

Webometric Analysis of Library Websites of Higher Educational Institutes (HEIs) of India: A Study through Google Search Engine

Rupak Chakravarty* & Shalini Wasan**

*Department of Library & Information Science,
Punjab University, Chandigarh-160 017
E-mail: *rupak@pu.ac.in. **shalini.wasan@gmail.com*

ABSTRACT

Library websites play an important role in dissemination of information of the institution and library resources. It acts as a trustworthy mirror of the institute. To evaluate the library website performance webometric tools and indicators are required. The present study calculated web impact factor (WIF) and R-WIF (Revised WIF) of top ten library websites of HEIs (Higher Educational Institutes) of India and further correlated both the formulas with Spearman's Rank Correlation. It was found that WIF and R-WIF are correlated and associated which depicts that there is very less difference between the two ranking methods. The position of library websites of HEIs of India is same while evaluating through both the formulas. After a pilot survey, the list of top ten HEIs of India is extracted from ranking web of Universities. Google search engine was chosen for the study.

Keywords: Web presence, Higher Educational Institutes, link domain, link analysis, WIF (web impact factor), R-WIF (revised WIF), Spearman's Rank Correlation

1. INTRODUCTION

Web impact factor (WIF) and Revised web impact factor (R-WIF) are web versions of the impact factors (IF) published by the Institute of Scientific Information for Scientific Journals which is now known as Thomson ISI. The WIF was developed by Ingwersen¹ to measure the impact of web area by number of links it receives in a search engine. Ingwersen used Alta Vista search engine to count number of webpages whereas the present study used Google search engine because Alta Vista search engine has now merged with Yahoo search engine.

Bjornborn and Ingwersen² define webometrics as 'The study of the quantitative aspects of the constructions and use of information resources, structures and technologies on web drawing on bibliometric and infometric approaches.' According to Thelwell³ webometrics can also be defined as '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 the study'.

Thus, webometric tools and techniques aspire to generate new knowledge that guides to explain phenomena of what the institute required to regain or sustain for its existing position and further provides scope for improvement in the future. Webometric tools

and techniques provide in depth appraisal based upon both qualitative and quantitative information.

Library websites play a crucial role for exponential growth of information regarding library attributes. HEIs in India are spending huge amount of human as well as capital resources to get efficient and effective facilities for their users on their library websites but still users are not able to find out the required pages or material they were looking for.

The scope of the study confines to webometric study of top 10 library websites of HEIs of India. The ranking web of universities is considered for the list of top 10 library websites of HEIs of India.

2. LITERATURE REVIEW

Jeysankar & Valarmathi⁴ studied websites of ICMR institutes for their study. They calculated webpage size, WAVE Web AIM accessibility error (a web accessibility tool that assist in the evaluation of web), various search engines' performances, the difference between pages in various time intervals and number of rich files. The paper also presented the link network diagram of ICMR institutes using Pajek software. Madhusudhan & Prakash⁵ explored different characteristics through linked analysis of 16 IIT websites. All the IITs' had their own websites and all websites were working under homogeneous Domain Name System

(DNS), i.e., (ac.in). The comparisons of ranking of Indian Institutes of Technology (IITs') had been done using WISER, WIF (in link) and World Rank. The WISER ranking and WIF (in link) having correlation, i.e., +0.0558824 which implied that there was much association or closeness between two ranking methods. WIF (in link) rank gave much closer value to the World Rank. Hence, the IIT Bombay and IIT Madras secured first and second position respectively in WIF (in link) as well as in world rank whereas WIF (in link) position of IIT Rajasthan had one position better ranked than the world ranking, i.e., 15th position. The findings of the study would guide webmasters and library professionals so that they could overcome the limitations faced by the users and further improve their websites.

Baharum & Jaafar⁶ studied that users could form a schema (mental model help users to interact with the website) for the location of web objects on informational websites. The study was investigative and included logo, internal links, search, advertisement, external links, site title, login, language, content and calendar. 94 participants from 10 Asian countries filled the questionnaire. The study concluded that the advances of technology significantly affect users' schema for the location of the web objects layout within 4 years.

Jeyshankar & Maria⁴ discussed the websites of universities in Kerala. The paper further calculated the web impact factor of each kind of links for every websites in Kerala. Kerala Agricultural University was ranked 1st with 12700 webpages of all the universities in Kerala but the University of Calicut was in 2nd rank according to the web pages. It was also revealed that Kerala Agricultural University with 0.0002 % of self-link web impact factor occupied the last rank even though it had 12700 webpages because it had lowest number of external links (32). So, at the end authors concluded that to use the links of the universities could be act as parameters to some extent to show the websites webpages and the results should not specifically rests on a particular link.

