Scientometric Mapping on Webometrics: A Global Perspective

Samir Kumar Jalal
Deputy Librarian
Central University of Bihar, Camp Office: BIT
Patna, P.O-B.V. College, Patna, Bihar.
Email: dylibrarian@cub.ac.in

Abstract

Tries to focus on the quantitative growth and development of Webometrics Research through the publication output during the period of 1997-2008. Analyses the trend in Webometrics literature by selecting randomly 154 full text research articles published by researchers from around the globe during the period. Shows that the UK dominated in webometric research followed by USA. It also reflected that Prof. Mike Thelwall from University of Wolverhampton, UK achieved the credit of occupying top position in publication. The highly productive institution is University of Wolverhampton, UK. The most preferred journals for communication in webometrics were Scientometrics, JASIST and Journal of Documentation. Result shows that 125 (87.17%) publications with single country affiliation; 29 (18.83%) publications had collaboration with more than one country. In fact, Indian contribution to Webometrics Research is still in an infancy stages.

Key Words : Webometrics, Scientometric Mapping

Introduction

Scientometric mapping describes the pattern in terms of research output. Scientometric studies characterize the disciplines using the growth pattern and other attributes. Research publications are clearly one of the quantitative measures for the basic research activity in a country. It must be added, however, that what excites the common man, as well as the scientific community, are the peaks of scientific and technological achievement, not just the statistics on publications. There are also other kinds of research and technology development-mission oriented, industry-oriented, country-specific, etc. Progress in these cannot be obviously measured by counting only the number of publications. Scientometrics is a discipline which analyses scientific publications and citations appended to the papers to gain an understanding of the structure of science, growth of science at global level, performance of a country in a particular domain, performance of institutions, departments/divisions, and scientific eminence of an individual scientist. It also helps in knowing the information seeking behaviour of scientists and engineers by way of identifying they publish and what they cite.

Due to the tremendous growth of Internet, Web server and web-based resources during post 1990s, there has been much interest grown in the field of Internet research. Traditional bibliometric and informetric laws have been applied for measuring the web in order to assess the performance of the websites in terms of link analysis, web content analysis and web citation analysis. One important measure may be the use of Web Impact Factor to judge the quality of websites of host institutes recognized by valid domain names. According to Bjorneborn and Ingwersen, the definition of Webometrics is "the study of quantitative aspects of the construction and use of information resources structure and technologies on the web drawing on Bibliometrics and Informetrics approach". The calculation of WIF is an important part of Webometric research. Webometrics tries to measure the World Wide Web to get knowledge about number and type of hyperlinks and reference structure of the World Wide Web (WWW) and usage pattern.

Background works on Webometrics

The research on webometrics was way back to 1995 when a discussion on the potential for transferring information science techniques to the Internet was published in a French online journal by Marcia J Bossy. The first published article on information science hyperlink analysis by Larson on the performance of an investigation of link structures in academic web spaces. His article on 'bibliometrics of the World Wide Web: an exploratory analysis of the intellectual structure of cyberspace was presented at the American society of Information Science conference. Larson's objective was to assess the link structure of a topic on the web. Many information scientists have realized that the advanced features of search engines could be used for data collection and hyperlink analysis in webometric research. Hence, Rousseau's informetric analysis on the web and Rodriguer's web citation analysis gave an impetus to this field. It is a fact that Rodriguer has introduced the concept of information impact on the Internet in a Spanish documentation journal earlier. Almind and Ingwersen coined the term 'Webometrics', which is defined as the quantitative analysis of web related phenomenon. In other hand, Isidoro Aguillo brought out e-journal named 'Cybermetrics' which is synonymous to webometrics.

