Journal of Library and Information Sciences
June 2020, Volume 8, Number 1, pp. 83-99
ISSN 2374-2372 (Print) 2374-2364 (Online)

Published by American Research Institute for Policy Development
DOI: 10.15640/jlis.v8n1a10
URL: https://doi.org/10.15640/jlis.v8n1a10

Web Presence of Repositories of Indian Institutes of Technology: A Webometric Study

Dr. Babita Jaiswal[1] & Richa Arya[2]

Abstract

Links weave web documents together in a complex, structured hypertext corpus. Web links studies have been conducted by the scholars and information professionals all over the world since its introduction. The main objective of the study is to investigate the web presence of repositories of Indian Institutes of Technology (IITs) as a part of the World Wide Web. To investigate the web presence the few sub-objectives have been framed such as to identify IITs which are having their repositories; to find out the TLD, cc TLD and file formats use by the repositories; to find out total number of Web Pages, Link Web Pages, Self Link Web Pages, External Link Web Pages and Inlink Web Pages; to calculate Overall Web Impact Factor, Simple Web Impact Factor, Self Link Web Impact Factor, External Link Web Impact Factor and Inlink Web Impact Factor. Google covers more hyperlinks with advanced search facilities to count links possessed by the websites. Therefore, the present study has used Google to collect raw data, i.e., number of web pages, number of inlinks, self links and external links etc. Boolean operators are also used to form query syntaxes for retrieving data. SocSciBot web crawler is also employed. It is used to create network map of all the websites. URLs/websites of repositories are tested through command line textual queries like link, linkdomain, site, etc. Findings have been derived after analyzing the data.

Keywords: Webometric, Web Presence, IITs, Indian Institutes of Technology, Repositories

According to Cronin and McKim (1996) the web is reshaping the ways in which scholars communicate with one another. New kinds of scholarly and proto-scholarly publishing are emerging. Thanks to the Web, work-in-progress, broadsides, early drafts and refereed articles are now almost immediately sharable. A study of the web has been named webometrics by Almind and Ingwersen (1997). It is a new research field now passing through a necessary tentative and exploratory phase. Webometric analyses nature, structures and content properties of web sites and pages, as well as link structures are important to understand the virtual highway and their interconnections. According to Bjornborn & Ingwersen (2004) Webometric is 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. This definition covers quantitative aspects of both the construction side and the usage side of the web embracing four main areas of Webometric research: (1) Web page content analysis; (2) Web link structure analysis;(3) Web usage analysis (including log files of users' searching and browsing behavior); (4) Web technology analysis (including search engine performance). According to Thelwall (2009) webometric is concerned with measuring aspects of the web: web sites, web pages, parts of web pages, words in web pages, hyperlinks, web search engine results. The importance of the web itself as a communication medium and for hosting an increasingly wide array of documents, from journal articles to holiday brochures, needs no introduction. Given this huge and easily accessible source of information, there are limitless possibilities for measuring or counting on a huge scale the number of web sites, the number of web pages, the number of blogs or on a smaller scale e.g., the number of web sites in Australia, the number of web pages in The Library of Congress web site, the number of blogs mentioning Mr.Donald Trump before the presidential campaign.

Theoretical overview

Link structure

Web link research in Webometric research started with Larson, Rousseau and Almind & Ingwerson. Larson (1996) was one of the first information scientists to perform an exploratory analysis of the intellectual structure of cyberspace.

[1] Associate Professor, Department of Library & Information Science, Lucknow University, Lucknow (India)
[2] Assistant Professor, Department of Library and Information Science, Institute of Management and Technology, Lucknow (India)

In his classic Webometric article on sitations, that is inlinks, Rousseau (1997) analyses the patterns of distribution of web sites and incoming links. Almind and Ingwersen (1997) applied a variety of bibliometric like methods to the Nordic portion of the web in order to observe the kinds of page connections and define the typology of web pages actually found at national Nordic level. Links weave web documents together in a complex, structured hypertext corpus. Link structures represent implicit human annotations that can be exploited for knowledge discovery. For example, inferring web communities, identifying authoritative web pages, topic distillation, or improving search engine ranking algorithms. The creator of the WWW at CERN in 1989-1990, Berners-Lee, envisaged this development. (Bjornborn & Ingwersen, 2001) Links as a ranking factor are what allowed Google to start to dominate the search engine market back in the late 1990s. One of Google's founders, Larry Page, invented PageRank, which Google used to measure the quality of a page based in part on the number of links pointing to it. This metric was then used as part of the overall ranking algorithm and became a strong signal because it was a very good way of determining the quality of a page. It was so effective because it was based upon the idea that a link could be seen as a vote of confidence about a page, i.e., it wouldn't get links if it didn't deserve to. The theory is that when someone links to another website, they are effectively saying it is a good resource. (Moogan, 2014). Self Link, external links or outgoing link, in-links or incoming links, reciprocal link, back link are different types of links. The web link structure analysis provides hyperlinks or self-links between documents and records of user behavior, counts and analysis of outgoing links, out links or external links from web pages, links to web pages or links coming from the other websites called in-links or incoming links. An incoming link is similar to receiving a citation in a document. These links are also known as backward links or ingoing links.

Web Impact Factor

WIF is a quantitative indicator developed by Peter Ingwersen in 1998 which is based upon link frequencies. It is used to assess the impact of a website or other areas of the web. Prior to Ingwersen, Rodríguez i Gairín introduced the concept of information impact on the Internet in a Spanish documentation journal in 1997. But his study was not as influential as Peter Ingwersen (Noruzi, 2006). Almind & Ingwersen (1997) defined WIF as the ratio of links made to a website, to the number of pages at the website. According to him, there are three types of Web Impact Factor such as simple, overall and external WIFs. Revised Web Impact Factor, Self Link Web Impact Factor (SLWIF) and In-link Web Impact Factor (ILWIF) can also be calculated by obtaining the statistics from well-known search engines. Overall Web Imapct Factor (OWIF) can be calculated by dividing the total links to a website (all inlinks, outlinks, and selflink pages) with the total number of Web pages present in the website at a time. Simple Web Impact Factor (SWIF) can be calculated by adding the total number of inlinks and self link and divided by the total number of web pages published in the web site which are indexed by the search engine, not all web pages available in the web site. External Link WIF for each website is calculated by dividing the external links web pages with the total number of web pages. Self-Link WIF for each website is calculated by dividing the self links web pages with the total number of web pages. In-Link (Revised) WIF for each website is calculated by dividing the in-links web pages with the total number of web pages.

Mukhopadhyay (2004) stated that Ingwersen followed the same method, adopted by ISI (the Institute for Scientific Information, Philadelphia) for calculating Impact Factor. According to Noruzi (2006) JIF measures citations made in journals published during one time period to articles published in another period while the WIF is a "snapshot" of a search engine database at a specific time. Compared with the content of a journal paper, the content of a web resource lacks peer review and thus lacks quality control. The WIF is therefore not exactly the equivalent of the JIF. However, the WIF was inspired by the JIF. The Web Impact Factor provides quantitative tools to rank, evaluate, categorize and compare websites of Top Level Domain, Sub Level Domains and Host Level Domain.

Top Level Domain

A TLD is one of the domains at the highest level in the hierarchical domain name system of the internet. It is the last part of the website that we enter. Originally the TLD space was organized into three main group- countries (ccTLD), categories (gTLD) and multi organizations (iTLD). Countries are designated in the domain name system by their two-letter code set in the ISO 3166-1 standard. For example .in for India, .jp for Japan, .fr for France and so on.

Sub Level Domain

A sub domain is related to the main or root domain. It is an extension of the domain name that will take the browser to a particular place on a webpage.

Host Level or Site Level domain

Each file on the internet is located on a certain "host", a computer connected to the internet. A host can be used to set up individual domains for specific purposes such as a domain for web access. For example in www.lkouniv.ac.in host level domain represented for the University of Lucknow is .lkouniv.

Table 1: Search statements for searching a total number of web pages under the specified domain

Domain	Search statement	Example	Result
ccTop Level Domain (ccTLD)	domain:domainname	domain:in	Retrieve a total number of web pages under the ccTLD of India.
Sub Level Domain (SLD)	domain:subdomainname:domainname	domain:ac.in	Report the total number of web pages under ac.in domain. Here, ac is subdomain academic and in is India.
Host Level Domain (HLD) or Site Level Domain (SLD)	domain:url or site address	domain:iitr.ac.in	Retrieve the total number of pages on the server of IIT,Roorkee.