Tafaraji & Tahamtan⁷ studied the findings of webometric analysis of websites of medical universities of Iran. Majestic SEO software and search engines like Google, Yahoo and Bing. The number of webpages, in links and rich files were calculated for these 43 universities. The findings indicated a significance relationship between webometric universities ranking in Iranian Ministry of Health. The study further indicated that the use of rich files could be a better and more reliable view of Iranian medical universities ranking.

Krishna Kumar & Nirmala⁸ analysed websites of Research Council of India through webometrics.

Alexa Traffic Rank and the Global Rank were considered for the study. This study revealed an overall preview of the traffic and page ranks of Research Councils of India websites.

Shukla & Tripathi⁹ examined the extent of back links to different categories of web-pages of the library websites belonging to institutes of national importance and premier library websites of management institutions in India. The study concluded with the remarks that Yahoo site explorer had retrieved maximum number of back links. This indicated that Yahoo site explorer was more reliable than Google, All the Web and AltaVista in terms of back links

Moghaddam & Farshid¹⁰ studied that webometrics serves as a method for identifying the most accredited free e journals of Medical Sciences. It has been found that information collected from and about links between web-pages and websites can reflect real world phenomena and relationships between their organisations. Further, authors concluded with the remarks that there should be interlinking for free e-journals between the local bodies.

Jayshree & Ravichandran¹¹ explore URL citation analysis of eighty two technical and science education institutions from the website department of higher education, Ministry of Human Resource Development, Government of India's website. URL citation analysis is one of the methods for web impact assessment. Spearman and Chi-square analysis were used to highlight the relationship between the institution groups and website categories. Chi-square test reveals that the National Institutes of Technology had more impact in each category of websites.

Prabakaran & Lihitbar¹² aim to analysed the 17 Space Science Library websites in the world. The data was collected on number of webpages, external links, in links, self links and total links of the websites. Various web impact factors were calculated on the basis of the web links and ranked the websites. The study concluded that the need for ranking of library website in space research organisation was to serve the users by providing better services through their websites.

Since 2004, the ranking web or webometrics is the largest academic ranking of Higher Educational Institutions of the world. Cybermetrics Lab (Spanish National Research Council, CSIC) performed this exercise every 6 months for providing reliable, multidimensional, updated and useful information about the performance of Universities from all over the world based on their web presence and impact. After every 6 months previous reports are regenerated so that each website reflects the current state of dbase of all times. No fake universities are included by the ranking web of universities. Rather, it is found that users stop showing interest

in the library websites if they could not find their requirements. Thus, it is the need of the hour to evaluate the library websites through WIF and R-WIF. Further, Spearman's Rank Correlation has been used to correlate these formulas.

3. OBJECTIVES

The objectives of the present study are to:

- Find the number of self link pages, in link pages and external link pages of library websites of HEIs of India;
- Calculate WIF and R-WIF of library websites of top ten HEIs of India; and
- Calculate the correlation between WIF and R-WIF of library websites of HEIs of India through Spearman's Rank Correlation.

4. METHODOLOGY

For calculation of WIF and R-WIF of library websites of HEIs of India, an appropriate search engine is required. The search engine should be enormous, cover more hyperlinks with advanced search facilities to count links possessed by the websites. Google search engine possessed the above criteria and is safer and covers more hyperlinks as compared to other search engines. In present study, therefore, Google was used to determine the requirements of both the formulas. Utilising Boolean operators different formulations were used which resulted in different outcomes. Google search engine is used to collect data, i.e., number of webpages, number of In-links and number of self links.

- WP=Total number of webpages extracted in Google, for example, site: www.library.iitb.ac.in.
- SL=Number of self-links will be extracted in Google, for example, site: www. Library.iitb.ac.in AND link domain: www. library.iitb.ac.in.
- IL= Number of In-links will be extracted in Google for example, site: www. library. iitb.ac.in NOT link domain: www. Library. iitb.ac.in.
- Web Impact Factor (WIF) can be calculated by adding number of In-links and number of Self-links extracted in Google divided by total number of web pages extracted in Google, i.e.,

$$WIF = IL + SL/WP$$
- Revised Web Impact Factor (R-WIF) can be calculated by dividing number of In-links extracted in Google with total number of web pages extracted in Google, i.e.,

$$R-WIF = IL/WP$$
- Spearman's Rank Correlation-The relationship between WIF and R-WIF can be established through the correlation coefficient (r). In the present study, the Spearman's Rank Correlation has been used.

According to this method,

$$\text{Correlation Coefficient (r)} = \frac{N\Sigma XY - (\Sigma X) * (\Sigma Y)}{\text{Sqrt} ([N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2])}$$

5. ANALYSIS

The results were extracted on 25th April, 2015 from Google search engine.