A range of new term was proposed for the emerging research field webometrics in mid-1990s, for instance, Netometrics (Bossy, 1995); Webometry (Abraham); Internetometrics (Almind & Ingwersen); Webometrics (Almind & Ingwersen, 1997); Cybermetrics (journal started in 1997 by Aguillo); Web Bibliometry (Chakraborty, et. al.). The idea of measuring average link frequencies i.e. WIF was developed by Peter Ingwersen in 1998. Thelwall, Vaughan & Bjorneborn contributed a classical article on Webometrics to demonstrate basic concept, origin, scope and coverage of Webometrics and related reviews. Thomas & Willett described a webometric analysis of the linkages to websites associated with individual departments of Library and Information Science (LIS) in United Kingdom (UK) universities. Kim investigated motivations for creating links in electronic publications in order to find out the relationship between citations and scholarly e-journals.

Objectives

The main objective of the study is to present the growth of literature and make the quantitative assessment of role of by way of analyzing various features of research output such as geographical distribution of publications, publication productivity and domain-wise activity index, authorship and collaboration pattern, language-wise distribution of publications, institution-wise distribution of publications. Following are some of the objectives of the study -

- To measure the contributions of different Journals in the field of Webometrics research;
- To know year-wise growth of literature on Webometrics;
- To find out country-wise research output on Webometrics;
- To reveal the truth regarding the contribution of various Institutes in the field of Webometrics;
- To find out the author collaboration in the field of Webometrics;
- To find out the subject growth under its sub-domain.

Scope of the Study

The scope of webometrics is centered upon the following four broad areas. These are as follows:

- Webpage content analysis (includes automatic categorization of webpages and texts using different search engines and tools for web analysis)
- Web link structure analysis (includes the categorization of hyperlinks and inlinks, self-links and external links to a particular website, patterns of linking etc)
- Web usage analysis (which includes the exploitation of log files for users' searching and browsing behavior)
- Web technology analysis (includes the performance of search engines with respect to information retrieval and supporting webometric analysis)

Literature Review

Surulinathi, et al. made an attempt to analyze quantitatively the growth and development of Knowledge Management Research in India in terms of publication output as reflected in Scopus database. During 1999-2007 a total of 51 papers were published by the Knowledge Management researchers to various domains

Kademani; Anil Kumar; Vijay Kumar made an attempt to analyse quantitatively the growth and development of Science and Technology Research in India in terms of publication output as reflected in Science Citation Index (SCI) (1990-2004). Total of 182111 papers were published by the Indian scientists and engineers to various domains. The study also focused on the visualization of Indian contribution to various micro-domains also. It was found that multi-authored papers were 163887 (90%).

Patra & Mishra made a systematic analysis of the rise in bioinformatics literature growth as available from NCBI PubMed using standard bibliometric techniques. Bradford's law of scattering was used to identify core journals and Lotka's law employed to analyze author's productivity pattern. Study also explored publication type, language and the country of publication. Twenty core journals were identified and the primary mode of dissemination of information was through journal articles. Authors with single publication were more predominant (73.58%) contrary to that predicted by Lotka's law. The study provides useful information to scientists wishing to undertake work in this area.

Pouris reports the findings of a scientometric analysis of nanoscale research in South Africa during the period 2000-2005 using the ISI databases. The article identifies trends over time, major institutional contributors, journals in which South African authors publish their research, international collaborators and performance in comparison to four comparator countries (India, Brazil, South Korea and Australia). The major findings of the investigation are as follows: nanoscale research in South Africa is driven by individual researchers interests up to date and it is in its early stages of development; the country's nano-scale research is below what would one expect in light of its overall publication output; the country's nano-research is distributed to a number of Universities with sub-critical concentration of researchers.

Lee described results of a scientometric study of the Institute of Molecular and Cell Biology (IMCB) to evaluate the research performance of IMCB in the first ten years since its establishment. The findings indicate that in the ten years, IMCB produced 395 research papers, 33 book chapters, 24 conference papers, and 4 monographs, graduated 46 PhDs and 14 MScs, and filed 10 patents. In its quest to become world-class,

IMCB researchers have been very selective in where they publish - 95.6% of the articles were published in ISI journals. The articles received an average of 25 to 35 citations per article, and the percentage of uncited articles is 11.6%. Four articles received more than 200 citations, and 18 received between 100 to 200 citations. It is found from the literature survey that there is no work done on the growth of literature in the field of webometrics.