Chukwunonso et al. (2013) pointed out that the advent of the Internet and the World Wide Web has changed the face of technology especially in the way we communicate and interact. The increasing use of this technology especially in teaching and learning has attracted enormous research efforts. These web-based technologies contribute to collaborative learning by enabling people to interact with each other from different locations. As Alexander & Tate (1999) described that over the past decade, the Internet or the World Wide Web has established itself as the key infrastructure for information administration, exchange, and publication. Wang & Kitsuregawa (2002) reported that search engines are the most commonly used tool to retrieve that information.

Since the mid-1990s increasing efforts have been made to investigate the nature and properties of the World Wide Web by applying modern informetric methodologies to its space of contents, link structures, and search engines. Web-links studies have been conducted by scholars and information professionals all over the world since its introduction. Studies conducted by Sujitha & Jeyshankar (2013), Jeyshankar & Sujitha (2014). Beerappa & Shesadri (2016), Borgman & Furner(2002), Björneborn & Ingwersen (2001), Thelwall (2002a and 2002b) Mukhopadhyay(2004), Björneborn (2004), Tang and Thelwall (2004), Noruzi,(2005) Qiu, Chen, & Wang (2004) and Nwagwu & Omoverere (2008) are notable among them.

Indian Institutes of Technology

The IIT system has twenty three Institutes of Technology. The first of these to be instituted are at Kharagpur (1951), Mumbai (1958), Chennai (1959), Kanpur (1959), Delhi (1961), Guwahati (1994) and Roorkee (1847, Joined IITs in 2001). The IITs are known for their rigorous admission process, which attracts high-achieving Indian students. The academic and research programs of these institutes are comparable to those of leading institutions worldwide. Each IIT may have its own statement of its vision, mission, and core values, but all share a common theme. These statements are essentially excellence driven, holistic in scope, covering education, research, and outreach, emphasizing national relevance yet global understanding, nurturing academic freedom, creativity, innovation, integrity, and the overall development of their students. Quality in all its endeavors is the hallmark of the IIT system. Keeping in view the glorious history of IITs as centre of learning the present study is an initiative to identify the web presence of one of the major component of any modern academic institution i.e. repositories.

Objectives

The main objective of the study is to investigate the web presence of repositories of Indian Institutes of Technology (IITs) as a part of the World Wide Web. To investigate the web presence the following sub-objectives have been framed:

- To identify IITs having their repositories.
- To find out profile details of repositories such as name, URL, etc.
- To find out the TLD, cc TLD and file formats used by the repositories.
- To find out the total number of Web Pages, Link Web Pages, Self Link Web Pages, External Link Web Pages, and Inlink Web Pages.
- To calculate Overall Web Impact Factor, Simple Web Impact Factor, Self Link Web Impact Factor, External Link Web Impact Factor and Inlink Web Impact Factor.
- To establish the ranking of repositories by evaluating their Web Impact Factors.
- To find out the link pattern among the repositories' websites of IITs.

The results of investigations can be of particular value to the government, governing bodies of IITs system and website developers.

Methods and tools

The following procedures were used to study the repositories of Indian Institutes of Technology:

1. Selection of IITs

There are 23 IITs in India. Among them 13 IITs have created their repositories. Therefore, the present study measured repositories of 13 IITs. Table 2, 3 and 4 provides information about the number of IITs, IITs having repositories and profile of repositories respectively.

Table 2: List of IITs

S.N.	Name	Year	URL
1	IIT (BHU) Varanasi	1919	http://www.iitbhu.ac.in/
2	IIT (ISM) Dhanbad	1926	https://www.iitism.ac.in/
3	IIT Kharagpur	1951	http://www.iitkgp.ac.in/
4	IIT Bombay	1958	http://www.iitb.ac.in/
5	IIT Madras	1959	https://www.iitm.ac.in/
6	IIT Kanpur	1959	http://www.iitk.ac.in/
7	IIT Delhi	1963	http://www.iitd.ac.in/
8	IIT Guwahati	1994	http://www.iitg.ac.in/
9	IIT Roorkee	2001	https://www.iitr.ac.in/
10	IIT Roper	2008	http://www.iitrpr.ac.in/
11	IIT Gandhinagar	2008	https://www.iitgn.ac.in/
12	IIT Bhubaneswar	2008	https://www.iitbbs.ac.in/
13	IIT Jodhpur	2008	http://www.iitj.ac.in/
14	IIT Patna	2008	https://www.iitp.ac.in/
15	RAITH (Hyderabad)	2008	http://raiith.iith.ac.in/
16	IIT Indore	2009	https://www.iiti.ac.in/
17	IIT Mandi	2009	http://www.iitmandi.ac.in/
18	IIT Palakkad	2015	https://iitpkd.ac.in/
19	IIT Tirupati	2015	http://www.iittp.ac.in/
20	IIT Bhilai	2016	http://www.iitbhilai.ac.in/
21	IIT Goa	2016	http://www.iitgoa.ac.in/
22	IIT Jammu	2016	http://www.iitjammu.ac.in/
23	IIT Dharwad	2016	http://www.iitdh.ac.in

Table 3: List of IITs having their repositories

S. N.	IIT
1	Indian Institute of Technology, Delhi
2	Indian Institute of Technology,Roper
3	Indian Institute of Technology Gandhinagar
4	Indian Institute of Technology, Kharagpur
5	Indian Institute of Technology, Bhubaneswar
6	Indian Institute of Technology, Indore
7	Indian Institute of Technology, Jodhpur
8	Indian Institute of Technology,Patna
9	Indian Institute of Technology, Bombay
10	Indian Institute of Technology, Madras
11	Indian Institute of Technology,Guwahati
12	Indian Institute of Technology, (Hyderabad)
13	Indian Institute of Technology, Roorkee

Table-4: Profile of Repositories

S. N.	Name of IIT	Repositories	URL
1	IIT Delhi	C L IITD	eprint.iitd.ac.in/
2	IIT Roper	CL IITRPR DSpace R	http://dspace.iitrpr.ac.in/
3	IIT Gandhinagar	D R IITGN	https://www.iitgn.ac.in/
4	IIT Kharagpur	IDR (D L KGP)	http://www.idr.iitkgp.ac.in
5	IIT Bhubaneswar	IDR@IITBBS	idr.iitbbs.ac.in/
6	IIT Indore	IITI IDR	dspace.iiti.ac.in/
7	IIT Jodhpur	IITJ IR	ir.iitj.ac.in/
8	IIT Patna	IR of IITP	http://idr.iitp.ac.in/jspui/
9	IIT Bombay	IR of IITB	dspace.library.iitb.ac.in/
10	IIT Madras	IRepose IITM	http://irepose.iitm.ac.in/
11	IIT Guwahati	LBCL D R	gyan.iitg.ernet.in/
12	RAIITH Hyderabad	RAIITH	http://raiith.iith.ac.in
13	IIT Roorkee	S B IITR	shodhbhagirathi.iitr.ac.in/

2. Webometric tools

For conducting a webometric study selection of search engines is required for counting the number of pages on the website and pages linked to the websites. Smith (1999a) stated that a search engine should have a large database, covering as much of the Web as possible. Presently, Google has the largest databases and search commands both for links and for the number of pages at a web site. Jalal, Biswas and Mukhopadhyay (2010), Beerappa and Sheshadri(2016),Chakravarty & Wasan(2015) made their study by using Google. As Google covers more hyperlinks with advanced search facilities to count links possessed by the websites the present study also used Google to collect raw data, i.e., number of webpages, number of in-links, self links and external links etc.. Boolean operators are also used to form query syntaxes for retrieving data. SocSciBot web crawler is also employed. It is used to create a network map of all the websites. URLs/websites of repositories are tested through command line textual queries like link, linkdomain, site, etc.

3. Collection of Data

An excel sheet was created to record the data. The following data were collected and recorded on the excel sheet:

- Number of Webpages
- Number of Link pages

- Number of Self Link Web Pages
- Number of Inlink Web Pages
- Number of External Link Web Page

3.1 Query syntaxes

Query syntaxes which are being used in webometric studies for retrieving the data are given in the table 5.

