Need to Know**’. They framed four different questions for searching the databases. The results were analysed. Only the top10 items of each search set are examined for relevance, full text availability, full text effort, currency, and overlap. The results are as follows in the form of bar graphs¹ which are self explanatory:



Best Relevant Results: **ASP MT** (77% R)
 Worst Relevant Results: **GS** (51% R, plus most NR, most MR)

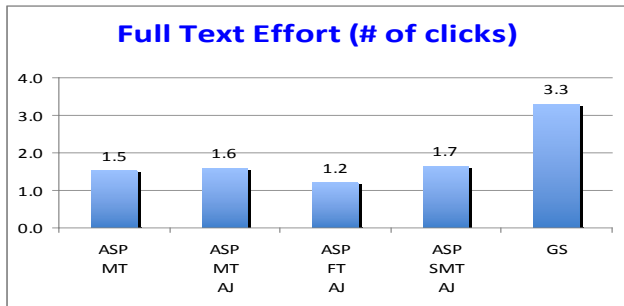
Fig 1: Relevance



Most FT Availability: **ASP FT AJ** (85%)
 Second best: **ASP MT** (70%)
 Least in FT Availability: **GS, ASP SMT AJ** (64%)

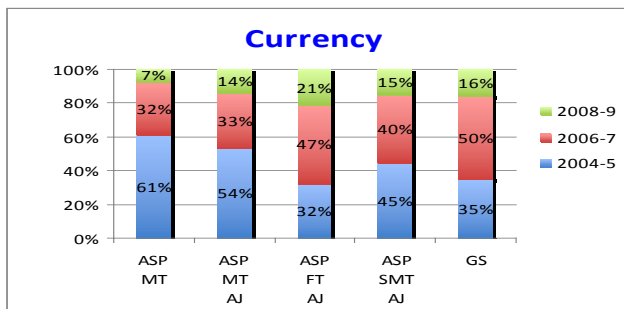
Fig 2: Full text availability

¹ The abbreviations and their full form shown in the results: ASP=Academic Search Premier; MT= Metadata Search; FT= Full Text Search; SMT=Smart Text Search; MT/AJ= Metadata Search for Academic Journals; FT/AJ = Full Text Search for Academic Journals; SMT/AT= Smart Text Search for Academic Journals; GS = Google Scholar



Least effort needed: **ASP FT AJ** (1.2 clicks)
 Most effort needed: **GS** (3.3 clicks)

Fig 3: Full Text Effort



2004-05: **ASP MT** (61%)
 2006-07: **GS** (50%)
 2008-09: **ASP FT AJ** (21%), although neither system retrieves many

Fig. 4: Currency

The authors conclude as follows:

- **ASP outperforms GS in terms of:**
 - Higher Relevance (especially in metadata-only searches)
 - More FT availability and easier access to FT
 - More effective advanced searching
- **GS outperforms ASP in terms of:**
 - “Newer” items (slight advantage when limiting by date)
 - GS retrieves more items, plus more ASP items are indexed by GS than the other way around.
 - Covers some types of materials not readily found in library databases (books, grey literature, materials in institutional repositories)
- **Top 10 results:**
 - Similar searches in ASP and GS produce very different top 10 results.
- **Recommendations:**
 - ASP is a good primary tool, GS is a good supplement.
 - Searchers may want to use both systems to have the best of both worlds.

4.2 Case Study 2

Aguillo's study is on 'Is Google Scholar useful for bibliometrics? A webometric analysis. The results of the study were reported in *Scientometric journal* in 2011. Without looking into the methodology used in the study, we will just see the results which would be sufficient for our purpose here. Some of the results of the study are as follows:

- Google Scholar was not designed as a direct competitor to the other citation databases, being this extra feature (citation counts and links) mainly oriented to improve the searching experience.
- It is really a huge database and Google is clearly intending to enlarge its coverage, not only by adding additional sources but by collecting every type of scientific material available from the public web.
- Our suggestion is that the use of Google Scholar for bibliometric or evaluation purposes should be **done with great care**, especially regarding the items not overlapping with those present in the Scopus or WoK citation databases.
- However, the recent launching of a new service called Google Scholar Citations and the huge update and revamping of Microsoft Academic Search is changing the level of commitment of these engines to the citation analysis, especially for personal description and evaluation purposes.
- The possibilities open to authors to correct errors, modify profiles and combine results, in a typically Web 2.0 fashion, makes these new offerings a serious and free competence to ResearcherID (ISI Thomson) or Scopus Author Identifier services.

