ABHIYANTRIKI
An International Journal of Engineering & Technology
(AIJET)
Vol. 2, No. 1 (January, 2015) http://www.aijet.in/ eISSN: 2394-627X

Conceptual Approach to Webometrics: An Exploratory Study

Vanlalfeli
Professional Assistant
UGC - Academic Staff College
Mizoram University, Aizawl, India

Akhandanand Shukla*
Assistant Professor
Department of Library & Information Science
Mizoram University, Aizawl, India

Abstract

Paper reviews the development of webometrics researches conducted by scholars and depicts the origin of webometrics from bibliometrics. Relationship between metrics branches such as bibliometrics, scientometrics, informetrics, webometrics and cybermetrics has been established and discussed the boundary of each. Webometrics techniques like web link structure analysis, web page content analysis, web usage analysis, and web technology analysis elaborated. Further webometric terminology and diagrams used for understanding of webometrics terms has been explained thoroughly.

***Keywords**: Webometrics, Web Impact Factor, Web Link Analysis, Domain Names.*

****Author for correspondence*** akhandanandshukla@gmail.com

1. Introduction

A number of quantitative methods have been adopted by Library and Information managers in order to evaluate library resources and services more objectively and effectively. These quantitative methods in Library and Information Science (LIS) includes metric studies which are known as Bibliometrics in 1960's, Scientometrics in 1970's and Informetrics in the midst of 1980's. In present day with the advent of information technology, the concepts of Webometrics and Cybermetrics have emerged during the 1990's (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. (Babu & Jeyshankar, 2009). The research field of webometrics has its roots in library and information science particularly in Scientometrics and Bibliometrics. Because of this birthright, 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 Websites 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 i.e., 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).

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). It is 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" (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 et al., 2010; Frías, 2010, Shukla & Tripathi, 2009).

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. (Frías, 2010). In the above definition given by Bjorneborn & Ingwersen, the term structure denotes the features of the website. 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 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. 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). 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.

Webometrics was started by the realization that the web is an enormous document repository with many of these documents being academic-related (Almind & Ingwersen, 1997).

Webometric research includes link analysis, web citation analysis, search engine evaluation and purely descriptive studies of the web (Thelwall, 2000). One of the most visible outputs of webometrics is the ranking of world universities based upon their web sites and online impact (Aguillo, 2006).

3. Relationship between Bibliometrics, Scientometrics, Informetrics, Webometrics and Cybermetrics

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: Lotka's Law (Productivity of authors in terms of scientific papers), Bradford's Law (Scattering of articles over different journals), and 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: 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 (Walia & Kaur, 2008; Sangam, 2008).

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, 2008; Holmberg, 2009).

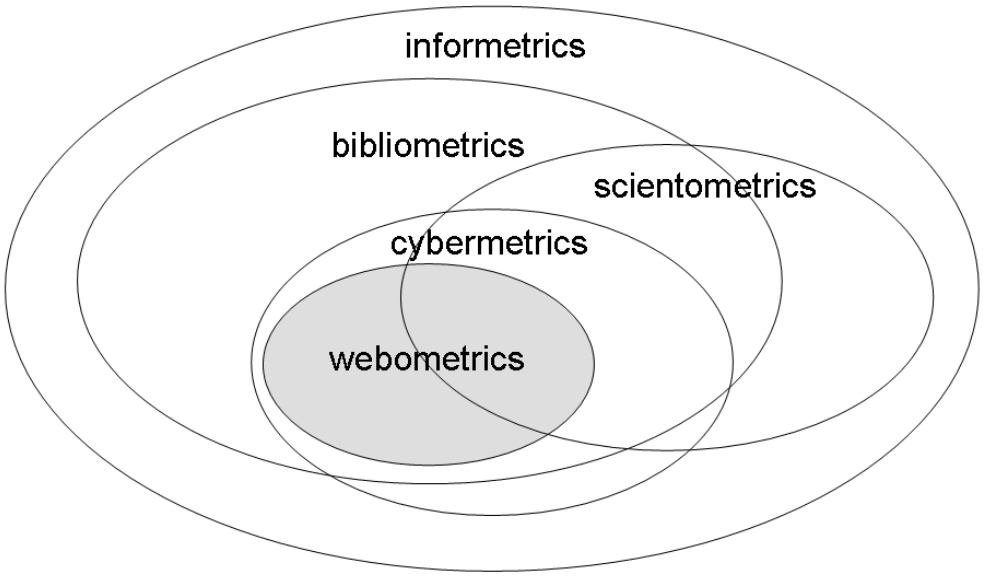


Fig. 1: Infor-, Biblio-, Sciento-, Cyber-, and Webo-metrics (Björneborn & Ingwersen, 2004)