Table-5: Query syntax

S.N.	Query syntaxes	Search results
1	domain: URL	No. of web pages at the websites under the URL
2	Link domain: URL	No. of link pages linking to the websites under the URL
3	link domain: URL AND domain: URL	No. of self-link pages from the same website (hyperlink)
4	link domain: URL AND NOT domain: URL	No. of web pages not underthe URL but provide a hyperlink to URL hence called external link pages
5	link domain: URL NOT domain: URL	No. of links incoming from other websites (In-links)

These query syntaxes are designed to retrieve information about links available on the web. Ayoub et.al. (2019), Ahmad et.al.(2018), Khan & Idrees (2015), Prabakaran & Lihitkar (2015), Jalal, Biswas & Mukhopadhyay (2010), Noruzi (2006) & Mukhopadhyay(2004) have used these query syntaxes to collect webometric data. The present study also used these syntaxes to collect data. The specific search syntaxes which are used for different repositories are given in Table 6.

Table-6: Search Syntax used to collect data about repositories

S.N.	Name of IR	domain: URL	Link domain: URL	link domain: URL AND domain: URL	link domain: URL AND NOT domain: URL	link domain: URL NOT domain: URL
1	C L IITD	domain:eprint.iitd.ac.in	Linkdomain:eprint.iitd.ac.in	link domain: eprint.iitd.ac.in AND domain: eprint.iitd.ac.in	link domain:eprint.iitd.ac.in AND NOT domain: eprint.iitd.ac.in	link domain: eprint.iitd.ac.in NOTdomain: eprint.iitd.ac.in
2	CL IITR PR Dspace R	domain:dspace.iitrpr.ac.in	Linkdomain:dspace.iitrpr.ac.in	linkdomain:dspace.iitrpr.ac.in AND domain: dspace.iitrpr.ac.in	linkdomain:dspace.iitrpr.ac.in AND NOT domain: dspace.iitrpr.ac.in	link domain: dspace.iitrpr.ac.in NOTdomain: dspace.iitrpr.ac.in
3	D R IITGN	domain:repository.iitgn.ac.in	Linkdomain:repository.iitgn.ac.in	linkdomain:repository.iitgn.ac.in AND domain: repository.iitgn.ac.in	linkdomain:repository.iitgn.ac.in AND NOT domain: repository.iitgn.ac.in	linkdomain:repository.iitgn.ac.in NOTdomain: repository.iitgn .ac.in
4	IDR(DLKGP)	domain: idr.iitkgp.ac.in	Linkdomain:idr.iitkgp.ac.in	link domain: idr.iitkgp.ac.in AND domain: idr.iitkgp.ac.in	link domain: idr.iitkgp.ac.in AND NOT domain: idr.iitkgp.ac.in	link domain: idr.iitkgp.ac.in NOT domain: idr.iitkgp.ac.in
5	IDR@IITBBS	domain: idr.iitbbs.ac.in	Linkdomain:idr.iitbbs.ac.in	link domain: idr.iitbbs.ac.in AND domain: idr.iitbbs.ac.in	link domain: idr.iitbbs.ac.in AND NOT domain: idr.iitbbs.ac.in	link domain: idr.iitbbs.ac.in NOTdomain: idr.iitbbs.ac.in
6	IITI IDR	domain: dspace.iiti.ac.in	Linkdomain:dspace.iiti.ac.in	link domain: dspace.iiti.ac.in AND domain: dspace.iiti.ac.in	link domain: dspace.iiti.ac.in AND NOT domain: dspace.iiti.ac.in	link domain: dspace.iiti.ac.in NOTdomain: dspace.iiti.ac.in
7	IITJ IR	domain: ir.iitj.ac.in	Linkdomain:ir.iitj.ac.in	link domain: ir.iitj.ac.in AND domain: ir.iitj.ac.in	link domain: ir.iitj.ac.in AND NOT domain: ir.iitj.ac.in	link domain: ir.iitj.ac.in NOT domain: ir.iitj.ac.in
8	IR of IITP	domain: idr.iitp.ac.in	Linkdomain:idr.iitp.ac.in	link domain: idr.iitp.ac.in AND domain: idr.iitp.ac.in	link domain: idr.iitp.ac.in AND NOT domain: idr.iitp.ac.in	link domain: idr.iitp.ac.in NOT domain: idr.iitp.ac.in
9	IR of IITB	domain:dspace.library.iitb.ac.in	Linkdomain:dspace.library.iitb.ac.in	linkdomain:dspace.library.iitb.ac.in AND domain: dspace.library.iitb.ac.in	link domain: dspace.library.iitb.ac.in AND NOT domain: dspace.library.iitb.ac.in	linkdomain:dspace.library.iitb.ac.in NOT domain: dspace.library.iitb.ac.in
10	IRepose IITM	domain:irepose.iitm.ac.in:	Linkdomain:irepose.iitm.ac.in	linkdomain:irepose.iitm.ac.in: AND domain: irepose.iitm.ac.in:	link domain: irepose.iitm.ac.in: AND NOT domain: irepose.iitm .ac.in:	link domain: irepose.iitm.ac.in: NOT domain: irepose.iitm.ac.in:
11	LBCL D R	domain: gyan.iitg.ernet.in	Linkdomain:gyan.iitg.ernet.in	linkdomain:gyan.iitg.ernet.in AND domain: gyan.iitg.ernet.in	link domain: gyan.iitg.ernet.in AND NOT domain: gyan.iitg.ernet.in	link domain: gyan.iitg.ernet.in NOTdomain: gyan.iitg.ernet.in
12	RAIITH	domain: raiith.iith.ac.in	Linkdomain:raiith.iith.ac.in	link domain: raiith.iith.ac.in AND domain: raiith.iith.ac.in	link domain: raiith.iith.ac.in AND NOT domain: raiith.iith.ac.in	link domain: raiith.iith.ac.in NOTdomain: raiith.iith.ac.in
13	S B IITR	domain:shodhbhagirathi.iitr.ac.in	Linkdomain:shodhbhagirathi.iitr.ac.in	linkdomain:shodhbhagirathi.iitr.ac.in AND domain: shodhbhagirathi.iitr.ac.in	linkdomain:shodhbhagirathi.iitr.ac.in AND NOT domain: shodhbhagirathi.iitr.ac.in	linkdomain:shodhbhagirathi.iitr.ac.in NOT domain: shodhbhagirathi.iitr. ac.in

4. Calculation of Web Impact Factor (WIF)

S.N.	Web Impact Factor	Formula
1	OWIF(Overall Web Impact Factor)	Total links to a website (all inlinks, Outlinks and selflink pages) / Total number of Web pages present in the website at a time
2	SWIF (Simple Web Impact Factor)	Number. of in-Link & self- links Web Pages / Total Number of Web Pages
3	SLWIF (Self-Link Web Impact Factor)	Number of Self-Link Web Pages / Total Number of Web Pages
4	ELWIF (External-Link Web Impact Factor)	Number of External Link Web Pages / Total number of Web Pages
5	ILWIF (In-Link Web Impact Factor)	Number of In-Link Web Pages / Total Number of Web Pages

Table 7: Formula for calculating Web Impact Factor

The WIF can be regarded as a tool for measuring the accuracy of web search engine performance, web site organization, linking, and structuring of pages. Formulas used for calculation of WIF are given in the table 7. In the present study WIF is calculated by keeping the fact in mind that Web Impact Factor depends on the coverage of search engines. According to Nourzi (2006) the impact factor of any website will be proportional to the search engines' coverage. Furthermore, the web site sets in search engines' databases are not constant but may vary in composition from month to month. So, the results of analyses of the Web by search engines can only be regarded as rough indications rather than definite conclusions.

5. Data Analysis

5.1 General Information about IITs' Repositories

Table-8: Software adopted by IRs

Software	Frequency	Percentage
Dspace	12	92
Eprints	1	8
Total	13	100

Table 8 shows software used by IITs to create Institutional Repository. Open-source digital library packages are gaining popularity nowadays. The Institutional Repositories in India are rapidly growing with the help of open-source software like DSpace, Greenstone, Eprint, Nitya, HTML etc. It is found that 12(92%) IITs use DSpace for maintaining IR. It is because DSpace customizes the application to fit their needs, as it is a free and open-source. DSpace can manage and preserve all types of digital content and frequently used by educational, government, private and commercial institutions. RAITH, Hyderabad use Eprint software.