4.3 Case Study 3

Meho and Yang have conducted a study on 'Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science Versus Scopus and Google Scholar'. The suggestions made in the study are quite useful in the present context of this chapter. Among many, only some of them are reproduced here:

- The study found that the addition of Scopus citations to those of WoS could significantly alter the ranking of authors.
- The study also found that GS stands out in its coverage of conference proceedings as well as international, non-English language journals, among others. Google Scholar also indexes a wide variety of document types, some of which may be of significant value to researchers and others.
- The use of Scopus and GS, in addition to WoS, reveals a more comprehensive and accurate picture of relationship of LIS with other fields.

5. Conclusion on data sources for bibliometric studies

No one source is suitable in all kinds of situations. Judicious selection of the sources for data collection needs to be made for more meaningful results. Use of multiple sources – one

complementing the others – is the best strategy that can be adopted. Each database has own strengths and weaknesses. The researcher should know them in advance to arrive at meaningful inferences from the data collected from those sources. It is likely that more databases may evolve in future.

6. Software/Tools for Bibliometric Analyses

Quantification is important in all aspects of life. Even in scholarly world, the academic and research activities need to be measured. Bibliometrics has become a dominant tool for measuring the value of research activity. Collecting and analyzing huge amount of data for bibliometrics 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 bibliometricians. In this section, a bird's-eye-view is presented about the software/tools available for bibliometric 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. Only popular products have been discussed. They have been discussed in alphabetical order of their names to avoid any bias.

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 bibliometric 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 bibliometrics.

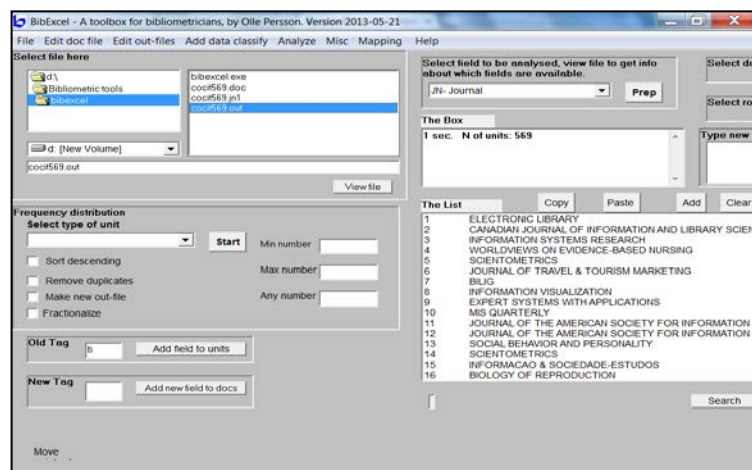


Fig 5: Screenshot of BibExcel

- 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.

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.