From figure 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 (Holmberg, 2009). Björneborn & Ingwersen (2004) 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 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 (Thelwall, Vaughan & Björneborn, 2005).

4. Webometrics and its Techniques

Björneborn & Ingwersen (2004) 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).

Web Link Structure Analysis

Link analysis is the quantitative study of hyperlinks between web pages (Thelwall, 2007). 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 (Henzinger, 2000).

i) This study provides ***hyperlinks*** between documents and records of user behaviour.
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***. An incoming link is similar to receiving a citation in a document. These links are 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 (Jeyshankar & Sujithai).

The link relations between the web nodes have been described in the following fig. 2 below.

Letters A-H represents different web node levels such as web pages, web directories, websites, or top level domains of countries or generic sectors and, the arrows represent the linking pattern of these nodes among one another (Thelwall, Vaughan & Björneborn, 2005).

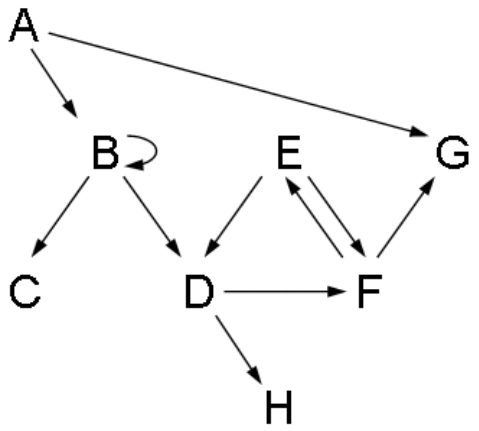


Fig. 2: Basic Webometric Link Terminology (Björneborn, 2004).

- 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 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. 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).

The two *co-linked* web nodes C and D in figure 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-link* 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 Websites. 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; Thelwall, Vaughan, & Björneborn, 2005).

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 (Walia & Kaur, 2008; Mukherjee, 2011).

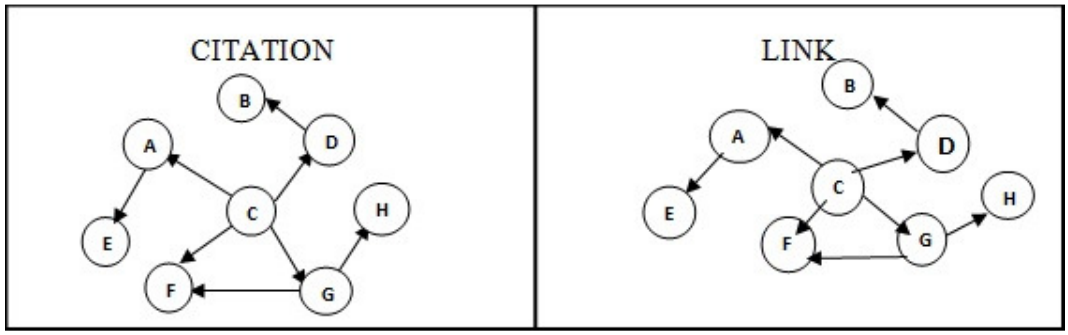


Fig. 3: Functional Relational between Citation and Link (Mukherjee, B., 2011)

Table 1: Observation on Citation and Links (Mukherjee, B., 2011)

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.** Webpage A has 2 links, one pointed to it and other emerged from it.
3. Paper C and G are bibliographically coupled.	**3.** Webpage C and G are linked to a common page.

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 Webpages 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 Websites and pages and about authors of them (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 Websites are mainly created for navigational purposes, links between Websites may be a rich source of information about the content and use of the Websites connected to each other with hyperlinks. Links between Websites 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; Holmberg, 2009).

