

## Webometrics Ranking of Institutional Digital Repositories of India: A Case Study

**Paulomi Saha**

Librarian

Kolkata Teachers' Training Institute  
Kankinara, West Bengal, India  
E-mail paulomisaha.com@gmail.com

**Abstract** - This study deals with the evaluation of the web presence and webometric analysis of the institutional repositories of India. The calculation of the webpages (size), visibility of the websites of the IDRs, the presence of qualitative rich files of the IDRs and the impact of the IDRs in Google scholar is measured with the help of search engines-Google and Google Scholar and with the required Webometric Ranking Indicators (WRI). The Alternative Document Model (ADM) is obtained from calculating the aggregate of all the above mentioned WRI and based on this ADM rank the final webometric rank of the IDRs is obtained, whereas the IDRs are extracted based on high, medium and low 'activity' of the IDRs. Presently, the first 20 IDRs based on High, Medium and Low 'activity' of the IDRs each categorized in the search parameter of Registry of Open Access Repositories (ROAR) is taken for the webometric study out of total 123 repositories of India as recorded by ROAR. This paper also shows a comparative study between maximum and minimum responsive IDRs in respect of every WRI along with their ADM and final rank and also based on the High, medium and low activity of the IDRs. Webometric study is significant to measure the impact of universities, IDRs and library websites so that information professionals, library professionals, researchers, students can get a clear idea to access the qualitative websites to fulfil their information need.

**Keywords:** Webometrics, Institutional Digital Repository, Impact Factor, Web Impact Factor, Ranking, Google, Google Scholar, ROAR, India.

### Introduction:

Web based information resources have a great role to play in academic and research activities. Keeping this fact in view, now days, information professionals are largely depending upon the web based information sources. Huge amount of data in every subject stream is present in different websites. Everybody today would like to be on internet because of the wealth of information that lies there to be exchanged. And, with its global reach and millions of users, the internet is now world's biggest electronic library, which contain a vast amount of information. Now internet users faces problem regarding the retrieval of the desired information. (Thanuskodi, 2012) Many attempts were made and solution came as 'webometrics'<sup>1</sup>. It aims to measure and evaluate impact factor of the websites of institutions, institutional repositories and universities by using webometric indicators tools in order to analyse the visibility of the websites, the presence of rich files in the websites, the total webpages of the institutional websites in order to enhance the efficiency of the websites through optimizing web content and re-designing. The webometric analysis can be performed through the number of Webpages, number of rich files, number of in-links and self-links. An institutional repository needs to be in continual service and their aim should be to recognize

---

<sup>1</sup> <https://en.wikipedia.org/wiki/Webometrics>

that they are making commitments of disseminating information for the long term. The main aim of creating and maintaining IDR is to bring like “hidden shells which contain pearls” to the users from inside as well as outside the institutions who are in search of related scholarly contents”(Lynch, 2004).Academicians, researchers, students or any kind of users can access these scholarly contents across the world through open access system over internet. The web presence of any institution can be measured with the help of search engine’s advanced facilities. Users can take the help of Yahoo, AltaVista and Google whenever and wherever required for the webometric analysis. In the present study Google and Google Scholar are used to evaluate the web activity of the institutional repository of India.

**Objectives:**

- To evaluate the Institutional Repositories of India extracted from ROAR repository on the basis of their websites’ activity.
- To choose the appropriate search engines which can fulfil the objectives of the study.
- To evaluate the total number of web pages, in-links, external links, rich files of IDRs in Google search engine and impact of IDRs in Google scholar by the help of perfect choice of indicators for ranking the repositories of India.
- To measure the alternative document model (ADM) rank by measuring the webometric indicators for ranking (WRI).

**Research Questions:**

- What are the search parameters of the database for evaluating and ranking the impact factor of IDRs of India?Which database or databases will provide the complete list of IDRs of India?What are the tools or indicators for evaluating the impact factor of IDRs and rank them according to their results?
- Which search engine should be selected for the study that will provide all necessary data that needed?Are Google and Google scholar reliable tools for measurement of links to the institutional repository?

**Research methods:**

**Choice of open access repository:**

As per the objective of the study, the URLs of the Institutional Digital Repositories of India are collected from Registry of Open Access Repositories (ROAR) database. The URLs of the IDRs are collected based on the activity of the IDRs of India. The activity of the IDRs measured from- High activity, medium activity and low activity of the IDRs.

**Selection of search engine:**

The Google and Google Scholar search engine is used to obtain link counts of all the IDRs in the present study.