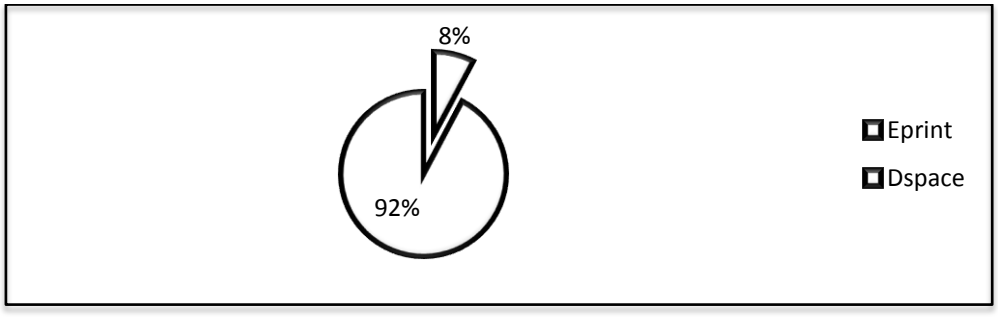

Figure-1: Software used by IR

Table-9: Generic Top-level domain used by IRs

GTLD	Frequency	Percentage
.ac	12	92
.ernet	1	8
Total	13	100

Table-10: URL analysis of Institutional Repository websites

S.N.	Name of IIT	URL	Top-level domain	
			Generic TLD (gTLD)	Country code TLD (ccTLD)
1	C L IITD	eprint.iitd.ac.in/	.ac	.in
2	CLIITR PR DSpace R	http://dspace.iitrpr.ac.in/	.ac	.in
3	D R IITGN	https://www.iitgn.ac.in/	.ac	.in
4	IDR (D L KGP)	http://www.idr.iitkgp.ac.in	.ac	.in
5	IDR@IITBBS	idr.iitbbs.ac.in/	.ac	.in
6	IITI IDR	dspace.iiti.ac.in/	.ac	.in
7	IITJ IR	ir.iitj.ac.in/	.ac	.in
8	IR of IITP	http://idr.iitp.ac.in/jspui/	.ac	.in
9	IR of IITB	dspace.library.iitb.ac.in/	.ac	.in
10	IRepose IITM	http://irepose.iitm.ac.in/	.ac	.in
11	LBCL D R	gyan.iitg.ernet.in/	.ernet	.in
12	RAIITH	http://raiith.iith.ac.in/	.ac	.in
13	S B IITR	shodhbhagirathi.iitr.ac.in/	.ac	.in

A uniform resource locator (URL), which is also known as Web address, is reference to a web resources. The first part of the URL consists of the protocol, the second specifies the domain names which is followed by directory and file name. For studying the URLs of the IITs Institutional repositories websites, the domain names have been taken into account. Table 9 and 10 reveals most frequently used TLDs of IITs IR websites. The TLD .ac.in is used by 12 (92%) IR websites,whereas LBCL Digital Repository of IITG has used .ernet.in. The TLDs are further divided into generic TLD (gTLD) and country code TLD (ccTLD). The TLD .ac stands for academic domain, whereas .ernet stands for Education and Research Network. The ccTLD .in, which stands for the country India, has been used by all of the IITs library websites (i.e., 100%).

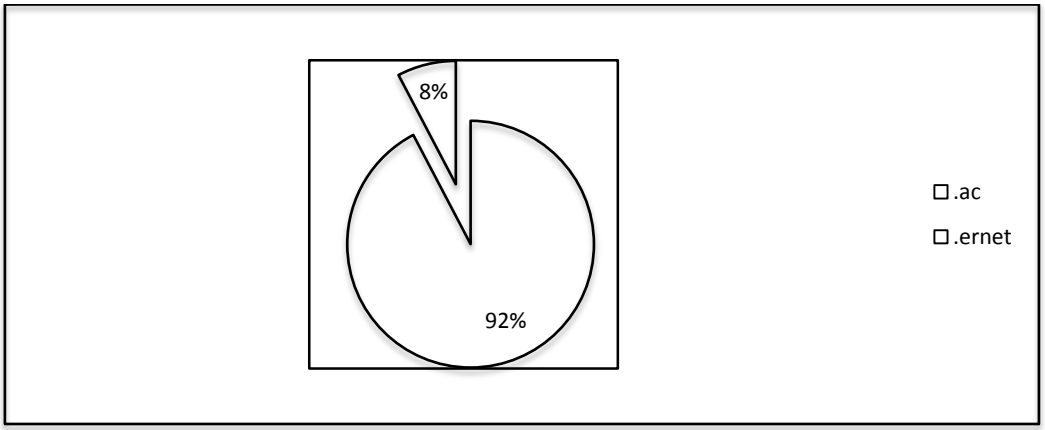

Figure- 2: URL analysis of Institutional Repository websites

Table-11: Types of File Formats Supported by IITs Repository

		File Supported								
S. N.	**Name**	**HTML**	**PDF**	**DOC**	**Excel**	**PPT**	**JPG/JPEG**	**PNG**	**GIF**	**Total**
1	C L IITD	×	√	√	×	×	×	×	×	2
2	CL IITR Dspace R	✓	√	√	×	×	×	×	×	3
3	D R IITGN	√	√	√	×	√	×	×	×	4
4	IDR (DL KGP)	×	√	√	×	×	×	×	×	2
5	IDR@IITBBS	×	√	×	×	×	×	×	×	2
7	IITI IDR	√	√	×	×	×	×	×	×	2
8	IITJ I R	√	√	×	×	×	×	×	×	2
6	IR of IITP	√	√	×	×	×	×	×	×	2
9	IR of IITB	×	√	√	×	×	×	×	×	2
10	IRepose IITM	√	×	×	×	×	×	×	×	1
11	LBCL D R	√	×	×	×	×	×	×	×	1
12	RAIITH	√	√	×	×	×	×	×	×	2
13	S B IITR	√	√	×	×	×	×	×	×	2
Total		9	11	5	0	1	0	0	0	27

A file format is a standard way wherein information is encoded for storage in a computer file. The format of a file is based on the end of its name, i.e., the letters following the final period. This portion of the filename is known as the filename extension. For example, HTML documents are identified by names that end with .html (or .htm), pdf files (.pdf) and GIF images by .gif. Table 11 shows the type of file formats used by repositories. After analyzing all the Web Pages inside the repository of IITs it is found that document file formats (.html, .pdf, .doc, .ppt), image file formats (.jpg/.jpeg, .png, .gif, etc.) are used to display the information about the repository. Table shows that D.R. IITGN uses highest number of file format i.e. 4 followed by C L IITR DSpace that uses 3 file formats. The most popular file format is PDF as 11 repositories use this format. HTML is used by 9 repositories. Other file formats are not very popular. No repository uses JPG/JPEG, PNG and GIF and MS Excel file.

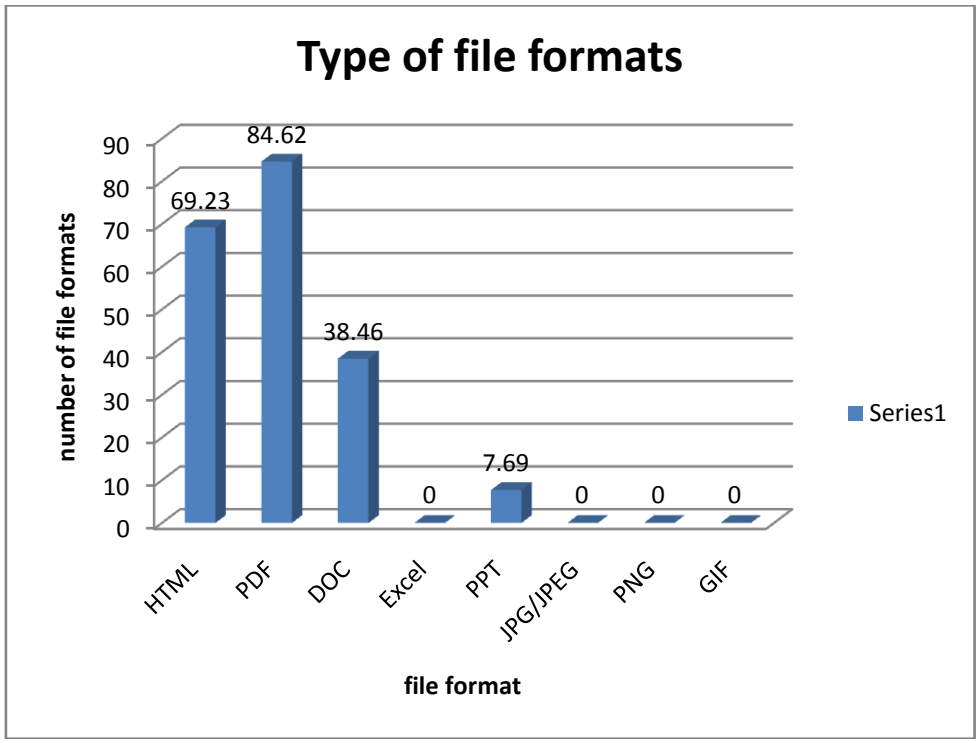

Figure-3: File formats supported by Repositories

5.2 Link Analysis

Table-12: Web Pages (WP), Link Web Pages (LP) and Overall Web Impact Factor (OWIF)

