LIBRARY WEBSITES OF INDIAN INSTITUTES OF TECHNOLOGY (IITs): A WEBOMETRIC STUDY

A Dissertation submitted in partial fulfillment of the requirement for the Degree of Master of Philosophy in Library & Information Science

Submitted by

Vanlalfeli

MZU Registration No. 2289 of 2007-2008

M. Phil. Registration No.: MZU/ M.Phil. /132 of 21.05.2013

Supervisor

Dr. Akhandanand Shukla

Assistant Professor

DEPARTMENT OF LIBRARY & INFORMATION SCIENCE

MIZORAM UNIVERSITY

2013

CHAPTER - 1

INTRODUCTION

1.1 Introduction

Library and Information Science (LIS) is concerned with how different information resources and information structures (interrelations of information resources) are generated, organized, distributed and utilized by different users in different contexts (Björneborn, 2004). Libraries disseminate knowledge and information among users and are local gateways to national and global knowledge (National Knowledge Commission, 2007). The libraries have applied advanced technologies, mainly computer and information and communication technology (ICT). By the application of ICT, library services become faster and easy accessible (Shukla, 2009).

There has been a revolutionary symbiosis between computer and communication technologies over a decade. The invention of the World Wide Web (WWW), a part of the Internet, the mother of networks has practically webbed the information globally less than one roof. There has been a shift in navigational approaches from syntactical to semantic (i.e., from words to concepts). An ever increasing number of research institutes, universities and business organizations are currently providing information themselves such as their articles, publications, reports, catalogues and other resources on the Internet in general and the WWW in particular. This is now becoming the accepted method of disseminating and sharing information resources in hypermedia. Information Science research has also changed, with much research into how the new technologies are being used, particularly e-mail and the web. In addition to user studies there have been attempts to extract new kinds of information from the Web (Rao, 2010). Being a global document network initially developed for scholarly use and now inhabited by a diversity of users, the web constitutes an obvious research area for metric studies in the field of Information Science (Berners-Lee & Cailliau, 1990).

1.2 The World Wide Web

The World Wide Web (WWW) was created in 1990 by CERN physicist Tim Berners-Lee. The WWW, or the Web, is a network of interconnected sites and hyperlinks between them which contains enormous source of information. It is a collection of millions of files stored on thousands of computers (called Web Servers) all over the world. These files represent text documents, pictures, video, sounds, programs, interactive environments, and any other kind of information that has ever been recorded in computer files (Young, 1999).

The terms Internet and World Wide Web are used interchangeably without much distinction. The Internet is a worldwide, publicly accessible series of interconnected computer networks that transmit data using the standard Internet Protocol (IP). It consists of millions of smaller domestic, academic, business, and government networks, which together carry various information and services, such as electronic mail, online chat, file transfer and the interlinked web pages and other resources of the WWW. In contrast, the Web is one of the services that run on the Internet (Wikipedia).

The WWW is a collection of web pages connected to each other using hyperlinks (Jalal, 2010). Web page is combined with formatting instructions of Hypertext Markup Language (HTML) and Extensible Markup Language (XML) used to write on the WWW (Jeyshankar, Sujitha & Valarmathi, 2012) whose main distinction is to provide hypertext that will navigate to other web pages via links (Wikipedia).

1.2.1 Elements of the Web

The following are the hardware, software and protocols that construct the Web.

→ **Clients and Servers:** A Web Server is a computer connected to the Internet that runs a program and takes responsibility for storing, retrieving, and distributing some of the web's files. A web client or Web browser is a computer that requests files from the web. (The word client generally refers to a program that causes a computer to request files – a Web browser, for example). When a client computer wants access to one of the files on the Web, the network directs the request to the Web server that is responsible for that file. The server then retrieves the file from its storage media and sends it to the client computer that requested it (Young, 1999).

→ **URL and Transfer Protocol:** Viewing a web page on the World Wide Web normally begins either by typing the Uniform Resource Locator (URL) of the page into a web browser or by following a hyperlink to that page or resource. A uniform resource locator is also known as web address, is a specific character string that constitutes a reference to a resource. In most web browsers, the URL of a web page is displayed on top inside an address bar. An example of a typical URL would be "http://www.iitbhu.ac.in/library/". A URL is technically a type of Uniform Resource Identifier (URI), but in many technical documents and verbal discussions, URL is often used as a synonym for URI. (Comer, 1997; Goel, 2010; Wikipedia).

The first part of a URL specifies the transfer protocol, the method that a computer uses to access this file. Most web pages are accessed with the Hypertext Transfer Protocol (the language of web communication), which is why Web addresses typically begin with http (or its secure version, https or shttp). The Hypertext Transfer Protocol (HTTP) is an application protocol for distributed, collaborative, hypermedia information systems. HTTP resources are identified and located on the network by Uniform Resource Identifiers (URIs) or, more specifically, Uniform Resource Locators (URLs) by using the http or https URI schemes.

The next part of the address denotes the host name or domain name of the web server. The domain name or literal numeric IP address gives the destination location for the URL. The domain google.com, or its numeric IP address 72.14.207.99, is the address of Google's website. The domain name portion of a URL is not case sensitive since DNS ignores case:

http://en.example.org/ and HTTP://EN.EXAMPLE.ORG/ both open the same page (Comer, 1997; Young, 1999; Wikipedia).

→ **Hypertext, Hyperlinks and HTML**: Tim Berners-Lee, in 1990, developed the "Hyper Text Transfer Protocol" (HTTP), the "Hyper Text Markup Language" (HTML), the first web browser, the first HTTP Server and the first web page. On August 1991, the web was made open for public (Shukla, 2009).

Hypertext is text with hyperlinks. A software system for viewing and creating hypertext is a hypertext system, and to create a hyperlink is to hyperlink (or simply to link). A user following hyperlinks is said to navigate or browse the hypertext. The Web is a growing organism and one of the most important characteristics of the Web is that a web page has ability to link to other web pages through hyperlinks (Noruzi, 2005).

A hyperlink (or link) is a reference to data that the reader can directly follow, or that is followed automatically. A hyperlink points to a whole document or to a specific element within a document. By clicking on the hyperlinks, users can move from one web page to another located on different servers (Jalal, 2010). Hyperlinks are often used to implement reference mechanisms, such as tables of contents, footnotes, bibliographies, indexes (Wikipedia). Since 1996, hyperlinks have been studied

extensively by applying existing bibliometric techniques to the Web (Ingwersen, 1998). Understanding the hyperlink structure is fundamental to understanding the Web connectivity structure, because hyperlinks have been used in web indexing and information retrieval, as well as page ranking. If the Web were a car, hyperlinks would be the engine, because without them, we are not going anywhere (Noruzi, 2005). The research field of webometrics tries to create new knowledge from this hidden information embedded in the hyperlink structure on the Web and to understand what kinds of real world phenomena and relationships it may represent (Holmberg, 2009).

Hypertext Markup Language (HTML) is the main markup language for creating web pages and other information that can be displayed in a web browser. It is the universal language of the web. HTML is written in the form of HTML elements consisting of tags enclosed in angle brackets (like <html>), within the web page content. URIs and hyperlinks in Hypertext Markup Language (HTML) documents form webs of inter-linked hypertext documents. The purpose of a web browser is to read HTML documents and compose them into visible or audible web pages. The browser does not display the HTML tags, but uses the tags to interpret the content of the page. HTML elements form the building blocks of all websites. HTML allows images and objects to be embedded and can be used to create interactive forms (Nair, 2002; Noruzi, 2005; Young, 1999).

→ **Web Pages and Web Sites:** A Web page is an HTML document that is stored on a web server and that has a URL to that it can be accessed via the Web.

A Web site is a collection of web pages belonging to a particular person or organization. Typically, the URLs of these pages share a common prefix, which is the address of the *home page* of the site. The home page is the front door of the site, and is set up to help viewers find whatever is of interest to them on that site. The URL of the home page also serves as the URL of the Web site. The content of Web pages is a visible kind of information, but the Web also contains information of a more hidden kind, like the connections that hyperlinks create between different Web sites and what those connections and hyperlink networks represent (Young, 1999; Nair, 2002).

→ **Web Browser:** In the 1990s, using a browser to view web pages and to move from one web page to another through hyperlinks came to be known as ‘browsing,’

‘web surfing,’ or ‘navigating the web’ A web browser is a program that a computer runs to communicate with web servers on the Internet, which enables it to download and display the web pages. A web browser displays a web page on a monitor or mobile device. Web browsers coordinate web resources centered on the written web page, such as style sheets, scripts and images, to present the web page (Young, 1999; Wikipedia). A web browser can have a Graphical User Interface (GUI), like Internet Explorer, Mozilla Firefox, Chrome and Opera, or can be text-based, like Lynx or Links. The request of the user extracts the page content according to its HTML instructions onto the users' terminal where information can be made use of (Jalal, 2010).

1.3 Webometrics

Webometrics is defined as “the study of quantitative aspects of the construction and use of information resources; structures and technologies on the web, drawing on bibliometric and informetric approaches” (Bjorneborn & Ingwersen 2004). This definition covers the quantitative aspects of both the construction side and usage side of the web which embraces the core areas of webometrics study which are:

a) Web page content analysis
b) Web link structure analysis (e.g. Hyperlink, Self link and External link)
c) Web usage analysis (e.g. exploiting log files for users searching and browsing behavior); and
d) Web technology analysis (including search engine performance) (Walia & Kaur, 2008).

Web page content analysis is a kind of subject analysis based on the content of the website. Web link structure study means citation analysis that provides links to other web page/sites. Web usage analysis is part of a more general user and usage research, and Web technology analysis refers to information systems evaluation (described in detail in Chapter 3).

Webometric analysis generally concentrates on the analysis of websites mainly in the performance of the institutions web domains because they are stable and well-defined institutions on the Web for a long time. Institutional websites are increasingly used for a wide variety of purposes, such as uploading the prospectus, library catalogue, promote achievement of individuals, research groups, new publications, etc.

Therefore, there is a necessity and desire to know about library websites of Institutes of Technology. In this context, it is logical to investigate measures of the effectiveness of Indian Institutes of Technology library websites, both to study the communication activity that they represent and to build useful evaluation metrics (Kothainayak & Gopalakrishnan, 2011).

1.4 Websites

> *"A website is an important medium to convey the aim, educational courses run, scholarships, infrastructure, facilities and many more information to the potential students, general public, faculty, professionals, organizations etc."* (Shukla, 2009).

A website is hosted on a computer system known as a web server, also called an HTTP server accessible via a network such as the Internet or a private local area network through an Internet address known as a Uniform resource locator. It is a set of related web pages served from a single web domain. All publicly accessible websites collectively constitute the World Wide Web. Websites are written in, or dynamically converted to, HTML (Hyper Text Markup Language) and are accessed using a software interface classified as a user agent. Web pages can be viewed or otherwise accessed from a range of computer-based and Internet-enabled devices of various sizes, including desktop computers, laptops, PDAs and cell phones (Comer, 1997; Young, 1999; Nair, 2002).

Websites have many functions and can be used in various fashions; a website can be a personal website, a commercial website, a government website or a nonprofit organization website and are typically dedicated to a particular topic or purpose. Any website can contain a hyperlink to any other website, so the distinction between individual sites, as perceived by the user, can be blurred.

The websites collectively constitute the World Wide Web, in which the people around the world look for their information regardless of the time and place (Jeyshankar, Sujitha & Valarmathi, 2012). An institution, company or an individual tells us how to get to their Web site by giving us the address of their home page. From the home page, we can get to all the other pages on their site (Rosa, 2005).

For example, the Web site for Indian Institute of Technology, Delhi (IITD) has the home page address of http://www.iitd.ac.in (The home page address actually includes a specific file name like *index.html* but in some cases, when a standard default name is set up, users don't have to enter the file name.) The home page address leads to thousands of pages (but a Web site can also be just a few pages). Since *site* implies a geographic place, a Web site can be confused with a Web server. A server is a computer that holds the files for one or more sites. A very large Web site may be spread over a number of servers in different geographic locations. A synonym and less frequently used term for Web site is "Web presence." That term seems to better express the idea that a site is not tied to specific geographic location, but is "somewhere in cyberspace." However, "Web site" seems to be used much more frequently (Comer, 1997; Young, 1999; Kumar, Babu, Gopalakrishnan & Rao, 2010; Wikipedia).

Websites comes under two categories namely: Static Website and Dynamic Website.

(a) Static Website: A static website is one that has web pages stored on the server in the format that is sent to a client web browser. It is primarily coded in Hypertext Markup Language (HTML). This type of website usually displays the same information to all visitors. A static website will generally provide consistent, standard information for an extended period of time (Goel, 2010; Wikipedia).

(b) Dynamic Website: A dynamic website is one that changes or customizes itself frequently and automatically, based on certain criteria. Dynamic websites can have two types of dynamic activity: Code and Content. Dynamic code is invisible or behind the scenes and dynamic content is visible or fully displayed (Young, 1999; Comer, 1997; Wikipedia).

Since the evolution of WWW, academic and research institutes have web spaces to disseminate knowledge and information globally (Shukla, 2009) and almost all these institutes have websites for themselves as well as for their libraries (Jeyshankar, Sujitha & Valarmathi, 2012). Through the WWW, different library services are being provided. Library Websites are one source of showcasting the knowledge generated by the academic institutions (Jalal, 2010). These academic institutions library websites are mostly used and visited from different disciplines especially people who

are engaged in research activities. Therefore, the web is measured to check the effectiveness of various services provided over it (Shukla, 2009) which is an area of the webometric study.

The Indian Institutes of Technology (IITs) are a group of autonomous public engineering institutes of India. The IITs are governed by the Institutes of Technology Act, 1961 and Institutes of Technology (Amendment Act, 2012) which has declared them as "Institutions of national importance", and lays down their powers, duties, framework for governance etc. All the IITs have libraries for the use of their students. The electronic libraries allow students to access on-line journals and periodicals. The IITs have taken an initiative along with Ministry of Human Resource Development (MHRD) to provide free online videos of actual lectures of different disciplines under National Program on Technology Enhanced Learning (NPTEL). This initiative is undertaken to make quality education accessible to all students (Wikipedia). Therefore in order to disseminate quality information among the IITs, library websites play an important role. The present study has been conducted so that the problems related to these library websites and improvements to be made can be recognized easily through webometric studies.

1.5 Measurement of the Web

In order to measure the web various studies have been conducted. The most popular types of measurement of web is search engine performance, recall and precision of search engines, search engine's query analysis, indexing of search engine's databases, behavior of search engines, website usage analysis, website content analysis, URL analysis, depth of the web, qualitative analysis of web documents etc. Search engines can be used as an important tool for measuring the web in the field of webometrics. Besides this, search engines are normally used as information retrieval tool for variety of purposes (Shukla, 2009).

1.5.1 Search Engines

According to Alan Poulter, "A www search engine is defined as retrieval service, consisting of a database(s) describing mainly resources available on the www, search software and a user interface also available via www" (Kannan & Abilash, 2011). Search engines are mechanisms that aid user to search the entire Internet for relevant information. A Web search engine employs a search agent (also called a spider) that

goes out and looks for information on Web pages which is indexed and stored in a huge database (Young, 1999). They use the automatic process to update, modify and maintain the references to web sites and web pages. They provide keyword searching capability, based on the indexing of text contained within a document and deliver a list of WWW link (URLs) that contain the key word entered in the search statement (Kannan & Abilash, 2011).

A crawler based search engine is made of three major parts: the crawler, the indexer and the query engine (Brin & Page, 1998). A web crawler (also known as a Web spider or Web robot) as defined by Kobayashmi and Takeda, is a program or automated script which browses the World Wide Web in a methodological, automated manner. The process is called web crawling or spidering. Most of the search engines use web crawler as a means of providing up-to-date data. Web crawlers are mainly used to create a copy of all the visited pages for later processing by a search engine that will index the downloaded pages to provide fast searches (Hatua, 2011). The indexer is to extract words from the pages forwarded to it by crawlers, and to turn each document into a list of words. This list of words is then inverted into 'inverted index' or 'inverted file', and it facilitates the work of the query engines. The search engines may also decide to save other information about documents such as metadata describing their contents, URLs of page linking to this page. The query engine is responsible for receiving request and providing answers to the users. The query engines consult the inverted file and the database of additional information to match the query. The capability to handle various advanced techniques such as phrase search, time specific search etc. of various search engines allow it to display more results.

According to Oppenheim, et.al. (2000), search engines are divided into four categories: robot (tool that automatically collects and indexes documents, e.g. Google, Altavista), directories (that rely mainly on human to identify and categorise resources, e.g. Yahoo), meta-search engines (that does not crawl individual site rather collect results of various search engines and display them, eg. Metacrawler, Vivisimo) and software tools (software for searching the Web that is installed in the user's computer, e.g. Copernic). Geography- specified and subject-specified search tools (e.g. SOSIG, BUBUL, Infolibrarian etc.) are also included (Mukherjee, 2011).

1.5.2 Application of Search Engine in Webometric Research

In the present study, search engine have been used for measuring the web. The main areas of application of search engine in webometric research are described below:

➔ **URL Analysis:** URL analysis is the study of analyzing the structure and properties of URLs embedded in interlinks web pages (i.e. incoming links or outgoing links) between one website/domain to another website/domain or in any single website. Thus we can understand that URL analysis is the study of characteristic and behavior of web page address of websites. The URL analysis includes study of top level domains (TLD), country code TLDs (ccTLD), generic TLDs (gTLD) sites/domains, sub-sites / sub-domains, file types of web pages, structure of URLs etc. (Shukla, 2009). At present, most of the search engines (e.g. Altavista, Google, AlltheWeb etc.) provide some in-built command (link:page address) to extract inlink towards a particular Web page but not the whole Web site. Some of the search engines (Microsoft based MSN), however claim that they have a different string command to excavate total inlinked page of whole Website (linkdomain:web address). But the results obtained from these must be used with caution because no search engine is able to fetch all pages of the web and most of the search engines display duplicate results to attract user towards their coverage (Mukherjee, 2011).

➔ **Link Analysis:** The web consists of number of pages interconnected with hyperlinks. These hyperlinks are inserted not for navigational purposes, but to point pages with similar and relevant contents. These hyperlinks are usually known as "*outlink*" of the hypertext document. The total number of hypertext documents in any website / domain can be measured by using inbuilt text command (site: website address) in search engines. Using this command along with some advanced command one can extract the total page, file format etc of the whole site. They add structure to the collection and it is important to explore the ways these links are employed for characterization of the Web. There are various inbuilt text commands available by which one can measure the total number web pages, i.e., *selflinks*, *outlink*, *inlinks* etc. (Shukla, 2009; Mukherjee, 2011). This measurement helps the search engines to rank the pages.

- **Web Impact Factor (WIF):** WIF is the quantitative indicator for calculating webometric activities. Ingwersen (1998) developed the idea of Web

Impact Factor (WIF) based on Journal Impact Factor (JIF). Web Impact Factor is the measurement of average link frequencies and is based on hyperlinks and their citations. According to Shukla (2009),

> "*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*."

This means, the numerator is the number of link pages made to a website and the denominator is a measure of the size of that very website. This very idea was adopted from the journal impact factor (JIF), originally proposed by Eugene Garfield in 1972. The WIF is extemporaneous result of search engines' database at a specific time (Shukla, 2009). It provides quantitative tools for ranking, evaluating, categorizing, and comparing websites, top-level domains and sub-domains (Walia & Kaur, 2008).

There are three types of links i.e., inlinks, outlinks, and self-links. Links coming into a site from another site is *inlinks* (also known as *backlinks*), links outgoing from a site to another site is *outlinks*, and links coming from the same site is *self-links*. Outlinks are used to pointing external sources and are also known as Outbound links. The *self-links* are made for navigational purposes only. More the Backlinks (also known as Inbound links) to a website reveals the impact of the website in that field (Shukla & Tripathi, 2009).

The Web Impact Factor (WIF) is of three types:

a) **Overall WIF**: It is calculated by the total number of combining *inlinks* pages and *self-links* pages

b) **Inlinks (Revised) WIF**: It is calculated by the total number of *inlinks* pages coming from outside; and

c) **Self-link WIF**: It is calculated by the total number of *self-links* pages of the website. (Shukla, 2009).

Calculation of Web Impact Factor (WIF)

The web impact factor (WIF) can be calculated by using the following formula given by Ingwersen (1998) where:

A= total links to a website (all inlinks and self-link pages)

B= inlinks to the website (subset of A)

C= self-links within the same website

D= total number of web pages present in the website at a time.

A. Calculation for Overall WIF:

A = total links to a website (all inlinks and self-link pages)

D = total number of web pages present in the website at a time

Overall WIF= A/D

B. Calculation for Inlink (Revised) WIF:

B = inlinks to the website

D = total number of web pages present in the website at a time

Revised WIF=B/D

C. Calculation for Self-link WIF:

C = self-links within the same website

D = total number of web pages present in the website at a time

Self-link WIF= C/D

➔ **Web Citation Analysis:** Formerly, Web link analysis is used to be assumed as successor of citation analysis. But later, it is being observed that a web page can be linked (inlinks) from different websites or different web contents. Inlinks may come from journal articles, conference papers, research reports etc. (Mukherjee, 2011). The citation analysis is the measurement of scholarly communication of documents and web citation analysis is the measurement of scholarly communication of documents on the web parlance. The term web citation describes the number of web links that any online article or online journal has received during a given time. Search engines Google, Yahoo!, AltaVista, and AlltheWeb etc. are used in the web link analysis. However, Google Scholar is used in extraction of web citation from scholarly resources. As mentioned in Google Scholar's Support for Scholarly Publishers website "Google and Google Scholar have boosted the worldwide visibility and accessibility of content. Working with publishers of scholarly information to

index peer reviewed papers, theses, preprints, abstracts, and technical reports from all disciplines of research and make them searchable on Google and Google Scholar." (Mukherjee, 2007; Shukla, 2009).

➔ **Web Content Analysis:** Content analysis is considered as a scholarly methodology in the humanities by which texts are studied as to authorship, authenticity, and their meaning. Content analysis of web documents is known as web content analysis which is now prevalent in the area of web to check the authenticity and quality of information available over the Internet. Recall and precision efficiency of search engines is also one aspect of web content analysis (Shukla, 2009).

1.6 Significance of the Study

There are number of webometric studies which have been conducted in library and information science perspective in India and in the world. In the case of IITs library websites, few webometric studies have been conducted. So, the study is an attempt to fill up the gap. Therefore, the study is an attempt to investigate the effectiveness of library websites in comprehensive manner by analyzing the URL analysis, backlink counts, calculating Web Impact Factor (WIF), link pattern and search engine's performance. Thus, present study will help to show the current status of library websites by webometric analysis. URL analysis and backlink counts will help librarians and webmasters to improve upon their library website's URL structure and web significance among all IITs library websites. Link pattern among library websites will help to understand the significance of the information as well as usefulness of the web page. Search engine's performance will help to understand the reliability of search engine in terms of webometric studies in case of IITs library websites. The comparative ranking may help librarians and webmasters to improve upon the library web interface.

1.7 Scope of the Study

The present study is confined to library websites of Indian Institutes of Technology (IITs). IITs are the apex educational institutions for imparting education and research in engineering. Presently there are sixteen (16) IITs in the country which are governed by The Institutes of Technology Act, 1961 and The Institute of Technology (Amendment Act of 2012). In this act IITs are declared as "Institutions of national importance" (MHRD). The lists of sixteen (16) Indian Institutes of Technology (IITs)

and their year of establishment are given in the following Table 1.1 and the coded form along with each institute's website are given in Table 1.2.

