

A Webometric Analysis of Institutional Repositories using Dspace in India

Mrs. Roma Asnani, Scientific & Technical Assistant (LS)

Major technological changes have been noticed in scholarly communication and exchange in scientific communication since last few years. Web presence and appearance of scholarly content on the web has gained a special significance. This webometrics study is based on web impact factor of institutional repositories using Dspace in India with the help of OpenDOAR database for data collection, Google search engine and Pajek, network design software through WebCrawler Socscibot 4 for network analysis was used as a graphical tool for creating inlinks topology.

Introduction

Open access and high availability of scholarly content led to completely new area of research that deals with qualitative and quantitative measures of scholarly communication and publication. Open access journals and institutional repositories are two important component of open access publishing system. While open access journals publish scholarly content following traditional tools of peer review and make them available free of cost for the user, institutional repositories host digital documents of an institution that are organized, preserved, and made accessible to worldwide user community. Lynch (2004) has pointed out that an institutional repository needs to be a service with continuity behind it ... Institutions need to recognize that they are making commitments for the long term". The main aim of creating and maintaining IR 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. Researchers can access these scholarly contents across the world through open access system over Internet.

The science of webometrics (also called cybermetrics) deals with measuring the World Wide Web to analyse number and types of hyperlinks, structure of the World Wide Web and usage patterns. Björneborn and Ingwersen (2004), defines webometrics as "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." The term webometric was first coined by Almind and Ingwersen (1997).

One relatively straightforward measure is the "Web Impact Factor" (WIF) first introduced by Ingwersen (1998). According to Alastair G Smith (2013), web impact analyses are based on counting and analysing the URLs returned by search engines in response to queries designed to match documents or ideas with a purpose to produce indicators of the extent to which the documents or ideas are mentioned online, either to directly access web impact or to indirectly assess predominantly offline impact through measurement of the online component of that impact.

Literature Review

Webometric studies is based on the use of large-scale search engines, such as Google, Yahoo, Altavista, Exalead, Blekko and Bing that allow measurements to be made of the total number of pages in a web site and the total number of backlinks to the web site.

Ingwersen (1998), analysed WIF of seven small and medium scale national and four large web domains as well as six institutional web sites. He stated that data

isolation method makes use of sets of inverted but logically identical Boolean set operations and their mean values in order to generate the impact factors associated with internal- (self-) link web pages and external-link web pages.

Alireza Noruzi (2006), stated that Web Impact Factors can be calculated as a way of comparing the attractiveness of web sites or domains on the Web and concluded that while the WIF is arguably useful for quantitative intra-country comparison, application beyond this (i.e., to inter-country assessment) has little value.

Amit Kumar Das (2008) examined 28 institutional digital repositories listed in OpenDoar using altavista search engine and measured the visibility of the website and ranked them through web impact factor external links.

Aguillo (et al.) (2010) pointed out that webometrics indicators can be categorised into three groups namely activity related (documents deposited by author, institution, subject), usage related (visits, visitors, downloads), and visibility related (citations and web links) with the help of search engines. He also shown snapshot of top 300 repositories from a total of 592 worldwide, with a strong presence of US, German and British institutional repositories and the leadership of the large subject repositories through Ranking Web of World Repositories website.

Alastair G Smith (2013) explored webometric measures that could be used to evaluate the impact of institutional

repositories, using institutional repositories at tertiary institutions in Australia as a case study. URL citation inlinks (occurrences of the repositories' URL in the text of web pages), as found through Google searches, as well as links from the general web, links made from other Australasian academic institutions and from Wikipedia.

Objectives

The objectives of this study are as follows:

- To find out visibility and the performance of Indian Institutional Repositories using Dspace;
- To know the relationships between the number of web pages, inlinks and selflinks ; and
- To calculate the web impact factor.

Data Collection

This webometric study is based upon link counts of Indian IRs using Dspace. The data required for analysis was collected from OpenDOAR. OpenDOAR contains data for 67 IRs across the country including 42 IRs that are using Dspace. Google search engine was chosen for getting results from 31 IRs that are active working properly, remaining 11 institutional repositories that are not working properly, were left out of this study. The results required for WIF calculations were obtained from the google search engines using advance queries. Search results were collected for each IR keeping various search conditions constant for each IR within a single day (23-12-2013). The methodology followed can be stated into the following four steps:

Search Command	Results
site: http://ir.inflibnet.ac.in/	Total number of Web Pages
site: http://ir.inflibnet.ac.in/ NOT linkdomain: http://ir.inflibnet.ac.in/	Total number of external links or inlinks
site: http://ir.inflibnet.ac.in/ AND linkdomain: http://ir.inflibnet.ac.in/	Total number of self-links
linkdomain: http://ir.inflibnet.ac.in/	Total number of links

Methodology

Web Impact Factor (WIF) is the web versions of impact factor. The WIF provides quantitative tools for ranking, evaluating, categorizing, and comparing web sites. There are three types of link. External links (inlinks or backlinks) i.e. links coming into a site from other sites, self-links i.e. links within the same site and total links i.e. total number of links to a site. There are three types of WIF: External link WIF, self-link WIF and total WIF.