S.N.	Name of IR	Total Number of Web Pages (A)	Links Web Pages(B)	OWIF(A/B)	OWIF (A/B*100)%	Rank
1	C L IITD	9,840	8650	1	87.9	6
2	CL IITR Dspace R	7,990	7790	0.974	97.4	2
3	D R IITGN	7,440	6560	1	88.1	5
4	IDR (DL KGP)	30,000	9	0.0003	0.03	12
5	IDR@IITBBS	345	308	0.892	89.2	4
6	IITI IDR	33,400	28000	1	83.8	7
7	IITJ IR	34,100	11900	0	34.897	11
8	IR of IITP	5,450	2500	0.458	45.8	9
9	IR of IITB	187000	124000	0.663	66.3	10
10	IRepose IITM	28,700	5	0.0002	0.02	13
11	LBCL D R	20,600	19500	1	94.6	3
12	RAIITH	4,630	3420	1	73.8	8
13	S B IITR	27,700	27200	0.981	98.1	1

"Web impact factor (WIF) is the number of outside web pages linking to a website which is divided by the number of web pages in that very website at a time". (Ingwersen, 1998). Table 12 shows that the overall web impact factor of the repositories. Based on WIF, each repository is ranked. It is clear from the data that SB IITR has got first rank with the highest 98.1% OWIF. CL IITR DSpace with 97.4% OWIF and LBCL D R with 94.6% OWIF secured second and third rank respectively. IRepose IITM has the lowest WIF at 0.02%.

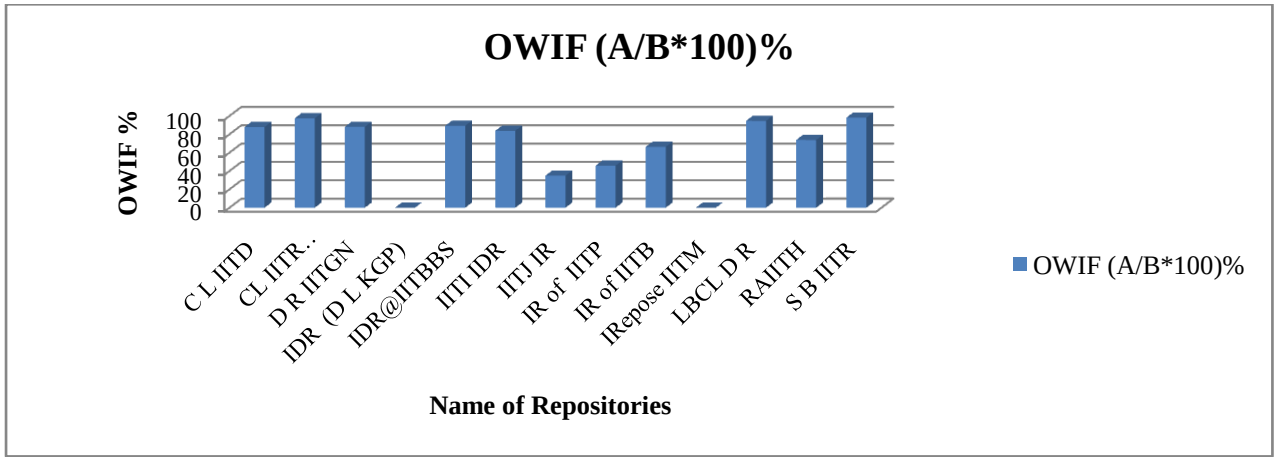

Figure-4: Overall Web Impact Factor

Table-13: Link Pages (LP) and Simple Web Impact Factor (SWIF)

S.N.	Name of IR	Total Number of Web Pages (A)	Self-Link Web Pages (C)	In-Link Web Pages (E)	SWIF (self+in/TNWP)	SWIF (C+E/A*100)%	Rank
1	C L IITD	9,840	9,040	8,730	1.805	18.05	8
2	CL IITR Dspace R	7,990	7780	7760	1.944	19.44	6
3	D R IITGN	7,440	9,190	6,350	2.088	20.88	4
4	IDR (D L KGP)	30,000	29600	29400	1.966	19.66	5
5	IDR@IITBBS	345	359	277	1.843	18.43	7
6	IITI IDR	33,400	22,700	16,500	1.173	11.73	12
7	IITJ IR	34,100	23,300	16,100	11.588	115.88	1
8	IR of IITP	5,450	4,610	4480	1.67	16.7	9
9	IR of IITB	187000	150000	135000	1.524	15.24	10
10	IRepose IITM	28,700	10	7	0.0006	00.6	13
11	LBCL D R	20,600	23900	20,800	2.169	21.69	3
12	RAIITH	4,630	2,830	2,740	1.203	12.03	11
13	S B IITR	27,700	30900	30700	2.223	22.23	2

Table 13 shows Simple Web Impact Factor of the repositories. It is clear from the table that IITJ IR has got first rank with the highest 115.88 % SWIF. SB IITR with 22.23% SWIF and LBCL D R with 21.69% SWIF secured second and third rank respectively. On the other hand, RAIITH 12.03%, IITI IDR 11.73% and IRepose IITM are lowest in rank.

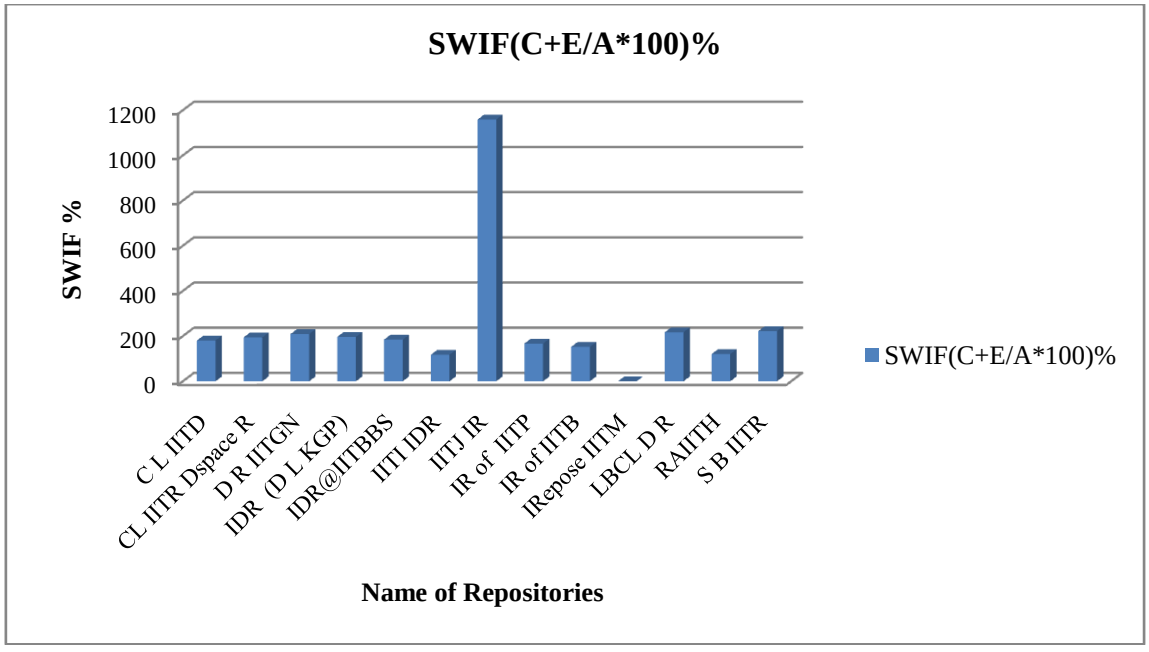

Figure-5: Simple Web Impact Factor

Table-14: Self Link Web Pages (SLWP) and Self Link Web Impact Factor (SLWIF)

