

May 2019

A webometric Analysis of Select Knowledge Portals of National Repute in India

Tabasum Maqbool

University of Kashmir, tabasum.lis@gmail.com

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

Part of the [Scholarly Communication Commons](#), and the [Scholarly Publishing Commons](#)

Maqbool, Tabasum, "A webometric Analysis of Select Knowledge Portals of National Repute in India" (2019). *Library Philosophy and Practice (e-journal)*. 2323.

<https://digitalcommons.unl.edu/libphilprac/2323>

A webometric Analysis of Select Knowledge Portals of National Repute in India

By

Dr Tabasum Maqbool

Chief Librarian , Institute of Advanced Studies in Education

Srinagar, Jammu & Kashmir (India) Pin - 190001

Email: tabasum.lis@gmail.com

ABSTRACT

The purpose of the study is to understand the performance of knowledge portal of national repute in India on webometric scale and shall also assist admins of the select sites to take corrective measures if needed and also visualize a better and more efficient web presence. The study has selected the reputed knowledge portals of national importance for assessing their impact and performance in terms of webometrics. The investigator selected Alexa web ranking portal (<https://www.alexa.com/>) for harvesting the required data for select web portals. Besides, to gauge the revised web impact factor (RWIF) Developer Shed, an Open Site Explorer service, was employed that measure the total number of indexed pages by different search engines. the harvested data has been analyzed, refined, tabulated and interpreted to ferret out meaningful findings. The study found Shodhganga is having best rank at global and Indian levels and National Digital Library of India (NDLI) has highest RWIF. The highest estimated daily unique pageviews per visitor is for egyankosh (6.91 pages) strangely it also has third biggest bounce rate (56.00%). The Swayam portal has second highest daily unique pageviews per visitor (6.66) and has least bounce rate among the select sites. It is found that Swayam is most popular site with daily time on site of 09:06 minutes followed by National Digital Library of India (06:48). The Swayam e-learning web portal is overwhelmingly used by people in India (97.30%) followed by National Digital Library of India (96.30%). The study shall offer an insight to admins of knowledge portals about different aspect that may need to be addressed for more efficient and effect web visibility.

Keywords: webometrics, cybermetrics, web science, Web Impact Factor, Revised Web Impact factor, Knowledge portals, higher education,

INTRODUCTION

The primary role of library is to disseminate of knowledge to the prospective seekers. But at the same time, it can procure each and everything that is available out there. Therefore, it always does apply different management and statistical methods to take wise decisions to offer quality services to its clientele. Library professionals developed different metrics to measure information sources and services. In this direction propounded different concepts like Librametry, Informetrics, bibliometrics, Scientometrics and now applying what is called webometrics to see the long-term utility, viability and significance of different resources and

portals for the user community and particularly by doing so visualizing the development of a nation and human society at large.

'Webometrics is pertaining to gauging various fragments of the websites, website pages, parts of web pages, words in pages, hyperlinks, Search engine results. The utility of the web itself as a communication medium and for organizing wide cluster of records'. The term webometrics was coined by **Almind and Ingwersen (1997)**, which recognized the web as an important source for gaging documents & information. Later on, 'information scientists realised that several significant web metrics could be calculated employing the noval search characteristics of one of the prime search engines of the day: AltaVista' (**Ingwersen, 1998; Rodríguez i Gairín, 1997**). Its 'link search proficiencies offered extensive link measurements to be employed for the first time. This was demonstrated in a study that involved a measure of the web pages in each Scandinavian country that connected to the pages of each other Scandinavian country' (**Ingwersen, 1998**).