Web Page Content Analysis

A number of webometric investigations have focused not on websites 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 (Thelwall, 2007).

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 (Jeyshankar & Sujithai).

Web Usage Analysis

It is a new type of statistical analysis of the Science and Technical Information (STI) in the web context. 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. The data recorded in server logs reflects the access of a website 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. The other is the number of times an information source is ordered; in this case we are in face of e-commerce transactions. Web usage analysis covers (Jeyshankar & Sujithai):

- 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.

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's performances (Google, AltaVista, Yahoo etc.) which incorporate more technology than other websites. 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 (Jeyshankar & Sujithai):

- 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.

5. Webometric Terminology and Diagrams

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 website, or the website 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, 2002; 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 (Comer, 1997; Wikipedia).

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 (Wikipedia).
- 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 (Wikipedia).

Types of Domain Name: Domain names are organized in sub-ordinate 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).

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; Wikipedia):

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; Wikipedia).

Internationalized Country Code Top-Level Domains (IDN ccTLD): 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; Wikipedia).

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 (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 (.).

Websites, Subsites & Sub-subsites

Zooming in on a single website may reveal several subunits in the shape of subsites, sub-subsites, and so on, as indicated by hierarchically derivative domain names. In figure 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 (Björneborn & Ingwersen, 2004).

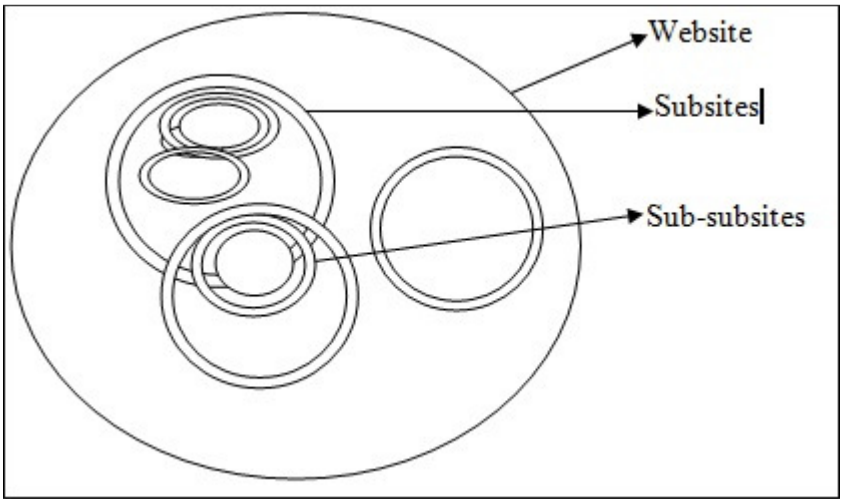
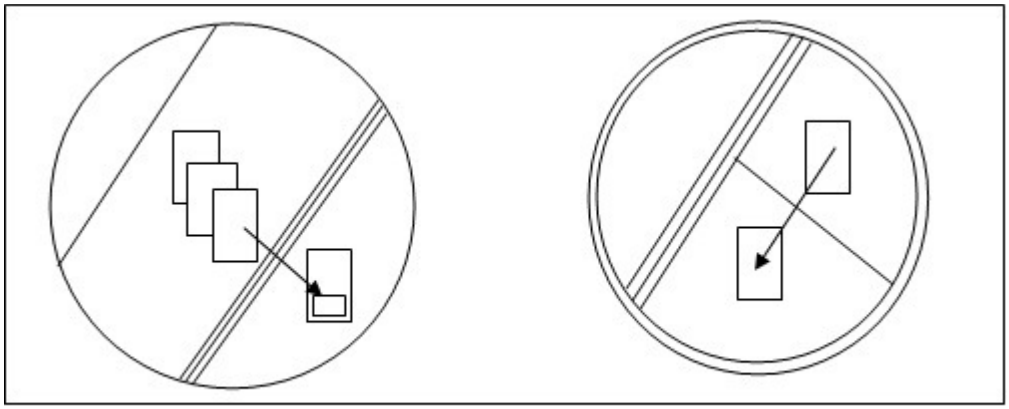


Fig. 4: Simplified Web node diagram of a Website containing subsites and sub-subsites (Björneborn & Ingwersen, 2004).

