

## The Webometrics

Prashant Goswami

Umesh Sharma

Anil Kumar Shukla

### Abstract

*It has been experienced that web based information resources have great role to play in academic and research activities. Keeping this fact in view information professionals are largely depending upon the web based information resources. A huge amount of data in every subject stream is available on different websites. But problem is only one that how to retrieve desired information. Many attempts were made and solution comes as 'webometrics'. In this paper an attempt has been made to describe this concept.*

**Keywords :** Webometrics, Bibliometrics, Information Resources

### 1. Introduction

"The web as a way of life" states that Internet users have come to rely on the vast amount of research and information content available, and often consult the web before making dramatic life changing decisions [1]. Web resources are apple of information professional's eye due to its value added services to meet their current and diversified information needs. In the World Wide Web (WWW), the web pages are the entities of information, with hyperlinks from them acting as citations. Quantitative analysis on the WWW is being carried out in the same way, as is tradition in citation databases. As information on web increases towards entropy, it's needed to apply some theory/ metrics (measurement) to develop new methods, modeling techniques and metaphors to examine this emerging complex network. Through webometrics (by Almind and Ingwersen in 1997)[2] study one can observe that how users actually react and use specific web document. The web is in out of control in growth, which means opportunities exist where good system architecture and diligent analysis can be applied for everyone's benefit. On the basis of the study and conception the definition of webometrics is, "The webometrics study is based on quantitative measurement – indirectly includes the qualitative aspect also – of structure, use of information resources and technologies on WWW drawing on bibliometric and infometric approach".[3]

In the above definition 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, it is helpful to know that how one page of website is attached with other page (via home page content, link structure and the like).

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 Net. The only method to search information on Net is via 'key word searching' through search engines or via 'specific URL'. Till today the 'voice

---

instruction technology' is not developed for searching on Net. Now the question arises that how to find those 'key words' (search terms) that are used, to search. We can have these words from log files. When ever 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.

The last phrase used in the definition is "drawing on bibliometric and infometric 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 infometric. The infometric 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's why the above phrase is used.

Here I include the term "qualitative aspect" in the above definition. Because while studying / measuring / analyzing the SUIIT part of website(s) one use to define the features of website. The SUIIT is the sdws (acronym) of Structure, Use of information resources and Technology.

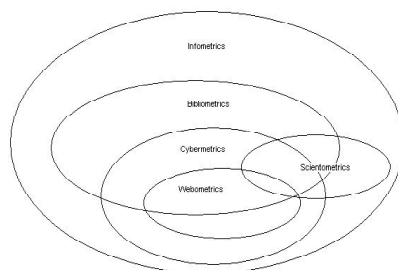
## **2. Relationship between 5 Metric Sciences [I.B.S.C.W.]**

Between these five metrics sciences bibliometrics, first came in existence in 1969. It is the statistical analysis method of publication pattern and the like. The other metric sciences – infometrics, scientometrics, cybermetrics and webometrics, are also statistical methods and having their origin on the basis of bibliometrics. But all of them are having their different subject area.

Before defining the relationship, it's essential to define all the terms. Here the terms are defined as the stub (in short).

1. Infometrics: the study of quantitative aspect of information in any form.
2. Bibliometrics: the study of quantitative aspect of production, dissemination and use of recorded information.
3. Cybermetrics: the study of quantitative aspect of Internet as a whole.

4. Scientometrics: the study of quantitative aspect of science as a discipline or economic activity.
5. Webometrics: the study of quantitative aspect of web/web site.



**Figure-1: Relationship diagram of 5 Metrics**

In the diagram the circle of Infometrics covers all other metrics circles, because according to stub (given above), 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.

Bjorneborn & Ingwersen 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.

In the diagram the circle of webometrics overlap the circle of bibliometrics, but within the boundaries of cybermetrics. Webometrics circle can't overlap the circle of cybermetrics because web is a part of cyberspace. But in the diagram the circle of webometrics ellipse lying outside the bibliometrics, because some aspect of webometrics (link structure, technologies and so on), dose not included in bibliometrics or it is beyond the boundaries of bibliometrics. The following point will be more helpful to understand the relationship between bibliometrics and webometrics.

### **3. Comparative Study of Bibliometrics and Webometrics**

The reason behind this comparative analysis is to describe the similarities and differences between the both. The webometrics is a new concept with increasing dimension and based on bibliometrics [a well established phenomenon], but some part of webometrics study is beyond the boundaries of bibliometrics. So the comparative analysis is based on the definition of webometrics.

---

The definition of webometrics includes four parts: 1. Web page content analysis, 2. Web link structure analysis, 3. Web usage analysis (e.g. exploiting log files for user's searching and browsing behavior), and 4. Web technology analysis (including search engine performance).

- i. If we elaborate the first point 'web page content analysis', we will find that it includes the content of home page of a particular website. It shows that what type of information is available on that web site. Like the very same we can also perform the study of books/literature through bibliometrics for their content analysis.
- ii. Second point is 'web link structure analyses. The link structure includes "inlinks" and "outlinks". Inlinks are the links, which are used to connect the pages of the same website. As the term binding is used for a book, like the very same the term inlink is used for the website. On the other hand outlinks are the link which connects a website with other, of the same subject. To better understand this term, we can compare it with the references/bibliographies given in a book. If the pages of a website are not properly linked, one can't properly access the website. Just like the above the links of the book can be said as the references or bibliographies and these may be measured as bibliometrically.
- iii. The third point is 'web usage analyses. Through this analysis anyone can know, how users are using the web. Log files are the good source for this analysis. Log files records the request sent to a web server by user, browsers and can be mined for useful information about how they are using the site. The following diagram is helpful to know the procedure of communication between web browser and web server. The principle drawback from a webometrics perspective is that log files typically cover one site/or all sites owned by the same server, and so they are generally not a good data source for studying all web server all together. Like the very same we can also do 'library usage analysis' or the like though bibliometrics. In the library users search materials/information via asking questions, via catalogue, through library software or the like. After collecting this type of information one can generate information, that how much a library searching materials are authentic and the defaults can quickly be removed.
- iv. The fourth point is 'web technology analysis' (search engine performance). Technology is a term, which denotes the quality. The above point includes search engine performance, 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 it's self is a very broad phenomenon but as for as webometrics study is concerned, it is measurable and a useful study tool for web based study. But the bibliometrics study is totally based on mathematical formula and quantitative statistical method, and the web technology is totally a different phenomenon. So this part of webometrics study is beyond the boundaries of bibliometrics.

Thus it can be said 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.

### References

1. Sangam, S.L. and Munavalli, S.B. IT literature: A webometrics study. Library progress (international), vol. 24 (1), 2004. p. 15-21.2.Webometrics – Ten-year expansion.  
<http://eprints.rclis.org/archive/00006264/>
3. Goswami, Prashant. Webometric Study of Universities Websites of Uttar Pradesh. MPHIL Dissertation, Vikram University, Ujjain. 2006-07. p. 30.

### About authors

**Mr. Prashant Goswami**, Librarian, Government Nehru Degree College, Ashok Nagar (M.P.)

**Mr. Umesh Sharma**, Semi Professional Assistant, JRR Sanskrit University, Jaipur, (Raj.)

**Mr. Anil Kumar Shukla**, Librarian, NCET, Bhopal (M.P.)