**Choice of indicators for ranking:**

**Table-1: Ranking indicators for webometric ranking**

Indicators	Search engine used	Syntax of the indicators
<b>Size (S):</b> The total webpages of an IDR domain shown in the search engine is the size of that particular IDR.	In the present study Google search engine is used.	site:domain name. e.g.:site:shodhganga.inflibnet.ac.in/

<p><b>Visibility (V):</b> Visibility is determined by the impact of the IDRs online. It is the counting of all the external inlinks that the IDR web domain receives from the third party.</p>	<p>Google search engine is used.</p>	<p>site:url NOT linkdomain:url. e.g.:site:http://shodhganga.inflibnet.ac.in/ NOT linkdomain:http://shodhganga.inflibnet.ac.in/</p>
<p><b>Rich Files (R):</b> The followings are the file formats which are selected after evaluation of their relevance to academic and publication activities - Adobe Acrobat (.pdf), Adobe PostScript (.ps), Microsoft Word (.doc) and Microsoft Powerpoint (.ppt).</p>	<p>Google is used to extract the data of these file formats and to merge the result.</p>	<p>site:urlfiletype:pdf site:urlfiletype:ps site:urlfiletype:doc site:urlfiletype:ppt e.g.:site:http://shodhganga.inflibnet.ac.in/filetype:pdf</p>
<p><b>Google scholar (Sc):</b> Google scholar provides the number of papers and citations for each academic domain, these results from the scholar database represent papers, reports and other academic items in Google Scholar.</p>	<p>Google scholar is used here.</p>	<p>url of the IDR e.g.:shodhganga.inflibnet.ac.in/</p>

### 0.1. Alternative document model (ADM):

ADM is a method of summation of web content information of institutions or databases into units for counting purposes. It can work on any web pages, directory, domain etc. which are concern with the webometric analysis. Web relationship between the sites can be measured with the help of ADM count.

In the present study ADM count is used by aggregating the ranking of IDRs based on the WRI (Size, Visibility, Rich files and Google Scholar) for webometric ranking of the IDRs and a final rank is given to the IDRs according to the Alternative document model ranking. The lowest the ADM ranking, the top will be the final webometric ranking of the IDRs. The four ranks (Visibility, Size, Rich files and Google Scholar) are combined according to a formula where each one has a different weight:

$$\text{Webometric Rank (WR)} = 4 \times \text{Rank V} + 2 \times \text{Rank S} + 1 \times \text{Rank R} + 1 \times \text{Rank Sc.}$$

Where, V=Visibility, S=Size, R=Rich files and Sc=Google Scholar.

### 1. Data collection:

Table 2 shows the webometric rank counting of all the first 20 institutional digital repositories in India based on the “High activity” of the IDRs obtained from the ROAR database. The table comprises the names of IDRs and the Webometric rank calculation(ADM rank) by aggregating all the ranks of the ranking indicators (size, visibility, Rich files and Google scholar) of the IDRs. See Annexure 1 for the ranking of the IDRs according to size, visibility, Rich files and the impact of the IDRs in Google Scholar.

Finally the webometric rank is given to the IDRs, the lowest the ADM rank, the highest is the final rank of the IDRs.

**Table-2: Webometric Rank(WR) counting of the IDRs of India retrieved from ROAR based on High Activity of the IDRs**

NAME OF THE IDRs	ADM RANK	FINAL RANK
Open Access Repository of ICRISAT	23	1
Eprints@CMFRI	32	2
IIT Bombay	40	3
Saurashtra University	41	4
Eprints@NML	47	5
Dyuthi at CUSAT	51	6
ePrints@MoES	53	7
Open Access Repository of Indian Theses	54	8
Ethesis@NIT Rourkela	56	9
Delhi Technological University	60	10
EPrints@IICB	62	11
Dspace@Bangalore Management Academy	68	12
Entrepo	72	13
Eprints@MDRF	74	14
Eprints@IARI	76	15
CSIR-IMMT	82	16
CSIR-AMPRI DRS	83	17
Digital Knowledge Repository of Central Drug Research Institute	92	18
DSpace @ GGSIPU	97	19
ePrints@ATREE	98	20

Table 3 shows the webometric rank counting of all the first 20 institutional digital repositories in India based on the “Medium activity” search parameter of the IDRs obtained from the ROAR database. The table comprises the names of IDRs and the Webometric rank calculation (ADM rank) by aggregating all the ranks of the ranking indicators (size, visibility, Rich files and Google scholar) of the IDRs. See Annexure 1 which comprises ranking the medium activity of the IDRs according to size, visibility, Rich files and the impact of the IDRs in Google Scholar.

Finally the webometric rank is given to the IDRs, the lowest the ADM rank, the highest is the final rank of the IDRs.