Although some websites 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 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 (Thelwall, Vaughan, & Björneborn, 2005).

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 5a. More complex and numerous

linkages within a site or sub-sites can be illustrated by combinations of elements in fig. 5, showing links between pages located either at different directory levels (fig. 5a) or in sibling directories at the same level (fig. 5b) in the website file hierarchies.



5a 5b

Fig. 5: Simplified Web node diagrams of a website and a sub-site with links between different directory levels including page sub-elements (Bjorneborn & Ingwersen, 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).

6. 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

[1] 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.

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

[3] 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.

[4] 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.
[5] Bar-Ilan, J. (2000). Evaluating the stability of the search tools HotBot and Snap: a case study. *Online Information Review*, 24(6), 439-449.
[6] Björneborn, L., & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics*, 50(1), 65-82.
[7] 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.
[8] 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.
[9] 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.
[10] 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.
[11] Borgman, C. & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, 36(1), 2-72.
[12] Comer, E. D. (1997). *Computer networks and Internet*. New Jersey: Prentice-Hall Inc.
[13] *Country code top-level domain name*. (n.d.). Retrieved on September 18, 2013, from http://en.wikipedia.org/wiki/country_code_top_level_domain_name
[14] Cronin, B. (2001). Bibliometrics and beyond: some thoughts on web-based citation analysis. *Journal of Information Science*, 27(1), 1-7.
[15] *Domain Name*. (n.d.). Retrieved on September 18, 2013, from http://en.wikipedia.org/wiki/domain_name
[16] *Domain Name System*. (n.d.). Retrieved on September 18, 2013, from http://en.wikipedia.org/wiki/domain_name_system
[17] Frías, E. R. (2010). *Application of webometric techniques to the study of accounting and financial variables* (Doctoral Dissertation). University of Granada. Retrieved on March 12, 2013, from http://estebanromero.com/wp-content/uploads/2010/02/thesis-English-version-4-3-2010-DEF.pdf
[18] Goel, M. K. (2010). *World Wide Web*. New Delhi: Rajat Publications.
[19] 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 March 11, 2013, from citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.32.4054 &rep=rep1&type=pdf#page=5
[20] Holmberg, K. (2009). *Webometric network analysis: mapping cooperation and geopolitical connections between local government administrations on the web* (Doctoral Dissertation). Åbo Akademi University Press. Åbo Akademi University. Retrieved on March 15, 2013, from http://urn.fi/URN
[21] *Hyperlink*. (n.d.). Retrieved on October 29, 2013, from http://en.wikipedia.org/wiki/hyperlink
[22] Ingwersen, P. (1998). The calculation of Web Impact Factors. *Journal of Documentation*, 54(2), 236-243.
[23] Jeyshankar R., & Sujithai, M. (n.d). Webometrics tools and techniques: an overview. Available at http://dc370.4shared.com/doc/NOdc4LqQ/preview.html
[24] Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10-25.

[25] 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.

[26] *Link*. (n.d.). Retrieved on September 18, 2013, from http://en.wikipedia.org/wiki/Link

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

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

[29] 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

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

[31] 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 August 12, 2013, from http://www.oecd.org/sti/ieconomy/37730629.pdf

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

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

[34] Sangam, S. L. (2008). Areas in the field of scientometrics and informetrics. *In*: Koganuramath, M.., Kumbhar, B. D., & Kademani, B. S. (Eds), *Library and information science profession in the knowledge society*. New Delhi: Allied Publishers. 265-271.

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

[36] 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.

[37] 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.

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

[39] Thelwall, M. (2000). Web impact factors and search engine coverage. *Journal of Documentation*, 56(2), 185-189.

[40] Thelwall, M. (2007). Bibliometrics to webometrics. Journal of Information Science, 34(4), 1-18.

[41] Thelwall, M., Vaughan, L., & Bjorneborn, L. (2005). Webometrics. Annual Review of Information Science and Technology, 39(1), 81-135.

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

[43] 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.

[44] Walia, P. K., & Kaur, P. (2008). Webometric analysis of library associations' websites of India. IASLIC Bulletin, 53(3), 131-143.

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

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