Methodology

The study tries to concentrate in finding the trend in growth of research on webometrics in last 12 years (1997-2008) from all over the world. We have consulted Library and Information Science Abstract (LISA) database and Information Science Abstract (ISA) database for the purpose of collecting data. The fulltext articles have been collected during the said time span from only scholarly journals. We have chosen sample of 154 scholarly full-text articles, which are published from reputed global journals covering various aspects of webometrics in particular during the period 1997-2008. The study is limiting to the random sample of 154 scholarly research articles published on and after 1997 because it is known from the background study that the first article published in 1997 and the term webometrics is coined in 1997. On the other hand, there may be many Internet free articles or short communication, editorial materials, books, letters, conference proceedings articles, Internet documents etc., which are not taken into consideration due to short time span. Then all the required data have been extracted manually and plotted in the excel sheet in order to represent data in the tabular and graphical form to know the exact status of the research on webometrics.

Data Analysis & Results

Year-wise distribution of Productivity

During 12 years (1997-2008) there were so many literature published on webometrics in reputed journals, conference proceedings and reviews and electronic publications. We have randomly taken 154 scholarly articles published during the period in which it is found that highest number of articles (26) published in the year 2004. The year 1997 is having only 3 articles as it may be for the inception of webometric research. The average number of publications published per year was 12.83.

Table-1 gives year-wise growth of publications published during the period under study.

Table-1: Year-wise Distribution of Literature on Webometrics

Yr. of Pub	# of articles Pub	Percentage
2008	11	7.14
2007	7	4.55
2006	15	9.74
2005	6	3.90
2004	**26**	**16.88**
2003	15	9.74
2002	20	12.99
2001	19	12.34
2000	14	9.09
1999	10	6.49
1998	8	5.19
1997	3	1.95
Total	**154**	**100**

Preferred Journals for Communication of Webometric Research

The distribution of publications (154) were spread over 48 journals. The leading journals preferred by the information scientists were: Scientometrics with 37 papers, Journal of American Society for Information Science and Technology (JASIST) with 16 papers, Journal of Documentation (JOD) with 13 papers, Journal of Information Science (JIS) with 12 papers, and Cybermetrics with 10 papers. Table-2 provides journal-wise scattering of publications. Only five journals have published more than 50 percent of the publications in the field of webometrics.

Table-2 : Top 10 Journals on Webometrics

Name of Journals	# Articles	Percentage
Scientometrics	37	24.03
JASIST	16	10.39
Jl. of Documentation	13	8.44
Jl. of Information Science	12	7.79
Cybermetrics	10	6.49
Inf Processing & Mgmt	6	3.9
ASLIB Proceedings	4	2.6
Internet Research	3	1.95
JASIS	3	1.95
Nature	3	1.95
Online InfReview	3	1.95

Country-wise Research Output on webometrics

The papers on webometric research are spreaded over across 21 countries. Fig-1 provides the countries publishing journals in webometric research. Among the top ranking journals publishing the papers are from UK with 83(35.17 %) publications followed by the USA with 59 (25.00%) publications in journals, Israel with 13 (5.51%) publications. Only 5 countries have published more than 75 percent of the articles in the field.