S.N.	Name of IR	Total No. Web Pages (A)	Self-Link Web Pages (C)	SLWIF(A/C)	SLWIF (A/C*100)%	Rank
1	C L IITD	9,840	9,040	0.918	91.8	7
2	CL IITR Dspace R	7,990	7780	0.973	97.3	6
3	D R IITGN	7,440	9,190	1.235	123.5	1
4	IDR (D L KGP)	30,000	29600	0.986	98.6	5
5	IDR@IITBBS	345	359	1.04	104	4
6	IITI IDR	33,400	22,700	0.679	67.9	11
7	IITJ IR	34,100	23,300	0.683	68.3	10
8	IR of IITP	5,450	4,610	0.845	84.5	8
9	IR of IITB	187000	150000	0.802	80.2	9
10	IRepose IITM	28,700	10	0.0003	0.03	13
11	LBCL D R	20,600	23900	1.16	116	2
12	RAIITH	4,630	2,830	0.611	61.1	12
13	S B IITR	27,700	30900	1.115	111.5	3

Table 14 shows the Self Links Web Impact Factor of the institutional repository. It is clear from the data that DR IITGN has secured first rank with 9190 self link web pages and its web impact factor is 123.5%. LBCL DR with 116% SLWIF secured second rank while SB IITR secured third rank with 111.5% SLWIF. IRepose IITM with 10 self link web pages got lowest rank.

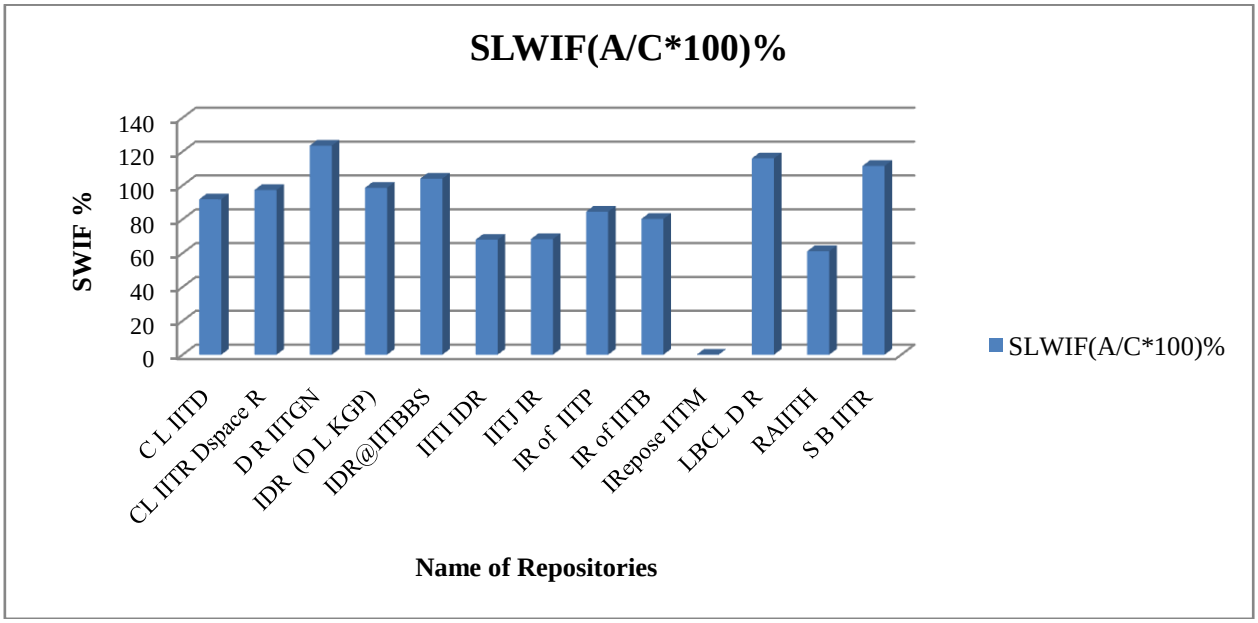

Figure-6: Self–Links Web Impact Factor

Table-15: External Link Web Pages and External Web Impact Factor

S.N.	Name of IR	Total No. Web Pages (A)	External Link Web Pages (D)	ELWIF(A/D)	ELWIF (A/D*100)%	Rank
1	C L IITD	9,840	8,830	0.897	89.7	5
2	CL IITR Dspace R	7,990	7780	0.973	97.3	3
3	D R IITGN	7,440	6,390	0.858	85.8	7
4	IDR (D L KGP)	30,000	29700	0.99	99	2
5	IDR@IITBBS	345	329	0.953	95.3	4
6	IITI IDR	33,400	24,500	0.733	73.3	9
7	IITJ IR	34,100	19,600	0.574	57.4	12
8	IR of IITP	5,450	4490	0.823	82.3	8
9	IR of IITB	187000	128000	0.684	68.4	10
10	IRepose IITM	28,700	6	0.0002	0.02	13
11	LBCL D R	20,600	18,200	0.883	88.3	6
12	RAIITH	4,630	2,740	0.591	59.1	11
13	S B IITR	27,700	30800	1.112	111.2	1

Table 15 reveals that S B IITR has 30800 external web pages and got first rank with the 111.2% EWIF. IDR (D L KGP) with 99% EWIF secured second rank. Table reveals RAIITH, IITJ IR, and IRepose IITM are lowest in rank.

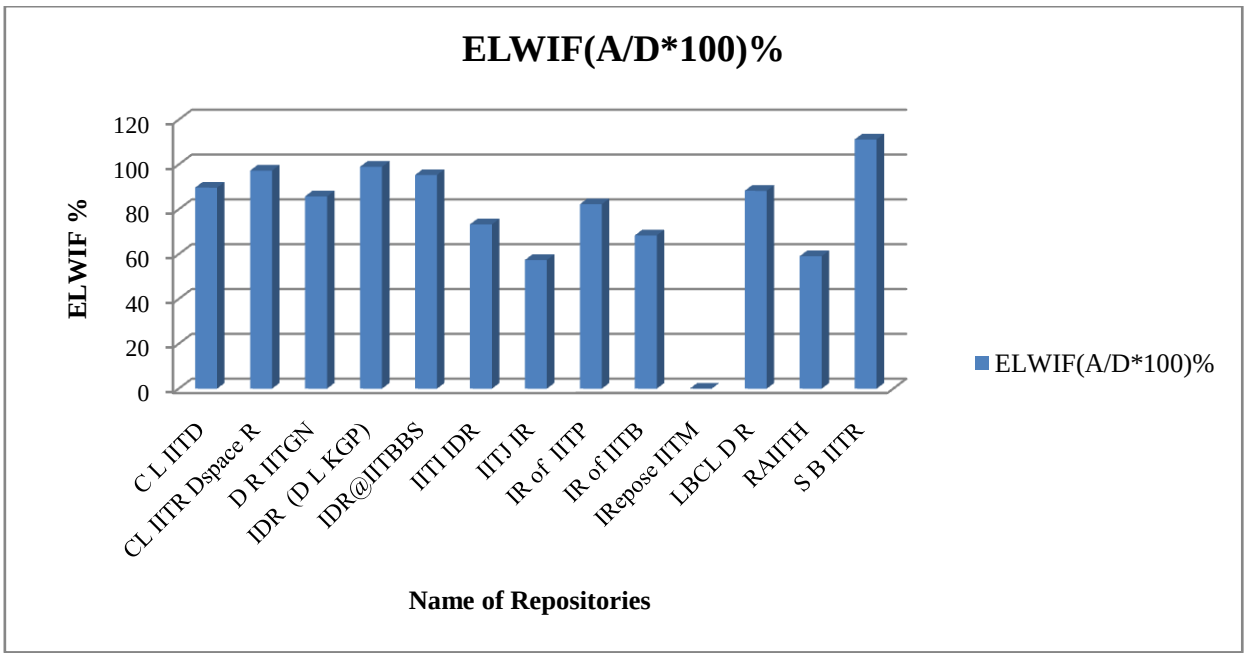

Figure-7: External-Links Web Impact factor

Table-16: In-link Web Pages (ILWP) and In-Link Web Impact Factor (WIF)

