SOFTWARE TOOLS FOR BIBLIOMETRICS.

Dr. Siddiqui Eraj
Librarian
Dr. Rafiq Zakeria College for Women
Dr. B. A. M. University,
Aurangabad, Maharashtra, India
Email: siddiquieraj@gmail.com

Abstract: *This paper gives us an overview of Bibliometrics and its online tools which are available for analysis of books, articles, or other publications*

Key words: Bibliometrics; Google Scholar; Web of Science; Scopus.

Defination:- "Bibliometrics is a statistical analysis of books, articles, or other publications."

Oxford Dictionaries

Introduction

Bibliometrics is a set of methods to quantitatively analyze academic literature. Citation analysis and content analysis are commonly used bibliometric methods. While bibliometric methods are most often used in the field of library and information science, bibliometrics have wide applications in other areas. Many research fields use bibliometric methods to explore the impact of their field, the impact of a set of researchers, or the impact of a particular paper.

Usage

Historically bibliometric methods have been used to trace relationships amongst academic journal citations. Citation analysis, which involves examining an item's referring documents, is used in searching for materials and analyzing their merit. Citation indices, such as Institute for Scientific Information's Web of Science, allow users to search forward in time from a known article to more recent publications which cite the known item.

Data from citation indexes can be analyzed to determine the popularity and impact of specific articles, authors, and publications. Using citation analysis to gauge the importance of one's work, for example, is a significant part of the tenure review process. Information scientists also use citation analysis to quantitatively assess the core journal titles and watershed publications in particular disciplines; interrelationships between authors from different institutions and schools of thought; and related data about the sociology of academia.

Some more pragmatic applications of this information includes the planning of retrospective bibliographies, "giving some indication both of the age of material used in a discipline, and of the extent to which more recent publications supersede the older ones;" indicating through high frequency of citation which documents should be archived; comparing the coverage of secondary services which can help publishers gauge their achievements and competition, and can aid librarians in evaluating "the effectiveness of their stock". There are also some limitations to the value of citation data. They are often incomplete or biased; data has been largely collected by hand (which is expensive), though citation indexes can also be used; incorrect citing of sources occurs continually; thus, further investigation is required to truly understand the rationale behind citing to allow it to be confidently applied.

Bibliometrics are now used in quantitative research assessment exercises of academic output which is starting to threaten practice based research. The UK government is considering using bibliometrics as a possible auxiliary tool in its Research Excellence Framework, a process which will assess the quality of the research output of UK universities and on the basis of the assessment results, allocate research funding.

Other bibliometrics applications include: creating thesauri; measuring term frequencies; as metrics in scientometric analysis, exploring grammatical and syntactical structures of texts; measuring usage by readers.

History

The term bibliometrics was coined by Alan Pritchard in a paper published in 1969, titled Statistical Bibliography or Bibliometrics? He defined the term as "the application of mathematics and statistical methods to books and other media of communication".

Although citation analysis is not new (the Science Citation Index began publication in 1961), before it ould be calculated by computers it was done manually and so was time-consuming. Automated algorithms are making it much more useful, versatile, and widespread. This led to the creation of the new field of computational bibliometrics. The first such algorithm for automated citation extraction and indexing was by CiteSeer. Google's PageRank is based on the principle of citation analysis. Patent citation maps are also based upon citation analysis (in this case, the citation of one patent by another).

Online Tools

1. **Google Scholar Citations**

Google Scholar is a freely accessible web search engine that indexes the full text of scholarly literature across an array of publishing formats and disciplines. Released in beta in November 2004, the Google Scholar index includes most peer-reviewed online journals of Europe and America's largest scholarly publishers, plus scholarly books and other non-peer reviewed journals. While Google does not publish the size of Google Scholar's database, third-party researchers estimated it to contain roughly 160 million documents as of May 2014.

"KNOWLEDGE LIBRARIAN"- AN INTERNATIONAL PEER REVIEWED BILINGUAL E-JOURNAL OF LIBRARY AND INFORMATION SCIENCE
Volume: 01, Issue: 01, Sept. – Oct. 2014 eISSN NO. 2394-2479

History

Google Scholar arose out of a discussion between Alex Verstak and Anurag Acharya, both of whom were then working on building Google's main web index.

In 2006, in response to release of Microsoft's Windows Live Academic Search, a potential competitor for Google Scholar, a citation importing feature was implemented using bibliography managers (such as RefWorks, RefMan, EndNote, and BibTeX). Similar features are also part of other search engines, such as CiteSeer and Scirus.

In 2007, Acharya announced that Google Scholar had started a program to digitize and host journal articles in agreement with their publishers, an effort separate from Google Books, whose scans of older journals do not include the metadata required for identifying specific articles in specific issues.

In 2011, Google removed Scholar from the toolbars on its search pages, making it both less easily accessible and invisible to users who were not already aware of its existence.

