

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

2019

WEBOMETRICS RESEARCH METHODS ADOPTED IN LIBRARY AND INFORMATION SCIENCE: AN OVERVIEW

murugan andham

murugatheni1972@gmail.com

murugan andham

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>



Part of the [Library and Information Science Commons](#)

andham, murugan and andham, murugan, "WEBOMETRICS RESEARCH METHODS ADOPTED IN LIBRARY AND INFORMATION SCIENCE: AN OVERVIEW" (2019). *Library Philosophy and Practice (e-journal)*. 2869.

<https://digitalcommons.unl.edu/libphilprac/2869>

WEBOMETRICS RESEARCH METHODS ADOPTED IN LIBRARY AND INFORMATION SCIENCE: AN OVERVIEW

Muruganandham.G.,

*Librarian, Meenakshi Medical College Hospital & Research Institute, Kanchipuramai

E-mail: murugatheni1972@gmail.com

Abstract

At the moment, the technologies and applications have made the internet an important medium of communication and information dissemination. It provides a much enhanced interface between the user and information. The web set the foundation for global communication of computer users. Increasing the importance of web is clearly seen in the form of references or citations in the present day publications. Studies of the Web have been named webometrics or cybermetrics as in the electronic journal of that name. The web in particular has made all kinds of information easier to access and has encouraged people to make information publically available that they previously would not have thought to do. It was a huge advantage for researchers and students because the information needed to study a phenomenon might be freely available online, saving time in the data collection phase of a project and making larger scale studies possible. webometrics research is having some impact in other disciplines in the sense of attracting citations. These include the cognate disciplines of computer science, Sociology, Physics and Complex science. For some years bibliometric studies have identified webometrics as one of the largest information science fields. Nevertheless, most published webometrics research is relatively theoretical and, as a new research field, seems unlikely to survive unless it is useful in some way. In general it is very important for academics to identify hot research topics and avoid unpromising ones and so methods are needed to help this identification process. Moreover, there is a significant amount of citation of webometrics research by disciplines outside information science, including computing, sociology, physics and complex science. Nevertheless, it seems that the potential user base for current webometric techniques is wider still, creating a need for awareness-raising.

Keywords: Webometrics, Research Methods, bibliometric, Library and Information Science, Computer Science.

Introduction

Nowadays, the technologies and applications have made the internet an important medium of communication and information dissemination. It provides a much enhanced interface between the user and information. The web set the foundation for global communication of computer users. Increasing the importance of web is clearly seen in the form of references or citations in the present day publications. Studies of the Web have been named webometrics or cybermetrics as in the electronic journal of that name. The web in particular has made all kinds of information easier to access and has encouraged people to make information publically available that they previously would not have thought to do. It was a huge advantage for researchers and students because the information needed to study a phenomenon might be freely available online, saving time in the data collection phase of a project and making larger scale studies possible.

The simplest way to investigate a web phenomenon is probably to read what the participants have written, look at their pictures and videos, or listen to their recordings, as appropriate. For example, reading the comments on a YouTube video may give insights into the audience reaction to the video or the issues within it.

Thelwall (2009)¹ defined webometric is "the study of Web-based content with primarily quantitative methods for social science research goals using techniques that are not specific to one field of study". According to Bjerneborn and Ingsversan (2004)² webometrics is, "the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web drawing on bibliometric and informetric approaches".

Webometrics is concerned with measuring aspects of the web: web sites, web pages, parts of web pages, words in web pages, hyperlinks, web search engine results. The importance of the web itself as a communication medium and for hosting an increasingly wide array of documents, from journal articles to holiday brochures, needs no introduction. Given this huge and easily accessible source of information, there are limitless possibilities for measuring or counting on a huge scale (e.g., the number of web sites, the number of web pages, the number of

blogs) or on a smaller scale (e.g., the number of web sites in Ireland, the number of web pages in the CNN web site).

REVIEW OF LITERATURE

Jeyashankar and Ramesh Babu (2009)³ examined the Webometric study of Tamilnadu Universities. Their study revealed that some universities in Tamil Nadu had higher number of web pages, but correspondingly their link pages were very small in number and their websites fell behind in their simple, self link and external link web impact factor.

Elgohary (2008)⁴ studied the Webometric analyses of Arab Universities on the web and found Web Impact Factors of Arab universities as well as the relation to other variables such as language. The paper provided some measures to the universities and how they were viewed by the outside world.

Konnur & Rajani (2010)⁵ examined that webometric studies are researching to a new conquest by improving their service base especially with application of Information technology for access and delivery of e-content to their clients. Archana (2010)⁶ made an attempt to understand and analyze the presence and presentation of libraries of Engineering Colleges (EC) in Kerala in their respective websites, and found that websites of academic institutions are the prime sources of information about the institution. Libraries, being the main provider of information for the academics, need to be represented in the respective homepages with due importance but it is observed that even the highest scoring libraries satisfy only half of the criteria listed for analysis. It is high time for the library professionals associated with these institutions to be serious about the role of library websites in providing marketing and quality services.