Table 1.1: List of Indian Institutes of Technology (IITs) with Establishment Year

Sl. No.	Name of Institutes	Year of Establishment
1.	Indian Institute of Technology Kharagpur	1951
2	Indian Institute of Technology Bombay	1958
3	Indian Institute of Technology Madras	1959
4	Indian Institute of Technology Kanpur	1959
5	Indian Institute of Technology Delhi	1961 (1963)*
6	Indian Institute of Technology Guwahati	1994
7	Indian Institute of Technology Roorkee	1847 (2001)*
8	Indian Institute of Technology Ropar	2008
9	Indian Institute of Technology Bhubaneswar	2008
10	Indian Institute of Technology Hyderabad	2008
11	Indian Institute of Technology Gandhinagar	2008
12	Indian Institute of Technology Patna	2008
13	Indian Institute of Technology Jodhpur	2008
14	Indian Institute of Technology Mandi	2009
15	Indian Institute of Technology Indore	2009
16	Indian Institute of Technology (BHU) Varanasi	1916 (2011)*

Source: Sujithai, M & Jeyshankar, R., 2013

∗ Already established institutes were given the status of IIT in the year mentioned within the parenthesis.

Table 1.2: List of IITs (in alphabetical order) Websites

Sl. No.	Indian Institutes of Technology	(Code Name)	Website Address
1.	IIT (BHU) Varanasi	IITBHU	www.itbhu.ac.in
2.	IIT Bhubaneswar	IITBBS	www.iitbbs.ac.in
3.	IIT Bombay	IITB	www.iitb.ac.in
4.	IIT Delhi	IITD	www.iitd.ac.in
5.	IIT Gandhinagar	IITGN	www.iitgn.ac.in
6.	IIT Guwahati	IITG	www.iitg.ac.in
7.	IIT Hyderabad	IITH	www.iith.ac.in
8.	IIT Indore	IITI	www.iiti.ac.in
9.	IIT Kanpur	IITK	www.iitk.ac.in
10.	IIT Kharagpur	IITKGP	www.iitkgp.ac.in
11.	IIT Madras	IITM	www.iitm.ac.in
12.	IIT Mandi	IITMANDI	www.iitmandi.ac.in
13.	IIT Patna	IITP	www.iitp.ac.in
14.	IIT Jodhpur	IITJ	www.iitj.ac.in

15.	IIT Roorkee	IITR	www.iitr.ac.in
16.	IIT Ropar	IITRPR	www.iitrpr.ac.in

1.8 Statement of the Problem

Library websites disseminate knowledge and information among users and attract potential users by showing richness of collection/resources, variety of services and number of facilities etc. in their study, teaching and research. The library websites of Indian Institutes of Technology provide information for the students, faculty, research scholars etc. While some of them are found to function and update properly, others are not which can cause serious problems for the users of their websites. Therefore, the study may help users and librarians to understand the current position of IITs libraries of India.

Webometrics is a quantitative study of web-related phenomena. There are number of researches conducted based on webometric data in an academic arena. From the LIS perspective, there have been few researches so far, based on webometric analysis of websites of Library and Information Science departments (Thomas & Willet, 2000) and Library Associations of India (Walia & Kaur, 2008). There is still lack of research in webometric analysis of library websites of Indian Institutes of Technology (IITs). From the LIS perspective, it would thus be interesting to investigate webometric analysis of Indian Institutes of Technology library websites. Further students, faculties, institutions and public in general are interested in comparison of academic and research institutions. So, rankings of them will actually reflect the status and standing of such organization. Now all the IITs have their websites for libraries. So there is need to investigate as to how the websites of these IITs libraries are being used in order to disseminate information to their users. The present work is concerned with the evaluation of IITs library websites based on Web Impact Factor (WIF) analysis.

1.9 Objectives of the Study

The objectives of the present study are to:

- Study the URL analysis of IITs library websites.
- Calculate the web impact factor (WIF) of library websites of IITs and rank them as per the WIF.
- Find out the link pattern amongst the library websites of IITs.

- Evaluate the search engine's performance for webometric studies.

1.10 Methodology

The present study is designed to calculate the Web Impact Factor of the IITs library websites and evaluate the performance of search engine for webometric studies. Therefore, the survey and observation methods of research are being found appropriate and have been utilized for conducting the present study.

1.10.1 Selection of the Indian Institutes of Technology Library Websites

Indian Institutes of Technology (IITs) are the apex educational organizations for imparting education and research in engineering and declared as *Institutions of National Importance*. Libraries are one of the important component of any educational institution which plays a very vital role in study and research of individuals. Therefore, library websites of these Institutions of National Importance have been purposively selected as sample for webometric study. All of the IITs library have website through which they are disseminating knowledge and information to their users.

The web addresses of the IITs library are given in the following Table 1.3

Table 1.3: List of IITs (coded form) **and their library websites**

SN	IITs	Library Website / URL
1.	IITBHU	http://www.iitbhu.ac.in/library/
2.	IITBBS	http://www.iitbbs.ac.in/Library.php
3.	IITB	http://www.library.iitb.ac.in/
4.	IITD	http://library.iitd.ac.in/
5.	IITGN	http://www.iitgn.ac.in/library.htm
6.	IITG	http://www.iitg.ernet.in/rs/lib/public_html/index.html
7.	IITH	http://library.iith.ac.in/
8.	IITI	http://www.iiti.ac.in/Library/about_central_library.html
9.	IITK	http://library.iitk.ac.in/
10.	IITKGP	http://www.library.iitkgp.ernet.in/
11.	IITM	http://www.iitm.ac.in/library
12.	IITMANDI	http://www.iitmandi.ac.in/academics/lib/
13.	IITP	http://www.iitp.ac.in/index.php/services-and-amenities/central-library/about-iitp-library.html
14.	IITJ	http://www.iitj.ac.in/library/
15.	IITR	http://mgcl.iitr.ac.in/
16.	IITRPR	http://www.iitrpr.ac.in/library

1.10.2 Data Collection Method

The survey and observation methods of data collection are applied for collecting the library websites of IITs. The following steps are being followed:

Step 1: Collection of Library Websites / URLs

The URLs of libraries of IITs have been collected by visiting their library websites. The IITs which do not have separate single web page for library have been excluded from the study. The websites / URLs of libraries which are not easily extractable are also excluded from the study.

Step 2: Retrieving of Data

All the URLs/websites of library are tested through command line textual queries like link, linkdomain, site etc. inside the preferred search engine's database i.e. Google.

1.10.3 Data Collection Tools

In webometric studies, different types of data collection tools are being utilized. These tools are software programmes, search engines, online tools, personal web crawlers etc. In the present study, search engine Google is used for webometric data collection and analysis.

1.10.4 Time Duration of Data Collection

The data has been collected in three (3) rounds having gap of 15 days between each round. The months September and October have been chosen for the purpose randomly. The data has been collected without any fixed time frame and without any discrimination of morning, evening and night. The data collection hours and minutes (in Indian Standard Time) has been recorded while collecting from the search engine.

1.10.5 Data Collection & Analysis Techniques

By the use of preferred search engine, the following search expressions are used for data collection. The search expressions are:

- **Site/Domain Coverage:** This feature shows how many web pages search engine has indexed with a single domain name. For example: how many web pages does Google index from the site **www.iitbhu.ac.in/library**/

Search Engine	Query	Example
Google	site:	site:www.iitbhu.ac.in/library/

- **Sub-site/Sub-domain Coverage**: This feature shows how many web pages search engine has indexed for multiple sub-domains within a single domain name.

Search Engine	Query	Example
Google	site:	site:iitbhu.ac.in/library/

- **Links to single URL/web page:** This feature shows how many web pages are linked with the URL.

Search Engine	Query	Example
Google	link:	link:http://www.iitbhu.ac.in/library/

- **Links to a Domain:** This feature shows how many links are pointing to a website.

Search Engine	Query	Example
Google	linkdomain:	linkdomain:www.iitbhu.ac.in/library/

For analyzing the Web Impact Factor, formula given by Ingwersen (1998) is used which is shown in 1.5.2 of this chapter. The data obtained is being tabulated and ranked according to their effectiveness by the use of statistical package.

1.11 Chapterization

The study is presented in following chapters:

Chapter 1 – deals with the introduction regarding the WWW, webometrics, websites and the research design which covers the scope, significance, statement of the problem and research methodology adopted for the study.

Chapter 2 – deals with the review of related studies relevant in the field of webometrics.

Chapter 3 – deals with the concepts of webometrics in which the relationship between Bibliometrics, Scientometrics, Informetrics, Webometrics and Cybermetrics is illustrated through diagrams as well as the techniques used for webometric studies.

Chapter 4 – presents the analysis of data and findings of the study through tables and diagrams with suitable interpretation.

Chapter 5 – deals with the conclusion of the whole study and suggestions for library websites to improve them for future.

References:

Note: References are based on Publication Manual of American Psychological Association (6th ed.) with some modifications.

Almind, T. C., & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: Methodological approaches to webometrics. *Journal of Documentation, 53*(4), 404-426.

Asadi, M., & Shekofteh, M. (2009). The relationship between the research activity of Iranian medical universities and their web impact factor. *Electronic Library, 27*(6), 1026-1043.

Babu, B. R., & Jeyshankar, R. (2009). Websites of universities in Tamil Nadu: A webometric study. *Annals of Library and Information Studies, 56*(1), 69-79.

Babu, B. R., Jeyshankar, R., & Rao, P. N. (2010). Websites of Central Universities in India: A webometric analysis. *DESIDOC Journal of Library & Information Technology, 30*(4), 33-43.

Berners-Lee, T. & Cailliau, R. (1990). *World Wide Web: proposal for a hypertext project*. Available at http://www.w3.org/Proposal.html

Björneborn, L., & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics, 50*(1), 65-82.

Björneborn, L., & Ingwersen, P. (2004). Towards a basic framework for webometrics. *Journal of the American Society for Information Science and Technology, 55*(4), 1216-1227.

Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, *30*, 107-117. Retrieved 12 November, 2013, from CiteSeerX: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.42.3243

Comer, E.D. (1997). *Computer networks and internet*. New Jersey: Prentice-Hall Inc.

Goel, M. K. (2010). *World Wide Web.* New Delhi: Rajat Publications.

Ingwersen, P. (1998). The calculation of web impact factors. *Journal of Documentation, 54*(2), 236-243.

Indian Institutes of Technology. (n.d.). Retrieved April 4, 2013, from Wikipedia: http://en.wikipedia.org/wiki/Indian_Institutes_of_Technology

Islam, M. A., & Alam, M. S. (2011). Webometric study of private universities in Bangladesh. *Malaysian Journal of Library & Information Science, 16*(2), 115-126.

Islam, M. A. (2011). Webometrics study of universities in Bangladesh. *Annals of Library and Information Studies, 58*(2), 307-318.

Jalal, S. K., Biswas, S.C., & Mukhopadhyay, P. (2010). Web impact factor and link analysis of selected Indian universities. *Annals of Library and Information Studies, 57*(1), 109-121.

Jeyshankar, R., Sujitha, I. M., & Valarmathi, A. (2012). Web pages of ICMR institutes websites: A webometric analysis. *Global Advanced Research Journal of Library, Information and Archival Studies, 1*(1), 6-18.

Kannan, K., & Abilash, S. (2011). Information searching pattern of web search engine: a comparative study. *Library Progress (International), 31*(1), 35-42.

Kessler, M.M. (1963). Bibliographic coupling between scientific papers. *American Documentation, 14*(1), 10-25.

Kobayashi, M., & Takeda, K. (2000). Information retrieval on the web. *ACM Computing Surveys, 32*(2), 144-173.

Kumar, A. M. N., Babu, B. R., Gopalakrishnan, S., & Rao, P. N. (2010). Location of web objects and links in the web sites of universities in Tamil Nadu: a study. *SRELS Journal of Information Management, 47*(3), 325-338.

Hatua, S.R. (2011). Auto web browser: an alternative to Google. *IASLIC Bulletin, 56*(1), 3-10.

History of the World Wide Web. (n.d.). Retrieved on 12 March, 2013, from Wikipedia: http://en.wikipedia.org/wiki/History_of_the_World_Wide_Web

Larson, R.R. (1996). Bibliometrics of the World Wide Web: an exploratory analysis of the intellectual structure of cyberspace. In: Hardin, S. (Ed.), *Proceedings of the 59th Annual Meeting, ASIS 96*. Baltimore, pp. 71-79.

Lawrence, S. & Giles, C. L. (1998). Searching the World Wide Web. *Science, 280*, 98-100.

Ministry of Human Resource Development (MHRD), India. (n.d.). *Department of Higher Education, IIT*. Retrieved April 9, 2013, from http://mhrd.gov.in/iit_hindi

Mukherjee, B. (2011). Bibliometrics to webometrics: The changing context of quantitative research. *IASLIC Bulletin, 56*(2), 97-110.

Nair, R.R. (2002). *Internet for information services*. New Delhi: Ess Ess Publications.

National Knowledge Commission. (2007). *Libraries: Gateways to Knowledge*. New Delhi: National Knowledge Commission, Govt. of India.

Smith, A.G and Thelwall, Mike (2002). Web Impact Factors for Australian Universities. *Scientometrics, 54*(3), 363-380.

Noruzi, A. (2005). Web Impact Factors for Iranian Universities. *Webology, 2*(1).

Noruzi, A. (2005). Editorial: Fundamental differences between hyperlinks and citations. *Webology, 2*(2), editorial 4. Available at: http://www.webology.org /2005/v2n2/editorial4.html

Noruzi, A. (2006). The web impact factors: A critical review. *Electronic Library, 24*(4), 490-500.

Oppenhium, C., Morris, A., McKnight, C., & Lowley, S. (2000). The evaluation of search engines. *Journal of Documentation, 56*(2), 190-211.

Rosa, M. (2005). Website. *Whatis.com.* Available at: http://searchsoa.techtarget. com/definition/Web-site

Ratha, B., Joshi, L., & Naidu, G.H.S. (2012). Webometric study of IIT libraries websites. *DESIDOC Journal of Library & Information Technology, 32*(3), 249-254.

Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265-269.

Smith, A. G. (1999). A tale of two web spaces: comparing sites using web impact factors. *Journal of Documentation, 55*(5), 577-592.

Smith, A. G. & Thelwall, M. (2002). Web impact factors for Australasian universities. *Scientometrics, 54*(3), 363-380.

Spink, A., & Jansen, B. J. (2004). *Web search: public searching of the web*. Netherlands: Kluwer Academic Publishers.

Shukla, A., & Tripathi, A. (2009). Webometric analysis of institutes of national importance in India. *IASLIC Bulletin, 54*(3), 165-180.

Shukla, A. (2009). *Evaluating Indian Academic and Research Library Websites based on Web Impact Factor Analysis* (*Doctoral dissertation).* Varanasi, India: Department of Library & Information Science, Banaras Hindu University.

Sujithai, M., & Jeyshankar, R. (2013). Web page analysis of Indian Institute of Technologies' (IITs) websites: A webometric study. *International Journal of Digital Library Services, 3*(1), 55-65.

Tague-Sutcliffe, J. (1992). An introduction to informetrics. *Information Processing & Management, 28*(1), 1-3.

Thanuskodi, S. (2011). Webometric analysis of private engineering college websites in Tamil Nadu. *J Communication, 2*(2), 73-81.

Thanuskodi, S. (2012). A webometric analysis of selected institutes of national importance websites in India. *International Journal of Library Science, 1*(1), 13-18.

Thelwall, M. (2000). Web Impact Factors and Search Engine Coverage. *Journal of Documentation, 56*(2), 185-189.

Thelwall, M. (2011). A comparison of link and URL citation counting. *Aslib Proceedings, 63*(4), 419-425.

Thomas, O., & Willet, P. (2000). Webometric analysis of departments of librarianship and information science. *Journal of Information Science, 26*(6), 421-428. doi: 10.1177/01 6555150002600605

Turnbull, D. (1996). *Bibliometrics and the World Wide Web* (Technical Report FIS-12-19-1996-1). Faculty of Information Studies, University of Toronto. Available at http://www.ischool.utexas.edu/~donturn/research/bibweb.html

Vijayakumar, M. (2012). Webometric analysis of university websites in Srilanka. *International Journal of Information Dissemination and Technology, 2*(3), 155-159.

Walia, P. K., & Kaur, P. (2008). Webometric analysis of Library Associations' websites of India. *IASLIC Bulletin, 53*(3), 131-143.

Web browser. (n.d.). Retrieved 29 October, 2013, from Wikipedia: http://en.wikipedia.org/wiki/web_browser

Web page. (n.d.). Retrieved 29 October, 2013, from Wikipedia: http://en.wikipedia.org/wiki/web_page

Web search engine. (n.d.). Retrieved 29 October, 2013, from Wikipedia: http://en.wikipedia.org/wiki/Search_engines

World Wide Web. (n.d.). Retrieved 17 October, 2013, from Wikipedia: http://en.wikipedia.org/wiki/World_Wide_Web

Young, M. L. (1999). *Internet: the complete reference, Millennium Edition*. New Delhi: Tata McGraw-Hill Education Private Limited.

CHAPTER-2

REVIEW OF LITERATURE

2.1 Introduction

Review of literature helps in identifying substantive, theoretical, methodological, conceptual issues and addressing them appropriately in the context of the present study. Hence, in this chapter, studies which are available have been reviewed which would facilitate the identification of the gaps, if any, in the earlier works and the present study may help to fill them.

2.2. Webometrics and its development

Turnbull, D. (1996). *Bibliometrics and the World Wide Web.* (Technical Report FIS-12-19-1996-1). Faculty of Information Studies, University of Toronto.

The study explained bibliometrics as a standard method of information analysis and discussed its application to measuring information on the World Wide Web. It highlights that there is a need to apply theory from other disciplines (such as Information Studies) to develop new methods, modeling techniques and metaphors to examine the emerging complex network as information on the Web increases toward different measurements. The study shows the relevance of using the operational methods of bibliometrics to produce quantitative data and analysis to improve information systems use.

Björneborn, L., & Ingwersen, P. (2001). Perspectives of Webometrics. *Scientometrics*, *50*(1), 65-82.

The study describes areas of webometric research that demonstrate interesting progress and space for development. It has been discussed that there are major areas which have not been employed for conducting webometric studies. The problems with measuring Web Impact Factors are discussed. The study concludes by giving an outline of new directions of webometrics for performing knowledge discovery and issue tracking on the Web, partly based on bibliometric methodologies used in bibliographic and citation databases.

Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for Webometrics. *Journal of the American Society for Information Science and Technology*, *55*(14), 1216-1227.

The study defined Webometrics within the framework of Informetrics and Bibliometrics, as belonging to Library and Information Science, and as associated with Cybermetrics as a generic subfield. The researchers developed a consistent and

detailed link typology and terminology; and made a clear distinction among different Web node levels when using the proposed conceptual framework. A novel diagram notation to investigate link structures between the Web nodes in webometric analyses have been proposed by the researchers. The study stated that citation analysis and link analysis are not analogous to each other.

Thelwall, M., Vaughan, L., & Bjorneborn, L. (2006). Webometrics. *Annual Review of Information Science and Technology, 39*(1), 81-135.

The study concentrates on link analysis and also covers other aspects of webometrics, including Web log file analysis. The study is divided into four parts; basic concepts and methods; scholarly communication on the Web; general and commercial Web use; and topological modeling and mining of the Web. The first part of the study is concerned with basic concepts and methods of data collection and processing in Webometric studies. The second part, scholarly communication on the Web, is concerned with using link analysis to identify patterns in academic or scholarly Web spaces. The third part reviews link analysis studies that have used techniques similar to those applied to academic Web spaces. The fourth part covers mathematical approaches to modeling the growth of the Web or its internal link structure, mostly the product of computer science and statistical physics research. It concludes with detailed interpretations of small-world linking phenomena, an information science contribution to topological modeling of the Web.

Thelwall, M. (2007). Bibliometrics to Webometrics. *Journal of Information Science, 34*(4), 1-18.

The study reviews the distance that bibliometrics has travelled since 1958 by comparing early bibliometrics with current practice, and by giving an overview of a range of recent developments, such as patent analysis, national research evaluation exercises, new visualization techniques, new applications, new online citation indexes, and developments related to the creation of digital libraries. Webometrics, a modern, fast-growing offshoot of bibliometrics is reviewed in detail. Finally, future prospects are discussed with regard to both bibliometrics and webometrics.

Goswami, P., Sharma, U., & Shukla, A. K. (2008). The Webometrics. *6th International CALIBER-2008, University of Allahabad, Allahabad, February 28-29 & March 1, 2008*, 656-660.

This study focuses on the concept of webometrics and its development. The relationship between the five metrics i.e, Bibliometrics, Informetrics, Scientometrics, Cybermetrics and Webometrics has been discussed in detail. A comparative study between Bibliometrics and Webometrics has been highlighted also. It has been concluded that on the basis of origin, the bibliometrics is the base of all other metrics. But every type of metrics is having some unique features, which differentiate between all of them.

Thelwall, M. (2010). Webometrics: emergent or doomed? *Information Research*: *An International Electronic Journal*. *15*(4), colis713.

The study reveals how most published Webometric research is relatively theoretical and, as a new research field, seems unlikely to survive unless it is useful in some way. The study uses citation analysis and a survey of webometricians to assess the extent to which webometrics has found applications outside of its parent discipline. The results suggest that there has been a turn towards applied webometrics with several externally-financed studies being contracted. Moreover, there is a significant amount of citation of Webometric research by disciplines outside information science, including computing, communication science and health; and concludes that there is still progress to be made.

Mukherjee, B. (2011). Bibliometrics to Webometrics: the changing context of quantitative research. *IASLIC Bulletin, 56*(2), 97-110.

The study focuses on the concept of bibliometrics and other related terms. The study divided the era of quantitative research in two paradigm i.e, pre World Wide Web and post World Wide Web. Under pre World Wide Web era the concept and development of terms Librametry, Bibliometrics, Scientometrics and Informetrics and in post World Wide Web era the concepts of Webometrics and Cybermetrics have been discussed. The relationship of the terms *Webometrics* and *Cybermetrics* has been diagrammatically presented. The various domains of Webometric research and tools used for webometric researches have also been

discussed. It has been concluded that webometrics have established an important independent domain in quantitative research.

2.3. Webometric Analysis

Thomas, O., & Willet, P. (2000). Webometric analysis of departments of librarianship and information science. *Journal of Information Science, 26*(6), 421-428.

This paper describes a webometric analysis of the linkages (or 'citations') to websites associated with departments of librarianship and information science (LIS). Some of the observed citation counts appear counter-intuitive, and there is only a very limited correlation with peer evaluations of research performance, with many of the citations being from pages that are far removed in subject matter from LIS. It has been concluded that citation data are not well suited to the quantitative evaluation of the research status of LIS departments, and that departments can best boost their Web visibility by hosting as wide a range of types of material as possible.

Chu, H., He, S., & Thelwall, M. (2002). Library and Information Science schools in Canada and USA: a webometric perspective. *Journal of Education for Library and Information Science, 43*(2), 110-125.

The in-links to 53 American Library Association (ALA) accredited library and information science (LIS) schools' websites have been analyzed. The sites that generated the majority of in-links for the LIS schools are in the .org, .edu, or .net domain. Such citations are also content-based as opposed to the directory-type references from sites in the .com domain. Links to the outside world from the 53 schools, to a certain extent, reveal their connectivity with other sites on the Web. Based on co-link data, the structure underlying the 53 LIS schools have been described. It has been suggested that webometric research must be conducted with caution because both the data source (i.e., web-based data) and data collection instrument (e.g., web search engines) have obvious deficiencies.

Nwagwu, W. E., & Agarin, O. (2008). Nigerian university websites: a webometric analysis. *Webology, 5*(4), Article 62.

Using the AltaVista search engine, data on web links was collected randomly from 1000 selected web pages of 30 Nigerian universities to study the

pattern and frequency of *outlinks* and *inlinks*. The websites have a total of 44,567 links, representing an average of 45 links per page. Of these, 81.2 % were *inlinks* from Nigerian university websites to other websites, while 18.8% were *outlinks* from other websites to Nigerian university websites. The first generation universities have a higher percentage of links by target web pages (52.6%) than the others, which returned below 50% of the web pages as target pages. The web pages of Nigerian universities did not target other Nigerian university web pages. Also, the websites seem to link more with non-academic websites than with academic websites. The result shows that that there exists a general low-level of utilization of the Web for sharing and disseminating of information produced by Nigerian universities.