Let,

- A = Total number of web pages to a particular site
- B = Number of external links (inlinks or backlinks) to a given site
- C = Number of self-links to a given site
- D = Total number of links to a site

Therefore, WIF external link = B/A , and WIF self-link = C/A ,
WIF simple (total links) = D/A ;

Result

The Web Impact Factor for institutional repositories using Dspace in India is given below:

Institutional Repositories	Web pages	External links	Self links	Total links	WIF External links	WIF Self links	WIF total links	Rank by WIF External links
DSpace at Indian Institute of Management Kozhikode	6,820	9,260	11,200	11,200	1.3578	1.6422	1.6422	1
Thapar University (DSpace@TU)	13,600	7,210	9,940	12,900	0.5301	0.7309	0.9485	2
Information & Library Network (Shodhganga: A resevier of Indian Theses)	2,70,000	1,40,000	21,700	1,74,000	0.5185	0.0804	0.6444	3
NISCAIR Online Periodical Repository (NOPR)	1,13,000	39,300	45,000	90,100	0.3478	0.3982	0.7973	4
National Institute of Technology, Rourkela (DSpace@NITR)	12,200	2,910	4,960	8,650	0.2385	0.4066	0.7090	5
Digital repository of Cochin University of Sci & Tech (Dyuthi)	33,400	6,920	9,290	9,290	0.2072	0.2781	0.2781	6
Information & Library Network (DSpace@INFLIBNET)	92,000	17,600	16,100	41,700	0.1913	0.1750	0.4533	7
Indian Institute of Science, Bangalore (edt@IISc)	25,100	4,590	6,150	6,170	0.1829	0.2450	0.2458	8
National Institute of Oceanography (DRS@nio)	48,30,00,000	8,39,00,000	8,55,00,000	12,20,00,000	0.1737	0.1770	0.2526	9
Indian Institute of Technology, Bombay (DSpace@IITB)	10,200	1,500	2,030	5,050	0.1471	0.1990	0.4951	10
Osmania University (OUDL)	76,900	8,980	10,800	10,800	0.1168	0.1404	0.1404	11
Indian Institute of Astrophysics (DSpace@ IIA)	3,98,000	41,800	47,500	78,100	0.1050	0.1193	0.1962	12