**Table-3: Webometric Rank(WR) counting of the IDRs of India retrieved from ROAR based on Medium Activity of the IDRs**

NAME OF THE IDRs	ADM RANK	FINAL RANK
OpenMED@NIC	19	1
Open Access Repository of ICRISAT	28	2
Open Access Repository of Indian Theses	32	3
IIT Bombay	37	4
Eprints@CMFRI	41	5
Saurashtra University	45	6
Eprints@NML	55	7
ePrints@MoES	65	8

IR@NIT Rourkela	69	9
EPrints@IICB	75	10
IR of DTU	76	11
Entrepo	86	12
Ethesis@NIT Rourkela	88	13
Eprints@IARI	91	14
Eprints@MDRF	120	15
Eprint@IMMT	126	16
Digital Knowledge Repository of Central Drug Research Institute	126	17
ePrints@ATREE	126	17
CSIR-AMPRI DRS	129	19
DSpace @ GGSIPU	131	20

Table 4 shows the webometric rank counting of all the first 20 institutional digital repositories in India based on the “Low activity” of the IDRs obtained from the ROAR database. The table comprises the names of IDRs and the Webometric rank calculation (ADM rank) by aggregating all the ranks of the ranking indicators (size, visibility, Rich files and Google scholar) of the IDRs. See Annexure 1 comprising the ranking the low activity of the IDRs according to size, visibility, Rich files and the impact of the IDRs in Google Scholar.

Finally, the webometric rank is given to the IDRs, the lowest the ADM rank, the highest is the final ranking of the IDRs.

**Table-4: Webometric rank(WR) counting of the IDRs of India retrieved from ROAR based on Low Activity of the IDRs**

NAME OF THE IDRs	ADM RANK	FINAL RANK
OpenMED@NIC	21	1
Open Access Repository of Indian Theses	27	2
IIT Bombay	28	3
Open Access Repository of ICRISAT	33	4
Eprints@CMFRI	52	5
Eprints@NML	53	6
Saurashtra University	57	7
ePrints@MoES	66	8
IR@NIT Rourkela	68	9
Delhi Technological University	75	10
EPrints@IICB	75	10
Ethesis@NIT Rourkela	76	12
Eprints@IARI	80	13
Entrepo	86	14
Eprints@MDRF	92	15
CSIR-IMMT	97	16
Digital Knowledge Repository of Central Drug Research Institute	97	16
ePrints@ATREE	100	18
DSpace @ GGSIPU	102	19
CSIR-AMPRI DRS	103	20

## 2. Data Interpretation:

Almost all the IDR's responses on syntaxes used by the webometric ranking indicators (webpages, visibility, rich files and Google Scholar) in Google and Google scholar search engines. The visibility of maximum IDR's websites based on the activity of IDR's as classified in ROAR database- High, medium and low activity shows the same value and thus more or less owns the same rank and the value of maximum IDR's are low. Among the rich files the file formats which shows maximum visibility in the IDR's are pdf and html file formats. Ps and doc file format provides very little resource. The IDR's of medium and low activity search parameters of ROAR are the same and thus extract more or less same result in compare to that of High activity search parameter of the aforesaid database. The top 5 IDR's showing maximum amount of responses regarding webpages, visibility, rich files and effect of IDR's in Google Scholar based on High, medium and low activity of the IDR's listed in the ROAR database are the following:

- Open Access Repository of ICRISAT
- IIT Bombay
- Open Access Repository of Indian Theses
- Eprints@CMFRI
- OpenMED@NIC

But there are also some repositories which exists in the database but does not show much response towards the webometrics ranking indicator syntaxes. They are the following:-

- DSpace @ GGSIPU
- ePrints@ATREE
- CSIR-AMPRI DRS
- CSIR-IMMT
- Digital Knowledge Repository of Central Drug Research Institute
- The reasons behind the poor response of some IDR's might be the changes in websites day by day or may be the non-existence of the website or may be the websites changed without being updated in the ROAR database, thus much needed information are unable to obtain.

## 3. Conclusion:

Webometrics study is useful for several purposes. First, it can be used for fast pilot studies to identify areas for follow-up systematic bibliometric analyses. Second, it can be used to assess the extent to which researchers are successful in publicising their work online, given that this is an important activity. Third, it can be used for relational analyses of communication in disciplinary or geographic areas of science. Fourthly, its methods can help the analysis of Web 2.0 and online repositories for social sciences and humanities research goals. Fifthly, institute websites used by the students to find out a particular course offered, students may look for the semester examination schedule and results, institute websites are used for selection for their admission, some students may download application form and prospectus. Sixthly, faculty members may search for job vacancy and benefits. Finally, Institutional repositories' websites are increasingly used for wide variety of purposes like attracting new students, online library catalogue. (Thelwall, 2007)



Webometrics is a great challenge in the field of social sciences and humanities for the researchers to develop some concrete theories and methodologies which will help to build a solid framework for exploring the complex nature and functionality of the Web in future and also utilizing the benefits of webometric research.