Walia, P. K., & Kaur, P. (2008). Webometric analysis of Library Associations' websites of India. *IASLIC Bulletin*, *53*(3), 131-143.

This paper examines the websites of selected Library Association Websites of India and is confined to websites available on the public domain, on the WWW. The study includes analysis of hyperlinks and calculation of Web Impact Factor (WIF) for these websites. Google, Yahoo, AltaVista, AllTheWeb, etc have been used for data collection. The finding shows that DLA and SIS-India is found to have the maximum Web Impact Factor while KLA showed the least impact on the web.

Shukla, A., & Tripathi, A. (2009). Webometric analysis of institutes of national importance in India. *IASLIC Bulletin, 54*(3), 165-180.

The study reports the present scenario of *backlinks* structure of Institutes of National Importance of India by examining the extent of the backlinks given by different domains to these Institutes. Percentage of Deep Link Ratio, pattern of page pointing and link type relationship is also examined in the study. Data is collected using software programme called 'backlink analyzer'. Institutes of National Importance are also ranked on two different bases. From the study it has been clear that Indian Institutional websites attract more citations from commercial web domains than educational or any other Web domains.

Jeyshankar, R., & Babu, B. R. (2009). Websites of universities in Tamil Nadu: a webometric study. *Annals of Library and Information Studies*, *56*(1), 69-79.

The authors took a webometric study of 47 universities websites under which 27 were Government Universities and 18 were Private Universities in Tamil Nadu. The study calculates and analyses the number of web pages, link pages, self-link pages and external link pages of the universities websites. The Simple Web Impact Factor (WIF), self link WIF and external WIF are being calculated and ranked according to the WIF. Alta Vista search engine is used for collection of data. The results of the study shows that though the link pages are very small in number some universities have higher number of web pages and websites fall behind in their simple, self link and external link web impact factor.

Babu, B. R., Jeyshankar, R., & Rao, P. N. (2010). Websites of central universities in India: a webometric analysis. *DESIDOC Journal of Library & Information Technology*, *30*(4), 33- 43.

This study examines 40 central universities websites in India. Investigates domain systems of the websites, analyses the number of web pages and link pages and calculates the simple web impact factor, self link web impact factor, external link web impact factor and revised web impact factor for Central Universities in India and ranks the websites as per the WIF. It also develops a novel network diagram showing link structures between web nodes in webometric analysis. This study warns against taking the analogy between citation analysis and link analysis too far.

Thanuskodi, S. (2011). Webometric analysis of private engineering college websites in Tamil Nadu. *J Communication, 2*(2), 73-81.

The study calculated and compared the number of web pages, inlinks, and WIF of private engineering colleges in Tamil Nadu. It covers active exclusive websites, compared and then ranked these universities according to Webometric indicators. AltaVista is used for data collection. The findings showed that, in general, the private engineering colleges in Tamil Nadu did not have much impact on the web and were not known at the International level, evident by the Webometric indicators law.

Islam, A., & Alam, S. (2011). Webometric study of private universities in Bangladesh. *Malaysian Journal of Library & Information Science, 16*(2), 115-126.

The study examines 44 private university websites in Bangladesh. The number of web pages and link pages, the overall Web Impact Factor (WIF) and Absolute Web Impact Factor (WIF) are being identified and calculated. In a cross-sectional study, all the websites were analyzed and compared using AltaVista search engine. The study revealed that some private universities in Bangladesh have higher number of web pages but their link pages are very small in number, thus the websites fall behind in their Overall WIF, self link, external links and Absolute WIF. It is found that these universities did not have much impact factor on the web and were not known internationally.

Islam, A. (2011). Webometrics study of universities in Bangladesh. *Annals of Library and Information Studies*, *58*(2), 307-318.

This paper examines all the university websites in Bangladesh using Webometrics. The data for the study is obtained using AltaVista search engine for ranking the websites based on Webometric indicators. The study includes finding the simple WIF, self link WIF and external WIF of the universities in Bangladesh. It is found that some universities in Bangladesh have higher number of web pages but their link pages are fewer and websites fall behind in their Web Impact Factor.

Vijayakumar, M. (2012). Webometric analysis of university websites in Sri Lanka. *International Journal of Information Dissemination and Technology*, *2*(3), 155-159.

This study investigates the impact of 19 Sri Lankan universities. Inlink and self-link WIFs for all websites were based on AltaVista. It is found that the universities of Sri Lanka are possessing varied domains for their home pages namely .ac.net and .lk but most of them (89.47%) prefer the sub level domain like .ac.

Pechnikov, A. A., & Nwohiri, A. M. (2012). Webometric analysis of Nigerian university websites. *Webology*, *9*(1), Article 96.

The investigation conducted reveals a weak connectivity in the set of official websites of Nigerian universities. However, the connectivity becomes stronger when all the university websites are taken into account. It increases significantly with the addition of the only found web communicator to the university websites under National Universities Commission which approves the establishment of higher educational institutions in Nigeria and all academic programmes run by

them. It has been suggested that all universities should switch to the use of .edu.ng as their top-level domains.

Thanuskodi, S. (2012). A webometric analysis of selected institutes of national importance websites in India. *International Journal of Library Science*, *1*(1), 13-18.

The study focuses on the webpage content analysis of 10 selected institutes of national importance libraries in India. It is found that general information about homepage features are more in Indian Institutes of Technology and least in Indian Statistical Institute and Indian Institute of Science. The websites are up to date only in few institutes remaining websites do not mentions time or date in the homepage.

Aguillo, I. F. (2012). Is Google Scholar useful for bibliometrics? a webometric analysis. *Scientometrics*, *91*, 343-351.

This study identifies the usefulness of Google Scholar database for bibliometric analysis, and especially research evaluation. Instead of names of authors or institutions, a webometric analysis of academic web domains is performed. The bibliographic records for 225 top level web domains (TLD), 19,240 university and 6,380 research centers institutional web domains have been collected from the Google Scholar database. About 63.8% of the records are hosted in generic domains like .com or .org, confirming that most of the Scholar data come from large commercial or non-profit sources. One-third of the other items (10.6% from the global) are hosted by the 10,442 universities, while 3,901 research centers amount for an additional 7.9% from the total. The individual analysis show that universities from China, Brazil, Spain, Taiwan and Indonesia are far better ranked than expected. In some cases, large international or national databases, or repositories are responsible for the high numbers found. The study found that Google Scholar lacks the quality control needed for its use as a bibliometric tool; the larger coverage it provides consists in some cases of items not comparable with those provided by other similar databases.

Jeyshankar, R., Sujitha, I. M., & Valarmathi, A. (2012). Web pages of ICMR institutes websites: a webometric analysis. *Global Advanced Research Journal of Library, Information and Archival Studies*, *1*(1), 6-18.

For this study, the ICMR institutes in India have been taken and their link structure has been analyzed. It also concentrates on the classification of websites by webpage size, WAVE Web AIM Accessibility Error, various search engine performances, and the difference between pages in various time intervals and number of rich files has been calculated. It also presents the Link – network diagram of ICMR institutes using Pajek software.

Sujithai, M., & Jeyshankar, R. (2013). Web page analysis of Indian Institute of Technologies' (IITs) websites: a webometric study. *International Journal of Digital Library Services, 3*(1), 55-65.

The study analyzed the web pages of Indian Institute of Technology websites retrieved by commercial search engine. The result of the study shows that among the four web pages (Link Web Page, Self Link Web Page, External Link Web Page and Inlink Web Page) of IIT websites, the external link pages stand-in important to increase the number of web pages and provides more than other link pages.

2.4. Link Analysis

Henzinger, M. (2000). Link analysis in web information retrieval. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 23*(3)**,** 1-7.

The analysis of the hyperlink structure of the web has led to significant improvements in web information retrieval. This survey describes two successful link analysis algorithms and the state-of-the art of the field. The study reveals that the main use of link analysis is currently in ranking query results. Other areas where link analysis has been shown to be useful are crawling, finding related pages, computing web page reputations and geographic scope, prediction link usage, finding mirrored host, categorizing web pages, and computing statistics of web pages and of search engines. However, it has been concluded that research of the hyperlink structure of the web is just at its beginning and a much deeper understanding needs to be gained.

Park, H. W., & Thelwall, M. (2006). Web-science communication in the age of globalization. *New Media Society, 8*(4), 629-650.

This article examines the connectivity structure of links between university websites in 25 Asian and European countries as a case study of an inter-regional and intra-regional web phenomenon. The five most linked-to universities in each nation-state were selected and network analysis techniques were used. The

results suggested that the UK has a high impact on the formation of link mediated academic networks in Asia and Europe. Universities' websites in Asia are more heavily connected to European universities than linked to each other. The overall findings were indicative of globalization rather than regionalism, but a better characterization might be globalization with regional imbalances and individual high performing countries.

Xing, Y., & Chu, H. (2006). *Hyperlinking to academic websites: salient features examined*. Unspecified. (Unpublished), 1-16. Retrieved from http://eprints.rclis.org

This paper explores features of inlinks (incoming links) as opposed to that of citations so that better understanding can be achieved with regard to the limitations and implications in using links for evaluative webometric research. A total of 446 randomly selected cases of hyperlinking to 15 medical schools' websites were analyzed and then classified into a revised version of a taxonomy created in a previous study for identifying linking motivations. The classification of the linking data was accomplished within the context of linking and linked sites as well as based on reasons for hyperlinking. This research shows that only 5% and 7% of all the inlinks analyzed were made for reasons relating respectively to teaching/learning and research whereas 88% of the hyperlinks the target sites received were created for motivations relevant to service and general nature. These findings demonstrate that inlinking is not the same as citing since inlinks exhibit features considerably different from that of citations in several aspects. Inlink counts alone cannot serve as quality indicators for scholarly and evaluation purposes. Other factors (e.g., authors and intellectual contents of linked entities) have to be considered in evaluative, link-based webometric research.

Mandl, Thomas. (2007). The impact of web site structure on link analysis. *Internet Research, 17*(2), 196-206.

This paper analyzes whether the number of links pointing to a web page is biased by the structure of web sites. By web-design mining methods, two collections of web pages are extracted and the in-links counts are determined by querying web search engines. The paper finds that the structure bias and pages on a higher hierarchical level are likely to receive more links than other pages. The paper

shows that the structure bias of in-links should be considered by link analysis measures used in search engines.

Onyancha, O. B. (2007). A webometric study of selected academic libraries in Eastern and Southern Africa using link analysis. *South African Journal of Libraries and Information Science*, *73*(1), 25-39.

This study sought to audit and maps the selected university libraries' websites of Eastern and Southern Africa in order to measure the libraries' web structures, content, and visibility/presence. The study's focus areas included the number of web pages, in-links, out-links, location of library links on universities' websites, and the most popular link(s) targeted by these universities. The most popular sites were examined in order to determine the institutions/organizations/sites with which these libraries are linked and/or partner. Data was extracted by crawling the web using SOCSCIBOT and was analyzed using the SOCSCIBOT toolkit, which consists of Matrix and Pajek tools. The result of the study shows that libraries in Eastern and Southern Africa are well aware of the benefits and opportunities of the Internet and the WWW. Despite the digital divide and technological barriers in the third world, librarians in the two regions have been advanced in the construction of library websites. The ranking of these universities shows that South African university libraries performed better than Botswana, Kenya, Tanzania, Uganda and Zimbabwe. It has been suggested that librarians should become more involved in the construction of library websites and that libraries need to regularly update their websites in order to keep up with the current proliferation of Internet-based resources increasingly becoming freely available.

Yang, B., & Qin, J. (2008). Data collection system for link analysis. *IEEEXPLORE: Third International Conference on Digital Information Management (ICDIM), 2008.* 247-252.

The study exploited several possible ways to meet the needs of link analysis in Webometrics and developed a prototype (Link Discoverer) that collects data from both real-time links and search engines. The prototype consists of two parts: a crawling part for collecting real-time link data from a given domain or site and a search engine part for harvesting link data from search engines by using specific search commands. An experiment has been conducted to

evaluate the performance of Link Discoverer on link analysis. The results show that the Link Discoverer's functions can well satisfy the needs for link analysis. This study contributes to data collection methods and selection strategy in Webometrics and makes recommendations for improvement.

Yi, K., & Jin, T. (2008). Hyperlink analysis of the visibility of Canadian library and information science school web sites. *Online Information Review*, *32*(3), 325-347.

This paper probes into the external visibility of the web sites of seven ALA-accredited Canadian Library and Information Science schools. The number of inlinks to the schools' web sites is used as an indicator of the visibility of all or some portions of the LIS web sites. Inlinks pointing to the LIS school web sites were collected using the AlltheWeb search engine. Four content clusters were identified by which to group the content of all the inlinked LIS school web pages. These clusters were LIS, research, home page and resources. The most visible cluster was the LIS cluster and the least visible was the research cluster. In the study, the ranking of visible clusters, topics and web pages from the LIS web sites were also identified.

Jeyshankar, R. (2011). Link analysis and Web Impact Factor of Indian Nationalised Banks. *International Journal of Information Dissemination and Technology*, *1*(3), 171-179.

The study has been demonstrated that several versions of the metric of web can produce results that correlate with the various WIFs ratings of 27 nationalized banks websites in India showing that, despite being a measure of a purely Internet phenomenon, the results are susceptible to a wider interpretation. This study shows topology frame work / link network of Reserve Bank of India and linking the different nodes of nationalized banks websites in India.

Thelwall, M. (2011). A comparison of link and URL citation counting. *Aslib Proceedings*, *63*(4), 419-425.

This paper compares link counts with URL citation counts in order to assess whether the latter could be a replacement for the former if the major search engines withdraw their advanced hyperlink search facilities. The study covers 15 case studies and the results show a high degree of correlation between the two but with URL citations being much less numerous, at least outside academic world and

business. Significant differences between results indicate that the difference between link counts and URL citation counts will vary between webometric studies.

2.5 Web Impact Factor:

Ingwersen, P. (1998). The calculation of Web Impact Factors. *Journal of Documentation, 54*(2), 236-243.

This case study reports the investigations into the feasibility and reliability of calculating impact factors for web sites, called Web Impact Factors (WIF). The study analyses a selection of seven small and medium scale national and four large web domains as well as six institutional web sites taken from the web during a month. The data isolation and calculation methods are described. The results demonstrate that Web-IF's are calculable with high confidence for national and sector domains whilst institutional Web-IF's should be approached with caution. The 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. Their logical sum is assumed to constitute the workable frequency of web pages linking up to the web location in question. The logical operations are necessary to overcome the variations in retrieval outcome produced by the AltaVista search engine.

Smith, A. G. (1999). The impact of web sites: a comparison between Australasia and Latin America. *Paper presented at INFO'99, Congreso Internacional de Informacion, Havana, 4-8 October 1999.*

The study uses the concept of Web Impact Factors for comparing a sample of educational and research institutions in Australia, New Zealand, Central America, the Caribbean, and South America. Overall, the web sites for Australasian institutions have a higher external WIF than the web sites for Latin American institutions. While specific features of sites can affect the institution's Web Impact Factor, there is a small correlation between the proportion of English language pages at an institution's site and the institution's WIF. This indicates that for linguistic reasons, Latin American sites may not receive the attention that they deserve from the World Wide Internet. This raises the possibility that information may be ignored due to cultural, linguistic and geographic barriers and this should be taken into account in the development of the global Internet.

Thelwall, M. (2001). Results from a Web Impact Factor crawler. *Journal of Documentation*, *57*(2), 177-191.

In this paper, a specially made web crawler designed specifically for the calculation of reliable WIFs is presented. This crawler has been used to calculate WIFs for a number of UK universities, and the results of these calculations are discussed. The principal findings are that with certain restrictions, WIFs can be calculated reliably, but do not correlate with accepted research rankings due to the variety of material hosted on university servers. Changes to the calculations to improve the fit of the results to research rankings are proposed, but there are still inherent problems undermining the reliability of the calculation. It concludes that these problems still apply if the WIF scores are taken on their own as indicator of the general impact of any area of the Internet, but with care would not apply to online journals.

Thelwall, M. (2002). A comparison of sources of links for academic Web Impact Factor calculations. *Journal of Documentation*, *58*(1), 68-78.

This paper addresses which is the best domain for counting backlinks WIFs for British Universities calculated from different source domains are compared, primarily the .edu, .ac.uk and .uk domains, and the entire Web. The results shows that all four areas produce WIFs strongly with research ratings, but that none produced incontestably superior figures. It was also found that the WIF was less able to differentiate in more homogeneous subsets of universities, although positive results are still possible

Li, X. (2003). A review of the development and application of the Web Impact Factor. *Online Information Review*, *27*(6), 407-417.

This paper reviews how link based metric has been developed, enhanced and applied. Not only has the metric itself undergone improvement but also the relevant data collection techniques have been enhanced. WIFs have also been validated by significant correlations with traditional research measures. Bibliometric techniques have been further applied to the Web and patterns that might have otherwise been ignored have been found from hyperlinks. This paper concludes that there are still problems with data reliability and the interpretation of results that provide a significant challenge for future researchers.

Noruzi, A. (2005). Web Impact Factors for Iranian universities. *Webology*, *2*(1).

In this study, the author investigates the web impact factors (WIFs) among the Iranian universities. Alta Vista search engine is used for collection of data. The WIFs for Iranian universities were calculated by dividing link page counts by the number of pages found in AltaVista for each university at a given point in time. These WIFs were then compared, to study the impact, visibility, and influence of Iranian university websites. Overall, Iranian university web sites have a low inlink WIF. While specific features of sites may affect an institution's WIF, there is a significant correlation between the proportion of English-language pages at an institution's site and the institution's backlink counts. This indicates that for linguistic reasons, Iranian (Persian-language) websites may not attract the attention they deserve from the World Wide Web. This raises the possibility that the information may be ignored due to linguistic and geographic barriers, and this should be taken into account in the development of the global web.

Noruzi, A. (2006). The web impact factors: a critical review. *The Electronic Library*, *24*(4), 490-500.

This paper reviews how the WIF has been developed and applied. The WIF's advantages and disadvantages, data collection problems, and validity and reliability of WIF results have been discussed. It has been suggested that Web Impact Factors can be calculated as a way of comparing the attractiveness of web sites or domains on the Web. It is 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.

Boell, S. K., Wilson, C. S., & Cole, F. T. H. (2008). Usage of different web impact factors for ranking Australian Universities. *Collnet Journal of Scientometrics and Information Management*, *2*(2), 57-70.

This study describes how search engines can be employed for automated, efficient data gathering for Webometric studies using well defined query specific URLs in search engines (predictable URLs). It then compares the usage of staff-related Web Impact Factors to Web Impact Factors for a ranking of Australian universities, showing that rankings based on staff-related WIFs correlate much better with an established ranking from the Melbourne Institute than commonly used WIFs.

It also compares WIF data for Australian Universities provided by Smith (2002) for a longitudinal comparison of the WIF of Australian Universities over the last decade. It shows that size-dependent WIF values declined for most Australian universities over the last ten years, while staff dependent WIFs shows a moving trend**.**

Asadi, M., & Shekofteh, M. (2009). The relationship between the research activity of Iranian medical universities and their web impact factor. *Electronic Library, 27*(6), 1026-1043.

The study examines the relationship between researches in Iranian medical universities and their Web Impact Factor. The Altavista search engine was used for determining the number of pages, in-links and self-links of 42 web sites of Iranian medical universities. Tehran, Iran and Gilan medical universities had the first to third grade in the number of web sites' pages. Kerman, Kermanshah, Fasa and Qom had the highest grades of WIF in each group. Also, the WIF of university web sites was counted and shows that Hormozgan, Shiraz, Isfahan and Tehran had the highest grade in each group.

Jalal, S. K., Biswas, S. C., & Mukhopadhyay, P. (2010). Web Impact Factor and link analysis of selected Indian universities. *Annals of Library and Information Studies, 57*(1), 109-121.

The study explores the effectiveness of web impact factors (WIFs) for Indian universities' websites. The study also aims to find out the link patterns among the selected universities. The data have been collected using AltaVista and Google search engines. Personal web crawler i.e. SocScibot is used to generate link data in order to develop/form micro-link topology under study. The findings show that all the NITs are closely related in the topology framework whereas nodes are not linked significantly for the case of state universities and central universities.

Jati, H. (2011). Web Impact Factor: a webometric approach for Indonesian universities. *International Conference on Informatics for Development (ICID, 2008) C1*, 74-77.

This study explores the web impact factors for Indonesian universities. Counts of links to the web sites of Indonesian universities were calculated from the output of Yahoo search engine. The WIFs for Indonesian universities were calculated by dividing link page counts by the number of pages found in Yahoo for each

university at a given point in time. These WIFs from the approach is then compared, to study the impact, visibility, rich content and influence of Indonesian university web sites. The study concluded that Universitas Sriwijaya had the highest universities number of webpages ranked in with respect to number of web pages. Universitas Bina Nusantara had the highest impact on web with far more the number of inlinks as compared to the number of web pages on its website. Universitas Muhammadiyah Malaang was ranked on the top for presence of rich files from the contribution of .doc and .pdf file type.

Kumar, S.S., & Rabindra, M. (2012). Web Impact Factor and webometric study of National Institute of Technology (NITs) of India. *Pearl: A Journal of Library and Information Science*, *6*(1), 47-55.

The paper attempts to study websites of National Institute of Technology of India using web impact factors. Altavista search engine was used for collecting the required data. The simple web impact factor, self link web impact factor, external web impact factor and revised web impact factor of the NITs under the scope of study have been calculated. The study was made during August–September 2010. The study revealed that some NITs have higher number of web pages but their link pages are very small in numbers and websites fall behind in their simple, self link and external link web impact factor. This paper covers only 21 NITs out of 30 due to the absence of their official website. The websites of any institutions are regarded as the prime gateway of information. The paper gives a clear insight into the websites of Indian NITs and their impact factors. It can be useful for evaluation of the websites of other institutions.

Maharana, R. K., Panda, K.C., & Sahoo, J. (2012). Web Impact Factor (WIF) and link analysis of Indian Institute of Technologies (IITs): a webometric study. *Library Philosophy and Practice (e-journal)*, *789*, 1-11.

This paper examines and explores the web impact factor through a webometric study of the present 16 Indian Institutes of Technology (IIT) of India. Identifies the domain systems of the websites; analyzes the number of web pages and link pages, and calculates the simple web impact factor (WIF), self link web impact factor and external web impact factor of all the IIT. Also reflects that some IITs have higher number of web pages, but correspondingly their link pages are very small in

number and websites fall behind in their simple, self link and external link web impact factor.

Arif, A., & Ismail, N. A. (2013). Web Impact Factor for Malaysian public universities. *International Journal of Future Computer and Communication, 2*(3), 151-154.

This paper evaluates public universities in Malaysia based on the webometric perspective. Web Impact Factor evaluation has been carried out by using Majestic SEO and Google Search engine. It is found that Universiti Malaysia Sabah have the highest Average WIF followed by Universiti Malaysia Kelantan and in the third place is Universiti Pertahanan Nasional Malaysia.

Rassi, M., Jalali, D. A., Mousavizadeh, Z., Abbasgholi, N., & Haghighian, R. A. (2013). Assessment of visibility and Web Impact Factor (WIF) of Food Science & Technology Institutes websites based on webometrics scales. *Iranian Journal of Nutrition Sciences & Food Technology, 7*(5), 449-459.