S.N.	Name of IR	Total No. Web Pages (A)	In-Link Web Pages (E)	ILWIF(A/E)	ILWIF (A/E*100)%	Rank
1	C L IITD	9,840	8,730	0.887	88.7	5
2	CL IITR Dspace R	7,990	7760	0.971	97.1	4
3	D R IITGN	7,440	6,350	0.853	85.3	6
4	IDR (D L KGP)	30,000	29400	0.98	98	3
5	IDR@IITBBS	345	277	0.802	80.2	7
6	IITI IDR	33,400	16,500	0.494	49.4	10
7	IITJ IR	34,100	16,100	0.472	47.2	11
8	IR of IITP	5,450	4480	0.822	82.2	7
9	IR of IITB	187000	135000	0.722	72.2	8
10	IRepose IITM	28,700	7	0.0002	0.02	12
11	LBCL D R	20,600	20,800	1.009	100.9	2
12	RAIITH	4,630	2,740	0.591	59.1	9
13	S B IITR	27,700	30700	1.108	110.8	1

Table 16 shows the In-links Web Impact Factor of repositories. SB IITR has got the first rank with the 110.8% ILWIF. LBCL D R and IDR (D L KGP) with 100.9% and 98% ILWF secured second and third rank respectively. IITJ IR with 47.2% ILWIF is lowest in rank order.

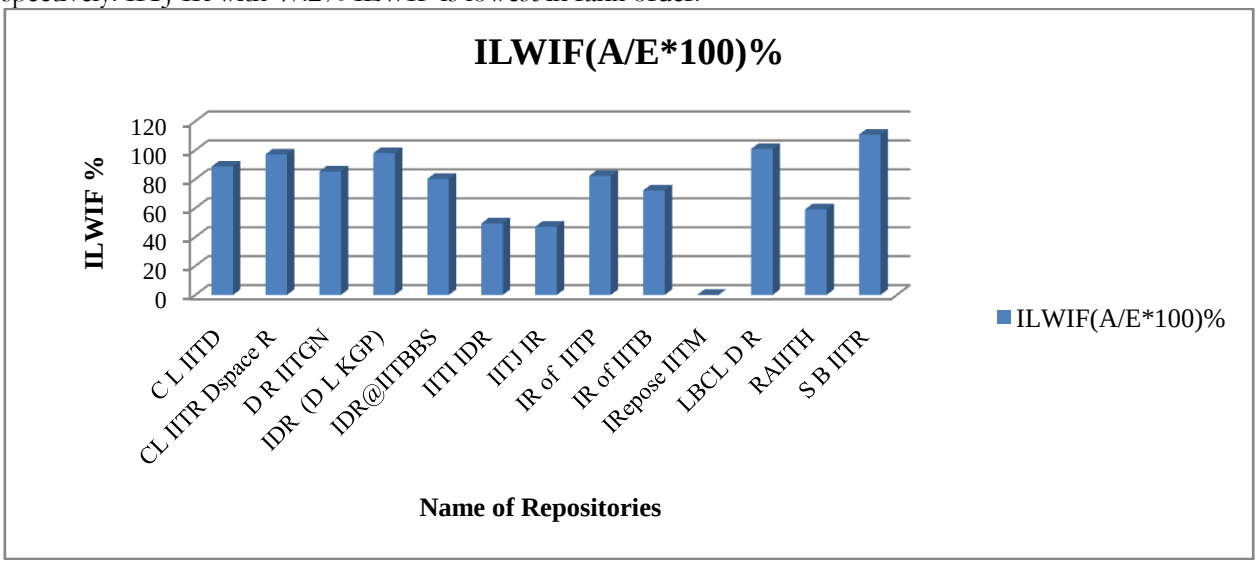

Figure-8: In-links Web Impact Factor

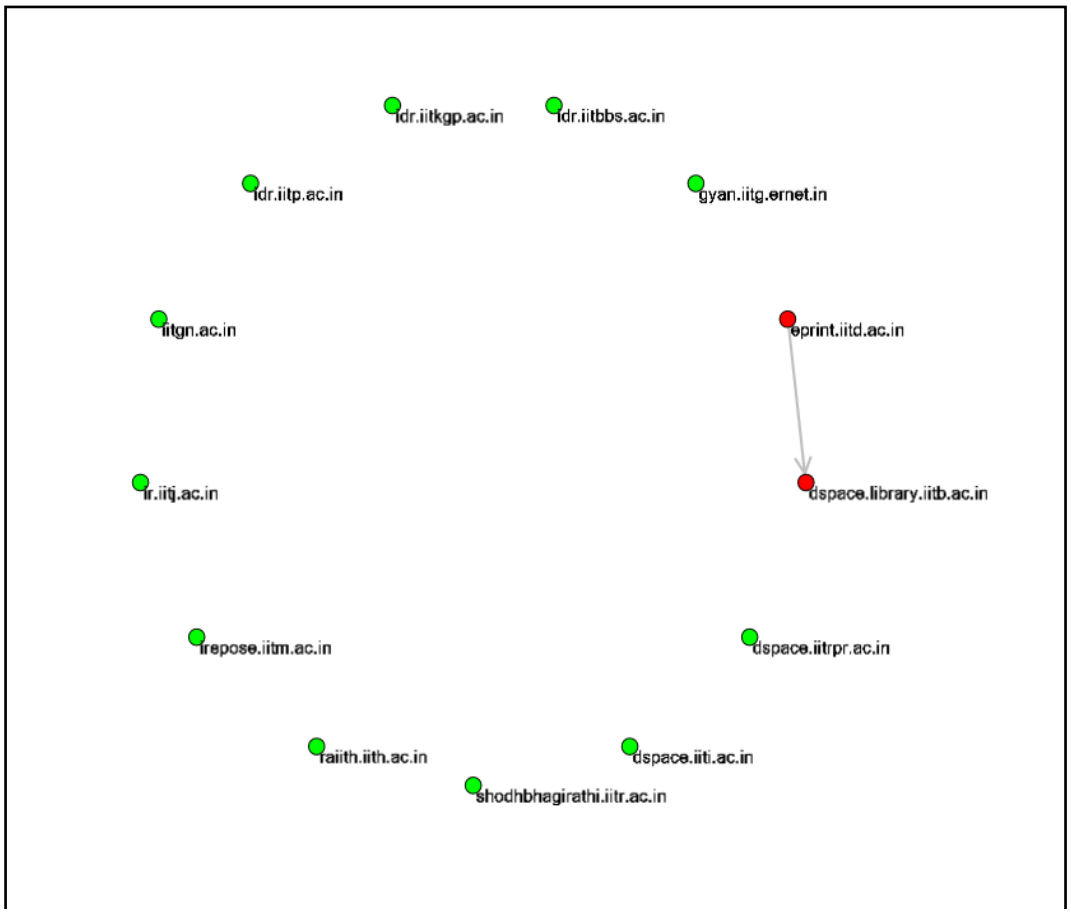

Figure 9: Network map of 13 repositories

The diagram shows the networking map of all the 13 repositories and reveals only IIT Delhi repository has an external-link to IIT Bombay repository.

Findings

- DSpace open source software is one of the most popular software as 92% repositories use DSpace software for maintaining IR.
- The Top Level Domains of IITs Institutional repositories websites are .ac.in and .ernet.in. The TLD .ac.in is used by 91% repositories.
- 84.62% IIT Repositories use PDF files followed by HTML 69.23% and DOC files 38.46% and MS-PowerPoint 7.69%.
- D.R. IITGN uses highest number of file formats i.e. 50% followed by C L IITR DSpace which uses 37.5% file formats.
- SB IITR has the highest OWIF (Over All Web Impact Factor) followed by CL IITR DSpace and LBCL D R.
- IITJ IR has secured the highest rank with 115.88 % SWIF (Simple Link Web Impact Factor).
- DR IITGN has the highest 123.5% SLWIF (Self-link Web Impact Factor) followed by LBCL DR with 116% SLWIF and SB IITR with 111.5% SLWIF.
- SB IITR secured first rank with the 1121.2% ELWIF (External Link Web Impact Factor) followed by IDR (D L KGP) with 99% and CL IITR DSpace R with 97.3% ELWIF.
- SB IITR with 110.8% ILWIF (In-link Web Impact Factor) secured first rank while LBCL D R with 100.9% and IDR (D L KGP) with 98% ILWIF secured second and third rank respectively.