The potential to research web links employing AltaVista proved mainly significant in starting webometric research. Due to the fact that hyperlinks are anatomically like academic citations in a way that these refer from a source document to a target document (**Borgman & Furner, 2002; Garfield, 1970; Moed, 2005a; Nicolaisen, 2007**). The similarity of links and citations, together with universities being early adopters of the web, resulted in the emergence of a number of important natural research goals. These attempted to assess whether hyperlinks could be used in similar ways to academic citations, the validity of using link counts and AltaVista data in research, and the best ways of counting links (**Thelwall, 2001b**). In parallel with this link analysis strand, other information scientists investigated the reliability and coverage of search engines and changes in the content of the web or individual collections of web pages (**Bar-Ilan, 2004**). These three types of web-based measurement research came to be collectively known as webometrics.

REVIEW OF LITERATURE

A reasonable amount of research has been conducted on different facets of webometrics. The study has made an effort to identify some of these that are very much relevant to the current study have been discussed below.

'An investigation was conducted to measure the WIF of 99 Arab universities belonging to 20 Arab countries revealed a sturdy correlation among external links and web presence' (**Elgohary, 2008**). A separate study on WIF was conducted on 'medical science universities found websites that have highest number of web pages and external links has low WIF compared to sites with limited web pages and external links obtained higher WIF' (**Aminpour, Kabiri, Otraj, & Keshtkar, 2009**). Likewise **Jeyshankar and Babu (2009)** studied the 'WIF of the Tamil Nadu university websites in India, found that the 'websites that have numerous web pages and fewer in-links obtained low WIF'.

Jalal, Biswas and Mukhopadhyay (2010) conducted a study to 'rank the central universities in India in tune with various webometric indicators'. **Zahedi, Shirazi and Dehghani (2010)** identified that 'there is a substantial correlation among the number of external links and the impact factor'. Similarly, a study analyzed the relation of web presence and web link patterns among South Asian countries' (**Vijayakumar, Kannappanavar, & Santosh Kumar, 2012**). In a

separate study **Madhusudhan and Prakash (2013)** have explored 'if there is a connection between the WISER, WIF (inlinks) and World Ranking methods'. **Lihitkar (2015)** has 'stated that a website with larger number of rich files and high number of results from Google Scholar obtains higher in ranking'. While **Gupta and Walia (2016)** have surveyed the 'web presence of ten African national libraries' websites. The ensuing research observed that the ranking based on WIF was not much reliable and it was biased towards the small number of web pages and in-links'. **Abu and Jayasekara (2017)** conducted a study to 'evaluate the ten leading MOOC providers using Alexa web databank. They observed Udemy as the most famous MOOC provider globally and Iversity' the least. The maximum bounce rate was found for 'Codecademy' and the lowermost for Open2study, while FutureLearn have the maximum daily page views per visitor and Coursera have the least'.

STATEMENT OF THE PROBLEM

India is one of the fast-emerging economies of the world with growing higher education sector. It has second largest population in the world and as such expectations and future education and research requirement shall be quite huge. Given this background it would be in the fitness of things to gauge the performance of some of the renowned and reputed higher education web portals disseminating knowledge without charging any fee. In this context study identify 5 reputed web portals of national repute for the study. The idea was to measure the different metrics pertaining to webometrics. This would help us to understand the performance of these portal on webometric scale and shall also assist admins of the select sites to take corrective measures if needed and also visualize a better and more efficient web presence.

SCOPE

The scope of the study is limited to following higher education knowledge web portals of national repute and importance.

1. Shodhganga (shodhganga.inflibnet.ac.in)

It is a national electronic repository of Theses that are submitted to different central and state universities in India. So far more than two hundred thousand Ph.D theses are available in different subject streams for researcher world over for furthering research and scholarship.

2. National Digital Library of India (<http://ndl.iitkgp.ac.in>)

This is a very significant initiative of Indian Institute of Technology and fully supported by ministry of human resources development, India. The aim of this portal is to harvest maximum amount of open access (OA) literature from different OA repositories globally and offer one stop shop for knowledge seekers and researchers especially from India.

3. Egyankosh (egyankosh.ac.in)

The Egyankosh is an inevitable knowledge portal for students of distance mode in India. The repository is hosted by Indra Gandhi National Open University, New Delhi. It is serving its registered students with study material and audio-visual resources since a decade now. But from last few years it has put its resources open

for all and as such has become important sources of knowledge for students of higher education in India and beyond.