Thanuskodi (2011)⁷ studied of Webometric Analysis of Private Engineering College Websites in Tamil Nadu. He reported that the websites is updated only in few engineering colleges remaining websites did not mentions time or date in the homepage. He further stated that webometric techniques are still in experimental stage in testing whether the classical bibliometric methods applied to the web are reliable and feasible means of comparing the analysis of websites.

Islam and Alam (2011)⁸ studied and analysed websites of private universities in Bangladesh. Their study revealed that some private universities in Bangladesh have higher number of web pages, but their link pages are very small in number, thus the websites fell behind in their Overall WIF, self link, external links and Absolute WIF. Finally, it was shown that these universities did not have much impact factor on the web and were not known internationally.

Similarities

Webometrics displays several similarities to informetric and scientometric studies and the application of common bibliometric methods. For instance, simplistic counts and content analysis of web pages are like traditional publication analysis; counts and analyses of outgoing links from web pages, here named outlinks, and of links pointing to web pages, called inlinks, can be seen as reference and citation analyses, respectively.

Outlinks and inlinks are then similar to references and citations, respectively, in scientific articles. However, due to its dynamic and distributed nature, the Web often demonstrates web pages simultaneously linking to each other. In this case not possible in the traditional paper-based citation world. The coverage of search engines of the total Web can be investigated in the same way as the coverage of domain and citation databases in the total document landscape and possible overlaps between engines detected. Since the Web consists of contributions from anyone who wishes to contribute, the quality of information or knowledge value is opaque due to the lack of kinds of peer reviewing; but citation-like link analyses may reveal clusters of sites to be reviewed. Patterns of Web search behaviour can be investigated as in traditional information seeking studies.

Issue tracking on the Web is carried out and knowledge discovery attempts are made, similar to common data or text mining in administrative or textual (bibliographic) databases. Since the Web is an information space quite different from the common scientific or professional databases, the similarities mentioned above may sometimes be superficial.

WEBOMETRICS RESEARCH METHOD

Webometrics, the quantitative study of Web-based phenomena was driven rapidly forward by a number of pioneering researchers and investigations to establish itself as apparently the largest coherent field within library and information science.

Three main directions for early research:

- link impact evaluation,
- link relationship mapping and
- search engine results analysis,

But from the perspective of digital library evaluation the main method is link analysis.

Link impact evaluation studies typically identified methods to count the number of hyperlinks to sets of academic-related Web sites or pages (e.g., universities, departments, open access journal Web sites) and evaluated whether the number of links found could reasonably be used to indicate the research impact of the target site or page.

Link relationship mapping investigations typically chart a coherent set of academic Web sites in two dimensions, either as a network diagram or as coordinates in a two-dimensional space using multi-dimensional scaling. The similarity measures used to construct the diagrams are either the number of links between pairs of Web sites or co-in link counts: the number of Web pages linking to both target pages and sites.

Link analysis can be used to evaluate a web site by comparing it to its competitors or to similar web sites and can also be used to identify missed audiences for a site. Links can reveal information about web sites because each link to a web site may be created to direct visitors to it. The link author believes that the target site is important or useful. For example, the course pages for an archaeology degree may contain links to the New Library of Alexandria for its images of ancient Egyptian artefacts. From the opposite perspective, discovering all the links to the New Library of Alexandria web site would give useful insights into who was using it and why. Of course, most people using a web site will not create a link to it but a link analysis can still give indicators about likely users and uses. Links can reveal information about where a web site is used because many links originate from web sites with a top-level domain that identifies

their geographic origin (e.g., .uk for UK, .es for Spain). Links can also reveal why a web site is used through reading the web pages hosting the links.

Search engine results analysis had the objective to assess the comprehensiveness and consistency of commercial search engine results, either because of their use for the raw data in much webometrics research or because of their wider importance as an information retrieval tool by Web users.

Webometrics research included many studies of visualisations and covered an increasing variety of topics, such as longitudinal quantitative studies of Web pages, Web citation analysis, blog analysis, business Web sites, digital library analysis, information retrieval factors, social network site investigations and theoretical contributions. Despite the bibliometric studies suggesting the importance of webometrics to information science, it may soon be surpassed by other emerging topics, such as the h-index or Web 2.0, and so it is a suitable time to conduct a review of the field in order to assess whether it is delivering enough useful knowledge to safeguard its future within library and information science.

Objectives:

- to point to selected areas of webometric research that demonstrate interesting progress and space for development as well as to some currently less promising areas.
- to identify the reasons for introducing Webometrics in library and information science
- to explore various components of Webometrics used in library and information science

Limitations

- The definition of webometrics selected, based on researchers more than content, would not necessarily be universally accepted as the best choice.
- webometrics research is having some impact in other disciplines in the sense of attracting citations. These include the cognate disciplines of computer science, Sociology, Physics and Complex science.