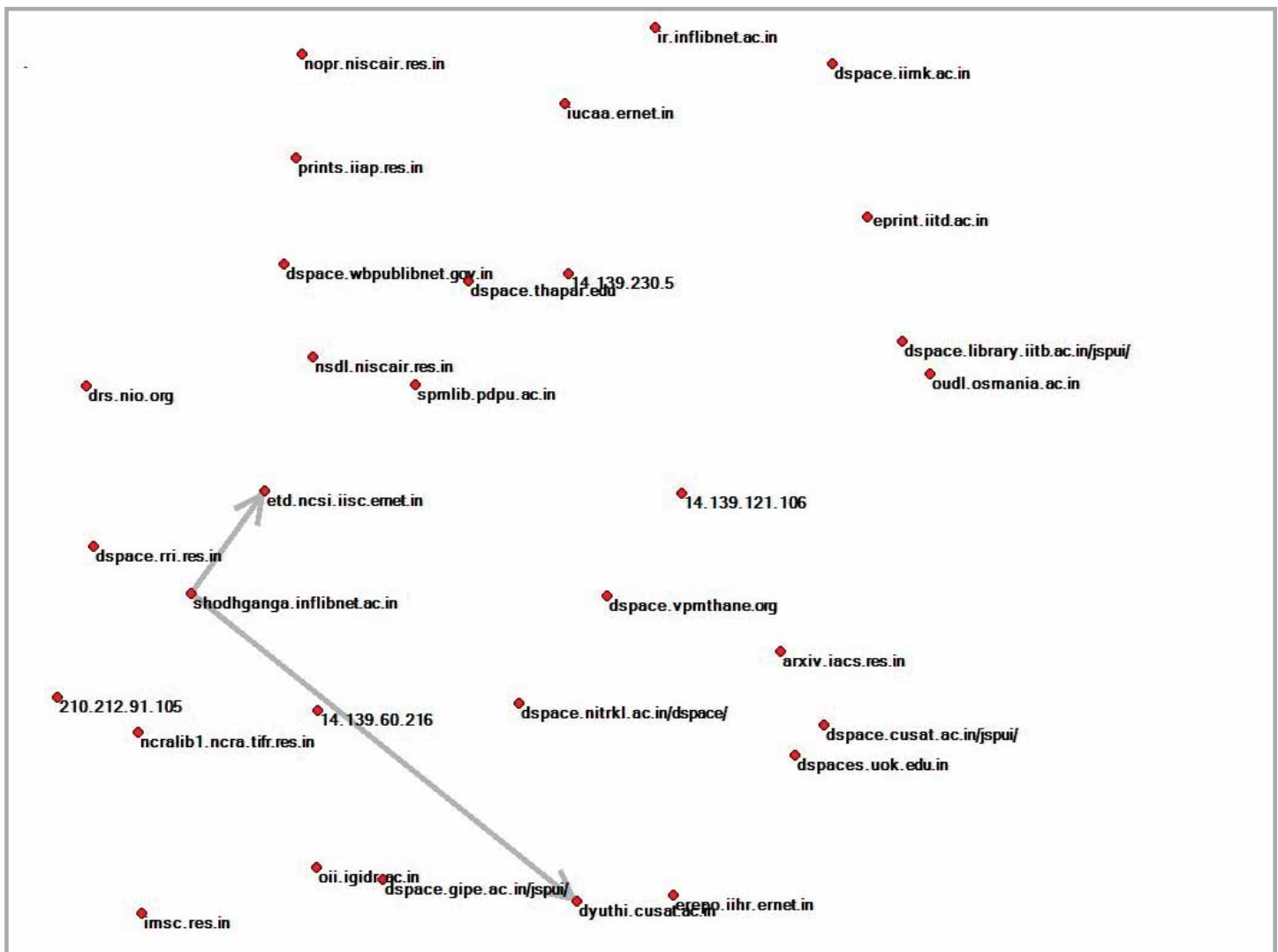
Indian Association for the Cultivation of Science (IACS Institutional Repository)	3,150	146	262	289	0.0463	0.0832	0.0917	13
Digital repository of West Bengal Public Library Network	1,15,000	5,310	7,600	7,610	0.0462	0.0661	0.0662	14
Indian Institute of Technology, Bombay (DSpace @ NCRA)	11,600	354	671	3,150	0.0305	0.0578	0.2716	15
Cochin University of Science & Technology (DigitalLibrary@CUSAT)	41,100	1,130	2,160	3,040	0.0275	0.0526	0.0740	16
Indira Gandhi Institute of Development Research (Kautilya@igidr)	8,960	225	254	329	0.0251	0.0283	0.0367	17
Indian Institute of Technology, Delhi (EPrints@IITD)	24,500	525	1,420	1,480	0.0214	0.0580	0.0604	18
Indian Institute of Horticultural Research (E-Repository@IIHR)	14,900	268	400	403	0.0180	0.0268	0.0270	19
Maharaja Sayajirao University of Baroda (DSpace @ M S University)	3,990	68	92	22,100	0.0170	0.0231	5.5388	20
University of Kashmir (KNoor)	5,790	72	177	254	0.0124	0.0306	0.0439	21
Inter-University Centre for Astronomy and Astrophysics (DSpace @ IUCAA)	28,700	295	379	956	0.0103	0.0132	0.0333	22
Gokhale Institute of Politics and Economics (Dhananjayarao Gadgil Library)	52,900	517	807	3,700	0.0098	0.0153	0.0699	23
National Science Digital Library(NSDL)	61,900	294	368	8,350	0.0047	0.0059	0.1349	24
Pandit Deendayal Petroleum University (DeepBlue Knowledge Repository@PDPUI)	1,360	6	7	9	0.0044	0.0051	0.0066	25
Raman Research Institute (RRI Digital Repository)	3,60,000	1,560	1,940	19,200	0.0043	0.0054	0.0533	26
Guru Gobind Singh Indraprastha University (DSpace @ GGSIPU)	6,280	9	10	21,800	0.0014	0.0016	3.4713	27
Aryabhata Research Institute of Observational Sciences (ARIES, Digital Repository)	8,900	5	6	22,000	0.0006	0.0007	2.4719	28
Institute of Mathematical Sciences (DSpace@IMSC)	7,360	3	268	570	0.0004	0.0364	0.0774	29
Vidya Prasarak Mandal - Thane	115	0	0	54	0.0000	0.0000	0.4696	30
Central Drug Research Institute (DKR@CDRI)	2	0	0	0	0.0000	0.0000	0.0000	31