The purpose of this study is link analysis of food science & technology institutes websites based on webometrics scales to propose a pattern website for National Nutrition & Food Technology Research Institute in Iran. Visibility, Web Impact Factor (WIF), self link, and total link of websites of the food science & technology institutes via webometrics have been studied in the research. Cluster analysis and Multi- dimensional scales studied the main clusters and drawing two-dimensional map of website based on co-links. Services provided by the websites have been investigated through their inlinks and visibility to identify the most effective elements of Food Research Institutes in Iran. The statistical population is 63 active websites searched and downloaded by Google search engine in the period of one month. The data were collected using webometrics formula via Yahoo search engine & Yahoo Explorer during December 2010 to January 2011. The results indicated that the FOI website has the highest impact factor among the sites research sample with WIF of 72, while the NIFA received the highest number of inlinks having the highest visibility. The Cluster analysis results of the website showed that websites had collaborated in five major clusters in the Web environment but seven website have remained independent. Multi- dimensional analysis map also confirmed the existence of five categories. This study showed that there has been a relation

between inlinks with types of presented services and the content of websites. To conclude an appropriate model has been presented for the website of National Nutrition & Food Technology Research Institute.

References:

Aguillo, I. F. (2012). Is Google Scholar useful for bibliometrics? A webometric analysis. *Scientometrics, 91*, 343-351.

Arif, A., & Ismail, N. A. (2013). Web impact factor for Malaysian public universities. *International Journal of Future Computer and Communication, 2*(3), 151-154.

Asadi, M., & Shekofteh, M. (2009). The relationship between the research activity of Iranian medical universities and their web impact factor. *Electronic Library, 27*(6), 1026-1043.

Babu, B. R., Jeyshankar, R., & Rao, P. N. (2010). Websites of central universities in India: a webometric analysis. *DESIDOC Journal of Library & Information Technology, 30*(4), 33- 43.

Boell, S. K., Wilson, C. S., & Cole, F. T. H. (2008). Usage of different web impact factors for ranking Australian Universities. *Collnet Journal of Scientometrics and Information Management, 2*(2), 57-70.

Björneborn, L., & Ingwersen, P. (2001). Perspectives of Webometrics. *Scientometrics, 50*(1), 65-82.

Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for Webometrics. *Journal of the American Society for Information Science and Technology, 55*(14), 1216–1227.

Chu, H., He, S., & Thelwall, M. (2002). Library and Information Science schools in Canada and USA: a webometric perspective. *Journal of Education for Library and Information Science, 43*(2), 110-125.

Goswami, P., Sharma, U., & Shukla, A. K. (2008). The Webometrics. *6th International CALIBER-2008, University of Allahabad, Allahabad, February 28-29 & March 1, 2008*, 656-660. Retrieved on 11 March, 2013, from http://ir.inflibnet.ac.in/bitstream/handle/1944/1318/70.pdf

Henzinger, M. (2000). Link analysis in web information Retrieval. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 23*(3)**,** 1-7. Retrieved on 11 March, 2013, from citeseerx.ist.psu.edu/viewdoc /download?doi=10.1.1.32.4054&rep=rep1&type=pdf#page=5

Ingwersen, P. (1998). The calculation of Web Impact Factors. *Journal of Documentation, 54*(2), 236-243.

Islam, A. (2011). Webometrics study of universities in Bangladesh. *Annals of Library and Information Studies, 58*(2), 307-318.

Islam, A., & Alam, S. (2011). Webometric study of private universities in Bangladesh. *Malaysian Journal of Library & Information Science, 16*(2), 115-126.

Jalal, S. K., Biswas, S. C., & Mukhopadhyay, P. (2010). Web impact factor and link analysis of selected Indian Universities. *Annals of Library and Information Studies, 57*(1), 109-121.

Jati, H. (2011). Web impact factor: a webometric approach for Indonesian Universities. *International Conference on Informatics for Development (ICID, 2011),C1,* 71-74. Retrieved 27 May, 2013, from *staff.uny.ac.id/.../C1_74-77*

Jeyshankar, R. (2011). Link analysis and Web Impact Factor of Indian nationalised banks. *International Journal of Information Dissemination and Technology, 1*(3), 171-179.

Jeyshankar, R., & Babu, B. R. (2009). Websites of universities in Tamil Nadu: A webometric study. *Annals of Library and Information Studies, 56*(1), 69-79.

Jeyshankar, R., Sujitha, I. M., & Valarmathi, A. (2012). Web pages of ICMR Institutes websites: a webometric analysis. *Global Advanced Research Journal of Library, Information and Archival Studies, 1*(1), 6-18.

Kumar, S.S., & Rabindra, M. (2012). Web impact factor and webometric study of National Institute of Technology (NITs) of India. *Pearl: A Journal of Library and Information Science, 6*(1), 47-55.

Li, X. (2003). A review of the development and application of the Web Impact Factor. *Online Information Review, 27*(6), 407-417.

Maharana, R. K., Panda, K.C., & Sahoo, J. (2012). Web Impact Factor (WIF) and link analysis of Indian Institute of Technologies (IITs): a webometric study. *Library Philosophy and Practice (e-journal), 789*, 1-11.

Mukherjee, B. (2011). Bibliometrics to Webometrics: the changing context of quantitative research. *IASLIC Bulletin, 56*(2), 97-110.

Mandl, Thomas. (2007). The impact of web site structure on link analysis. *Internet Research, 17*(2), 196-206.

Noruzi, A. (2006). The web impact factors: a critical review. *The Electronic Library, 24*(4), 490-500.

Noruzi, A. (2005). Web Impact Factors for Iranian universities. *Webology, 2*(1).

Nwagwu, W. E., & Agarin, O. (2008). Nigerian university websites: a webometric analysis. *Webology, 5*(4), Article 62.

Onyancha, O. B. (2007). A Webometric study of selected academic libraries in Eastern and Southern Africa using link analysis. *South African Journal of Libraries and Information Science, 73*(1), 25-39. Retrieved 11 March, 2013, from http://uir.unisa.ac.za/handle/10500/5234?show=full

Park, H. W., & Thelwall, M. (2006). Web-science communication in the age of globalization. *New Media Society, 8*(4), 629-650.

Pechnikov, A. A., & Nwohiri, A. M. (2012). Webometric analysis of Nigerian university websites. *Webology, 9*(1), Article 96.

Rassi, M., Jalali, D. A., Mousavizadeh, Z., Abbasgholi, N., & Haghighian, R. A. (2013). Assessment of visibility and web impact factor (WIF) of Food Science & Technology Institutes websites based on webometrics scales. *Iranian Journal of Nutrition Sciences & Food Technology*, *7*(5), 449-459.

Smith, A. G. (1999). The impact of web sites: a comparison between Australasia and Latin America. *Paper Presented at INFO'99, Congreso Internacional de Informacion,* Havana, 4-8 October 1999. Available at: http://www.vuw.ac.nz/staff/alastair_smith/publns/austlat/

Smith, Alastair G. (1999). A tale of two Web spaces:comparing sites using web impact factors. *Journal of Documentation. 55*(5), 577-592.

Smith, A.G and Thelwall, Mike (2002). Web Impact Factors for Australian Universities. *Scientometrics, 54*(3), 363-380.

Shukla, A., & Tripathi, A. (2009). Webometric analysis of institutes of national importance in India. *iaslic bulletin*, *54*(3), 165-180.

Sujithai, M., & Jeyshankar, R. (2013). Web page analysis of Indian Institute of Technologies' (IITs) websites: a webometric study. *International Journal of Digital Library Services, 3*(1), 55-65.

Thanuskodi, S. (2012). A webometric analysis of selected institutes of national importance websites in India. *International Journal of Library Science, 1*(1), 13-18.

Thanuskodi, S. (2011). Webometric analysis of private engineering college websites in Tamil Nadu. *J Communication, 2*(2), 73-81.

Thelwall, M. (2011). A comparison of link and URL citation counting. *Aslib Proceedings*, *63*(4), 419-425

Thelwall, M. (2010). Webometrics: emergent or doomed? *Information Research, 15*(4). colis713. Available at http://InformationR.net/ir/15-4/colis713.html

Thelwall, M., Vaughan, L., & Bjorneborn, L. (2006). Webometrics. *Annual Review of Information Science and Technology*, *39*(1), 81-135.

Thelwall, M. (2002). A comparison of sources of links for academic Web Impact Factor calculations. *Journal of Documentation*, *58*(1), 68-78.

Thelwall, M. (2001). Results from a Web Impact Factor crawler. *Journal of Documentation*, 5*7*(2), 177-191.

Thomas, O., & Willet, P. (2000). Webometric analysis of departments of librarianship and information science. *Journal of Information Science*, *26*(6), 421-428.

Turnbull, D. (1996). *Bibliometrics and the World Wide Web.* (Technical Report FIS-12-19-1996-1). Faculty of Information Studies, University of Toronto. Available at http://www.ischool.utexas.edu/~donturn/research/bibweb.html

Vijayakumar, M. (2012). Webometric analysis of university websites in Srilanka. *International Journal of Information Dissemination and Technology*, *2*(3), 155-159.

Walia, P. K., & Kaur, P. (2008). Webometric analysis of Library Associations' websites of India. *IASLIC Bulletin*, *53*(3), 131-143.

Thelwall, M. (2007). Bibliometrics to Webometrics. *Journal of Information Science*, *34*(4), 1-18.

Xing, Y., & Chu, H. (2006). Hyperlinking to academic websites: salient features examined. Unspecified. (Unpublished), 1-16. Retrieved 27 May, 2013, from http://eprints.rclis.org/10398/ 1/Hyperlink-Xing%26Chu.pdf

Yi, K., & Jin, T. (2008). Hyperlink analysis of the visibility of Canadian library and information science school web sites. *Online Information Review*, *32*(3), 325-347.

Yang, B., & Qin, J. (2008). Data Collection System for Link Analysis. *IEEEXPLORE: Third International Conference on Digital Information Management (ICDIM), 2008*, 247-252.

CHAPTER- 3

WEBOMETRIC CONCEPTS

3.1 Introduction

Library and Information managers have adopted a number of quantitative methods in recent years in order to evaluate library resources and services more objectively and effectively. Several new terms have appeared for quantitative studies in Library and Information Science which includes metric studies. These metric studies are known as Bibliometrics in 1960's, Scientometrics in 1970's and Informetrics in the midst of 1980's. And now with advent of information technology the concept of Webometrics and Cybermetrics emerged during the 1990's. This has now emerged as a research front in its own right in Library and Information Science (Sangam, 2008). The metric studies are used for measuring scholarly communication; identify research trends and growth of knowledge; identify users of different subjects; estimate comprehensiveness of secondary periodicals; forecast past, present and future publishing trends; identify authorship and its trends in documents on various subjects; predict productivity of publishers or individual authors etc. (Jeyshankar & Babu, 2009).

The research field of webometrics has its roots in library and information science and in particularly in Scientometrics and Bibliometrics. Because of this heritage, its research methods are very similar to bibliometric methods but applied on the World Wide Web to extract and analyze information from the hyperlink structure and content on the Web. The original idea of webometrics was that hyperlinks between Web sites may be a valuable source of information, in a similar way that citations are a valuable source of information about use and visibility of scientific articles and about the connections between different authors and different articles. Both links on the Web and citations in scientific articles are used in a similar way: to reference something, some other source of information and because of this similarity both may be useful for researching various relationships between the researched documents, authors or organizations (Holmberg, 2009; Frías, 2010).

3.2 Development of Webometrics

Bibliometric tools and techniques have been used to provide an understanding of the dynamics of disciplines, developing and research funding. Bibliometrics has been defined as research of the quantitative aspects of production, distribution and use of all saved information (Tague-Sutcliffe, 1992). Bibliometrics has also been described as the research field that studies scholarly communication, publishing, and the development of literature (Borgman, 2000). Bibliometrics literally means "book

measurement", but in fact bibliometrics has been used to measure different aspects in all kinds of (usually) written documents. Measured quantities have usually been frequencies of different occurrences, like words in a text or authors of a collection of scientific articles. Analogous to this definition, webometrics would mean "measurement of the Web", which is quite appropriate (Holmberg, 2009).

Since 1996, increasing efforts have been made to investigate the web as a significant scholarly medium for science and scholarship by applying bibliometric techniques. Terms applied to this new area of study include "webometrics" (Almind & Ingwersen, 1997). Almind and Ingwersen have defined the discipline and gave its name, although the basic issue was identified by Rodriguez Gairin in 1997 and was pursued in Spain by Aguillo in 1998. Larson is also a pioneer with his early exploratory link structure analysis and also Rousseau with the first pure informetric analysis of the Web (Walia & Kaur, 2008). Webometrics is a quantitative study of web-related phenomena and webometric studies can be applied to web with commercial search engines providing the raw data and seems to be widely accepted by the research community together with the term Cybermetrics. (Babu, 2010; Frías, 2010).

Björneborn & Ingwersen (2004) defined both terms by limiting the research areas of Webometrics and Cybermetrics.Webometrics is defined as "the study of quantitative aspects of the construction and use of information resources; structures and technologies on the web, drawing on bibliometric and informetric approaches" while Cybermetrics does the same but on the whole Internet. Hence, Cybermetrics is more focused on the study of non web based Internet phenomena, e.g. emails, chat, newsgroup studies, etc (Goswami, Sharma, & Shukla, 2008; Frías, 2010).

In the above definition given by Bjorneborn & Ingwersen, the term structure denotes the features of the website. It means that the information that may be provided by the website, make a well defined flow chart for that. The simple mean of structure is "plan for – process of making". In other words the term structure may be defined as how one page of website is attached with other page (via home page content, link structure and the like) (Frías, 2010).

The second term 'use of information resource' denotes the meaning that how the particular information resource is being used, or how to search information on the web. The only method to search information on web is via 'key word searching'

through search engines or via ‘specific URL’. We can have the ‘key words’ (search terms) from log files whenever we give any term to search, that all are saved in log files of that particular web server, which are used for searching. These terminologies (search terms) causes for the websites updation.

As raw material is essential for finished product, like that technology is essential to develop a website. The term technology includes different types of high level languages, protocols, browsers, web servers, special software and many more that are essential to develop a website. As much the technology will be used, that much the website will be having features. It means the quality web site is the result of good technology. So it can be said that through webometrics study, the websites (search engine or other) can be ranked. It is an important study; to measure the web site, because via quality/standard web site, the information can be searched and retrieved quickly (Comer, 1997; Goswami, Sharma, & Shukla, 2008).

The last phrase used in the definition is “drawing on bibliometric and informetric approach”. The concept of webometrics is based on bibliometrics, because like the bibliometrics study, one can measure the different quantitative aspect of the web in webometrics study. Secondly it is based on informetric. The informetric study is such type of study, which measures the quantitative aspect of any type of information and through webometrics study one can get the information about web (web site). That is why the above phrase is used.

Here the term “qualitative aspect” is included in the above definition because while studying /measuring / analyzing the SUiT part of website(s) one uses to define the features of website. The SUiT is the acronym of Structure, Use of information resources and Technology (Goswami, Sharma, & Shukla, 2008).

Webometrics, then, describes counting or measuring web resources in mathematical value. It defines the extent of web usage for research. Since the web allows documents to be linked together, the measurement of these links forms the fabric of webometrics and is defined as web “measured on the basis of web characteristics or presence on the Internet” (Samuel, 2008). Therefore webometrics is based on two indicators:

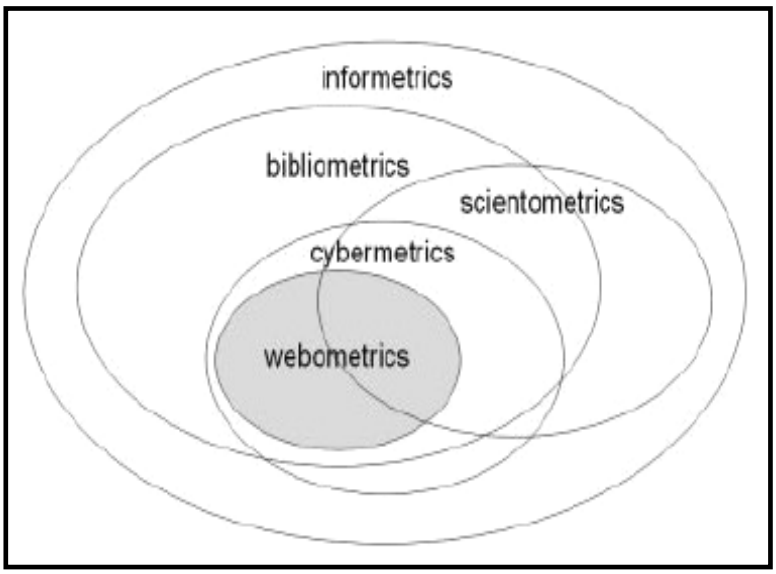
- Volume of published materials of institutions/individual on the web, and

- The visibility and impact of the web pages measured by the citation (site citations or links they receive)

Webometrics was started by the realization that the web is an enormous document repository with many of these documents being academic-related. Webometric research includes link analysis, web citation analysis, search engine evaluation and purely descriptive studies of the web (as discussed in 1.5.2 in Chapter 1). One of the most visible outputs of webometrics is the ranking of world universities based upon their web sites and online impact.

3.2.1 Relationship between Bibliometrics, Scientometrics, Informetrics, Webometrics and Cybermetrics

Figure 3.1**:** The relationship between the LIS fields of infor-/biblio-/sciento-/cyber-/webo-/metrics. (The sizes of the overlapping ellipses are made for sake of clarity only)



Source: Walia & Kaur (2008)

Before defining the relationship, it is essential to describe in all the terms briefly. The terms are described here as follows:

➔ **Bibliometrics:** The term 'bibliometrics' was coined by British Scientist Alan Pritchard in 1969. It can be defined as the application of mathematics and statistical methods to books and other media of communication (Pritchard, 1969). Bibliometric include studies of the growth of the literature in some subject, how much literature is contributed by various individuals, groups, or organisations or countries; how much

exists in various languages; how the literature on some subject becomes out of date (studies of obsolescence). Another important bibliometric study includes what sources author cite, citation studies and geographical distribution of documents. (Sangam, 2008; Rao, 2010). The bibliometric studies or research is conducted by applying three laws which are:

(a) Lotka's Law: Productivity of authors in terms of scientific papers.
(b) Bradford's Law: Scattering of articles over different journals.
(c) Zipf's Law: Frequency of occurrence of words in text.

➔ **Scientometrics:** The term Scientometrics originated as a Russian term for the application of quantitative methods to the history of science. The term was introduced and came into prominence with founding of the journal named 'Scientometrics' by T. Braun in 1977. Tague-Sutcliffe (1992), defines Scientometrics as "the study of the quantitative aspects of science as a discipline or economic activity". It is also used as a generic term for a system of knowledge, which endeavours to science and technology studies. Thus, Scientometrics is a part of the sociology of science and has application to science policy making. It involves studies in History of Science, growth of science and scientific institutions, behaviour of science and scientists and science policy and decision- making (Sangam, 2008).

➔ **Informetrics:** The term 'informetrics' was first proposed by Otto Nacke of West Germany in 1979. An FID Committee with broadly defined objectives in the provision of research and technical data subsequently gave this name. Tague-Sutcliffe (1992), defines informetrics as "the study of the quantitative aspects of information in any form, not just records or bibliographies, and in any social group not just scientists". Brookes (1991), characterised informetrics as a "generic term that embraces both bibliometrics and Scientometrics". The major study areas in informetrics include citation studies where impact factor, h-index and co-citation are also used. Three laws which are Lotka's Inverse square law, Zipf's law of word occurrence and Bradford's law of Scattering are used for research in the area of informetrics (Walia & Kaur, 2008; Sangam, 2008).

➔ **Webometrics:** As described in 3.1 and 3.2 in this Chapter, the research field of webometrics encompasses the nature and properties of the World Wide Web by

applying modern informetric methodologies. The term webometrics is a coinage from two modern English words, “web” and “metric”. The word web is a short of World Wide Web. The Dictionary of Science defined web as: “a hypermedia system… that allows users to view and retrieve information from “documents” containing links’. On other hand, metrics has to do with counting or measurement. Webster’s Comprehensive Dictionary of English Language defined metrics as “the mathematical theory of measurement.” Webometrics is the new discipline that intends to apply Bibliometrics, Scientometrics, Informetrics and Cybermetric techniques to the process of scientific communication, which takes place on Web in order to know and describe them from a quantitative point of view.

➔ **Cybermetrics:** Cybermetrics is one of the recently emerging fields in the line of metric studies. It is mainly concerned with the computer-science-based approach. Cybermetrics is proposed as a generic term for “the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the whole Internet drawing on bibliometric and informetric approaches” (Björneborn & Ingwersen 2004). Cybermetrics thus encompasses statistical studies of discussion groups, mailing lists, and other computer-mediated communication on the Internet including the www. Besides covering all computer-mediated communication using Internet applications, this definition of cybermetrics also covers quantitative measures of the Internet backbone technology, topology and traffic. The breadth of coverage of cybermetrics implies large overlaps with proliferating computer-science-based approaches in analyses of web contents, link structures, and web usage and web technologies (Thelwall, Vaughan, & Björneborn 2005; Sangam, 2010; Holmberg, 2009).

From Figure 3.1 the relationship between Informetrics, Bibliometrics, Scientometrics, Cybermetrics and Webometrics are observed and clearly shows how Webometrics associates with Bibliometrics and overlaps Scientometrics to an extent. There are different conceptions of Informetrics, Bibliometrics and Scientometrics. The circle of Informetrics covers all other metrics circles because it is a quantitative aspect of any type of information. The part, which overlaps the circle of bibliometrics, of Scientometrics, shows the politico-economical aspects of Scientometrics. The economic aspect of science shows the impact of scientific research over the society (Goswami, Sharma, & Shukla, 2008; Holmberg, 2009).

Björneborn & Ingwersen (2005) have proposed a differentiated terminology distinguishing between studies of the web and studies of all Internet applications. They use 'webometrics' for study of web and 'cybermetrics' for study of Internet applications. Some part of cybermetrics ellipse lying outside the bibliometrics. It is because some activities in cybermetrics normally are not recorded, but communicated synchronously as in chat rooms. The circle of Webometrics overlaps the circle of bibliometrics, but within the boundaries of cybermetrics. Webometrics circle cannot overlap the circle of cybermetrics because web is a part of cyberspace. But in Figure 3.1 the circle of webometrics ellipse lying outside the bibliometrics, because some aspect of webometrics (link structure, technologies and so on), does not include in bibliometrics or it is beyond the boundaries of bibliometrics. Webometrics is partially covered by Scientometrics, as many scholarly activities in science today are web-based. As ideas rooted in Bibliometrics, Scientometrics and Informetrics have contributed to the emergence of webometrics; ideas in webometrics might now contribute to the development of these embracing fields (Walia & Kaur, 2008; Goswami, Sharma & Shukla, 2008).

3.3 Webometrics and its Techniques

Björneborn & Ingwersen (2005) definition of webometrics covers the quantitative aspects of both the construction side and usage side of the web which embraces the four main areas of webometrics research: link structure analysis, web page content analysis, web usage analysis and web technology analysis (Björneborn, 2004).

3.3.1 Web Link Structure Analysis

Link analysis is the quantitative study of hyperlinks between webpages. The use of links in bibliometrics was triggered by Ingwersen (1998). It has been used successfully for deciding which web pages to add to the collection of documents (i.e., which pages to *crawl*), and how to order the documents matching a user query (i.e., how to *rank* pages). It has also been used to categorize web pages, to find pages that are related to given pages, to find duplicated web sites, and various other problems related to web information retrieval.

> i) This study provides ***hyperlinks*** between documents and records of user behaviour. To be precise, *hypertexts* (i.e., collections of documents connected by hyperlinks).