In 2012, an individual Google Scholar page feature was added. Individuals, logging on through a Google account with a bona fide address usually linked to an academic institution, can now create their own page giving their fields of interest and citations. Google Scholar automatically calculates and displays the individual's total citation count, h-index, and i10-index. Examples are here and here. Some academic journals have their own Google Scholar Page; see e.g. the International Journal of Internet Science. Top citations in a field of interest can also be accessed, e.g. here.

A feature introduced in November 2013 allows logged-in users to save search results into the "Google Scholar library", a personal collection which the user can search separately and organize by tags.

Volume: 01, Issue: 01, Sept. – Oct. 2014 eISSN NO. 2394-2479

Google Scholar offers "Scholar Citations profiles", public author profiles that are editable by authors themselves. According to Google, "three quarters of Scholar search results pages show links to the authors' public profiles" as of August 2014.

Google Scholar's advertising slogan – "Stand on the shoulders of giants" – is taken from a quote by Isaac Newton and is a nod to the scholars who have contributed to their fields over the centuries, providing the foundation for new intellectual achievements.

Features and specifications

Google Scholar allows users to search for digital or physical copies of articles, whether online or in libraries. It indexes "full-text journal articles, technical reports, preprints, theses, books, and other documents, including selected Web pages that are deemed to be 'scholarly.'" Because many of Google Scholar's search results link to commercial journal articles, most people will be able to access only an abstract and the citation details of an article, and have to pay a fee to access the entire article. The most relevant results for the searched keywords will be listed first, in order of the author's ranking, the number of references that are linked to it and their relevance to other scholarly literature, and the ranking of the publication that the journal appears in.

Using its "group of" feature, it shows the available links to journal articles. In the 2005 version, this feature provided a link to both subscription-access versions of an article and to free full-text versions of articles; for most of 2006, it provided links to only the publishers' versions. Since December 2006, it has provided links to both published versions and major open access repositories, but still does not cover those posted on individual faculty web pages; access to such self-archived non-subscription versions is now provided by a link to Google, where one can find such open access articles.

Through its "cited by" feature, Google Scholar provides access to abstracts of articles that have cited the article being viewed. It is this feature in particular that provides the citation indexing previously only found in Scopus and Web of Knowledge. Through its "Related articles" feature, Google Scholar presents a list of

Volume: 01, Issue: 01, Sept. – Oct. 2014 eISSN NO. 2394-2479

closely related articles, ranked primarily by how similar these articles are to the original result, but also taking into account the relevance of each paper.

As of July 2013, Google Scholar is not yet available to the Google AJAX API.

Google Scholar’s legal database of US cases is extensive. Users can search and read published opinions of US state appellate and supreme court cases since 1950, US federal district, appellate, tax and bankruptcy courts since 1923 and US Supreme Court cases since 1791. Google Scholar embeds clickable citation links within the case and the How Cited tab allows lawyers to research prior case law and the subsequent citations to the court decision. The Google Scholar Legal Content Star Paginator extension inserts Westlaw and LexisNexis style page numbers in line with the text of the case.

Ranking algorithm

While most academic databases and search engines allow users to select one factor (e.g. relevance, citation counts, or publication date) to rank results, Google Scholar ranks results with a combined ranking algorithm in a "way researchers do, weighing the full text of each article, the author, the publication in which the article appears, and how often the piece has been cited in other scholarly literature". Research has shown that Google Scholar puts high weight especially oncitation counts and words included in a document's title. As a consequence the first search results are often highly cited articles.

2. Scopus

Scopus is a bibliographic database containing abstracts and citations for academic journal articles. It covers nearly 21,000 titles from over 5,000 publishers, of which 20,000 are peer-reviewed journals in the scientific, technical, medical, and social sciences (including arts and humanities). It is owned by Elsevier and is available online bysubscription. Searches in Scopus also incorporate searches of patent databases.

Since Elsevier is the owner of Scopus and is also one of the main international publishers of scientific journals, an independent and international Scopus Content Selection and Advisory Board was established to prevent a potential conflict of interest in the choice of journals to be included in the database and to maintain an open and transparent content coverage policy, regardless of publisher. The board consists of scientists and subject librarians.

A 2008 study compared PubMed, Scopus, Web of Science, and Google Scholar and concluded:

"PubMed and Google Scholar are accessed for free. Scopus offers about 20% more coverage than Web of Science, whereas Google Scholar offers results of inconsistent accuracy. PubMed remains an optimal tool in biomedical electronic research. Scopus covers a wider journal range but it is currently limited to recent articles (published after 1995) compared with Web of Science. Google Scholar, as for the Web in general, can help in the retrieval of even the most obscure information but its use is marred by inadequate, less often updated, citation information."

Evaluating ease of use and coverage of Scopus and the Web of Science (WOS), a 2006 study concluded that "Scopus is easy to navigate, even for the novice user. The ability to search both forward and backward from a particular citation would be very helpful to the researcher. The multidisciplinary aspect allows the researcher to easily search outside of his discipline" and "One advantage of WOS over Scopus is the depth of coverage, with the full WOS database going back to 1945 and Scopus going back to 1966. However, Scopus and WOS complement each other as neither resource is all inclusive. ".