### References:

1. Aguillo, I.F., Ortega, J.L., Fernández, M. &Utrilla, A.M. Indicators for a webometric ranking of open access repositories. *Scientometrics*, Vol. 82( 3), 477-486. Doi:10.1007/s11192-010-0183-y
2. Almind, T.C. &Ingwersen, P. (1997). Informetric analyses on the world wide web: methodological approaches to 'webometrics'. *Journal of Documentation*, Vol. 53, Iss: 4, 404 – 426. Doi:<http://dx.doi.org/10.1108/EUM0000000007205>
3. Asnani, R. (2013). A webometric analysis of institutional repositories using Dspace in India. *INFLIBNET Newsletter*, Vol.20 , No.4, 22-27.
4. Bjerneborn, L. &Ingwersen, P. (2001). Perspective of webometrics. *Scientometrics*, Vol. 50(1), 65-82. Doi:10.1023/A:1005642218907
5. Bjerneborn, L &Ingwergwn, P (2004). Towards a basic framework for Webometrics. *Journal of American Society for Information and Technology*, Vol. 55(14), 16-27. Doi: 10.1002/asi.20077
6. Brookes, B. C. (1990). Biblio-, Sciento-, Infor-metrics??? What are we talking about? *Informetrics*. 89/90, 31-42.
7. Chandra, R., Nugroho, A.P. &Saleh, F. (2014). Evaluating open access paper repository in higher education for ASIAN region. *International Conference on Internet Studies*. Retrieved from <http://arxiv.org/abs/1502.04069>.
8. Chakravarty, R. &Wasan, S. (2015). Webometric analysis of library websites of higher educational institutes (HEIs) of India: a study through Google search engine. *DESIDOC Journal of Library & Information Technology*, Vol. 35, No. 5, 325-329 Doi:10.14429/djlit.35.5.8788
9. Jalal , S. K. , Biswas , S. C. , &Mukhopadhyay , P. S. (2009). Bibliometrics to webometrics. *Indian journals*, Vol., 15(1), 3-20. Retrieved from <http://www.indianjournals.com/>
10. Jalal , S. K. , Biswas , S. C. , &Mukhopadhyay , P. S. (2010). Web impact factor and link analysis of selected Indian universities. *Annals of Library and Information Studies*, Vol.57, 109-121. Retrieved from <http://hdl.handle.net/10760/16164>
11. Nkiruka, A.V., Thomas, A.E. (2015). Availability and utilization of institutional repositories as indicators to institutional web ranking. *European Journal of Computer Science and Information Technology* Vol.3, No.2, 29-40. Retrieved from <http://www.eajournals.org>
12. Noruzi, A. (2006). Web presence and impact factors for middle eastern countries. *Online*, 30(2), 22-28. Retrieved from <http://hdl.handle.net/10760/7264>.
13. Smith , A. G. (2012). Webometric evaluation of institutional repositories. Retrieved from <http://researcharchive.vuw.ac.nz/xmlui/bitstream/handle/10063/2506/paper.pdf?sequence=1> on 30.12.2015
14. Thanuskodi, S. (2012). A webometric analysis of selected institutes of national importance websites in India. *International Journal of Library Science*, 1(1), 13-18. doi : 10.5923/j.library.2012010103
15. Thelwall, M., Vaughan, L., &Bjerneborn, L. (2003). Webometrics. *ARIST*, Vol., 39, 81-135. Doi: 10.1002/aris.1440390110

16. Thelwall, M. (2007). Bibliometrics to Webometrics. *Journal of Information Science*, Vol. 34 (4), 1–18, Doi: 10.1177/0165551506
17. Thomas, O. & Willett, P. (2000). Webometric analysis of departments of librarianship and information science. *Journal of Information Science*, Vol., 26(6), 421-28. Retrieved from <http://jis.sagepub.com/>
18. Vijayakumar, M. ,Kannappanavar , B. U. & K. T. , S. K. (2012). Webometric analysis of web presence and links of SAARC countries. *DESIDOC Journal of Library & Information Technology*, Vol.32(1), 70-76. Retrieved from <http://publications.drdo.gov.in/ojs/index.php/djlit/article/view/1409>
19. Webometrics: Methods and tools. Retrieved from Sodhganga Database from website: [http://shodhganga.inflibnet.ac.in/bitstream/10603/203204/10/10\\_chapter%202.pdf](http://shodhganga.inflibnet.ac.in/bitstream/10603/203204/10/10_chapter%202.pdf)
20. Wells, P. (2009). Institutional repositories: investigating user groups and comparative evaluation using link analysis. Retrieved from <https://www.researchgate.net/>