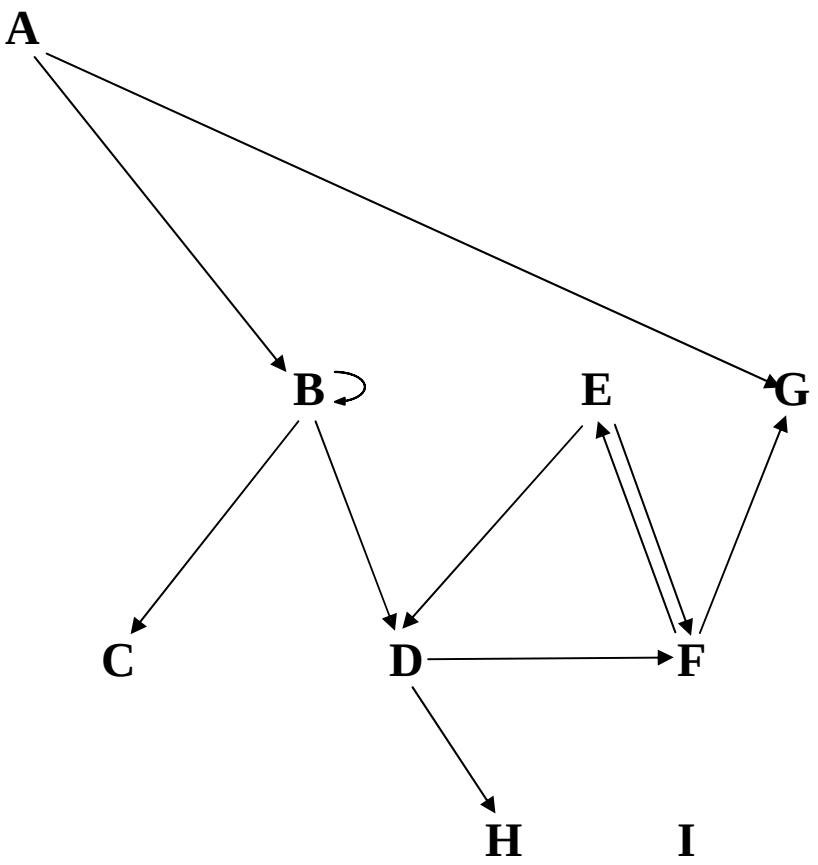
ii) This study provides counts and analysis of outgoing links from web pages, here named ***outlinks or external links***.

iii) This study provides links to web pages or links coming from the other websites called ***inlinks or incoming links.*** The concept covers all links from other websites pointing to a certain webpage or website. An incoming link is similar to receiving a citation in a document. These links also known as backward links.

iv) ***Reciprocal Link:*** If two web pages or two websites have a link pointing to each other, we define the link as a reciprocal link.

The link relations between the web nodes have been described in the following Fig. 3.2 below.

Figure 3.2: Basic network terminology



Source: Thelwall, M., Vaughan, L. & Björneborn, L. (2005)

The letters may represent different Web node levels, for example, Webpages, Web directories, Web sites, or top level domains of countries or generic sectors. And, the arrows represent the linking pattern of these nodes among one another.

- B has an *inlink* from A
- B has an *outlink* to C
- B has a *selflink*
- E and F are *reciprocally* linked
- A has a *transversal outlink* to G: functioning as a shortcut
- H is *reachable* from A by a directed *link path*
- I has neither in- nor *outlinks;* I is *isolated*
- B and E are *co-linking* to D; B and E have *co-outlinks*
- C and D are *co-linked* from B; C and D have *co-inlinks*

The terms *outlink* and *inlink* are commonly used in Computer Science-based Web studies. The term *outlink* implies that a directed link and its two adjacent nodes are viewed from the source node providing the link, analogous with the use of the term *reference* in bibliometrics. A corresponding analogy exists between the terms *inlink* and *citation*, with the target node as the spectator's viewpoint (Holmberg, 2009).

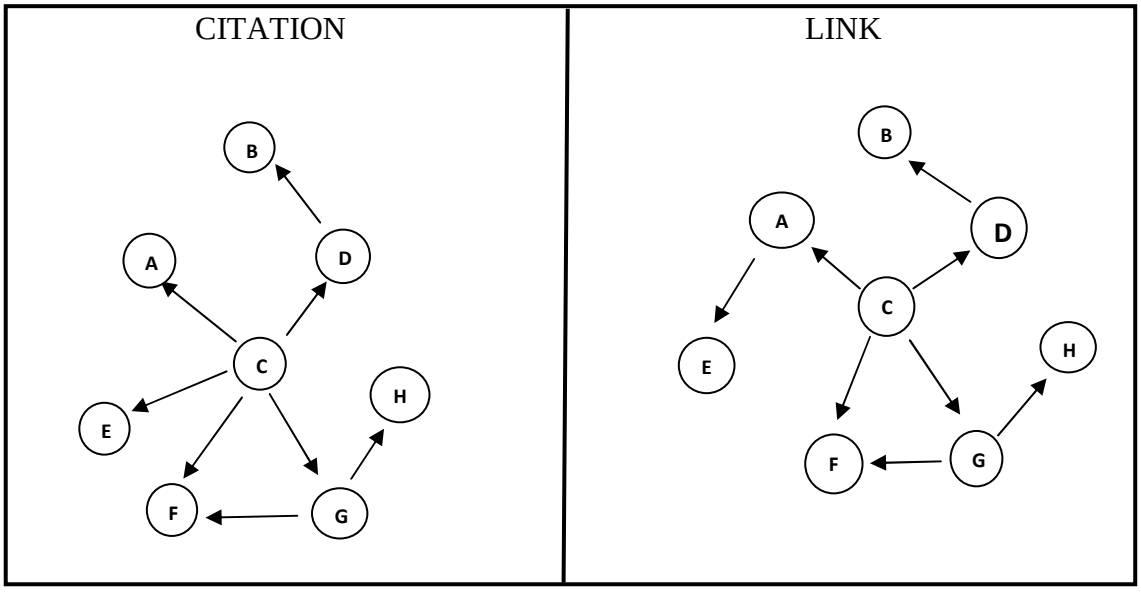
On the Web, *selflinks* are used for a wider range of purposes than self-citations in scientific literature. Page *selflinks* point from one section to another within the same page. Site *selflinks* (also known as *internal links*) are typically navigational pointers from one page to another within the same site. Most links on the Web connect web pages containing cognate topics. However, some links may break a typical linkage pattern in a web node neighbourhood and connect dissimilar topical domains. Such (loosely defined) *transversal links* function as cross-topic shortcuts and may affect so-called small-world phenomena on the Web (Thelwall, Vaughan, & Björneborn, 2005; Holmberg, 2009).

The two *co-linked* web nodes C and D in Figure. 3.2 with *co-inlinks* from the same source node are analogous to the bibliometric concept of *co-citation* pointed out by Small (1973). Correspondingly, the two *co-linking* nodes B and E having *co-outlinks* to the same target node are analogous to a *bibliographic coupling* has been proposed by Kessler (1963). *Co-links* is proposed as a generic term covering both concepts.

Co-inlinking is based on co-citation analysis (Small, 1973): if a single document cites two other documents, these documents are likely to have similar content, and the more often they are cited together, the stronger the similarity is assumed to be. Co-inlinking may reveal something about the external view on the relationships between the studied set of Web sites. Co-outlinking is based on bibliographic coupling where two documents are assumed to be similar if they both cite to a third document (Kessler, 1963; Holmberg, 2009).

Webometrics is a concept which deals with Web based phenomena, using methods originally designed for bibliometric analysis of scientific journal article citation pattern. The hope that web links could be used to provide information similar to that extracted from journal citations has been a key factor in stimulating much Webometric research. The co-citation coupling as used to establish subject similarities between two documents are also visualized in two web documents. These similarities of citation and links can be visualized in the following Figure 3.3 (Walia & Kaur, 2008; Mukherjee, 2008).

Figure 3.3 Functional Relational between Citation and Link



Source: Mukherjee, B. (2011)

Table 3.1: Observation on citation and links

OBSERVATIONS	
1. Paper C has 4 references.	1. Web page C has 4 links on it.
2. Paper A has 2 citations. One is citing references and other is cited references.	2. Paper A has 2 links, one pointed to it and other emerged from it.
3. Paper C and G are bibliographically coupled.	3. Paper C and G are linked to a common page.

Source: Mukherjee, B. (2011)

Citations are recognitions of previous work and they refer readers to other articles and other information sources in a similar way that Web links refer visitors to other Web pages and Web information sources. The similarities between the use and structure of citations and Web links gave some scientists the idea of using Web links as sources of information about Web sites and pages and about authors of them (e.g. Almind & Ingwersen, 1997; Ingwersen, 1998). Cronin (2001) wrote that "the Web affords bibliometricians rich opportunities to apply and adapt their techniques to new contexts and content" and according to Bar-Ilan (2000) there are a lot of valuable and freely available data hidden in the Web and that the Web has an excellent potential to serve also as a bibliographic database. In fact, the Web has been used to study scholarly communication much in the same way that scientific literature has been used for the same purpose (Holmberg, 2009).

While links within Web sites are mainly created for navigational purposes, links between Web sites may be a rich source of information about the content and use of the Web sites connected to each other with hyperlinks. Links between Web sites indicate that information at one location is at least known or valued at another location, demonstrating some kind of connection between the sites and their authors. Links could also mean that the target page is somehow useful for the source page of the link. In a way, a link may therefore tell more about the source page than the target page (Walia & Kaur, 2008; Shukla, 2009; Holmberg, 2009).

3.3.2 Web Page Content Analysis

A number of webometric investigations have focused not on web sites but on academic publications; using the web to count how often journal articles are cited. Content analyses have shown that links between academic websites tend to be created

for scholarly or educational reasons, a partial similarity with citation analysis. Vaughan and Shaw (2003) discussed online citations which are relatively trivial, for example appearing in journal contents lists rather than in the reference sections of academic articles. If this can be automated then it would give an interesting alternative to the ISI citation indexes.

i) This study provides hits on the systematic organization of web based information sources.

ii) To enable the users to reduce their time in the choice of right sources.

iii) The study compares the efficiency of search engines in retrieving the required information sources.

iv) The study will be useful for students, researchers, scientists who seek information through www.

v) Simplistic counts and content analysis of web pages are like traditional publication analysis.

3.3.3. Web Usage Analysis

It is a new type of statistical analysis of the Science and Technical Information (STI) in the web context is produced. A web server log is an important source for performing web usage analysis and bibliometrics because it explicitly records the browsing behaviour of the site visitors. A set of server sessions is the necessary input for any Web usage analysis. The data recorded in server logs reflects the access of a Web site by multiple users. These log files can be stored in various formats. The Web server stores query data. Query data is generated by online visitors while searching for records (web pages) relevant to their information needs. There are two web usage factor i.e., web users' information retrieval and Web customers' orders. The first is a web usability factor, and the other is a web customer order factor. These factors can be considered for evaluating online information sources by the observation of the information displayed by users, and the documents ordered by user customers. A situation is the number of times an information source is used or displayed by online users. This is a well known situation in information retrieval. The other is the number of times an information source is ordered; in this case we are in face of ecommerce transactions. Web usage analysis covers:

- Log files for users searching

- Browsing behaviour
- Log analysis for security applications
- Web usage pre-processing
- Novel techniques for discovery and analysis of Web usage patterns
- Integrating semantics and domain knowledge in Web usage mining and analysis
- Reliability and consistency of Webometrics
- Integration of click stream data with back-end data and related metrics
- Intelligent summarization/explanation of changes in Web usage metrics (Comer, 1997; Goswami, Sharma, & Shukla, 2008)..

3.3.4. Web Technology Analysis – Search Engine Performance

The fourth one is 'web technology analysis' (search engine performance). Technology is a term, which denotes the quality. It includes different search engine performances (Google, Altavista, Yahoo etc.), because search engines are such type of websites, which incorporates more technology than other web sites. The result of search engine comes as the big list of URL's of different website of a particular subject. In short it can be said that technology itself is a very broad phenomenon but as far as webometrics study is concerned, it is measurable and a useful study tool for web based study. The search engines performance determines following information:

- Measuring the search engines.
- Total number of hits retrieved.
- Number of relevant hits retrieved.
- The content of the page like what is the page all about etc.
- Ranking of search engine (Comer, 1997; Goswami, Sharma, & Shukla, 2008).

3.4. Webometric Terminology and Diagrams

3.4.1. Domain name

A domain name is an identification string that defines a realm of administrative autonomy, authority, or control on the Internet. Domain names are used in various networking contexts and application-specific naming and addressing purposes. In general, a domain name represents an Internet Protocol (IP) resource, such as a personal computer used to access the Internet, a server computer hosting a web site, or the web site itself or any other service communicated via the Internet. Domain names

are formed by the rules and procedures of the Domain Name System (DNS). Any name registered in the DNS is a domain name (Goel**,** 2010; Nair, 2012; Wikipedia).

The Domain Name System (DNS) is a hierarchical distributed naming system for computers, services, or any resource connected to the Internet or a private network. It associates various information with domain names assigned to each of the participating entities. Most prominently, it translates easily memorized domain names to the numerical IP addresses needed for the purpose of locating computer services and devices worldwide. An often used analogy to explain the Domain Name System is that it serves as the phone book for the Internet by translating human-friendly computer hostnames into IP addresses. For example, the domain name www.example.com translates to the addresses 192.0.43.10. Unlike a phone book, the DNS can be quickly updated, allowing a service's location on the network to change without affecting the end users, who continue to use the same host name (Comer, 1997; Young, 1999).

3.4.1.1. Domain name syntax

A domain name consists of one or more parts, technically called labels that are conventionally concatenated, and delimited by dots, such as example.com.

- The right-most label conveys the top-level domain. For example, the domain name *www.example.com* belongs to the top-level domain *com*.

- The hierarchy of domains descends from the right to the left label in the name; each label to the left specifies a subdivision, or sub domain of the domain to the right. For example: the label *example* specifies a node *example.com* as a sub domain of the *com* domain, and *www* is a label to create www.example.com, a sub domain of *example.com*. This tree of labels may consist of 127 levels. The full domain name may not exceed a total length of 253 ASCII characters in its textual representation.

- A **hostname** is a domain name that has at least one associated IP address. For example, the domain names *www.example.com* and *example.com* are also hostnames, whereas the *com* domain is not. However, other top-level domains, particularly country code top-level domains, may indeed have an IP address, and if so, they are also hostnames (Young, 1999; Wikipedia).

3.4.1.2. Types of Domain Name

Domain names are organized in subordinate levels (sub-domains) of the DNS root domain, which is nameless. The first-level set of domain names are the top-level domains (TLDs), including the generic top-level domains (gTLDs) such as the prominent domains com, info, net and org, and the country code top-level domains (ccTLDs). Below these top-level domains in the DNS hierarchy are the second-level and third-level domain names that are typically open for reservation by end-users who wish to connect local area networks to the Internet, create other publicly accessible Internet resources or run web sites. The registration of these domain names is usually administered by domain name registrars who sell their services to the public (OECD, 2006; Comer; 1997; Wikipedia).

1) Top-Level Domain (TLD):

A top-level domain (TLD) is one of the domains at the highest level in the hierarchical Domain Name System of the Internet. The TLD names are installed in the root zone of the name space. For all domains in lower levels, it is the last part of the domain name, that is, the last label of a fully qualified domain name. For example, in the domain name www.example.com, the TLD is com. Responsibility of management of most TLDs is delegated to specific organizations by the Internet Corporation for Assigned Names and Numbers (ICANN), which operates the Internet Assigned Numbers Authority (IANA), and is in charge of maintaining the DNS root zone. IANA today distinguishes the following groups of top-level domains (OECD, 2006; Comer, 1997):

- **Country-Code Top-Level Domains** (ccTLD): A country code top-level domain (ccTLD) is an Internet TLD generally used or reserved for a country, a sovereign state, or a dependent territory. Two letter domains established for countries or territories. With some historical exceptions, the code for any territory is the same as its two-letter ISO 3166 code (Comer, 1997).

- **Internationalized Country Code Top-Level Domains** (IDN ccTLD): The limited set of ASCII characters permitted in the DNS prevented the representation of names and words of many languages in their native alphabets or scripts. To make this possible, ICANN approved the Internationalizing Domain Names in Applications (IDNA) system, by which user applications, such as web browsers, map *Unicode* strings into the valid DNS character set

using *Punycode*. In 2009 ICANN approved the installation of internationalized domain name country code top-level domains. The IDN ccTLDs are in non-Latin character sets (e.g., Arabic or Chinese) (Wikipedia).

- **Generic Top-Level Domains** (gTLD): TLDs with three or more characters. The core group of generic TLD consists of the *com*, *info*, *net*, and *org* domains. In addition, the domains *biz*, *name*, and *pro* are also considered generic; however, these are designated as restricted, because registrations within them require proof of eligibility within the guidelines set for each. Thus, domains edu, gov, int, and mil are now considered sponsored top-level domains, much like the many newly created themed domain names (e.g., jobs). The entire group of domains that do not have a geographic or country designation is still often referred to by the term generic TLDs (Comer, 1997; Young, 1999).

2) Second-level and lower level domains:

Below the TLDs in the domain name hierarchy are the second-level domain (SLD) names. These are the names directly to the left of .com, .net, and the other TLDs. As an example, in the domain example.co.uk, co is the second-level domain.

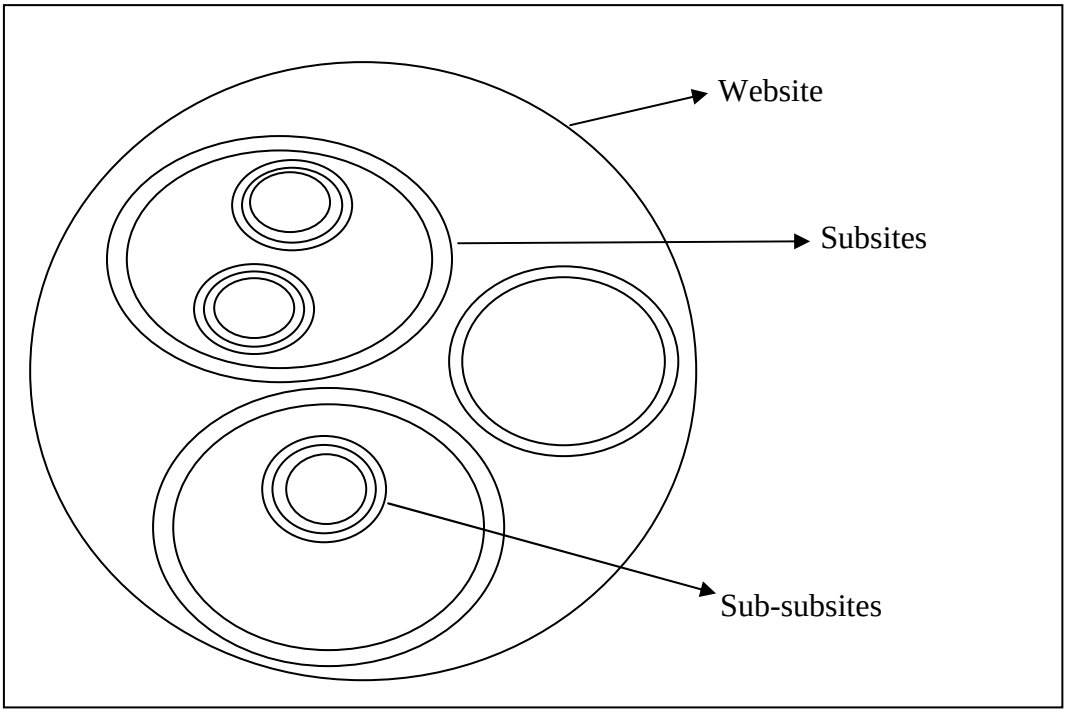
Next are third-level domains, which are written immediately to the left of a second-level domain. There can be fourth- and fifth-level domains, and so on, with virtually no limitation. An example of an operational domain name with four levels of domain labels is www.sos.state.oh.us. The www preceding the domains is the host name of the World-Wide Web server. Each label is separated by a full stop (dot). ‘sos’ is said to be a sub-domain of ‘state.oh.us’, and ‘state’ a sub-domain of ‘oh.us’, etc. In general, sub domains are domains subordinate to their parent domain (Young, 1999; Wikipedia).

Second-level (or lower-level, depending on the established parent hierarchy) domain names are often created based on the name of a company (e.g. bbc.co.uk), product or service (e.g. hotmail.com). Therefore, ftp.ifla.org might be an FTP server, www.ifla.org would be a World Wide Web server, and mail.ifla.org could be an email server, each intended to perform only the implied function. The hierarchical DNS labels or components of domain names are separated in a fully qualified name by the full stop (.).

3.4.2 Websites, Subsites & Sub-subsites

Zooming in on a single Web site may reveal several subunits in the shape of subsites, sub-subsites, and so on, as indicated by hierarchically derivative domain names. In Fig 3.4 subsites and sub-subsites are denoted as circles with double and triple borderlines, respectively. Subordinate sublevels would logically be denoted with additional number of borderlines. For the sake of simplicity, the diagram does not reflect actual numbers and sizes of elements.

Figure 3.4: Simplified Web node diagram of a Web site containing subsites and sub-subsites (Björneborn & Ingwersen, 2005).



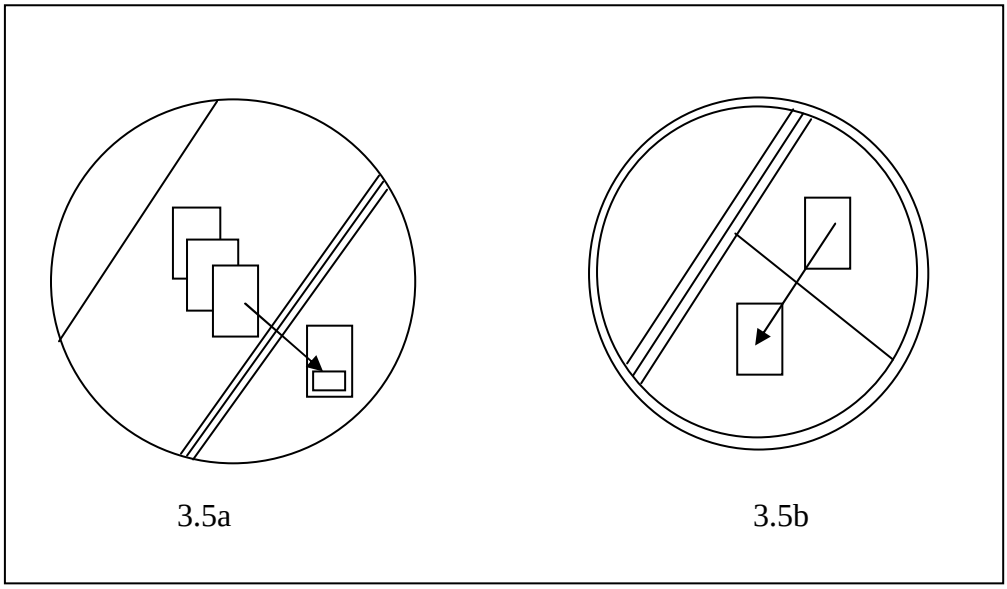
Source: Thelwall, M., Vaughan, L. & Björneborn, L. (2005)

Although some Web sites subdivide into derivative domain names, as noted, others locate the same type of subunits in folder directories. Obviously, such diverse allocation and naming practices complicate comparability in webometric studies. In Figure 3.5 directories, subdirectories, and so on, are denoted by one or more diagonal lines resembling URL slashes and reflecting the number of directory levels below the URL root level**.**

Web pages may consist of sub-elements, such as text sections, and frames. Additional bands illustrate such page sub-elements as in the targets of the page selflink *h* and the page outlink *i* from the two sibling Web pages in the same directory in Figure 3.5a.

More complex and numerous linkages within a site or subsites can be illustrated by combinations of elements in Figure 3.5, showing links between pages located either at different directory levels (Figure 3.5a) or in sibling directories at the same level (Figure 3.5b) in the Web site file hierarchies.

Figure 3.5.: Simplified Web node diagrams of a Web site and a subsite with links between different directory levels including page sub-elements (Bjorneborn & Ingwersen, 2005).