The WIF analysis evaluated the visibility and impact of institutional and academic web sites, as well as their competitive relations to other web sites. WIF External link measured the visibility of the website and could be ranked them through web impact factor external links. In the above table, Institutional Repositories of Indian Institute of Management Kozhikode, Thapar University and INFLIBNET Centre (Shodhganga) have good WIF external link for web visibility as compared to other repositories. National Institute of Oceanography has highest numbers of webpages, inlinks, selflinks and total links pages. The WIF provides a way to evaluate a web site's relative importance, especially when we compare it to others in the same field, Shodhganga: A reservoir of Indian Theses, Digital repository of Cochin University of Science & Technology (Dyuthi) and Indian Institute of Science, Bangalore (edt@IISc) are electronic theses and dissertation repositories. Shodhganga has highest

number of webpages, inlinks, selflinks and total links pages and web impact factor also highest as compared to other electronic theses and dissertation repositories.

Link mapping of Institutional Repositories

Pajek, Software available on SocSciBot (<http://socscibot.wlv.ac.uk/>) was used for generating inlink topology for 31 institutional repositories covered in the study. The network diagram showing links between 31 IRs covered in this study is given below. The link analysis point out that 3 repositories namely, Shodhganga, Dyuthi and edt@IISc are related to each other on similarity index. It may be noted all the 3 repositories host electronics version of theses and dissertation.

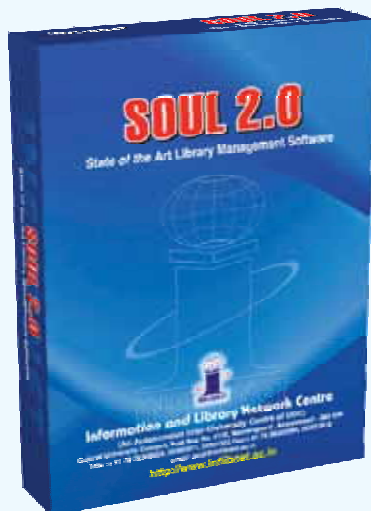


Conclusion:

Ease of access to web sites, site language, site updating and type of materials has all been identified as contributors to the WIF. In the WIF analysis, a good authority is a well-linked, popular page on a topic. The WIFs are always approximate and not absolute. The web pages, inlinks, selflinks, total links and web impact factors are not stable, because everyday webmasters may delete the old outlinks to several websites and others are linking to new ones. Noruzi (2006), stated that a website's ultimate success depends upon its quality and other competitive factors including its size, language, visibility, feasibility and popularity and not only its impact factor. Link analysis using suitable software can be used effectively to identify repositories that are hosting documents of similar nature.

References:

1. Ingwersen, Peter (1998).The calculation of web impact factors. In: Journal of Documentation 54 (2): 236–243. DoI: 10.1108/EUM0000000007167
2. Lynch, C. (2004), "An interview", RLG DigiNews, Vol.8 No.4. Available at:
http://www.rlg.org/en/page.php?Page_ID=19481#article0
3. Björneborn, Lennart and Ingwersen, Peter (2004). Toward a basic framework for webometrics". In: Journal of the American Society for Information Science and Technology 55(14): 1216–1227. DoI: 10.1002/asi.20077.
4. Noruzi, Alireza (2006). The Web Impact Factor: A Critical Review In: Electronic Library, 24 (4), 490-500
5. Das, Amit kumar (2008).Link analysis on institutional digital repositories in India: a webometrics study, In: 23rd National Seminar of IASLIC on Library Profession in Search of a New Paradigm to be held in Bose Institute, Kolkata during 10-13 December 2008, 755-764
6. Aguillo, Isidro F. (et al.) (2010). Indicators for a Webometric Ranking of Open Access Repositories. In: Scientometrics 82 (3), 477-486
7. Smith, A.G. (2013). Web based impact measures for institutional repositories. In: Proceedings of the ISSI 2013 conference Vienna, Austria, 15-19 July 2013
8. <http://socscibot.wlv.ac.uk/>



SOUL 2.0

STATE-OF-THE-ART INTEGRATED
LIBRARY MANAGEMENT SOFTWARE

<http://www.inflibnet.ac.in/soul>