4. Swayam (<https://swayam.gov.in/>)

It is an e-learning platform exclusively made for prospective Indian students who want to further their knowledge endeavours. The e-learning portal Swayam is developed by ministry of human resource development of India in technical collaboration with Microsoft. The platform offers hundreds of degree and certificate courses in different academic and technical disciplines.

5. Consortium for Educational Communication (cec.nic.in)

The Consortium for Educational Communication is inter-university center in India with a goal to disseminate education with aid of multimedia. In this context the center has developed numerous multimedia contents and has also launched television channels to reach the unreached. In order to reach maximum number of people and make use of new technology the center has launched its web portal to disseminate multimedia contents. So far it has hosted thousands of hours of multimedia content related with different science and social science subject streams.

OBJECTIVES

1. To measure the web ranking for select websites of national repute, hosting knowledge resources for higher education in India.
2. To quantify the revised web impact factor of select sites.
3. To measure average pageviews per visitor on daily basis and bounce rate of the select sites
4. To gauge average time spent by users daily on the select sites and major countries that access resources of select sites.

METHODOLOGY

The methodology of the study comprises of following phases

Phase – I

A thorough review of literature was done on the basis of knowledge and experience gained. The investigator decided to select Alexa web ranking portal (<https://www.alexa.com/>) for harvesting the required data for select web portals. Alexa Internet, Inc. is an American web traffic analysis company. Alexa provides web traffic data, global rankings, and other information on 30 million websites (Alexa, 2018). In tune with objectives of the study data was harvested from the Alexa web ranking portal. In order to achieve objective – 2 of the study data was partly obtained from other source that is discussed in phase – II.

Phase – II

The study was keen to find the pure web impact factors, (excluding self-links), for the select web sites. The WIF scrutiny for self-links doesn't give actual impact of site compared to inlinks,

since the most of self-links within a web site can be developed for browsing instead of approving and validating the content of target pages (Smith, 1999; Thelwall, 2000). Besides, bigger the web site the higher number of self-link pages is expected. While as on the contrary Inlinks from external sources/sites demonstrate appreciation and credibility obtained by the target pages. Therefore, revised impact factor (RWIF) takes into account only inlinks obtained by a websites divided by sum of web pages of a website. it is calculated as shown below:

$$\text{RWIF} = \frac{\text{A = inlinks (external backlinks) to the website}}{\text{B = number of web pages published in the website which are indexed by the search engine (not all web pages of the website)}}$$

To measure the 'total number of indexed pages by search engines, Developer Shed, an Open Site Explorer service, was employed. The Indexed Pages tool of the SEO CHAT application (<http://tools.seochat.com/tools/domainindexed-pages/>) that queries the major search engines and returns the total indexed page count for each URL' (Noruzi, 2005). To count the inlinks (external backlinks) for websites, alexa web databank (<https://www.alexa.com/>) was used to derive the required data. The 'search engine listed by SEO CHAT application that retrieved maximum indexed pages of select sites were considered for the study' that is in line with the argument put forth by Noruzi, (2005) that the search engine selected for undertaking the WIF study should have a large database, covering as much of the web as possible.

Accordingly, the harvested data has been analyzed, refined, tabulated and interpreted to ferret out meaningful findings that can help policy makers to fine tune the efforts wherever required to make these web portals of national repute crisper and appealing to the knowledge seekers.