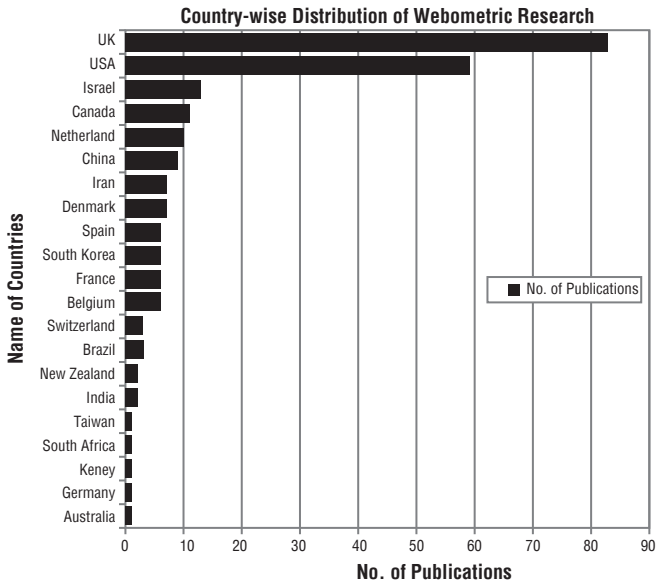

Fig-1: Country-wise Research Output on Webometrics

Contribution of Researcher on Webometrics

The following table explains the individual authors contribution in the field of webometrics. It is found that Prof. Mike Thelwall, University of Wolverhampton, UK made a remarkable contribution i.e. 58 (18.71%) articles, which is very far away from any other researchers in the field. Table-3 provides a list of top 22 most productive contributors in the field of webometrics research. The number of articles along with their institute and country affiliation is mentioned in the table. There were more than 150 researcher contributed in the field but those researchers (22, who have contributed 3 or more articles in the concerned field of study) are listed in table-3.

Table-3 : Contribution of Researcher on Webometrics

Rank	Name of Contributors	# of article	Institutions	Country
1	Thelwall, M	58	University of Wolverhampton	UK
2	Bar-Ilan, J.	12	The Hebrew University of Jerusalem	Israel
3	Wilkinson, D	10	University of Wolverhampton	UK
4	Li,X	7	University of Wolverhampton	UK
5	Harries, G	6	University of Wolverhampton	UK
6	Ingwersen, P	5	The Royal School of Librarianship	Denmark
7	Kousha, K	5	University of Tehran	Iran
8	Vaughan, L	5	University of Western Ontario	Canada
9	Barjak, F	4	University of Applied Sciences Northwestern Switzerland	Swizerland
10	Cothey, V	4	University of Wolverhampton	UK
11	Kretschmer, H	4	Royal Netherlands Academy of Arts and Sc.	Netherland
12	Aguillo, I	3	CINDOC	Spain
13	Björneborn, L	3	Royal School of Library and Information Sc.	Denmark
14	Henzinger, M.R	3	Compaq Western Research Laboratory	USA
15	Kim, H	3	Sejong University	South Korea
16	Kleinberg, J.M.	3	Cornell University	USA
17	Koehler, W	3	Valdosta state University	USA
18	Musgrove, P	3	University of Wolverhampton	UK
19	Park, H.W	3	Sejong University	South Korea
20	Price, L	3	University of Wolverhampton	UK
21	Smith, A.G	3	Victoria University of Wellington	New Zealand
22	Snyder, H	3	Indiana University	USA

Author Collaboration of Literature on Webometrics

The following figure-2 explains the multiple authorship patterns in the field of webometrics during the year 1998-2008. It is well known that nowadays research is carried out by group of researchers rather than by a single researcher. It has been observed from the following figure that highest number (69) of articles

contributed by single-authored and second highest articles (46) is written by two authors jointly. There are 125 (87.17%) publications with single country affiliation; 29 (18.83%) publications had collaboration with more than one country.