Conclusion

The web is playing a very significant role in the entire arena of industry and business, agriculture and particularly in education. For Libraries and repositories, WWW is a boon through which they offer a wide range of information services to their users. Therefore, Webometric has become an important segment in the field of Library and Information Science through which analysis of websites has been conducted by information professionals to find the best websites and to rank websites based on Web Impact Factor. Findings of the

Webometric study are helpful for librarians and webmasters to redesigning the library websites in a more interactive and informative way. This study is also an effort in this direction.

References

Ahmad, M., et al. (2018). Calculating Web Impact Factor for University Websites of Jammu and Kashmir: a Study. *International Journal of Science Technology and Management*, 7(5), 17-27. http://www.ijstm.com /images /short_ pdf/1527576299_187_IJSTM.pdf

Almind, T. & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: Methodological approaches to Webometrics. *Journal of Documentation*, 53(4), 404-426. https://doi.org/10.1108/EUM0000000007205

Ayoub,A. Et al.(2019) Global Visibility and Web Impact of Leading Universities of SAARC Nations. *Library Philosophy and Practice.1-15*. https://digitalcommons.unl.edu/libphilprac/2443

Beerappa, G. & Sheshadri, S. (2016).Web Link analysis of Selected Documentation Centres in India: A Webometric Study. *International Research: Journal of Library and Information Science,* 6(2), 367-380.. https://www.researchgate .net/publication /308067372_ Web_Link_

Bjornborn & Ingwersen (2004). Towards a basic framework for webometrics. *Journal of the American Society for Information Science & Technology,* 55(14), 1216-27. https://doi.org/10.1002/asi.20077

Björneborn, L. & Ingwersen, P. (2001). Perspectives of Webometrics. *Scientometrics,* 50(1), 65-82. ftp://obaluae.inf.puc-rio.br/pub/docs/Estudos_de_Metrias/Internet_Estudos/Aula%205%20Discuss %E3o-Perspective_of_Webometrics.pdf

Björneborn, L. (2004). Small world link structures across an academic web space: A library and information science approach. (Ph.D. dissertation, Royal School of Library and Information Science, Copenhagen). http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.115.2353&rep=rep1&type=pdf

Borgman,C.L.& Furner,J.(2002) Scholarly Communication and Bibliometrics *Annual Review of Information Science and Technology*, 36, 3-72. DOI:10.1002/aris.1440360102

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*, 35(5), 325-329. doi:10.14429/djlit.35.5.8788

Chukwunonso,F. et al.(2013)The impact of the internet and the world wide web on distance and collaborative learning [conference Paper]. ICCGI 2013: The Eighth International Multi-Conference on Computing in the Global Information Technology. France. https://www.researchgate .net /publication/ 258117974_The_Impact_of_the_Internet_and_the_World_Wide_Web_On_Distance_and_Collaborative _Learning

Cronin, B. & Mckim, G. (1996) Science and scholarship on the World Wide Web: A North American perspective. *Journal of Documentation*, 52(2), 163-171. https://doi.org/10.1108/eb026965

Google(2005). Google Technology: PageRank explained. http://www.google.com/technology/

Indian Institutes of Technology (IITs) (2020).https://www.encyclopedia.com/international/ encyclopedias-almanacs –transcripts-and-maps/indian-institutes-technology-iits

Ingwersen,P.(1998)The calculation of Web Impact Factors. *Journal of Documentation*, 54 (2), 236-243. https://doi.org/ 10.1108/ EUM0000000007167

Jalal, S.K. Biswas, S.C.& Mukhopadhyay, P.S.(2010). Web Presence of Selected Asian Countries: A Webometric Study. Collnet *Journal of Scientometrics and Information Management,* 4(2), 57-68. doi: 10.1080/ 09737766.2010.10700893

Jeyshankar,R. & Sujitha, I(2014). Link Analysis of websites of universities in Kerala: A webometric study. International Journal of Library and Information Studies,4.https://www.academia.edu/15422777/ LINK_ANALYSIS_ OF_ WEBSITES_OF_ UNIVERSITIES_ IN_KERALA_A_WEBOMETRIC_STUDY

Khan, A. & Idrees, H. (2015).Calculating Web impact factor for university websites of Pakistan. *The Electronic Library*, 33(5), 883-895. https://doi.org/10.1108/EL-01-2014-0022

Larson, R. (1996, October, 21-24). Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of cyberspace. In: Steve Hardin (Ed.), ASIS Proceedings of the 59th ASIS Annual Meeting. Baltimore, 33: 7178. http://sherlock.berkeley.edu/asis96/asis96.html

Moogan,P.(2014) The Beginner's Guide to Link Building. https://moz.com/beginners-guide-to-link-building

Mukhopadhyay, P. (2004).Measuring Web Impact Factors: a webometric Study based on the Analysis of Hyperlinks.http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.626.8019&rep=rep1&type=pdf

Noruzi, A. (2005). Web impact factor for Iranian universities. Webology, 2(1). http://www. Webology .org/2005/v2n1/a11.html

Noruzi, A. (2006). The Web Impact Factor: a critical review. The Electronic Library, 24(4), 490-500. https://doi.org/10.1108/02640470610689188

Nwagwu, W. E. & Omoverere, A. (2008). Nigerian University Websites: A Webometric Analysis. *Webology,* 5(4) http://www.webology.org/2008/v5n4/a62.html

Prabakaran, R. & Lihitkar, S. (2015, March, 12-14) Web Link Analysis and Web Impact Factors of Space Science Libraries' Websites. [Conference Paper].10th International CALIBER-2015..Himachal Pradesh. http://ir.inflibnet.ac.in:8080/ir/ViewerJS/#../bitstream/1944/1872/1/33.pdf

Qiu, J., Chen, J., & Wang, Z. (2004). An analysis of backlink counts and Web Impact Factors for Chinese University websites. *Scientometrics,* 60 (3), 463-473. https://doi.org/10.1023/B_SCIE_000003438

Rousseau, R. (1997). Sitations: An exploratory study. *Cybermetrics,*1(1). https://pdfs. Semanticscholar .org/3427/2ee2120ae224dc4caf457a77a86909636f15.pdf

Smith, A.G. (1999) A tale of two Web spaces: Comparing sites using Web Impact Factors. *Journal of Documentation*, 55(5), 577-592. https:// eric.ed. gov/?id=EJ608491

Sujithai, M. and Jeyshankar, R.(2013). Web Page Analysis of Indian Institue of Technologies" (IITs) Websites: A Webometric Study. *International Journal of Digital Library Services*, 3(1), 55-65 http://www.ijodls.in /uploads/3/6/0/3/3603729/55-65313.pdf

Tang, R., & Thelwall, M. (2004). Patterns of national and international web inlinks to US academic departments: An analysis of disciplinary variations. *Scientometrics,* 60(3), 475-485. https://doi.org/10.1023/B:SCIE.0000034388.70594.cc

Thelwall, M. (2000). Web Impact Factors and search engine coverage. *Journal of Documentation,* 56 (2), 185-189. https://doi.org/10.1108/00220410010803801

Thelwall, M. (2002a). A comparison of sources of links for academic Web Impact Factor calculations. *Journal of Documentation*, 58(1), 66-78. https:// doi. Org /10 .1108/00220410210425412

Thelwall, M. (2002b). An initial exploration of the link relationship between UK university websites. *ASLIB Proceedings*, 54(2), 118-126. https://doi.org /10.1108/00012530210435248

Thelwall, M.(2009). Introduction to Webometrics: Quantitative Web Research for the Social Sciences. *Synthesis Lectures on Information Concepts, Retrieval, and Services,* 116. Doi:10.2200/S00176ED1V01Y200903ICR004

Wang, Y. & Kitsuregawa, M. (2002). Link Based Clustering of Web Search Results.[Conference Paper]. In Wang,X. & Yu, G. & Lu, H.(Eds.) Advances in Web-age information management. DOI 10.1007/3-540-47714-4_21.