RESULTS AND DISCUSSION

Global and National Level Ranking

The study was keen to know both global and Indian level ranking of select websites of national repute for hosting knowledge resources and offering e-learning opportunities to users. The resources and services offered these portals are free to web and as such can be availed by anyone with access to world wide web and interested to consume information resources for the pursuits of higher education. Therefore, it is expected to be exploited by optimum number of users. From the data collected from alexa web it is found that Shodhganga has best global (4,945) and Indian (505) level ranking that reflects its strong utility for the knowledge seekers. This is followed by National Digital Library od India another knowledge resources platform fast become popular among students and scholars of higher education in India. It has Global rank of 7,129, while in India its rank is slightly below Shodhganga with rank of 508. Swayam is another very significant flagship program of ministry of human resources development (India) with the objective to take higher and technical education to doorsteps of every aspiring student in India and also help knowledge seekers worldwide by free access to video lectures

and tutorials. The global rank of Swayam is 22,396 while as at India level (1,518) it is fast catching up given its recent origin.

Egyankosh is a repository of tutorials offered to distance learners by Indra Gandhi National Open University (IGNOU). The study material is not only used by the registered students of the university, but any knowledge seeker can download them. This make it a very significant site for students seeking higher education. It was found that egyankosh is globally not as popular as shodhganga or national digital library but has a decent global (39,003) and national (3,022) level rank. The last one is Consortium for Educational Communication (CEC) which has a very ordinary global and Indian presence on the web. This humble rank can be attributed to recent origin and very limited resources hosted till date. Table 1

Table 1: Ranking at National and Global levels		
Name of the Portal	Rank in India	Global Rank
Shodhganga	505	4,945
National Digital Library of India (NDLI)	508	7,129
Swayam	1,518	22,396
Egyankosh	3,022	39,003
Consortium for Educational Communication (CEC)	22,622	2,61,542

Revised Web Impact Factors

The revised web Impact factor (RWIF) of select websites of national repute is calculated and it is revealed that National Digital Library of India (0.0727) has the highest RWIF followed by Consortium for Educational Communication (0.0180) and Egyankosh (0.0064). While as Swayam (0.0027) and Shodhganga (0.0011) have the least revised web impact factors. Though the revised web impact factors of select sites is very low with highest for National Digital Library of India. This could be due to two reasons either the contents hosted are considered relevant and qualitative by other education related sites globally or there is lack of awareness among other such players in India and globally, Table 2.

Table 2: Revised web Impact Factors of Select Web portals of National repute in India			
Name of the Portal	A = inlinks (backlinks) to the Website	B= Number of Webpages Published in the Website Indexed by SEO CHAT	Formula : RWIF = A/B
National Digital Library of India (NDLI)	2401	33000	0.07275757
Consortium for Educational Communication	302	16,700	0.01808383

Egyankosh	420	65,600	0.00640243
Swayam	119	42,900	0.00277389
Shodhganga	2,909	2,540,000	0.00114527

Pageviews and Bounce Rate

The highest estimated daily unique pageviews per visitor is for egyankosh (6.91 pages) while as strangely it also has third biggest bounce rate (56.00%). Bounce rate is percentage of visits to the site that consist of a single pageview. The Swayam portal has second highest daily unique pageviews per visitor (6.66) and has least bounce rate among the select sites. This augurs well for the site. It seems site is gluing users to its resources once they visit the site. National Digital Library of India has second least bounce rate (27.30%) and third best daily pageviews (4.97). The worst bounce rate among the select sites is true for Consortium for Educational Communication (76.70%) and Shodhganga (72.80%) and also least daily pageviews are also found in case of these two sites (1.7, 2.55 pages respectively). This is a very serious issue with these sites and need to resolve it for better visibility and impact on the web. Table 3.