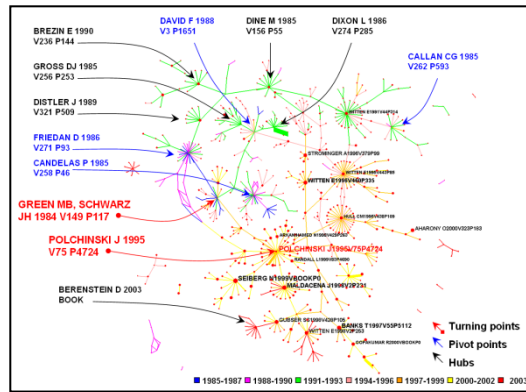


Fig.6: An illustrative example of visualization generated by CiteSpace

Eigenfactor Score: Journals, traditionally speaking, are the most valued source of information and communication for scientists. They continue to be so. Scholars like to publish their ideas in the important journals. Bibliometrics arguably provides some parameters for determining the importance of journals. One such parameter is citations. Eigenfactor is another parameter developed by Jevin West and Carl Bergstrom at the University of Washington. Journals are rated according to the number of incoming citations, with citations from highly ranked journals weighted to make a larger contribution to the eigenfactor than those from poorly ranked journals. The Eigenfactor score is intended to measure the importance of a journal to the scientific community, by considering the origin of the incoming citations, and is thought to reflect how frequently an average researcher would access content from that journal.



Fig 7: Eigenfactor

The Eigenfactor approach is thought to be more robust than the impact factor metric, which purely counts incoming citations without considering the significance of those citations. While the Eigenfactor score is correlated with total citation count for medical journals, these metrics provide significantly different information. For a given number of citations, citations from more significant journals will result in a higher Eigenfactor score.

Eigenfactor scores and Article Influence scores are calculated by eigenfactor.org. Eigenfactor scores are measures of a journal's importance. It can be used in combination with H-index to evaluate the work of individual scientists.

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 bibliometric 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. Bibliometric analysis uses the bibliographic information such as authors, titles, dates, author affiliations, references, etc., to measure and/or study various aspects. Some typical questions asked by bibliometricians that can be answered by HistCite analysis are:

- How much literature has been published in this field? When and in what countries has it been published? What countries are the major contributors to this field? What are the languages most frequently used by the items published in this field?
- What journals cover the literature of the field? Which are the most important?
- Who are the key authors in this field? What institutions do these authors represent?
- Which articles are the most important?
- How have the various contributors to the field influenced each other?

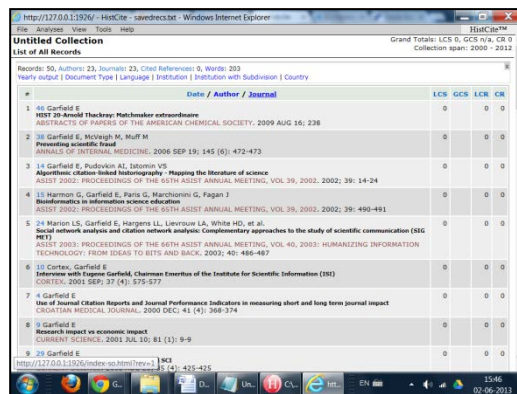


Fig. 8: Screen Shot of HistCite Analysis Window

HistCite can directly be integrated with Web of Knowledge (WoK) database of Thomson-Reuters, i.e., the data exported from Web of Knowledge can be read into HistCite. In order to do utilize this facility, the WoK search results have to be saved as a 'plain text' format. On the other hand, HistCite is not yet ready to read directly from other databases. Bibliographies from other sources can be manually entered into HistCite.

Pajek: It is software for analyses and visualization of huge networks with a large to very large number of vertices. Pajek, an unusual name in English, means a spider in Slovenian language. It was started in the year 1996 and developed into one of the most popular

software in the field of visualization. Parjek is very useful tool in areas like organic chemistry, genealogy, data mining, diffusion networks etc. It can also be used in bibliometrics to visualize the collaboration and citation networks. Pajek is developed by Vladimir Batagelj and Andrej Mrvar. Some procedures were contributed also by Matjaž Zaveršnik.

Publish or Perish: It is a popular software program among scholars that retrieves and analyzes academic citations. It is developed and maintained by A.W. Harzing. It is a valuable programme that combats many of the problems of interpreting Google Scholar outputs and allows academics to easily check their own or others' performance. It presents academic outputs quickly and computes excellent citation statistics about each author's work, including an overall 'times cited' score and times cited per year since publication. It uses Google Scholar to obtain the raw citations, then analyzes these and presents the following statistics:

- Total number of papers
- Total number of citations
- Average number of citations per paper
- Average number of citations per author
- Average number of papers per author
- Average number of citations per year
- Hirsch's h-index and related parameters
- Egghe's g-index
- The contemporary h-index
- The age-weighted citation rate
- Two variations of individual h-indices
- An analysis of the number of authors per paper.