Source: Thelwall, M., Vaughan, L. & Björneborn, L. (2005)

Naturally, any diagrammatic representation of large-scale hypertext structures will become too tangled to be of practical use, even less to be interpreted in any quantitative way. However, the proposed Web node diagrams with their simple and intuitive geometrical figures are intended to be used to emphasize and illustrate qualitative differences between investigated Web node levels. Moreover, the diagrams can illustrate important structural aspects of limited sub-graphs of a given Web space (Thelwall, Vaughan, & Björneborn, 2005).

Conclusion

From the above discussions it can be examined on how webometric studies can be conducted. With the presence of the World Wide Web different studies related to webometric research can be conducted as described above by analyzing the links and URLs. Webometrics in general, aims at designing and developing methodologies to measure visibility such as Web Impact Factor (WIF). The WIF provides a way to

evaluate a website's relative importance especially when we compare it to others in the same or a country's domain. It is computable in relation to a national sector and layer web segments or top level domains. It provides a quantitative indicator of website's long term influence; it reflects the ability of websites and webmasters to attract users. It may also provide novel insights into retrieval process on the web.

References:

Aguillo, I.F. et al., (2006). Scientific research activity and communication measured with cybermetrics indicators, *Journal of the American Society for Information Science and Technology*, 57(10), 1296–1302.

Almind, T. C., & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: Methodological approaches to webometrics. *Journal of Documentation, 53*(4), 404-426.

Asadi, M., & Shekofteh, M. (2009). The relationship between the research activity of Iranian medical universities and their web impact factor. *Electronic Library, 27*(6), 1026-1043.

Babu, B. R., & Jeyshankar, R. (2009). Websites of universities in Tamil Nadu: A webometric study. *Annals of Library and Information Studies, 56*(1), 69-79.

Babu, B. R., Jeyshankar, R., & Rao, P. N. (2010). Websites of Central Universities in India: A webometric analysis. *DESIDOC Journal of Library & Information Technology, 30*(4), 33-43.

Bar-Ilan, J. (2000). Evaluating the stability of the search tools HotBot and Snap: a case study. *Online Information Review, 24*(6), 439-449.

Björneborn, L., & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics, 50*(1), 65-82.

Björneborn, L. (2004). *Small-world link structures across an academic web space: a library and information science approach (Doctoral Dissertation).* Copenhagen, Denmark: Royal School of Library & Information Science.

Björneborn, L., & Ingwersen, P. (2004). Towards a basic framework for webometrics. *Journal of the American Society for Information Science and Technology, 55*(4), 1216-1227.

Brookes, B. C. (1991). Biblio-, sciento-, infor-metrics??? What are we talking about? In *Informetrics 89/90;* Selection of Papers Submitted for the Second International Conference on Bibliometrics, Scientometrics and Informetrics. London, Ontario.

Borgman, C.L. (2000). Scholarly communication and bibliometrics revisited. In Cronin, B. & Atkins, H.B. (eds.) *The web of knowledge – A festschrift in honor of Eugene Garfield*. ASIS, 2000.

Borgman, C. & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology, 36*(1), 2-72.

Comer, E.D. (1997). *Computer networks and internet*. New Jersey: Prentice-Hall Inc.

Country code top-level domain name. (n.d.). Retrieved September 18, 2013, from Wikipedia: http://en.wikipedia.org/wiki/*country_code_top_level_domain _name*

Cronin, B. (2001). Bibliometrics and beyond: some thoughts on web-based citation analysis. *Journal of Information Science, 27*(1), 1-7.

Domain Name. (n.d.). Retrieved September 18, 2013, from Wikipedia: http://en.wikipedia.org/wiki/domain_name

Domain Name System. (n.d.). Retrieved September 18, 2013, from Wikipedia: http://en.wikipedia.org/wiki/domain_name_system

Frías, E. R. (2010). Application of webometric techniques to the study of accounting and financial variables. Doctoral Dissertation. University of Granada. Retrieved 12 March, 2013, from http://estebanromero.com/wp-content/uploads/2010/02/thesis-English-version-4-3-2010-DEF.pdf

Goel, M. K. (2010). *World Wide Web.* New Delhi: Rajat Publications.

Goswami, P., Sharma, U., & Shukla, A. K. (2008). The Webometrics. *6th International CALIBER-2008, University of Allahabad, Allahabad, February 28-29 & March 1, 2008*, 656-660. Retrieved on 11 March, 2013, from http://ir.inflibnet.ac.in/bitstream/handle/1944/1318/70.pdf

Holmberg, K. (2009). *Webometric network analysis : mapping cooperation and geopolitical connections between local government administration on the web.* (Doctoral Dissertation). Åbo Akademi University Press. Åbo Akademi University. Retrieved 15 March, 2013, from http://urn.fi/URN:ISBN:978-951-765-511-8

Hyperlink. (n.d.). Retrieved 29 October, 2013, from Wikipedia: http://en.wikipedia.org/wiki/hyperlink

Ingwersen, P. (1998). The calculation of Web Impact Factors. *Journal of Documentation, 54*(2), 236-243.

Islam, M. A., & Alam, M. S. (2011). Webometric study of private universities in Bangladesh. *Malaysian Journal of Library & Information Science, 16*(2), 115-126.

Islam, M. A. (2011). Webometric study of universities in Bangladesh. *Annals of Library and Information Studies, 58*(2), 307-318.

Jalal, S. K., Biswas, S.C., & Mukhopadhyay, P. (2010). Web impact factor and link analysis of selected Indian universities. *Annals of Library and Information Studies, 57*(1), 109-121.

Jeyshankar, R., Sujitha, I. M., & Valarmathi, A. (2012). Web pages of ICMR institutes websites: A webometric analysis. *Global Advanced Research Journal of Library, Information and Archival Studies, 1*(1), 6-18.

Jeyshankar, R., & Babu, B. R (2009). Websites of universities in Tamil Nadu: A webometric study. *Annals of Library and Information Studies*; 56(1); 69-79.

Kannan, K., & Abilash, S. (2011). Information searching pattern of web search engine: a comparative study. *Library Progress (International), 31*(1), 35-42.

Kessler, M.M. (1963). Bibliographic coupling between scientific papers. *American Documentation, 14*(1), 10-25.

Kobayashi, M., & Takeda, K. (2000). Information retrieval on the web. *ACM Computing Surveys, 32*(2), 144-173.

Kumar, A. M. N., Babu, B. R., Gopalakrishnan, S., & Rao, P. N. (2010). Location of web objects and links in the web sites of universities in Tamil Nadu: a study. *SRELS Journal of Information Management, 47*(3), 325-338.

Kumar, A. M. N., Babu, B. R., Gopalakrishnan, S., & Rao, P. N. (2010). Location of web objects and links in the web sites of universities in Tamil Nadu: a study. *SRELS Journal of Information Management, 47*(3), 325-338.

Kumar, S.S., & Rabindra, M. (2012). Web impact factor and webometric study of National Institute of Technology (NITs) of India. *Pearl: A Journal of Library and Information Science, 6*(1), 47-55.

Kumar, S.S., & Rabindra, M. (2012). Web impact factor and webometric study of National Institute of Technology (NITs) of India. *Pearl: A Journal of Library and Information Science, 6*(1), 47-55.

Larson, R.R. (1996). Bibliometrics of the World Wide Web: an exploratory analysis of the intellectual structure of cyberspace. In: Hardin, S. (Ed.), *Proceedings of the 59th Annual Meeting, ASIS 96*. Baltimore, pp. 71-79.

Link. (n.d.). Retrieved September 18, 2013, from Wikipedia: http://en.wikipedia.org/wiki/Link

Mayr, P., & Tosques, F. (2005). *Google Web APIs: an Instrument for Webometric Analyses?* Available at: www.ib.hu-berlin.de/%7Emayr/arbeiten/ISSI2005_Mayr_Toques.pdf

Ministry of Human Resource Development (MHRD), India. (n.d.). *Department of Higher Education, IIT*. Retrieved April 9, 2013, from http://mhrd.gov.in/iit_hindi

Mukherjee, B. (2011). Bibliometrics to webometrics: The changing context of quantitative research. *IASLIC Bulletin, 56*(2), 97-110.

Nair, R.R. (2002). *Internet for information services*. New Delhi: Ess Ess Publications.

National Knowledge Commission. (2007). *Libraries: Gateways to Knowledge*. New Delhi: National Knowledge Commission, Govt. of India.

Noruzi, A. (2005). Editorial: Fundamental differences between hyperlinks and citations. *Webology, 2*(2), editorial 4. Available at: http://www.webology.org /2005/v2n2/editorial4.html

Noruzi, A. (2005). Web Impact factors for Iranian Universities. *Webology, 2*(1).

Noruzi, A. (2006). The web impact factors: A critical review. *Electronic Library, 24*(4), 490-500.

Organisation for Economic Co-operation and Development (OECD). (2006). *Evolution in the management of country code top level domain names (ccTLDs).* Report to Working Party on Communication Infrastructures and Services Policy (ICCP).1-49. Retrieved on 12 August, 2013, from, http://www.oecd.org/sti/ieconomy/37730629.pdf

Pritchard, A. (1969). Statistical Bibliography or Bibliometrics?. *Journal of Documentation, 25*(4), 348–349.

Rao, I. K.R. (2010). *Growth of literature and measurement of scientific productivity.* New Delhi: Ess Ess Publication.

Ratha, B., Joshi, L., & Naidu, G.H.S. (2012). Webometric study of IIT libraries websites. *DESIDOC Journal of Library & Information Technology, 32*(3), 249-254.

Samuel C. A. U. (2008). Information technology and web use characteristics of Nigerian private universities. *African Journal of Library, Archives and Information Science.* Available at: http://findarticles.com/p/articles/mi_7002 /is_2_18/ai_n31877465/

Sangam, S.L., & Keshava. (2008). Concept of Bibliometrics, Scientometrics and Informetrics.. In Amudhavalli, A. (2008). Dynamics in digital information systems. New Delhi: Ess Ess Publications. 301 - 313.

Sangam. S. L. (2008). *Areas in the field of Scientometrics and Informetrics*. In: Koganuramath, M. M., Kumar, B. D., & Kademi, B.S. (eds). Library and Information Science Profession in the Knowledge Society. New Delhi: Allied Publishers Pvt. Ltd. 265-271.

Sengupta. I. N. (2009). Bibliometrics, Informetrics, Scientometrics and Librametrics: An Overview. *Libri, 42*(2), 75-98, doi: 10.1515/libr.1992.42.2.75

Shukla, A., & Tripathi, A. (2009). Webometric analysis of institutes of national importance in India. *IASLIC Bulletin, 54*(3), 165-180.

Shukla, A. (2009). *Evaluating Indian Academic and Research Library Websites based on Web Impact Factor Analysis* (*Doctoral dissertation).* Varanasi, India: Department of Library & Information Science, Banaras Hindu University.

Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265-269.

Sujithai, M., & Jeyshankar, R. (2013). Web page analysis of Indian Institute of Technologies' (IITs) websites: A webometric study. *International Journal of Digital Library Services, 3*(1), 55-65.

Tague-Sutcliffe, J. (1992). An introduction to Informetrics. *Information Processing and Management, 28*(1), 1- 4.

Thanuskodi, S. (2011). Webometric analysis of private engineering college websites in Tamil Nadu. *J Communication, 2*(2), 73-81.

Thanuskodi, S. (2012). A webometric analysis of selected institutes of national importance websites in India. *International Journal of Library Science, 1*(1), 13-18.

Thelwall, M. (2000). Web Impact Factors and Search Engine Coverage. *Journal of Documentation, 56*(2), 185-189.

Thelwall, M., Vaughan, L., & Bjorneborn, L. (2006). Webometrics. *Annual Review of Information Science and Technology, 39*(1), 81-135.

Thelwall, M. (2007). Bibliometrics to Webometrics. *Journal of Information Science, 34*(4), 1-18.

Thelwall, M. (2011). A comparison of link and URL citation counting. *Aslib Proceedings, 63*(4), 419-425.

Thomas, O., & Willet, P. (2000). Webometric analysis of departments of librarianship and information science. *Journal of Information Science, 26*(6), 421-428. doi: 10.1177/01 6555150002600605

Vaughan, L., & Shaw, D. (2003). Bibliographic and web citation: What is difference?. *Journal of the American society for information sciences and technology, 54*(14),1313-1322.

Vijayakumar, M. (2012). Webometric analysis of university websites in Srilanka. *International Journal of Information Dissemination and Technology, 2*(3), 155-159.

Walia, P. K., & Kaur, P. (2008). Webometric analysis of Library Associations' websites of India. *IASLIC Bulletin*, *53*(3), 131-143.

Webometrics. (n.d.). Retrieved 29 October, 2013, from Wikipedia: http://en.wikipedia.org/wiki/webometrics

Web browser. (n.d.). Retrieved 29 October, 2013, from Wikipedia: http://en.wikipedia.org/wiki/web_browser

CHAPTER- IV

ANALYSIS OF DATA
&
PRESENTATION OF FINDINGS

4.1 Introduction

For conducting the present study, the World Wide Web was surfed to gather the URLs of websites from a total number of sixteen (16) IITs library websites. For this, search queries were made using the search engine "Google". The data has been collected in three rounds using the selected search engine for retrieving *inlinks*, *outlinks* and number of web pages of IITs library websites. To evaluate the chosen websites, the methodology used was investigative in nature. The data have been tabulated and analyzed based on the objectives of the present study.

4.2 URL Analysis of IITs Library Websites

A uniform resource locator (URL) also known as web address is a specific character string that constitutes a reference to a resource. The first part of the URL consists of the transfer protocol, the second specifies the domain names which is followed by directory and file name. For studying the URLs of the IITs library websites the domain names have been taken into account wherein Table 4.1 represents the most frequently used TLDs by the IITs for their library website.

Table 4.1: TLDs of IITs Library Websites

Name of IITs	**Uniform Resource Locator (URL)**	**Top Level Domain**	
		Generic Top Level Domain (gTLD)	**Country code Top Level Domain (ccTLD)**
IITBHU	http://www.iitbhu.ac.in/library/	.ac	.in
IITBBS	http://www.iitbbs.ac.in/Library.php	.ac	.in
IITB	http://www.library.iitb.ac.in/	.ac	.in
IITD	http://library.iitd.ac.in/	.ac	.in
IITGN	http://www.iitgn.ac.in/library.htm	.ac	.in
IITG	http://www.iitg.ernet.in/rs/lib/public_html/index.html	.ernet	.in
IITH	http://library.iith.ac.in/	.ac	.in
IITI	http://www.iiti.ac.in/Library/about_central_library.html	.ac	.in
IITK	http://library.iitk.ac.in/	.ac	.in
IITKGP	http://www.library.iitkgp.ernet.in/	.ernet	.in
IITM	http://www.iitm.ac.in/library	.ac	.in
IITMANDI	http://www.iitmandi.ac.in/academics/lib/	.ac	.in
IITP	http://www.iitp.ac.in/index.php/services-and-amenities/central-library/about-iitp-library.html	.ac	.in

IITJ	http://www.iitj.ac.in/library/	.ac	.in
IITR	http://mgcl.iitr.ac.in/	.ac	.in
IITRPR	http://www.iitrpr.ac.in/library	.ac	.in

From the illustration of Table 4.1, the study reveals the TLDs of IITs library websites which are .ac, .ernet, and .in. The TLD .ac.in is used by about 87.5% IITs library websites whereas only 12.50% have used .ernet.in. The TLDs are further divided into generic TLD (gTLD) and country code TLD (ccTLD). The gTLD .ac stands for academic domain which is used by 87.50% (14) IITs library websites. The gTLD .ernet stands for Education and Research Network is used only by IITKGP & IITG library websites (i.e. 12.50%). The ccTLD .in stands for the country India has been used by all of the IITs library websites (i.e. 100%). There are 37.5% (6) IITs library websites having sub-sites (e.g. http://**library**.iitk.ac.in/) whereas 31.25% (5) IITs library websites are following directory structure (e.g. http://www.iitj.ac.in/**library**/) and remaining 31.25% (5) IITs library websites are following the single page link structure attached with the IITs domain name (e.g. http://www.iitgn.ac.in/**library.htm**).

4.3 File Formats supported by IITs Library Websites

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), and GIF images by .gif. Many formats still use three-character extensions even though modern operating systems and application programs no longer have this limitation. Since there is no standard list of extensions, more than one format can use the same extension. The following table shows the type of file formats used by IITs library websites which are used for scholarly communication on their websites.

Table 4.2: Types of File Formats supported by IITs library websites

Name of IIT	FORMAT SUPPORTED									
	HTML	**PDF**	**MS-Word**	**MS- Excel**	**MS- PPT**	**JPG/ JPEG**	**PNG**	**GIF**	**WMA**	**Total Format Supported**
IITBHU	✓	×	×	×	×	✓	✓	×	×	3
IITBBS	✓	✓	✓	×	×	✓	×	✓	×	5
IITB	✓	✓	✓	×	×	✓	✓	✓	×	6
IITD	✓	×	×	✓	×	✓	✓	✓	×	5
IITG	✓	✓	✓	✓	×	×	×	×	×	4
IITGN	✓	✓	✓	✓	×	✓	×	✓	×	6
IITH	✓	✓	✓	✓	×	✓	✓	✓	×	7
IITI	✓	✓	✓	✓	×	✓	✓	✓	×	7
IITK	✓	✓	×	×	×	✓	×	✓	×	4
IITKGP	✓	✓	×	×	×	×	✓	✓	✓	5
IITM	✓	✓	✓	×	×	✓	×	✓	×	5
IITMANDI	✓	✓	×	×	×	✓	×		×	3
IITP	✓	×	×	×	×	✓	✓	✓	×	4
IITJ	✓	✓	×	×	✓	×	✓	✓	×	5
IITR	✓	✓	✓	✓	✓	✓	✓	✓	×	8
IITRPR	✓	✓	✓	×	×	✓	✓	✓	×	6
Percentage	100%	81.25%	56.25%	37.5%	12.5%	81.25%	62.5%	81.25%	6.25%	

After analyzing all the web pages (page wise) inside the IITs library websites, document file formats (.html, .pdf, .doc, .ppt), image file formats (.jpg/.jpeg, .png, .gif etc.), and audio/visual file formats (.wma) have been used to display the information about library. From the table 4.2, it has been found that all the IITs library websites are using HTML (100%) followed by PDF (81.25%), MS-Word (56.25%), MS-Excel (37.5%), and MS-PowerPoint (12.50%) under document file formats. In image file formats, JPG/JPEG (81.25%), PNG (62.5%), and GIF (81.25%) in IITs library websites. IITKGP is using only audio/visual (.wma) file format in the library website. In another way, IITR library website has used maximum file formats (8) to represent the information content of library followed by IIT Hyderabad and IIT Indore (7), IIT Bombay, IIT Gandhinagar and IIT Ropar (6), IIT Bhubaneswar, IIT Delhi, IIT Kharagpur, IIT Madras and IIT Jodhpur (5), IIT Guwahati, IIT Kanpur and IIT Patna

(4). The least file formats were used by IIT BHU and IIT Mandi (with 3 file formats) for their library websites.

4.4 Web Impact Factor (WIF) of IITs Library Websites

For calculating the WIF, the links of the IITs library websites were collected using the command textual queries site:URL and link:URL in the search engine "Google". By using these textual queries the number of web pages and number of links present in the library websites have been collected. Then, these links were analyzed based on the formula given by Ingwersen (as given in 1.5.2 in Chapter 1).

4.4.1 WIF & Ranking of IITs Library Websites in First Round

The first round of data was collected on 27th September, 2013 between 02.14 p.m. to 03.55 p.m. from the search engine by using command textual queries for collecting the number of web pages, the total links and the inlinks for the calculation of the WIF and Revised WIF. Table 4.3 illustrates the WIF of IITs library websites and their ranking based on first round Revised WIF data. The WIF and RWIF results are represented upto four digits after decimal point in each round.

Table 4.3: WIF of IITs Library Websites in First Round

Name of IIT (Coded Form)	**No. of Web Pages (A)**	**Total Links (B)**	**Inlinks (C)**	**WIF D=(B/A)**	**RWIF E=(C/A)**	**Rank** (based on RWIF)
IITBHU	62	6	1	0.0968	0.0161	5
IITBBS	Data not collected due to non availability of library website during the period of data collection					
IITB	23000	79	11	0.0034	0.00048	10
IITD	1820	9	1	0.0049	0.00054	9
IITG	1	5	1	5	1	1
IITGN	34	0	0	0	0	12
IITH	22	1	1	0.0455	0.0455	3
IITI	42	18	0	0.4286	0	12
IITK	12100	5	2	0.0004	0.0002	11
IITKGP	432	11	7	0.0255	0.0162	4
IITM	541	15	5	0.0277	0.0092	6
IITMANDI	169	2	1	0.0118	0.0059	7
IITP	1	32	1	32	1	1
IITJ	61	15	0	0.2459	0	12
IITR	501	5	1	0.0099	0.0019	8
IITRPR	6	4	1	0.6667	0.1667	2

(Data Collection Date & Time: 27/09/2013 between 02.14 p.m. – 03.40 p.m.)

From the Table 4.3, is has been observed that some IITs library websites have been indexed thoroughly in Google's database whereas some IITs library websites have not been indexed thoroughly (e.g. IITG, IITP, IITRPR). It has been observed that the libraries those having more number of indexed pages in Google retrieved very less number of links and *inlinks* comparatively. IITGN has no any links from other sources. From the Table 4.3, it can be inference that more number of indexed pages in Google leads to more number of *inlinks* to the websites. IITs library websites those having less number of indexed pages and at least single *inlink* leads to higher WIF than those having more number of indexed pages and more than one *inlink*. On the basis of Revised WIF (RWIF), it has been found that IITG and IITP have highest RWIF data (i.e. 1) and leads to 1st rank in first round of data. IITRPR, IITH, IITKGP, and IITBHU have obtained 2nd, 3rd, 4th and 5th rank respectively with the RWIF ranging between 0.1667 to 0.0161 in first round of data. IITGN has a number of 34 web pages but the total *inlinks* could not be established which resulted zero RWIF. Though IITI and IITJ have a good number of links but zero *inlinks* to their websites which resulted zero RWIF.

4.4.2 WIF & Ranking of IITs Library Websites in Second Round

The second round of data was collected on 12th October, 2013 between 12.30 p.m. to 01.05 p.m. Table 4.4 illustrates the WIF of IITs library websites and their ranking based on second round Revised WIF data.

Table 4.4: WIF of IITs Library Websites in Second Round

Name of IIT (Coded Form)	**No. of Pages (A)**	**Total Links (B)**	**Inlinks (C)**	**WIF D=(B/A)**	**RWIF E=(C/A)**	**Rank** (based on RWIF)
IITBHU	62	6	1	0.0968	0.0161	5
IITBBS	Data not collected due to non availability of library website during the period of data collection					
IITB	20600	78	11	0.0038	0.0005	10
IITD	1380	53	1	0.0384	0.0007	9
IITG	1	5	1	5	1	1
IITGN	33	0	0	0	0	12
IITH	22	1	1	0.0455	0.0455	3
IITI	46	18	0	0.3913	0	12
IITK	13300	5	3	0.0004	0.0002	11
IITKGP	428	9	6	0.0210	0.0140	6
IITM	523	72	10	0.1377	0.0191	4
IITMANDI	161	2	1	0.0124	0.0062	7

IITP	1	35	1	35	1	1
IITJ	70	16	0	0.2286	0	12
IITR	457	5	1	0.0109	0.0022	8
IITRPR	6	4	1	0.6667	0.1667	2

(Data Collection Date & Time: 12/10/2013 between 12.30 p.m. – 01.32 p.m.)