5.1 Ranking of Top Ten HEIs

Table 1 shows that ranking from both the formulas gives approximately same results. It was found that ranking of five library websites is same in both the formulas. These are website of Central library IIT Delhi (1st rank), website of J.R.D. Tata Memorial Library (2nd rank), website of Central library- Delhi University Library System (3rd rank), website of Central Library IIT Bombay (4th rank), website of Anna University Central library (9th rank). On the other hand, website of Central library IIT Madras ranked 5th while evaluating through WIF and ranked 10th while evaluating through R-WIF. Website of Central library IIM Bangalore ranked 6th while evaluating through WIF and ranked 8th while evaluating through R-WIF. Website of Scientific Information Resource Centre ranked 7th rank while ranking through WIF and ranked 6th while evaluating through R-WIF. Website of Central library IIT Kanpur ranked 8th while evaluating through WIF and ranked 7th while evaluating through R-WIF. Website of Library-SRM University ranked last while evaluating through WIF and ranked 5th while ranking through R-WIF.

5.2 Relationship between WIF and R-WIF

The relationship between WIF and R-WIF can be established through the correlation coefficient (r). The correlation coefficient (r) relates the strength and direction of linear relationship between two variables. The coefficient of determination represents the percent of the data that is the closest to the line of best fit. Correlation is always between -1.0 and +1.0. If the correlation is positive, we have a positive relationship. If it is negative, the relationship is negative. In the present study the Spearman's Rank Correlation has been used (Table 2).

According to this method,

$$\text{Correlation Coefficient (r)} = \frac{N\Sigma XY - (\Sigma X) * (\Sigma Y)}{\text{Sqrt} ([N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2])}$$

where, X=R-WIF, Y=WIF, Σ =Algebraic sum; N=Number.

Thus, the correlation coefficient (r) can be calculated as mentioned above. The values used in the formula are: Number (N) = 10; ΣX =6.40; ΣY = 15.43; ΣXY = 27.5; ΣX^2 =12.8; ΣY^2 =60.728

Table 1. Comparing ranking of WIF and R-WIF of top ten HEIs of India

S. No.	Library name	URL	WP	IL	SL	IL+SL	WIF	R-WIF
1.	Central Library, IIT Bombay	www.library.iitb.ac.in	3,46,000	2,04,000	3,05,000	509000	1.47 (4 th)	0.589 (4 th)
2.	J.R.D. Tata Memorial Library	www.library.iisc.ernet.in	1,06,000	97,900	1,61,000	258900	2.442 (2 nd)	0.923 (2 nd)
3.	Central library, IIT Madras	www.cenlib.iitm.ac.in	1,65,000	61	1,62,000	162061	0.982 (5 th)	0.00036 (10 th)
4.	Central library, IIT Kanpur	www.library.iitk.ac.in	10,50,000	80,900	1,41,000	221900	0.211 (8 th)	0.077 (7 th)
5.	Central library - Delhi University Library System	crl.du.ac.in/cl	48,10,000	41,40,000	41,80,000	8320000	1.729 (3 rd)	0.860 (3 rd)
6.	Central library, IIT Delhi	library.iitd.ac.in	15,100	49,400	55,400	104800	6.940 (1 st)	3.271 (1 st)
7.	Scientific Information Resource Centre	www.tifr.res.in/index.php	3,23,000	88,800	92,900	181700	0.562 (7 th)	0.274 (6 th)
8.	Library- SRM University	www.srmuniv.ac.in/aboutus/library	83,200	4,900	5,850	10750	0.129 (10 th)	0.058 (5 th)
9.	Central library, IIM Bangalore	www.iimb.ernet.in/library	1,92,000	64,300	94,700	159000	0.828 (6 th)	0.334 (8 th)
10.	Anna University Central Library	www.annauniv.edu/library	3,40,000	6,270	64,200	70470	0.207 (9 th)	0.018 (9 th)

Table 2. Correlation coefficient

S. No.	URL of library	X (R-WIF)	Y (WIF)	X ²	Y ²	XY
1.	www.library.iitb.ac.in	0.589	1.47	0.34	2.16	0.86
2.	www.library.iisc.ernet.in	0.923	2.442	0.85	5.95	2.25
3.	www.cenlib.iitm.ac.in	0.0003	0.982	0.00002	0.96	0.00029
4.	www.library.iitk.ac.in	0.077	0.211	0.0059	0.044	0.016
5.	crl.du.ac.in/cl	0.860	1.729	0.73	2.98	1.486
6.	library.iitd.ac.in	3.271	6.940	10.69	47.6	22.56
7.	www.tifr.res.in/index.php	0.274	0.562	0.07	0.31	0.0069
8.	www.srmuniv.ac.in/aboutus/library	0.058	0.129	0.004	0.0144	0.0512
9.	www.iimb.ernet.in/library	0.334	0.828	0.11	0.67	0.27
10.	www.annauniv.edu/library	0.018	0.207	0.000324	0.04	0.0036
	Total (Σ)	6.40	15.43	12.8	60.728	27.5