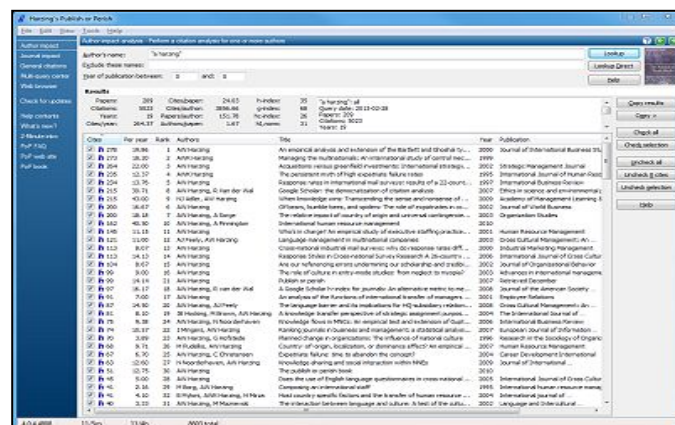


Fig 9: Screenshot of Publish or Perish

The results are available on-screen and can also be copied to the Windows clipboard (for pasting into other applications) or saved to a variety of output formats (for future reference or further analysis). Publish or Perish includes a detailed help file with search tips and additional information about the citation metrics.

Scholarometer: Scholarometer (previously Tenurometer) is called so as it provides service to scholars by computing citation-based impact measures. It is a social tool to facilitate citation analysis and help evaluate the impact of an author's publications. It is a browser extension/plugin presently compatible with Google Chrome and Mozilla Firefox. Being a platform independent tool, it runs on all systems that support Chrome and Firefox. It is an easy tool to use even by a non-expert. Scholarometer helps authors and academic administrators evaluate the impact of someone's research publications, citation-based impact measures. The figure shows how Scholarometer works.

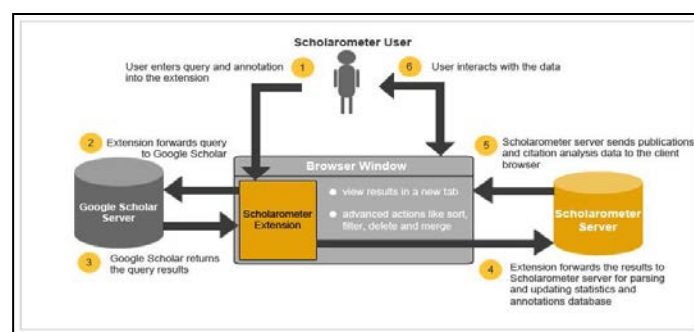


Fig.10: Scholarometer Process

Using Scholarometer, one can compute Hirsch's h-index, Egghe's g-index, and Schreiber's hm index. The latest version of Scholarometer can also calculate the new universal h-index (developed by Radicchi, Fortunato and Castellano).

One of the useful features of Scholarometer is that it allows filtering, sorting, deleting and live search to compute error free impact measures. For example, the user can merge multiple versions of the same paper; exclude papers by different authors with the same name, or other noisy data; filter papers by many criteria such as years, disciplines, name variations, and coauthors; and perform live search over the results. The impact measures are dynamically recalculated based on the user's manipulations.

Scholarometer users can save the finding into formats appropriate for local reference management software (e.g., EndNote), or for social publication sharing systems (e.g., BibSonomy). Currently, the system supports the following export formats: BibTex (BIB), RefMan (RIS), EndNote (ENW), comma-separated values (CSV), tab-separated values (XLS), and BibJSON.