Table 3: Daily Pageviews and Bounce Rate of Select Web portals of National repute in India		
Name of the Portal	Daily Pageviews per Visitor	Bounce Rate
Egyankosh	6.91	56.00%
Swayam	6.66	22.70%
National Digital Library of India (NDLI)	4.97	27.30%
Shodhganga	2.55	72.80%
Consortium for Educational Communication	1.7	76.70%

Daily Time on Site

Study was curious to know on which portal users spend maximum time. It is found that Swayam is most popular site with daily time on site of 09:06 minutes followed by National Digital Library of India (06:48), Egyankosh (04:00). The least time is spent on Shodhganga (02:58) and Consortium for Educational Communication (01:50). The Swayam turning out to be the most popular as for as time on site is concerned is not a surprise given the fact the swayam is an e-learning platform and host more multimedia content than text based one. But National Digital Library of India coming second in this race is a very good prospect given it host diverse open access resources. But one fails to understand why Consortium for Educational Communication don't get good score in this category given it host multimedia contents only. To decipher secret behind the low score study made an analysis of the site and found site doesn't have intuitive design and layout is quite dated and not enticing. Table 4.

Name of the Portal	Daily Time on Site
Swayam	09:06
National Digital Library of India (NDLI)	06:48
Egyankosh	04:00
Shodhganga	02:58
Consortium for Educational Communication	01:50

Visitors by Country

The Swayam e-learning web portal is overwhelmingly used by users in India (97.30%) followed by National Digital Library of India (96.30%), Consortium for Educational Communication (91.60%). While as Egyankosh is used by 87.20% of users that are from India. However, Shodhganga repository hosting more than 200000 ETDs have more diversity of users with 65.10% from India and rest from other parts of the globe.

After India, the second-best user base of National Digital Library of India is in USA (1.20%) while as second-best user bases of Egyankosh and Shodhganga are from Ethiopia (1.50%) and China (3%) respectively. Table 5

Name of the Portal	India	USA	China	Ethiopia
Swayam	97.30%	---	---	---
National Digital Library of India (NDLI)	96.30%	1.20%	---	---
Consortium for Educational Communication	91.60%	---	---	---
Egyankosh	87.20%	---	---	1.50%
Shodhganga	65.10%	---	3.00%	---

CONCLUSION

The Shodhganga ETD repository has best global and Indian level ranking that reflects its strong utility for the knowledge seekers. This is followed by National Digital Library of India another knowledge resources platform fast become popular among students and scholars of higher education in India. This augurs well for higher education and research in India. These sites usually host primary sources of information. Such information is essential for scientific and developmental growth of any nations. Especially when such information is available free to web.

Web impact factor like journal impact factor is considered one of the essential metrics to reveal the performance of websites, particularly when such sites are associated with education and research. The study gauged the revised web impact factor of select sites which offers true impact of a website. It doesn't take into consideration self-links of a website but purely takes into consideration inlinks (or the links from other sites that direct users to the site). Though the revised web impact factors of select sites is very low with highest for National Digital Library of India. This could be due to two reasons either the contents hosted are considered relevant and qualitative by other education related sites globally or there is lack of awareness among other such players in India and globally. In this backdrop admins and planners of these sites should introspect and come up with comprehensive strategies to claim the rightful place in the cyberspace.

The Swayam turning out to be the most popular as for as time on site is concerned is not a surprise given the fact the swayam is an e-learning platform and host more multimedia content than text based one. But National Digital Library coming second in this race is a very good prospect given it host diverse open access resources. But one fails to understand why Consortium for Educational Communication don't get good score in this category given it host multimedia contents only. To decipher secret behind the low score study made an analysis of the site and found site doesn't have intuitive design and layout is quite dated and not enticing.