Research areas bypassing webometrics

The Google Scholar searches for webometrics and link analysis studies not citing webometrics produced articles mainly from computer science and sociology. Some were isolated

studies but others formed coherent bodies of work. In addition, physics research known to the author but not found by the searches is discussed.

Computer Science

Many different Web-related computer science topics use link analysis but do not cite webometrics research. The topics include search engine ranking algorithms, Web information retrieval, question answering, link analysis, link mining, crawling, Web page classification, and spam filtering. This shows that there is a significant body of link analysis research that does not need webometrics. This could be because the computer science research is typically concerned with algorithm *construction* (e.g., *'Discovering authorities in question answer communities by using link analysis'*, *'Link analysis ranking: algorithms, theory, and experiments'*) whereas webometrics normally analyses the outputs of algorithms, giving a different perspective. In addition, there is some computer science research into traditional library and information science topics like scholarly communication (e.g., *'Semantic Web link analysis to discover social relationships in academic communities'*) and collaborative knowledge building (e.g., *'Link analysis for collaborative knowledge building'*).

Sociology

A type of link analysis has been successfully promoted by Richard Rogers, a sociologist at the University of Amsterdam. His variant focuses on *'issue networks'*, collections of networks around a topic of interest. His IssueCrawler software supports the generation of networks using hyperlink data from any (themed) set of Web sites entered by the user. Relatively little of this research seems to be published in recognised academic journals, however, perhaps due to its humanities-oriented nature. It is not clear why issue network research does not cite webometrics from a methodological perspective because much webometrics research relates to the accuracy of data from commercial search engines or the validity of using links in social science research. The reason may be because of the humanities focus on exploration and idea generation rather than a social science focus on data validity and methodological robustness.

Physics and complexity science

Much physics research is concerned with fitting simple mathematical models to large-scale data. In around 1999 a number of physicists noticed that links to Web pages did not have a standard Gaussian distribution, showing that they approximately fitted a power law (Adamic, Barabasi). webometrics research was probably not cited because it was not known or did not provide relevant methods due to its reliance upon commercial search engines (e.g., Ingwersen). In fact an early webometrics article seems to have been the first to observe power law phenomena in the Web, albeit on a small scale, but this went unnoticed or unrecognized.

This physics research was closely related to some computer science research into power laws. In fact a small computer science or physics research field of Web dynamics arguably began to subsequently form (e.g., book) with some participation by webometricians Judit Bar-Ilan who injected a search engine evaluation perspective. Most of the research was quite theoretical and abstract, however, often simulating Web link growth and attempting to build a simple mathematical model for it. This abstract approach perhaps rendered the more concrete webometrics research less relevant, and since commercial search engines were rarely used for raw data, the search engine evaluation webometrics was not found useful. Nevertheless, some webometrics contributions, such as the alternative document model concept for link counting could arguably have made a contribution to generating cleaner link data but was not used. Perhaps the computer scientists and physicists did not know of its existence and looked for relevant prior work mainly from within their own fields or did not regard webometrics as meeting other criteria for citing.

Conclusion

After years ago bibliometric studies have identified webometrics as one of the largest information science fields. Nevertheless, most published webometrics research is relatively theoretical and, as a new research field, seems unlikely to survive unless it is useful in some way. In general it is very important for academics to identify hot research topics and avoid unpromising ones and so methods are needed to help this identification process. Moreover, there is a significant amount of citation of webometrics research by disciplines outside information science, including computing, sociology, physics and complex science. However, it seems that

the potential user base for current webometric techniques is wider still, creating a need for awareness-raising.

REFERENCES

1. Thelwall, M. (2009), *“Introduction to webometrics: Quantitative Web research for the social sciences”*, New York, NY: Morgan & Claypool.
2. Bjerneborn, L. & Ingwersen, P. (2004), “Toward a basic framework for webometrics, *Journal of the American Society for Information Science and Technology*, **55**(14), 1216-1227.
3. Jeyashankar, R. and B. Rameshbabu (2009), “Websites of universities in Tamil Nadu: a webometric study,” *Annals of Library and Information Studies*, vol. 56, June, pp. 69–79.
4. Elgohary, A, (2008), “Arab universities on the web: a webometric study,” *The Electronic Library*, vol. 26, no. 3, pp. 374–386.
5. Konnur, P. and Rajani, S. (2010), *Academic Library Websites in Bangalore*, October 19-22.
6. Archana, S. (2010), “Web presence of the engineering college libraries in Kerala: An analysis of content”, *Kelpro Bulletin*, January 14, pp. 35-47.
7. S. Thanuskodi, 2011, “Webometric Analysis of Private Engineering College Websites in Tamil Nadu,” *J communication*, vol. 2, no. 2, pp. 73–81.
8. Islam A.M and S. M. Alam, (2011), “Webometric study of private universities in Bangladesh,” *Malaysian Journal of Library & Information Science*, vol. 16, no. 2, pp. 115–126.