Scholar h-index calculator: Scholar H-Index Calculator (hereafter called just the Calculator in this section) is an add-on for Google Chrome and Firefox which enhances Google Scholar results pages by showing a number of bibliometric data computed using the data appearing on video as input. Once installed, the Calculator works transparently when querying Google Scholar: as soon as you make a query, result pages are enriched with a number of useful data (e.g. the h-index computed on the basis of displayed data), and new functions are available.

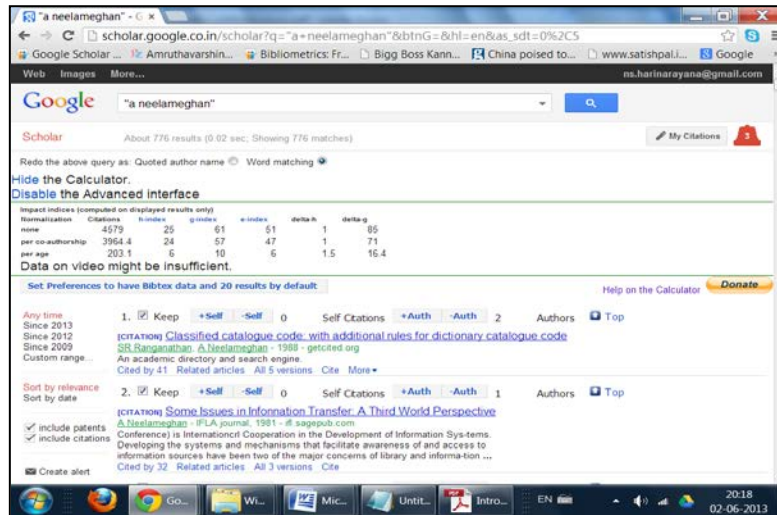


Fig.11: Scholar H-Index Calculator's Output for the author Prof. A. Neelameghan

Using the Calculator is quite easy. After the installation in the browser (Google Chrome or Firefox), the Google Scholar has to be used as usual. Once installed, the add-on displays on top of Google Scholar result pages, the corresponding h-index, g-index, e-index and other measures of impact for the submitted query. In comparison with other tools, the Calculator has many pleasant features. It is now possible to select or deselect a single paper; to manually increase or decrease the number of self citations; to manually increase or decrease the number of authors for a given paper; and to load and save data.

One interesting feature of the Calculator is that it provides the measures called delta-h and delta-g. These two values measure the minimum number of citations needed for incrementing the current h-index (g-index, respectively), by 1. delta-h and delta-g are the measure of how difficult would be for the author at hand to increase his/her h and g-index.

The latest version (2.3) has added features for normalization per author and normalization per age. In addition it also has a function called author list refinement.

7. Summary

Bibliometrics 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 bibliometric 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 bibliometrics 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.

Bibliometrics software and tools are used for bibliographic analyses. The popular bibliometric software/tools are: BibExcel, CiteSpace, Eigenfactor Score, HistCite, Pajek, Publish or Perish, Scholarometer, Scholar h-index Calculator and so on.

8. References

1. Aguillo, I. F. (2012). Is Google Scholar useful for bibliometrics? A webometric analysis. *Scientometrics*, 91(2), 343–351.
2. Elaine Bergman. (n.d.). Bibliometrics: From Garfield to Google Scholar. Technology. Retrieved from <http://www.slideshare.net/librarian68/upstate-ny-special-libraries-association-bibliometrics-presentation>
3. Hood, W. W., & Wilson, C. S. (2003). Informetric studies using databases: Opportunities and challenges. *Scientometrics*, 58(3), 587–608.
4. 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.
5. Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105–2125.
6. Merton, R. K. (1957). *Social Theory and Social Structure*. New York: Free Press.
7. Roemer, R. C., & Borchardt, R. (2012). From bibliometrics to altmetrics A changing scholarly landscape. *College & Research Libraries News*, 73(10), 596–600.
8. Stefaniak, B. (1987). Use of bibliographic data bases for scientometric studies. *Scientometrics*, 12(3), 149–161.