References

- Abu, K.S & Jayasekara, P. K. 2017. Webometric Analysis on Leading Course Providers in MOOC: A Study Based on Alexa Ranking. Retrieved from <http://ir.inflibnet.ac.in:8080/ir/bitstream/1944/2101/1/36.pdf>
- Almind, T. C., & Ingwersen, P. 1997. Informetric analyses on the World Wide Web: Methodological approaches to 'Webometrics'. *Journal of Documentation*, 53(4), pp. 404–426. doi:10.1108/EUM0000000007205
- Alexa. 2018. About Us: Information. Insight. Advantage. Retrieved from <https://www.alexa.com/about>
- Aminpour, F., Kabiri, P., Otraj, Z., & Keshtkar, A. A. (2009). Webometric analysis of Iranian universities of medical sciences. *Scientometrics*, 80(1), 253–264. <https://doi.org/10.1007/s11192-008-2059-y>
- Bar-Ilan, J. 2004. The use of Web search engines in information science research. *Annual Review of Information Science and Technology*, 38, pp. 231–288. doi:10.1002/aris.1440380106
- Borgman, C. L., & Furner, J. 2002. Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, 36, pp. 3–72. doi:10.1002/aris.1440360102
- Elgohary, A. 2008. Arab universities on the web: a webometric study. *The Electronic Library*, 26(3), 374–386. <https://doi.org/10.1108/02640470810879518>
- Garfield, E. 1970. Citation indexing for studying science. *Nature*, 227, pp. 669–671. doi:10.1038/227669a0

- Gupta, M., & Walia, P. 2016. WISER ranking of the African national libraries' websites. *Information Technologist*. 13(1)
- Ingwersen, P. 1998. The calculation of web impact factors. *Journal of Documentation*, 54(2), pp. 236–243. doi:10.1108/EUM0000000007167
- Jalal, S. K., Biswas, S. C., & Mukhopadhyay, P. 2010. Web Presence of Selected Asian Countries: A Webometric Study. *Collnet Journal of Scientometrics and Information Management*, 4(2), 57–68. <https://doi.org/10.1080/09737766.2010.10700893>
- Jeyshankar, R., & Ramesh Babu, B. 2009. Websites of universities in Tamil Nadu: a webometric study. *Annals of Library and Information Studies*, 56(June), 69–79.
- Lihitkar, S. 2015. WISER Ranking of Astronomy and Astrophysics Libraries' Websites. In *COLLNET2*. Delhi.
- Madhusudhan, M., & Prakash, S. 2013. Websites of Indian Institutes of Technology : A Webometric Study. *International Journal of Library and Information Studies*, 3(4), 93–107.
- Moed, H. F. 2005. *Citation analysis in research evaluation*. New York: Springer
- Nicolaisen, J. 2007. Citation analysis. *Annual Review of Information Science and Technology*, 41, pp. 609–641. doi:10.1002/aris.2007.1440410120
- Noruzi, A. 2005. Web impact factors for Iranian universities, *Webology*, Vol. 2 No. 1, Retrieved from www.webology.org/2005/v2n1/a11.html
- Noruzi, Alireza 2006. The web impact factor: a critical review, *The Electronic Library*, Vol. 24 Issue: 4, pp.490-500, <https://doi.org/10.1108/02640470610689188>
- Rodríguez i Gairín, J. M. 1997. Valorando el impacto de la información en Internet: AltaVista, el "Citation Index" de la Red. *Revista Española de Documentación Científica*, 20(2), pp. 175–181.
- Smith, A.G. 1999. A tale of two web spaces: comparing sites using web impact factors, *Journal of Documentation*, Vol. 55 No. 5, pp. 577-592. Retrieved from <https://www.learntechlib.org/p/89574/>.
- Thelwall, M. 2000. Web impact factors and search engine coverage", *Journal of Documentation*, Vol. 56 Issue: 2, pp.185-189, <https://doi.org/10.1108/00220410010803801>
- Thelwall, M. 2001. Extracting macroscopic information from web links. *Journal of American Society for Information Science and Technology*, 52(13), pp. 1157–1168. doi:10.1002/asi.1182
- Vijayakumar, M., Kannappanavar, B. U., & Santosh Kumar, K. T. 2012. Webometric analysis of web presence and links of SAARC countries. *DESIDOC Journal of Library and Information Technology*, 32(1), 70–76. <https://doi.org/http://dx.doi.org/10.14429/djlit.32.2.1409>
- Zahedi, Z., Shirazi, M., & Dehghani, L. 2010. A webometric analysis of ISI medical journals using Yahoo, AltaVista, and all of the web search engines. *J Assoc Inf Sci Technol*, 89–108.