Scopus also offers author profiles which cover affiliations, number of publications and their bibliographic data, references, and details on the number of citations each published document has received. It has alerting features that allows registered users to track changes to a profile and a facility to calculate authors' h-index.

Scopus can be integrated with ORCID.

3. **Web of Knowledge**

Web of Knowledge (formerly known as ISI Web of Knowledge) is an academic citation indexing and search service, which is combined with web linking and is provided by Thomson Reuters. Web of Knowledge covers the sciences, social sciences, arts and humanities. It provides bibliographic content and tools to access, analyze, and manage research information. Multiple databases can be searched simultaneously.

Overview

Web of Knowledge is described as a unifying research tool which enables the user to acquire, analyze, and disseminate database information in a timely manner. This is accomplished because of the creation of a common vocabulary, called ontology, for varied search terms and varied data. Moreover, search terms generate related information across categories.

Acceptable content for Web of Knowledge is determined by an evaluation and selection process based on the following criteria: impact, influence, timeliness, peer review, and geographic representation.

Search and analysis

Web of Knowledge employs various search and analysis capabilities. First, citation indexing is employed, which is enhanced by the capability to search for results across disciplines. The influence, impact, history, and methodology of an idea can be followed from its first instance, notice, or referral to the present day. This technology points to a deficiency with the keyword-only method of searching.

Second, subtle trends and patterns relevant to the literature or research of interest, become apparent. Broad trends indicate significant topics of the day, as well as the history relevant to both the work at hand, and particular areas of study.

Third, trends can be graphically represented.

Content

The combined databases 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
- Integrated and simultaneous searching across multiple databases

4. SCImagi Journal and Country Rank

SCImago Journal Rank (SJR indicator) is a measure of scientific influence of scholarly journals that accounts for both the number of citations received by a journal and the importance or prestige of the journals where such citations come from. The SJR indicator is a variant of the eigenvector centrality measure used in network theory. Such measures establish the importance of a node in a network based on the principle that connections to high-scoring nodes contribute more to the score of the node. The SJR indicator, which is inspired by the PageRank algorithm, has been developed to be used in extremely large and heterogeneous journal citation networks. It is a size-independent indicator and its values order journals by their "average prestige per article" and can be used for journal comparisons in science evaluation processes.

The *SJR indicator* is a free journal metric which uses an algorithm similar to PageRank and provides an alternative to the impact factor (IF), which is based on data from the Science Citation Index. Average citations per document in a 2-year period, abbreviated as Cites per Doc. (2y), is another index that measures

"KNOWLEDGE LIBRARIAN"- AN INTERNATIONAL PEER REVIEWED BILINGUAL E-JOURNAL OF LIBRARY AND INFORMATION SCIENCE
Volume: 01, Issue: 01, Sept. – Oct. 2014 eISSN NO. 2394-2479

the scientific impact of an average article published in the journal. It is computed using the same formula that journal impact factor (Thomson Reuters).

Rationale

If scientific impact is considered related to the number of endorsements, in the form of citations, a journal receives, then prestige can be understood as a combination of the number of endorsements and the prestige or importance of the journals issuing them. The *SJR indicator* assigns different values to citations depending on the importance of the journals where they come from. This way, citations coming from highly important journals will be more valuable and hence will provide more prestige to the journals receiving them. The calculation of the *SJR indicator* is very similar to the *Eigenfactor score*, with the former being based on theScopus database and the latter on the ISI Web of Science database.

Computation

The SJR indicator computation is carried out using an iterative algorithm that distributes prestige values among the journals until a steady-state solution is reached. The SJR algorithm begins by setting an identical amount of prestige to each journal, then using an iterative procedure, this prestige is redistributed in a process where journals transfer their achieved prestige to each other through citations. The process ends up when the difference between journal prestige values in consecutive iterations do not reach a minimum threshold value any more. The process is developed in two phases, (a) the computation of *Prestige SJR* (*PSJR*) for each journal: a size-dependent measure that reflects the whole journal prestige, and (b) the normalization of this measure to achieve a size-independent measure of prestige, the *SJR indicator*.

References

- Bibliometrics. (n.d.). Retrieved September 17, 2014, from http://www.oxforddictionaries.com/definition/english/bibliometrics
- Braun, T., Bujdosó, E., & Schubert, A. (1987). Literature of analytical chemistry: A scientometric evaluation. Boca Raton, Fla: CRC Press.
- Bibliometrics. (2014, September 16). Retrieved November 17, 2014, from http://en.wikipedia.org/wiki/Bibliometrics
- Google Scholar. (2014, September 16). Retrieved November 17, 2014, from http://en.wikipedia.org/wiki/Google_Scholar
- Scopus. (2014, November 16). Retrieved November 17, 2014, from http://en.wikipedia.org/wiki/Scopus
- Web of Knowledge. (2014, November 16). Retrieved November 17, 2014, from http://en.wikipedia.org/wiki/Web_of_Knowledge
- SCImago Journal Rank. (2014, October 27). Retrieved November 17, 2014, from http://en.wikipedia.org/wiki/SCImago_Journal_Rank