Literature Review

Bibliometric research methodologies of library and information science have always been used to provide tools for understanding the dynamics of disciplines, developing policy, and justifying research funding. Cronin & McKim (1996) [15] have pointed out that the Web is becoming a significant communication medium for science and scholarship, and bibliometric studies of scholarly publishing are being extended to the Web. A growing literature has emerged that applies bibliometric measures to cyberspace. Terms applied to this new area of study include "webometrics" (Almind & Ingwersen, 1997), [16] and "cybermetrics" (the title of a journal). Webometrics is defined as: "the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web, drawing on bibliometric and informetric approaches" and 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, 2004). [17]

A key to webometric studies has been the use of large-scale search engines, such as AltaVista and All the Web that allow measurements to be made of the total number of pages in a web site and the total number of back links to the web site. These search engines provide similar possibilities for the investigation of links between web sites/pages to those provided by the academic journals citation databases by the Institute of Scientific Information (ISI). But the content of the Web is not of the same nature and quality as the databases maintained by the ISI (Noruzi, 2006). [18]


Thanuskodi (2010b) [23] discussed the research output performance of social scientists on social science subjects. The analysis cover mainly the number of articles, authorship pattern, subject wise distribution of articles, average number of references per articles, forms of
documents cited, year wise distribution of cited journals etc. Yeoh and Kaur (2008) [24] analyses the publication output of Research in Higher Education for subject support in collection development in the light of growing interest in diversified domains of research in higher education. Consequently, analysis of 40 issues of publications revealed a diversified usage pattern of bibliographic reference sources by contributing researchers, with a cumulative total of citations being 8,374. A positive trend in research collaboration of contributing authors, and a steady growth in the use of reference sources, periodicals and web documents in the citations signify the trend of scholarly communication of research works in the electronic age. Similar to other disciplines of research findings, journals and books were the most cited source materials for researchers thrash out.

Verma, Tamrakar and Sharma (2007) [25] revealed that majority of the articles in the journal are two-authored and majority of the contributions are from New Delhi. Singh, Mittal and Ahmad [26] (2006) conducted a bibliometric study of literature on digital libraries. The important findings are that most articles (61 percent) are single-authored; author productivity is not in agreement with Lotka's Law,[27] except in one case where the number of articles is three; the maximum number of articles were published in 2003 with English being the most productive language; maximum articles were published in the journal D-lib Magazine; distribution of articles nearly follows Bradford's Law; and USA ranked first for maximum number of journals. Tiew (2000) found that 53% of articles contained journal self-citations, and a tendency is noticed for authors affiliated to the institution publishing the journal to cite the journal. Patra, Bhattacharya and Verma (2006) [28] analyzed the growth pattern, core journals and authors' distribution in the field of bibliometric using data from Library and Information Science Abstract (LISA) and found that the growth of literature does not show any definite pattern. Dhiman (2000) [29] has done ten year bibliometric study Ethnobotany Journal published during 1989-1998. In this paper examines year-wise, institution-wise, country-wise, authorship pattern, range of references cited and length of the articles.

Thelwall, Vaughan & Bjorneborn [30] state that the field of webometrics grew out of a realization that quantitative methods originally designed for bibliometric analysis of citation patterns of scientific journal articles could be applied to the Web by using commercial search engines to provide the raw data. Especially, AltaVista’s (http://www.altavista.com/) search
interface that allowed complex Boolean search strings including properties of links and URLs triggered this approach. As further outlined in the review on webometrics by Thelwall, Vaughan & Björneborn (ibid.), a considerable number of research papers have been published concerning scholarly communication on the Web, mostly originating in the hope that web links could be used to provide similar kinds of information to that extracted from journal citations.

Bossy [31] suggested how netometrics, as she called it, could supplement bibliometrics and scientometrics in observing “science in action” on the Internet, enabling account the cooperative aspect of science”. However, she made no empirical investigation of academic web spaces in the paper.

In his paper ‘Bibliometrics of the World Wide Web: an exploratory analysis of the intellectual structure of Cyberspace’. Larson [32] was one of the first information scientists to perform an investigation of link structures in academic web spaces. He used AltaVista in a co-citation analysis of a set of Earth Science related web sites and could produce clustering of web sites that had topical similarities.

Shortly after, Almind & Ingwersen [33], in a paper introducing the term webometrics, applied a variety of bibliometric-like methods to the Nordic portion of the Web in order to observe the kinds of page connections and define the typology of web pages found at national Nordic level.

In his article on ‘citations’ – using a term coined by McKiernan [34] for site in links – Rousseau [35] analyzed the patterns of distribution of web sites, site Inlinks and site self-links (‘self-citations’). Rousseau’s study operated with 343 web sites retrieved in AltaVista with the search string, informetrics OR bibliometrics OR scientometrics. The study showed that the distribution of TLDs (top level domains, such as .edu, .uk, .dk) for the investigated sites followed the ubiquitous power-law-like Lotka distribution. Similarly, Rousseau demonstrated that the distribution of inlinks to the 343 sites also followed a Lotka distribution.

Ingwersen [36] introduced the concept of the Web Impact Factor (WIF) for national domains and individual web sites with parallels to the Journal Impact Factors published by the Institute of
Scientific Information (ISI) for scientific journals receiving citations from scientific journals indexed in the ISI citation databases.

Thelwall [37] showed that the counts of inlinks to a set of 25 UK universities correlated significantly with their average research productivity using the five-yearly UK government Research Assessment Exercise (RAE) of research productivity.

Thomas & Willett [38] studied UK library and information science departments, finding no significant correlation between inlink counts and research ratings.

Smith & Thelwall [39] compared linking patterns between UK, Australian and New Zealand universities, and found that New Zealand was relatively isolated on the Web. A larger follow-up study mapped the interlinking between universities in the Asia-Pacific region showing that Australia and Japan were central web players in the region, with smaller countries attracting attention disproportionate to their size.