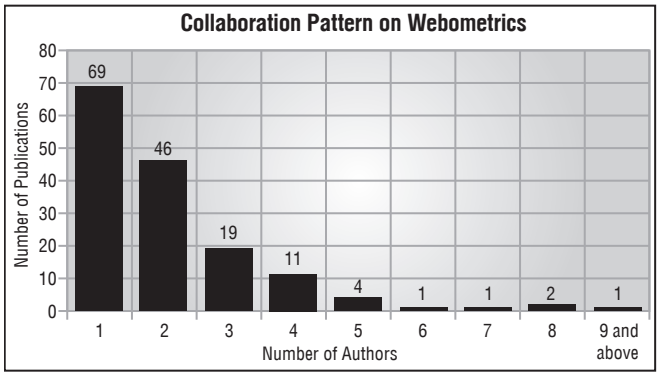

Fig-2: Multiple-Author-wise Distribution of Literature

Institution-wise Research Output on Webometrics

More than 68 research institutes around 21 countries are involved in webometric research. Table-4 provides 77 institutes are arranged according to their research productivities. University of Wolverhampton, UK topped the list with 83 (32.68%) publications, followed by Hebrew University of Jerusalem, Israel with 13 (5.12%) publications; University of Western Ontario, Canada with 10 (3.94%) publications.

Table-4: Top 10 Institution-wise Research Output on Webometrics

Name of Institutes	# of articles	% of Articles
University of Wolverhampton, UK	83	32.68
Hebrew University of Jerusalem	13	5.12
University of Western Ontario	10	3.94
Stanford University, USA	9	3.54
Royal School of Librarianship	8	3.15
University of Tehran, Iran	8	3.15
CINDOC	7	2.76
Carnegie Mellon University, USA	6	2.36
Royal Netherlands Academy of Arts and Sc.	6	2.36

Subject- Domain Wise Distribution

Keywords are one of the best Scientometric indicators to understand and grasp instantaneously the thought content of the papers and to find out the growth of the subject field. Analysis of the keywords appeared either on the title or abstract or full text articles either indexed by author of the articles and in case of authors of the concerned articles did not mentioned the keyword then author indexed them as the capacity of researcher in the field of webometrics. The basic purpose of such study is to know in which direction the webometric research is mostly grows. The high frequency keywords will enable us to understand the various aspects of webometric research under study.

Table-5: Main Subject- Domains in Webometric Research

Academic websites interlinking	Data Collection Methods	Link Pattern	Web link structure
Alternative Web Document Model	Departmental Linking Analysis	Link Motivations	Web Presence
Bibliometrics	Disciplinary Link Analysis	Ranking Problems	Web Sites
Citation Analysis	Geographic Trends	Search Engines	Web Usage
Classification Approach	Government Websites	Self-linked rates	Web Visibility
Co-citation Analysis	Hyperlink Analysis	Self-linking	WIF
Co-link Analysis	Impact Factor measures	University Links	
Commercial Links	Link Analysis	Web Indicators	

Conclusion

It has been found out from the study that during 1997-2008 webometrics has been grown a lot. More than 30 countries with about 150 authors, 80 institutes are involved in promoting the field. Although Indian contributions in webometric research in terms of publication output are not sufficiently high but at present Indian contributors (Mukhopadhyay; Jalal, Biswas & Mukhopadhyay; Walia & Kaur, Shah, Goshwami & Sharma) are working towards webometrics. The paper has considered only a sample of 154 research articles due to short time but there are other documentary sources like articles published in conference proceedings, PhD Thesis, software developed, books written in the concerned

field, research projects undertaken, which have not been considered here. Therefore, if comprehensive subject coverage is considered then more reasonable and accurate conclusion could have been drawn from this study.