From the Table 4.4, is has been found that some IITs library websites have been indexed thoroughly (e.g. IITB, IITD, IITK, IITM) in Google's database whereas some IITs library websites (e.g. IITG, IITP, IITRPR) have not been indexed thoroughly in Google's database till second round of data collection. It has been observed again that the libraries those having more number of indexed pages in Google, retrieved very less number of links and *inlinks* comparatively. IITGN has no any links from other sources. From the Table 4.4, it can be inference that more number of indexed pages in Google leads to more number of *inlinks* to the websites (in case of IITB, IITM, IITKGP). IITs library websites those having less number of indexed pages and at least single *inlink* leads to higher WIF (e.g. IITG, IITP, IITRPR, IITH) than those having more number of indexed pages and more than one *inlink*. On the basis of Revised WIF (RWIF), it has been found that again IITG and IITP have highest RWIF data (i.e. 1) and leads to 1st rank in first round of data. This is due to less no of inlinks with less number of indexed web pages. IITRPR, IITH, IITM, and IITBHU have obtained 2nd, 3rd, 4th and 5th rank respectively with the RWIF ranging between 0.1667 to 0.0161 in second round of data. IITGN again has an equivalent number (33) of web pages to the first round but the total *inlinks* could not be established which resulted zero RWIF. Though IITI and IITJ have a good number of *links* but zero *inlinks* to their websites which resulted zero RWIF. The ranking of IITs library websites are almost same to the first round due to similarity in data retrieved through search engine.

4.4.3 WIF & Ranking of IITs Library Websites in Third Round

The third round of data was collected on 27th October, 2013 between 05.11 p.m. to 05.20 p.m. Table 4.5 illustrates the WIF of IITs library websites and their ranking based on third round Revised WIF data.

Table 4.5: WIF of IITs Library Websites in Third Round

Name of IIT (Coded Form)	**No. of Pages (A)**	**Total Links (B)**	**Inlinks (C)**	**WIF D=(B/A)**	**RWIF E=(C/A)**	**Rank** (based on RWIF)
IITBHU	73	6	1	0.0822	0.0137	5
IITBBS	Data not collected due to non availability of library website during the period of data collection					
IITB	20400	78	5	0.0038	0.0002	10
IITD	1070	46	1	0.0429	0.0009	9
IITG	1	5	1	5	1	1
IITGN	33	0	0	0	0	12
IITH	24	1	1	0.0417	0.0417	3
IITI	51	18	0	0.3529	0	12
IITK	11500	5	2	0.0004	0.0002	11
IITKGP	453	8	4	0.0177	0.0088	6
IITM	521	89	10	0.1708	0.0192	4
IITMANDI	161	2	1	0.0124	0.0062	7
IITP	1	33	1	33	1	1
IITJ	40	14	0	0.35	0	12
IITR	549	5	1	0.0091	0.0018	8
IITRPR	6	4	1	0.6667	0.1667	2

(Data Collection Date & Time: 27/10/2013 between 05.01 p.m. – 05.40 p.m.)

From the Table 4.5, is has been found that IITB, IITD, and IITK have been indexed thoroughly in Google's database whereas IITG, IITP, and IITRPR have not been indexed thoroughly in Google's database till third round of data collection. It has been found again that library those having more number of indexed pages in Google, retrieved very less number of links and *inlinks* comparatively (e.g. IITB, IITD, IITK, IITR). IITGN has no links from other sources again till third round of data collection. From the Table 4.5, it can be inference that more number of indexed pages in Google leads to more number of *inlinks* to the websites (in case of IITB, IITM, IITKGP). IITs library websites those having less number of indexed pages and at least single *inlink* leads to higher WIF (e.g. IITG, IITP, IITRPR, IITH) than those having more number of indexed pages and more than one *inlink*. On the basis of Revised WIF (RWIF), it has been found that again IITG and IITP have highest RWIF data (i.e. 1) and leads to 1st rank in first round of data. This is due to less number of inlinks with less number of indexed web pages. IITRPR, IITH, IITM, and IITBHU have obtained again 2nd, 3rd, 4th and 5th rank respectively with the RWIF ranging between 0.1667 to 0.0136 in third rounds of data. IITGN again has an equivalent number (33) of web pages to the second round but the total *inlinks* could not be established which resulted zero RWIF.

Though IITI and IITJ have a good number of *links* but again zero *inlinks* to their websites which resulted zero RWIF. The ranking of IITs library websites are almost same to the first and second round of data collection due to similarity in data retrieved through search engine in all three rounds.

4.4.4 Average Revised WIF and Ranking of IITs Library Websites

The average Revised WIF of the three rounds of data from the study is presented in the Table 4.6. In order to calculate the average Revised WIF, the sum of all three rounds Revised WIF has been divided by three by using the following formula.

$$\textbf{Average Revised WIF} = \frac{\textbf{1st+2nd+3rd (Round Revised WIF)}}{\textbf{3}}$$

Table 4.6: Average Revised WIF and Ranking of IITs Library Websites

SN	Name of IIT (Coded Form)	1st Round RWIF	2nd Round RWIF	3rd Round RWIF	Average RWIF	Ranking (based on Average RWIF)
1.	IITG	1	1	1	1	1
2.	IITP	1	1	1	1	1
3.	IITRPR	0.1666667	0.1666667	0.1666667	0.166666667	2
4.	IITH	0.0454545	0.0454545	0.0416667	0.044191919	3
5.	IITM	0.0092421	0.0191205	0.0191939	0.015852154	4
6.	IITBHU	0.016129	0.016129	0.0136986	0.015318898	5
7.	IITKGP	0.0162037	0.0140187	0.00883	0.013017472	6
8.	IITMANDI	0.0059172	0.0062112	0.0062112	0.006113173	7
9.	IITR	0.001996	0.0021882	0.0018215	0.002001895	8
10.	IITD	0.0005495	0.0007246	0.0009346	0.000736223	9
11.	IITB	0.0004783	0.000534	0.0002451	0.000419113	10
12.	IITK	0.0001653	0.0002256	0.0001739	0.000188255	11
13.	IITGN	0	0	0	0	12
14.	IITI	0	0	0	0	12
15.	IITJ	0	0	0	0	12
16.	IITBBS	Data not collected due to non availability of library website during the period of data collection				

From Table 4.6, it has been observed that among all IITs library websites IIT Guwahati & IIT Patna ranked first having the highest average Revised WIF (i.e. 1). IIT Ropar and IIT Hyderabad ranked 2nd and 3rd with Revised WIF 0.166666667 and 0.044191919 respectively. IIT Madras and IIT BHU ranked 4th and 5th among all

respectively. IIT Kharagpur ranked 6th followed by IIT Mandi ranked 7th, IIT Roorkee ranked 8th, IIT Delhi ranked 9th, IIT Bombay ranked 10th, and IIT Kanpur ranked 11th. IIT Gandhinagar, IIT Indore and IIT Jodhpur have zero RWIF resulted zero average Revised WIF and all ranked 12th among all of the IITs library websites. IIT Bhubaneswar could not be included in the study due to lack of library websites during data collection. From the above ranking and RWIF data for three rounds, it has been inference that *inlinks* of the library websites remains constant during the whole study period and the results based on the Revised WIF proves that IITs library websites are having a very low visibility on the Web as measured from the link analysis.

4.5 Link Pattern among IITs Library Websites

The link pattern among IITs library websites have been analysed by collecting the data by visiting all the IITs library websites manually from each web page presents in the library websites. The following Table 4.7 is displaying the number of outlinks (outgoing) from one IIT library website to other IITs library websites and number of *inlinks* to one IIT library website from other IITs library websites.

Table 4.7: Link Pattern among IITs Library Websites

SN	**Names of IITs** (in coded form)	**No. of Outlinks to other IITs**	**No. of Inlinks from other IITs**
1.	IITB	0	4
2.	IITBBS	0	1
3.	IITBHU	0	1
4.	IITD	0	5
5.	IITG	0	4
6.	IITGN	0	1
7.	IITH	15	0
8.	IITI	0	1
9.	IITJ	1	1
10.	IITK	0	3
11.	IITKGP	0	3
12.	IITM	6	2
13.	IITMANDI	0	1
14.	IITP	0	1
15.	IITR	5	4
16.	IITRPR	3	1

From the Table 4.7 it has been found that IIT Hyderabad has outlinked to 15 other IITs library websites which highest in the number among all IITs library websites. Then after IIT Madras has 6 outlinks followed by IIT Roorkee (5), IIT Ropar (3) and IIT Jodhpur (1). Surprisingly, IITs library websites are not linking to each other except five IITs library websites. On the other side, except IIT Hyderabad, all the IITs library websites are having at least one *inlink* from other IITs library websites. It is amazing that most outlinks producing IIT library websites has no *inlinks* from any IIT library. IIT Delhi received 5 *inlinks* which is the highest among all IITs library websites. From the following Figure 4.1, we can see the link pattern among all IITs library websites.

Fig 4.1: Link Pattern among IITs Library Websites

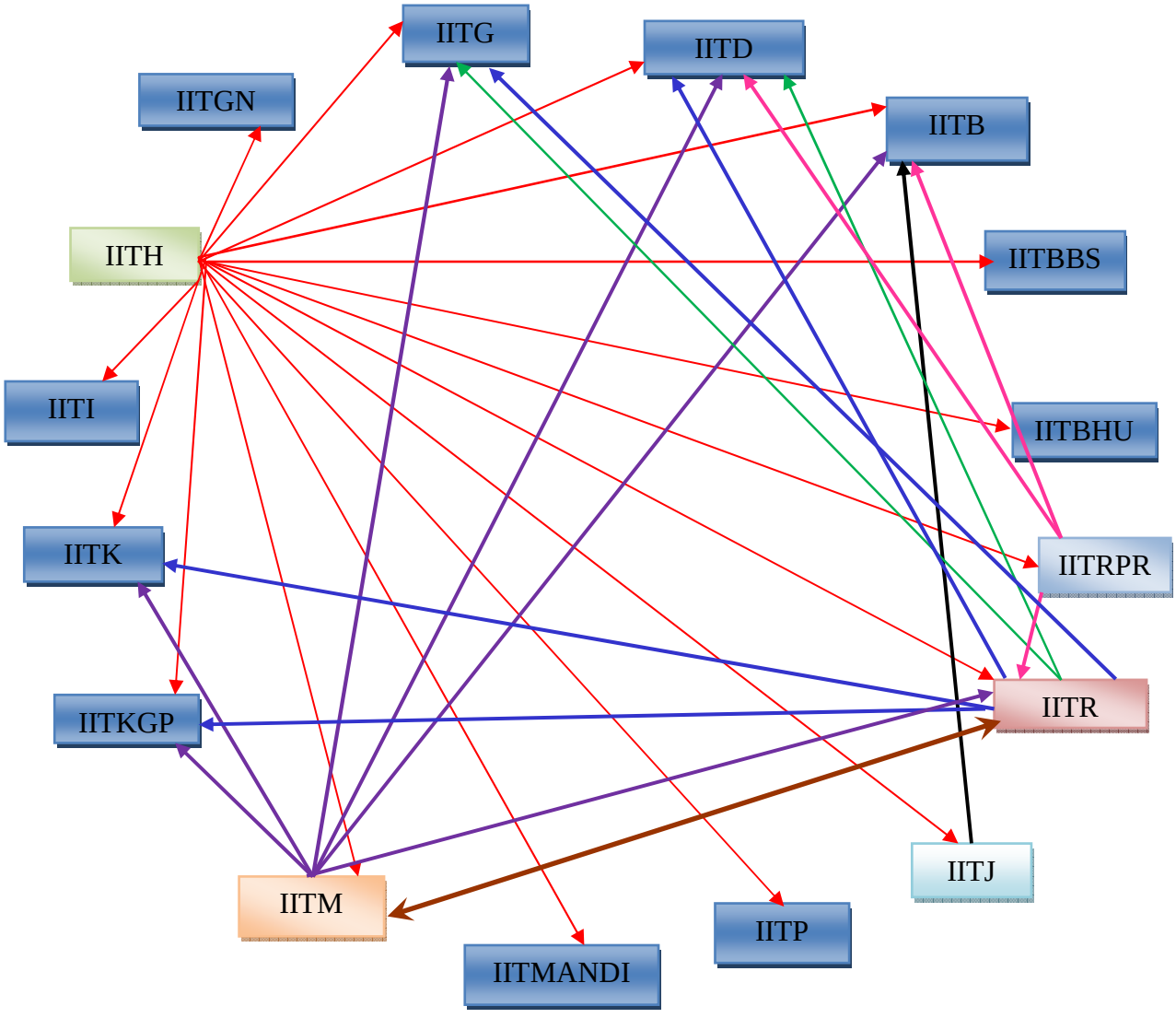


Figure 4.1 represents link pattern among IITs library websites. Maximum outlinks are originated from IITH to all other IITs library websites while none of the IITs library websites has linked to IITH. There is only one reciprocal link pattern among all library websites which is established between IITM and IITR. Majority of the library websites are having only one way link pattern. Rectangular dark bluish shades of IITs represent the group of IITs which do not have outgoing links to other IITs library websites. The IITs which have inlinks is depicted by arrow towards them. IITB and IITD library websites have established more number of incoming links among all of the library websites.

4.6 Search Engine Performance in Webometric Research

The search engine performance is one of the important criterions in webometric research for evaluating the reliability of the search engines taken for the study. Table 4.8 shows here the differences in performance of search result by using same search expression without gap after command and with gap after command. The search expressions used for this study are given below:

Search expression (without gap)	**Search expression (with gap)**
site:www.iitbhu.ac.in/library/	site: www.iitbhu.ac.in/library/

Table 4.8: Search Engine Performance in Webometric Data Collection
(with site command **site:URL** and **site: URL**)

Search Expression (without gap)	**No. of Hits**	**Search Expression** (with gap)	**No. of Hits**
site:www.iitbhu.ac.in/library/	65	site: www.iitbhu.ac.in/library/	101000
site:library.iitbbs.ac.in:……	1190000	site: library.iitbbs.ac.in:…….	276000
site:www.library.iitb.ac.in/	18800	site: www.library.iitb.ac.in/	5810000
site:library.iitd.ac.in/	952	site: library.iitd.ac.in/	215000
site:www.iitgn.ac.in/…….	1	site: www.iitgn.ac.in/.....	16000
site:www.iitg.ernet.in/lib/	33	site: www.iitg.ernet.in/lib/	21000000
site:library.iith.ac.in/	23	site: library.iith.ac.in/	109000
site:library.iiti.ac.in/	46	site: library.iiti.ac.in/	51400
site:library.iitk.ac.in/	9680	site: library.iitk.ac.in/	284000
site:www.library.iitkgp.ernet.in/	1730	site: www.library.iitkgp.ernet.in/	83100
site:www.cenlib.iitm.ac.in/	505	site: www.cenlib.iitm.ac.in/	112000
site:www.iitmandi.ac.in/……	162	site: www.iitmandi.ac.in/.....	42000
site:www.iitp.ac.in/index.php…….	1	site: www.iitp.ac.in/index.php...	587
site:library.iitj.ac.in/	19	site: library.iitj.ac.in/	52600
site:mgcl.iitr.ac.in/	510	site: mgcl.iitr.ac.in/	59300
site:www.iitrpr.ac.in/library	6	site: www.iitrpr.ac.in/library	80500

From Table 4.8, it has been observed that search expressions with gap are having more number of hits in comparison to search expression without gap. Though logically, search expressions without gap are correct. The command site:URL means the number of all the pages inside the URL. There is huge variation in data given by the same search engine for the same search expression. So, utmost care should be taken for such kind of study.

Table 4.9: Search Engine Performance in Webometric Data Collection
(with link command **link:URL** and **link: URL**)

Search Expression (without gap)	**No. of Hits**	**Search Expression** (with gap)	**No. of Hits**
link:www.iitbhu.ac.in/library/	6	link: www.iitbhu.ac.in/library/	74000
link: library.iitbbs.ac.in:….	2	link: library.iitbbs.ac.in:….	27200
link:www.library.iitb.ac.in/	77	link: www.library.iitb.ac.in/	5570000
link:library.iitd.ac.in/	46	link: library.iitd.ac.in/	22900
link:www.iitgn.ac.in/library.htm	5	link: www.iitgn.ac.in/library.htm	16000
link:www.iitg.ernet.in/lib/	6	link: www.iitg.ernet.in/lib/	8,34,00,000
link:library.iith.ac.in/	1	link: library.iith.ac.in/	18900
link:library.iiti.ac.in/	17	link: library.iiti.ac.in/	23900
link:library.iitk.ac.in/	4	link: library.iitk.ac.in/	119000
link:www.library.iitkgp.ernet.in/	43	link: www.library.iitkgp.ernet.in/	157000
link:www.cenlib.iitm.ac.in/	83	link: www.cenlib.iitm.ac.in/	102000
link:www.iitmandi.ac.in/……	2	link: www.iitmandi.ac.in/.....	35300
link:www.iitp.ac.in/index.php/…	30	link: www.iitp.ac.in/index.php/....	548
link:library.iitj.ac.in/	0	link: library.iitj.ac.in/	25800
link:mgcl.iitr.ac.in/	5	link: mgcl.iitr.ac.in/	112000
link:www.iitrpr.ac.in/library	4	link: www.iitrpr.ac.in/library	76300

Table 4.9 shows the same kind of result with link command for all the library websites. The command link:URL means the total number of web pages which have links to given URL.

4.7 Findings

The findings of the study are presented based on objectives:

1) From the URL analysis results, the study finds that most of the library websites are using the Top Level Domain (TLD) .ac.in where .ac (gTLD) is

used by the 15 IIT libraries and .ernet is used by IIT Kharagpur only. The country code TLD (ccTLD) .in is used by all of the IITs library websites.

2) Based on the type of file formats supported by the IITs library websites, IIT Roorkee used 8 kinds of file formats for representing the library information over the Web. IIT BHU and IIT Mandi are using very less number of file formats (3 file formats). IIT Kharagpur is only the library which is using audio/visual file format (WMA). All of the library websites under study are using HTML which is by default web page designing language. It has been seen that IITs library websites are not very much using the various file formats to display the knowledge and information content of the libraries.

3) Based on the three rounds of Revised WIF data, study finds that IITs library websites have more or less satisfied indexed web pages in search engine "Google" ranging from 0 to 26000. The newly established IIT Bhubaneswar library website has no any indexed web pages in the Google's database. This library websites couldn't be accessible during the whole study period due to which not included in RWIF calculation. IIT Bombay library website has the highest *inlinks* and *links* during all three rounds of data collection while overall ranking is 10^{th} among all. This is due to higher the number of indexed web pages than *inlinks*. IIT Guwahati and IIT Patna ranked 1^{st} among all 16 IITs library websites in each round due to equal number of *inlinks* with equal number of indexed web page which resulted highest RWIF in each round. In the third round of data collection, IIT Jodhpur appeared with one *inlink* and one indexed web page which resulted highest RWIF and 1^{st} rank with other two IITs library websites (IITG & IITP) among all 15 IITs library websites included in the RWIF calculation. IIT Jodhpur library website ranked 12^{th} in first and second round of data collection with zero *inlink*.

4) For the link pattern among all of IITs library websites, it has been found that there are 5 IITs library websites which are linking to others IITs library websites and IIT Hyderabad has the linked to rest of the IITs library websites. The remaining 11 IITs library websites does not have any outlink to other IITs library websites. Due to IIT Hyderabad library website's linking behaviour; all of the IITs

library websites have *inlinks* to their websites except IIT Hyderabad. IIT Delhi library websites has received maximum 5 *inlinks* from other IITs library websites. This link pattern shows the paths of links among the IITs library websites and clearly reflects the sparse linking among the IITs libraries websites under study. It has been inference that IITs library websites are not well linked to each other.

5) The study finds that the search engines performance are not always reliable for webometric studies. Before conducting webometric study by researchers, search engines and search expressions should be tested carefully. Till now there is no search engine which can give 100% satisfactory result but there is no other way to conduct such kind of webometric research. In a study conducted by Lawrence & Giles (1998), it has been found no search engine indexes more than 16% of the Web.

References:

Islam, A. (2011). Webometrics study of universities in Bangladesh. *Annals of Library and Information Studies, 58*(2), 307-318.

Jalal, S. K., Biswas, S. C., & Mukhopadhyay, P. (2010). Web impact factor and link analysis of selected Indian Universities. *Annals of Library and Information Studies, 57*(1), 109-121.

Jeyshankar, R., & Babu, B. R. (2009). Websites of universities in Tamil Nadu: A webometric study. *Annals of Library and Information Studies, 56*(1), 69-79.

Shukla, A. (2009). *Evaluating Indian Academic and Research Library Websites based on Web Impact Factor Analysis* (*Doctoral dissertation).* Varanasi, India: Department of Library & Information Science, Banaras Hindu University.

Shukla, A., & Tripathi, A. (2009). Webometric analysis of institutes of national importance in India. *IASLIC Bulletin, 54*(3), 165-180.

Sujithai, M., & Jeyshankar, R. (2013). Web page analysis of Indian Institute of Technologies' (IITs) websites: A webometric study. *International Journal of Digital Library Services, 3*(1), 55-65.

Vijayakumar, M. (2012). Webometric analysis of university websites in Srilanka. *International Journal of Information Dissemination and Technology*, *2*(3), 155-159.

Walia, P. K., & Kaur, P. (2008). Webometric analysis of Library Associations' websites of India. *IASLIC Bulletin, 53*(3), 131-143.

CHAPTER V

CONCLUSION & SUGGESTIONS

5.1 Conclusion

From the analysis and findings of the study, there exists a high percentage on the use of Top Level Domains (TLDs) by the IITs library websites in which the TLD .ac.in is used by 88.23%. Every IIT library website contains different file formats and from the selected file formats for the study the HTML file is used in all of IITs library websites for representing their content. The average Revised WIF of the three rounds of data has given the result that IITs library websites are not indexed in search engine's database sufficiently. IITs library websites which have highest number of indexed pages but less number of *inlinks* to the library websites leads to lower WIF and RWIF. Those library websites which do not have sufficient number of indexed pages with less number of *inlinks* have higher the WIF and RWIF value. There is no correlation found that higher the number of indexed web pages leads to higher WIF and RWIF. IITs library websites which were very new to the origin had higher WIF & RWIF than well established IITs library websites.

Further, the link distribution of IITs library websites has established the relationships between them. This study reflects that IITs library websites have higher number of web pages but correspondingly their linking behavior to others websites are very poor. The lower the presence of IITs library websites on the Web was due to the structural problems in library websites, limitation of access to the scientific resources, the instability of the web servers, less indexing of library websites in Web directories and lack of institutional repositories. The command textual query enables to give a clear understanding of the performance of the search engine and the main drawback is that they do not cover the entire Web or even the entire publicly indexable Web as Lawrence & Giles (1998) proves that only 16% of the Web is indexable. Search engine results have been found to be unreliable and to fluctuate, even over short periods of time also.

5.2 Suggestions

The study has investigated the IITs library websites from which the following suggestions are given:

1. Most of the IITs library websites are not structured properly and in some cases they are having only single web page for their library websites. This should be well structured and should be maintained by the webmasters and librarians.

2. The file formats are very important for giving out information to the users and since the IITs library websites supports these file formats less in number, it is suggested that all kind of file formats should be used for displaying the knowledge and information.

3. In order to increase the visibility of the IITs library websites, webmasters should develop and generate *outlinks* to more and more library websites.

4. IITs library websites link pattern shows how one library links to each other. The number of *outlinks* and *inlinks* are very less in number and the links are very sparse. Hence this needs improvement for each of the IITs library websites by providing links to each other which can be helpful for their users and this will result to increase in the number of *outlinks* and *inlinks*.

The title of the research has been worthy to undertake and hopes that it will pave a way for further webometric studies. This study can be useful for other researchers in webometrics area as well as the institutes under study so that more improvements can be made for the benefit of the library users through their library websites.

DECLARATION

I, Vanlalfeli, hereby declare that the dissertation entitled "**LIBRARY WEBSITES OF INDIAN INSTITUTES OF TECHNOLOGY (IITs): A WEBOMETRIC STUDY"** is the result of the work done by me, the contents of this dissertation did not form the basis for the award of any degree to me or to anybody else to the best of my knowledge and the dissertation has not been submitted by me for any research degree in any other University or Institute.