$$\begin{aligned}
 r &= \frac{10 \cdot 27.5 - (6.40 \cdot 15.43)}{\sqrt{[10 \cdot 12.8 - (12.8)^2] \cdot [10 \cdot 60.728 - (60.728)^2]}} \\
 &= \frac{275 - (98.75)}{\sqrt{[(115.2 \cdot 546.48)]}} \\
 &= \frac{176.25}{250.907} \\
 &= + 0.702
 \end{aligned}$$

Thus, value of *r* (plus sign) shows positive relationship between R-WIF and WIF which implied that there is much association or closeness between

R-WIF and WIF). In other words, there is very less difference and closeness between R-WIF and WIF.

6. CONCLUSIONS

Calculation of WIF and R-WIF through Google search engine for evaluating top ten library websites of HEIs of India and further correlating these formulas through Spearman's Rank Correlation shows positive and strong relationship between WIF and R-WIF. This closeness implies that there is much association between these two formulas. Also, it was found that ranking of half library websites of HEIs of India is same while evaluating through both the formulas whereas half of library websites secured different ranks while ranking through WIF and R-WIF.

REFERENCES

1. Peter, Ingwersen. Informetric analyses on the world wide web: Methodological approaches to webometrics. *Journal of Documentation*, 1998, **53**(4), 404-26.
2. Bjornborn & Ingwersen. Toward a basic framework for webometrics. *J. of the Amer. Soc. for Inf. Sci. & Tech.*, 2004, **55**(14), 1216-27.
3. Mike, Thelwall. Introduction to Webometrics: Quantitative web research for the social sciences. Morgan & Claypool, 2009, ISBN 978-1-59829-993-9.
4. Jeysankar, R.; Maria, Sujitha I. & Valarmathi, A. Web pages of ICMR institutes websites: A webometric analysis. *Global Adv. Res. J. of Lib. Inf. and Arc. Stud.*, 2012, **1**(1), 6-18.
5. Madhusudhan; Margam & Prakash, Shashi. Websites of Indian Institutes of Technology: A webometric study. *Inter. J. of Lib. and Inf. Stud.*, 2013, **3**(4).
6. Baharum, Aslina & Jaafar, Azizah. Users' expectation of web objects location: Case study of ASEAN countries. *Adv. in Visual Informatics, Lecture Notes in Comp. Sci.*, 2013, **8237**, 383-95.
7. Tafaraji, Roghaye & Tahamtan, Iman. Webometric analysis of Iranian medical universities according to visibility, size and rich files. *Webology*, 2014, **11**(1). <http://www.webology.org/2014/v11n1/a119.pdf>.
8. Krishna Kumar, P. & Nirmala, P.J. Webometric study of research councils of India. *E-Library Sci. Res. J.*, 2014, **2**(8), 1-6.
9. Shukla, A. & Tripathi. A back links analyses of institutes of national importance and premier management institutions libraries websites. *J. of Int. Acad. Res. for Multi.*, 2014, **2**(7), 560-75.
10. Moghaddam, Alireza Isfandyari & Danesh, Farshid. Webometric as a method for identifying the most accredited free electronic journal: The case of medical sciences. *The Electronic Library*, 2015, **33**(1), 75-87.
11. Jayshree, S. & Ravichandran R. Web impact assessment of identified higher education institutions in India. *Annals of Lib. and Inf. Stud.*, 2015, **62**(2), 7-18.
12. Prabakaran, R. & Lihitbar. Web link analysis and web impact factor of space libraries websites. In 10th International CALIBER, 2-14 March 2015.

About the Authors

Dr Rupak Chakravarty is working as Assistant Professor in the Department of Library & Information Science, Panjab University, Chandigarh. He is serving as Mendeley Advisor at Chandigarh region. His areas of interests are: Application of ICT in libraries, free open source software (FOSS), digital library, information sources, open access, institutional repositories (IRs), Library 2.0.

Ms Shalini Wasan is working as Sr. Librarian, Chandigarh College of Engineering and Technology, Chandigarh, India. She has published 21 papers in national and international journals. Her areas of interests are: Scholarly communication, institutional repository, information literacy, Web 2.0 and Web 3.0, RFID, knowledge management, national knowledge network, consortium, open source software, networking, and webometrics.