References

1. Abraham, R.H. Webometry: measuring the complexity of the World Wide Web. Visual Math Institute. University of California at Santa Cruz. (1996). Available at: http://www.ralph-abraham.org/vita/redwood/vienna.html
2. Aguillo, I.F. Cybermetrics: definitions and methods for an emerging discipline. (2002) Available at http://www.upmf-grenoble.fr/adest/seminaires/ISIDRO/cybermetrics.ppt
3. Almind, T.C. & Ingwersen, P. Informetric analysis on the World Wide Web: Methodological approaches to Webometrics. Centre for Informetric Studies, Royal School of Library & Information Science (CIS Report 2) (1996).
4. Almind, T.C., & Ingwersen, P. Informetric analyses on the World Wide Web: Methodological approaches to Webometrics, Journal of Documentation, 53(4) (1997) 404-26.
5. Bjorneborn, L & Ingwergwn, P. Towards a basic framework for Webometrics. Journal of American Society for Information and Technology,2004,55(14), 1216-27.
6. Bossy, M.J. The last of the litter: Netometrics, 1995. Retrieved from the Web November 10, 2007. http://biblio-fr.info.unicaen.fr/bnum/jelec/solaris/d02/2bossy.html.
7. Chakraborty, S., Joshi, M., Punera, K., & Pennock, D.M. The structure of broad topics on the web. Proceedings of the WWW2002 Conference. (2002). Retrieved from the Web dated November 10, 2007. http://www2002.org/CDROM/referred/338.
8. Ingwersen, P. The calculation of Web Impact Factors, Journal of Documentation,1998, 54(2), 236-43.
9. Jalal, S.K; Biswas, S.C & Mukhopadhyay, PS. Bibliometrics to Webometrics. Information Studies,2009, 15(1),3-20
10. Kademani, B. S; Sagar, A.; Anil Kumar; Vijay Kumar. Scientometric Mapping of Vacuum Research in Nuclear Science & Technology: A Global Perspective, Journal of Physics: Conference Series, 114 (2008)
11. Kademani, B. S; Anil Kumar; Vijay Kumar. Indian Science and Technology Research: A Scientometric mapping based on Science Citation Index, 2006. Available and retrieved dated Jan 08, 2009. http://eprints.rclis.org/archive/00007763/
12. Kim, H.J. Motivations for hyperlinking in scholarly electronic articles: a qualitative study. Journal of the American Society for Information Science, 2000, 51(10), 887-99.
13. Lee, L. A scientometric study of the research performance of the Institute of Molecular and Cell Biology in Singapore. Scientometrics, 2003, 56(1), 95-110
14. Mukhopadhyay, P.S. The calculation of Web Impact Factors for educational institutes of India: A Webometric analysis. In. Information Management in e-Libraries, 26-27 February 2002.
15. Patra, SK & Mishra, S. Bibliometric study of bioinformatrics literature. Scientometrics, 2006, 67(3), 477-489.
16. Pouris, A. Nanoscale research in South Africa: A mapping exercise based on Scientometrics. Scientometrics, 2007, 70 (3), 349
17. Rodríguez, G.J.M. Impact assessment of information on the Internet AltaVista, the citation index of the Web. (1997). http://kronosdoc.com/publicacions/altavis.html.
18. Rousseau, R. Sitation: an exploratory study. Cybermetrics, 1(1), paper-1. Retrieved from the Web November 8, 2000. http://cindoc.csis.cs/cybermetrics/articles/v1ilp1.html.
19. Shah, L; Goshwami, P. & Sharma, U. Webometric Study of University Websites of UttarPradesh and Rajasthan (India). Proceedings of fourth international conference on Webometrics, Informetrics and Scientometrics & ninth COLLNET meeting held at Berlin, 2008
20. Surulinathi M.; Amsavenin, N.; Maheswaran, K. and Srinivasaraghavan, S (). Scientometric Dimensions of Knowledge Management Research in India: A Study based on Scopus database. Sri Lankan Journal of Librarianship and Information Management, 2008, 2(2); 13-24
21. Thelwall, M., Vaughan, L., & Bjorneborn, L. Webometrics. Annual Review of Information Science and Technology,2005, 39, 81-135.
22. Thelwall, Mike. Web impact factors and search engine coverage. Journal of Documentation, 2000, 56(2), 85-189.
23. Thomas, O & Willett, P. Webometric analysis of departments of librarianship and information science. Journal of Information Science, 2000, 26(6), 421-28.
24. Walia, P.K & Kaur, P. Webometric Analysis of Library Associations' websites in India, IASLIC Bulletin,2008,53(3),131-143