Date: 19/12/2013 (VANLALFELI)

MIZORAM UNIVERSITY

(A Central University)
Department of Library and Information Science
Tanhril, Aizawl – 796 004

Gram – MZU, P.O. Box – 190, Phone – (0389) 2331607, Fax – 2331608
Email: akhandanandshukla@gmail.com

Dr. Akhandanand Shukla
Assistant Professor

CERTIFICATE

This is to certify that, Ms. **Vanlalfeli** has completed the dissertation entitled **"LIBRARY WEBSITES OF INDIAN INSTITUTES OF TECHNOLOGY (IITs): A WEBOMETRIC STUDY"** for awarding the degree of **Master of Philosophy** in **Library and Information Science** under my supervision. This is the candidate's original work and worthy of examination.

Tanhril, Aizawl
19th December, 2013

(Dr. Akhandanand Shukla)
Supervisor

Countersigned by:

Head of the Department
Dept. of Library and Information Science
Mizoram University

ACKNOWLEDGEMENT

First of all, I thank the Almighty God for all his guidance in the completion of my dissertation work.

Secondly, I sincerely thank my supervisor **Dr. Akhandanand Shukla** for all his support, guidance and encouragement right from the start of this research till the completion of the work.

I express my deepest gratitude to **Prof. R. K. Ngurtinkhuma,** Head of Department, **Prof. Pravakar Rath, Dr. S. N. Singh, Dr. R. N. Mishra, Dr. Manoj Kumar Verma, Dr. Lalngaizuali, Mr. Amit Kumar**, all the staffs and my colleagues of the Department of Library & Information Science.

Last but not least, I thank my parents and all my family members for their endless love, support and encouragement to complete this dissertation work.

Tanhril, Aizawl
19th December, 2013.

(Vanlalfeli)
Dept. of Library & Information Science,
Mizoram University.

PREFACE

Research activities are important for the development of a nation. These research activities are carried out in different areas especially in an academic arena. In Library and Information Science various researches on metric studies such as Librametrics, Bibliometrics, Scientometrics, Informetrics and Webometrics has been conducted. These metric studies play a vital role especially for measuring scholarly communication.

For the present study the research area of Webometrics has been selected. Webometric covers research of all network based communication using informetric or other quantitative. For conducting the present research, the Indian Institutes of Technology (IITs) library websites have been selected. These institutions are declared as *Institutes of National Importance* where teaching and research activities are carried out are influential for the promotion of the advanced technologies in national and international levels.

The present study is therefore, important because a library website is one source of disseminating information to the users on the web. The data for the study is collected by using search engine and the links which are present in the library websites are studied in detail using the Web Impact Factor (WIF) which helps in ranking of the institutions library websites. Hence, from this study it is hoped that the findings will prove to be inclusive and produced useful results for the improvements of the IITs library websites.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations

Term	Description
ASCII	American Standard Code of Information Interchange
ccTLD	Country Code Top Level Domain
CERN	European Organization for Nuclear Research
DNS	Domain Name System
FID	Federation Internationale de la Documentation
FTP	File Transfer Protocol
gTLD	Generic Top Level Domain
GUI	Graphical User Interface
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
IANA	Internet Assigned Numbers Authority
ICANN	Internet Corporation for Assigned Names and Numbers
ICT	Information and Communication Technology
IDN	ccTLD Internationalized Country Code Top Level Domain
IDNA	Internationalizing Domain Names on Applications
IIT	Indian Institute of Technology
IP	Internet Protocol
ISO	International Organization for Standardization
JIF	Journal Impact Factor
LIS	Library and Information Science
MHRD	Ministry of Human Resource Development
NPTEL	National Program on Technology Enhance Learning
PDA	Personal Digital Assistance
RWIF	Revised Web Impact Factor
SLD	Second – Level Domain
STI	Science and Technical Information
TLD	Top Level Domain
URI	Uniform Resource Identifier
URL	Uniform Resource Location
WIF	Web Impact Factor
WWW	World Wide Web
XML	Extensible Markup Language

.***Bibliography:***

[Based on Publication Manual of American Psychological Association (6th ed.) with some modifications]

Aguillo, I. F. (2012). Is Google Scholar useful for bibliometrics? A webometric analysis. *Scientometrics, 91*, 343-351. doi: 10.1007/s11192-011-0582-8

Aguillo, I.F. et al., (2006). Scientific research activity and communication measured with cybermetrics indicators, *Journal of the American Society for Information Science and Technology,* 57(10), 1296–1302.

Almind, T. C., & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: Methodological approaches to webometrics. *Journal of Documentation, 53*(4), 404-426.

Arif, A., & Ismail, N. A. (2013). Web impact factor for Malaysian public universities. *International Journal of Future Computer and Communication, 2*(3), 151-154.

Asadi, M., & Shekofteh, M. (2009). The relationship between the research activity of Iranian medical universities and their web impact factor. *Electronic Library, 27*(6), 1026-1043. Retrieved 3rd March, 2013, from http://dx.doi.org/10.1108/02640470911004101.

Babu, B. R., Jeyshankar, R., & Rao, P. N. (2010). Websites of Central Universities in India: A webometric analysis. *DESIDOC Journal of Library & Information Technology, 30*(4), 33-43.

Bar-Ilan, J. (2000). Evaluating the stability of the search tools HotBot and Snap: a case study. *Online Information Review, 24*(6), 439-449.

Berners-Lee, T. & Cailliau, R. (1990). *World Wide Web: proposal for a hypertext project*. Available at http://www.w3.org/Proposal.html

Björneborn, L. (2004). *Small-world link structures across an academic web space: a library and information science approach (Doctoral Dissertation).* Copenhagen, Denmark: Royal School of Library & Information Science.

Björneborn, L., & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics, 50*(1), 65-82.

Björneborn, L., & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics, 50*(1), 65-82.

Björneborn, L., & Ingwersen, P. (2004). Towards a basic framework for webometrics. *Journal of the American Society for Information Science and Technology, 55*(4), 1216-1227.

Boell, S. K., Wilson, C. S., & Cole, F. T. H. (2008). Usage of different web impact factors for ranking Australian universities. *Collnet Journal of Scientometrics and Information Management*; *2*(2); 57-70.

Borgman, C. & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, *36*(1), 2-72.

Borgman, C.L. (2000). Scholarly communication and bibliometrics revisited. In Cronin, B. & Atkins, H.B. (eds.) *The web of knowledge – A festschrift in honor of Eugene Garfield*. ASIS, 2000.

Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, *30*, 107-117. Retrieved 12 November, 2013, from CiteSeerX: http://citeseerx.ist.psu.edu/viewdoc/ summary?doi =10.1.1.42.3243

Brookes, B. C. (1991). Biblio-, sciento-, infor-metrics??? What are we talking about? In *Informetrics 89/90;* Selection of Papers Submitted for the Second International Conference on Bibliometrics, Scientometrics and Informetrics. London, Ontario..

Chu, H., He, S., & Thelwall, M. (2002). Library and Information Science Schools in Canada and USA: A webometric perspective. *Journal of Education for Library and Information Science*; 43(2); 110-125.

Comer, E.D. (1997). *Computer networks and internet*. New Jersey: Prentice-Hall Inc.

Country code top-level domain name. (n.d.). Retrieved September 18, 2013, from Wikipedia: http://en.wikipedia.org/wiki/*Country_code_top_level_domain _name*

Domain name. (n.d.). Retrieved September 18, 2013, from Wikipedia: http://en.wikipedia.org/wiki/ *Domain _name*

Cronin, B. (2001). Bibliometrics and beyond: some thoughts on web-based citation analysis. *Journal of Information Science, 27*(1), 1-7.

Easley, D. & Kleinberg, J. (2010). Link Analysis and Web Search. In Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Chapter 14. Cambridge University Press. Retrieved 18 October, 2013, from http://www.cs.cornell.edu /home /kleinber/networks-book/

Frías, E. R. (2010). Application of webometric techniques to the study of accounting and financial variables. Doctoral Dissertation. University of Granada. Retrieved 12 March, 2013, from http://estebanromero.com/wp-content/uploads/2010/02/thesis-English-version-4-3-2010-DEF.pdf

Goel, M. K. (2010). *World Wide Web.* New Delhi: Rajat Publications.

Goswami, P., Sharma, U., & Shukla, A. K. (2008). The Webometrics. *6th International CALIBER-2008, University of Allahabad, Allahabad, February 28-29 & March 1, 2008*, 656-660. Retrieved 11 March, 2013, from http://ir.inflibnet.ac.in/bitstream/handle/1944/1318/70.pdf

Hatua, S.R. (2011). Auto web browser: an alternative to Google. *IASLIC Bulletin, 56*(1), 3-10.

Henzinger, M. (2000). Link analysis in web information Retrieval. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 23*(3)**,** 1-7. Retrieved 11 March, 2013, from citeseerx.ist.psu.edu/viewdoc /download?doi=10.1.1.32.4054&rep=rep1&type=pdf#page=5

History of the World Wide Web. (n.d.). Retrieved 12 March, 2013, from Wikipedia: http://en.wikipedia.org/wiki/History_of_the_World_Wide_Web

Holmberg, K. (2009). *Webometric network analysis : mapping cooperation and geopolitical connections between local government administration on the web.* (Doctoral Dissertation). Åbo Akademi University Press. Åbo Akademi University. Retrieved 15 March, 2013, from http://urn.fi/URN:ISBN:978-951-765-511-8

Hyperlink. (n.d.). Retrieved 29 October, 2013, from Wikipedia: http://en.wikipedia.org/wiki/Hyperlink

Indian Institutes of Technology. (n.d.). Retrieved 4 April, 2013, from Wikipedia: http://en.wikipedia.org/wiki/Indian_Institutes_of_Technology

Ingwersen, P. (1998). The calculation of web impact factors. *Journal of Documentation, 54*(2), 236-243.

Islam, A. (2011). Webometrics study of universities in Bangladesh. *Annals of Library and Information Studies, 58*(2), 307-318.

Islam, A., & Alam, S. (2011). Webometric study of private universities in Bangladesh. *Malaysian Journal of Library & Information Science, 16*(2), 115-126.

Islam, M. A. (2011). Webometrics study of universities in Bangladesh. *Annals of Library and Information Studies, 58*(2), 307-318.

Jalal, S. K., Biswas, S. C., & Mukhopadhyay, P. (2010). Web impact factor and link analysis of selected Indian Universities. *Annals of Library and Information Studies, 57*(1), 109-121.

Jati, H. (2011). Web impact factor: a webometric approach for Indonesian Universities. *International Conference on Informatics for Development (ICID, 2011),C1*, 71-74. Retrieved 27 May, 2013, from *http://staff.uny.ac.id/.../C1_74-77*

Jeyshankar, R. (2011). Link analysis and Web Impact Factor of Indian nationalised banks. *International Journal of Information Dissemination and Technology, 1*(3), 171-179.

Jeyshankar, R., & Babu, B. R. (2009). Websites of universities in Tamil Nadu: a webometric study. *Annals of Library and Information Studies, 56*(1), 69-79.

Jeyshankar, R., Sujitha, I. M., & Valarmathi, A. (2012). Web pages of ICMR institutes websites: A webometric analysis. *Global Advanced Research Journal of Library, Information and Archival Studies, 1*(1), 6-18.

Kaliammal, A., & Tharmaraiselvi, G. (2008). Role of ICTs in Library and Information Science. Delhi: Authorpress.

Kannan, K., & Abilash, S. (2011). Information searching pattern of web search engine: a comparative study. *Library Progress (International), 31*(1), 35-42.

Kessler, M.M. (1963). Bibliographic coupling between scientific papers. *American Documentation, 14*(1), 10-25.

Kobayashi, M., & Takeda, K. (2000). Information retrieval on the web. *ACM Computing Surveys, 32*(2), 144-173.

Kumar, A. M. N., Babu, B. R., Gopalakrishnan, S., & Rao, P. N. (2010). Location of web objects and links in the web sites of universities in Tamil Nadu: a study. *SRELS Journal of Information Management, 47*(3), 325-338.

Kumar, A. M. N., Babu, B. R., Gopalakrishnan, S., & Rao, P. N. (2010). Location of web objects and links in the web sites of universities in Tamil Nadu: a study. *SRELS Journal of Information Management, 47*(3), 325-338.

Kumar, S.S., & Rabindra, M. (2012). Web impact factor and webometric study of National Institute of Technology (NITs) of India. *Pearl: A Journal of Library and Information Science, 6*(1), 47-55.

Kumar, S.S., & Rabindra, M. (2012). Web impact factor and webometric study of National Institute of Technology (NITs) of India. *Pearl: A Journal of Library and Information Science, 6*(1), 47-55.

Larson, R.R. (1996). Bibliometrics of the World Wide Web: an exploratory analysis of the intellectual structure of cyberspace. In: Hardin, S. (Ed.), *Proceedings of the 59th Annual Meeting, ASIS 96*. Baltimore, pp. 71-79.

Lawani, S. M. (2009). Bibliometrics: Its Theoretical Foundations, Methods and Applications. *Libri. 31*(1), 294-315. doi: 10.1515/libr.1981.31.1.294

Lawrence, S. & Giles, C. L. (1998). Searching the World Wide Web. *Science, 280*, 98-100.

Li, X. (2003). A review of the development and application of the Web impact factor. *Online Information Review, 27*(6), 407-417.

Link. (n.d.). Retrieved September 18, 2013, from Wikipedia: http://en.wikipedia.org/wiki/link

Maharana, R. K., Panda, K.C., & Sahoo, J. (2012). Web impact factor (WIF) and link analysis of Indian Institute of Technologies (IITs): A webometric study. *Library Philosophy and Practice (e-journal). 789*, 1-11. Retrieved 11 May, 2013, from http://digitalcommons.unl.edu/libphilprac/789.

Mandl, Thomas. (2007). The impact of web site structure on link analysis. *Internet Research, 17*(2), 196-206.

Mayr, P., & Tosques, F. (2005). *Google Web APIs: an Instrument for Webometric Analyses?* Available at: www.ib.hu-berlin.de/%7Emayr/arbeiten/ISSI2005 _Mayr_ Toques.pdf

Ministry of Human Resource Development (MHRD), India. (n.d.). *Department of Higher Education, IIT*. Retrieved April 9, 2013, from http://mhrd.gov.in/iit_hindi

Microsoft Research. (n.d.) Website structure understanding and its applications. Retrieved 4 May, 2013, from http://research.microsoft.com

Mukherjee, B. (2011). Bibliometrics to webometrics: The changing context of quantitative research. *IASLIC Bulletin, 56*(2), 97-110.

Nair, R.R. (2002). *Internet for information services*. New Delhi: Ess Ess Publications.

National Knowledge Commission. (2007). *Libraries: Gateways to Knowledge*. New Delhi: National Knowledge Commission, Govt. of India.

Noruzi, A. (2005). Editorial: Fundamental differences between hyperlinks and citations. *Webology, 2*(2), editorial 4. Available at: http://www.webology.org /2005/v2n2/editorial4.html

Noruzi, A. (2005). Web Impact factors for Iranian Universities. *Webology*, *2*(1).

Noruzi, A. (2006). The web impact factors: A critical review. *Electronic Library, 24*(4), 490-500.

Nwagwu, W. E., & Agarin, O. (2008). Nigerian University Websites: A Webometric Analysis. *Webology*, *5*(4), Article 62.

Onyancha, O. B. (2007). A Webometric study of selected academic libraries in Eastern and Southern Africa using link analysis. *South African Journal of Libraries and Information Science*, *73*(1), 25-39. Retrieved 11 March, 2013, from http://uir.unisa.ac.za/handle/10500/5234?show=full

Organisation for Economic Co-operation and Development (OECD). (2006). *Evolution in the management of country code top level domain names (ccTLDs).* Report to Working Party on Communication Infrastructures and Services Policy (ICCP). 1-49. Retrieved on 12 August, 2013, from, http://www.oecd.org/sti/ieconomy/37730629.pdf

Oppenhium, C., Morris, A., McKnight, C., & Lowley, S. (2000). The evaluation of search engines. *Journal of Documentation, 56*(2), 190-211.

Park, H. W., & Thelwall, M. (2006). Web-science communication in the age of globalization. *New Media Society*, *8*(4), 629-650.

Pechnikov, A. A., & Nwohiri, A. M. (2012). Webometric analysis of Nigerian university websites. *Webology*, *9*(1), Article 96.

Pritchard, A. (1969). Statistical Bibliography or Bibliometrics?. *Journal of Documentation, 25*(4), 348–349..

Rao, I. K.R. (2010). *Growth of literature and measurement of scientific productivity.* New Delhi: Ess Ess Publication.

Rassi, M., Jalali, D. A., Mousavizadeh, Z., Abbasgholi, N., & Haghighian, R. A. (2013). Assessment of visibility and web impact factor (WIF) of Food Science & Technology Institutes websites based on webometrics scales. *Iranian Journal of Nutrition Sciences & Food Technology, 7*(5), 449-459. Retrieved September 13, 2013, from http://nsft.sbmu.ac.ir/browse.php ?acode=A-10-1-348&slc_ang=en& sid=1

Ratha, B., Joshi, L., & Naidu, G.H.S. (2012). Webometric study of IIT libraries websites. *DESIDOC Journal of Library & Information Technology, 32*(3), 249-254.

Rosa, M. (2005). Website. *Whatis.com.* Available at: http://searchsoa.techtarget. com/definition/Web-site

Samuel C. A. U. (2008). Information technology and web use characteristics of Nigerian private universities. *African Journal of Library, Archives and Information Science.* Available at: http://findarticles.com/p/articles/ mi_7002 /is_2_18/ai_n31877465/

Sangam, S.L., & Keshava. (2008). Concept of Bibliometrics, Scientometrics and Informetrics. In Amudhavalli, A. (2008). Dynamics in digital information systems. New Delhi: Ess Ess Publications. 301 - 313.

Sangam.S.L. (2008). *Areas in the field of Scientometrics and Informetrics*. In Koganuramath, M. M., Kumar, B. D., & Kademi, B.S. (eds). Library and Information Science Profession in the Knowledge Society. New Delhi: Allied Publishers Pvt. Ltd. 265-271.

Sengupta. I. N. (2009). Bibliometrics, Informetrics, Scientometrics and Librametrics: An Overview. *Libri, 42*(2), 75-98, doi: 10.1515/libr.1992.42.2.75

Sharunizam, S., Gaby, H., Paul, G. (2012). Bibliometric and webometric methods for assessing research collaboration. *Library Review*, 61(8/9), 592 – 607 doi: 10.1108/00242531211292097

Shukla, A. (2009). *Evaluating Indian Academic and Research Library Websites based on Web Impact Factor Analysis* (*Doctoral dissertation).* Varanasi, India: Department of Library & Information Science, Banaras Hindu University.

Shukla, A., & Tripathi, A. (2009). Webometric analysis of institutes of national importance in India. *IASLIC Bulletin, 54*(3), 165-180.

.

Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265-269.

Smith, A. G., & Thelwall, M. (2002). Web impact factors for Australasian universities. *Scientometrics, 54*(3), 363-380.

Smith, A. G. (1999). A tale of two web spaces: comparing sites using web impact factors. *Journal of Documentation, 55*(5), 577-592.

Smith, A. G. (1999). The impact of web sites: a comparison between Australasia and Latin America. *Paper Presented at INFO'99, Congreso Internacional de Informacion*, Havana, 4-8 October 1999. Available at: http://www.vuw.ac.nz/staff/alastair_smith/publns/austlat/

Smith, A. G. (1999). A tale of two Web spaces:comparing sites using web impact factors. *Journal of Documentation. 55*(5), 577-592.

Spink, A., & Jansen, B. J. (2004). *Web search: public searching of the web*. Netherlands: Kluwer Academic Publishers.

Sujithai, M., & Jeyshankar, R. (2013). Web page analysis of Indian Institute of Technologies' (IITs) websites: A webometric study. *International Journal of Digital Library Services, 3*(1), 55-65.

Tague-Sutcliffe, J. (1992). An introduction to informetrics. *Information Processing and Management, 28*(1), 1- 4.

Thanuskodi, S. (2011). Webometric analysis of private engineering college websites in Tamil Nadu. *J Communication, 2*(2), 73-81.

Thanuskodi, S. (2012). A webometric analysis of selected institutes of national importance websites in India. *International Journal of Library Science, 1*(1), 13-18.

Thelwall, M. (2001). Results from a Web Impact Factor crawler. Journal of Documentation, *57*(2), 177-191.

Thelwall, M. (2002). A comparison of sources of links for academic Web Impact Factor calculations. *Journal of Documentation, 58*(1), 68-78. doi: 10.1108/002204102104245412

Thelwall, M. (2007). Bibliometrics to Webometrics. *Journal of Information Science, 34*(4), 1-18.

Thelwall, M. (2010). Webometrics: emergent or doomed? *Information Research, 15*(4). colis713. Available at http://InformationR.net/ir/15-4/colis713.html

Thelwall, M. (2011). A comparison of link and URL citation counting. *Aslib Proceedings, 63*(4), 419 - 425.

Thelwall, M., Vaughan, L., & Bjorneborn, L. (2006). Webometrics. *Annual Review of Information Science and Technology, 39*(1), 81-135.

Thomas, O., & Willet, P. (2000). Webometric analysis of departments of librarianship and information science. *Journal of Information Science, 26*(6), 421- 428. doi: 10.1177/01 6555150002600605

Turnbull, D. (1996). *Bibliometrics and the World Wide Web.* (Technical Report FIS-12-19-1996-1). Faculty of Information Studies, University of Toronto. Available at http://www.ischool.utexas.edu /~donturn/research/bibweb.html

Vaughan, L., & Shaw, D. (2003). Bibliographic and web citation: What is difference?. *Journal of the American society for information sciences and technology, 54*(14),1313-1322.

Vijayakumar, M. (2012). Webometric analysis of university websites in Srilanka. *International Journal of Information Dissemination and Technology, 2*(3), 155-159.

.

Walia, P. K., & Kaur, P. (2008). Webometric analysis of Library Associations' websites of India. *IASLIC Bulletin, 53*(3), 131-143.

Web browser. (n.d.). Retrieved 29 October, 2013, from Wikipedia: http://en.wikipedia.org/wiki/web_browser

Webometrics. (n.d.). Retrieved 29 October, 2013, from Wikipedia: http://en.wikipedia.org/wiki/Webometrics

Web page. (n.d.). Retrieved 29 October, 2013, from Wikipedia: http://en.wikipedia.org/wiki/Web_page

Web search engine. (n.d.). Retrieved 29 October, 2013, from Wikipedia: http://en.wikipedia.org/wiki/Search_engines

World Wide Web. (n.d.). Retrieved 17 October, 2013, from Wikipedia: http://en.wikipedia.org/wiki/World_Wide_Web

Xing, Y., & Chu, H. (2006). Hyperlinking to academic websites: salient features examined. Unspecified. (Unpublished), 1-16. Retrieved 27 May, 2013, from http://eprints.rclis.org/10398/ 1/Hyperlink-Xing%26Chu.pdf

Xue, S. (2004). Web usage statistics and web site evaluation: a case study of a government publications library web site. *Online Information Review, 28*(3),180-190.

Yang, B., & Qin, J. (2008). Data Collection System for Link Analysis. *IEEE;* 247-252.

Yang, B., & Qin, J. (2008). Data Collection System for Link Analysis. *IEEEXPLORE: Third International Conference on Digital Information Management (ICDIM), 2008*, 247-252.

Yi, K., & Jin, T. (2008). Hyperlink analysis of the visibility of Canadian library and information science school web sites. *Online Information Review, 32*(3), 325-347.

Young, M.L. (1999). *Internet: the complete reference*. New Delhi: Tata McGraw-Hill Education Private Limited.

Zeinolabedinio, M. H., Maktabifard, L. & Osareh, F. (2006). Collaboration analysis of world national library websites via webometric methods. *Proceedings on International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting, COLLNET 2006, SRDI-INIST-CNRS-LORIA Nancy, France, 10-12 May 2006*