The chapter critically analyzes the scholarly articles from journals, papers from conference proceedings, Web resources, technical reports, webpages and books from the perspective of finding out the knowledge gap in the core area of the concerned topic by analyzing wide spectrum of related primary documents.
## Contents: Review of Literature

<table>
<thead>
<tr>
<th>Section</th>
<th>Page No</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Introduction</td>
<td>25</td>
</tr>
<tr>
<td>2.2 Background Information</td>
<td>26</td>
</tr>
<tr>
<td>2.3 Bibliometrics</td>
<td>27</td>
</tr>
<tr>
<td>2.3.1 Evaluative Bibliometrics</td>
<td>28</td>
</tr>
<tr>
<td>2.3.2 Relational Bibliometrics</td>
<td>28</td>
</tr>
<tr>
<td>2.3.3 New Bibliometric Databases</td>
<td>28</td>
</tr>
<tr>
<td>2.3.4 Usage Data from Digital Libraries</td>
<td>29</td>
</tr>
<tr>
<td>2.4 Informetrics</td>
<td>29</td>
</tr>
<tr>
<td>2.5 Scientometrics</td>
<td>29</td>
</tr>
<tr>
<td>2.6 Cybermetrics</td>
<td>29</td>
</tr>
<tr>
<td>2.7 Webometrics</td>
<td>30</td>
</tr>
<tr>
<td>2.7.1 Webpage Content Analysis</td>
<td>30</td>
</tr>
<tr>
<td>2.7.1.1 Webpage Classification</td>
<td>31</td>
</tr>
<tr>
<td>2.7.1.2 General and Commercial Link Creation</td>
<td>31</td>
</tr>
<tr>
<td>2.7.1.3 Academic Web Link Creation</td>
<td>31</td>
</tr>
<tr>
<td>2.7.2 Web Technology Analysis</td>
<td>32</td>
</tr>
<tr>
<td>2.7.3 Web Usage Analysis</td>
<td>32</td>
</tr>
<tr>
<td>2.7.4 Web Link Structure Analysis</td>
<td>32</td>
</tr>
<tr>
<td>2.7.4.1 University Links</td>
<td>33</td>
</tr>
<tr>
<td>2.7.4.2 Commercial Links</td>
<td>34</td>
</tr>
<tr>
<td>2.7.4.3 Government’s Website Links</td>
<td>34</td>
</tr>
<tr>
<td>2.7.4.4 Political Party’s Website Links</td>
<td>34</td>
</tr>
<tr>
<td>2.7.4.5 Web Blog Links</td>
<td>34</td>
</tr>
<tr>
<td>2.7.4.6 Classification of Links</td>
<td>35</td>
</tr>
<tr>
<td>2.7.4.7 Motivation of Links</td>
<td>35</td>
</tr>
<tr>
<td>2.7.4.8 Link Analysis Algorithms</td>
<td>35</td>
</tr>
<tr>
<td>2.7.4.8.1 HITS</td>
<td>35</td>
</tr>
<tr>
<td>2.7.4.8.2 PageRank</td>
<td>36</td>
</tr>
<tr>
<td>2.7.4.9 Web Citation Analysis</td>
<td>36</td>
</tr>
<tr>
<td>2.7.4.9.1 Citations per Publication</td>
<td>36</td>
</tr>
<tr>
<td>2.7.4.9.2 Web Co-citation Analysis</td>
<td>36</td>
</tr>
<tr>
<td>2.7.4.9.3 Web Coupling</td>
<td>37</td>
</tr>
<tr>
<td>2.7.4.10 Alternative Document Models</td>
<td>37</td>
</tr>
<tr>
<td>2.7.5 Webometric Tools and Techniques</td>
<td>37</td>
</tr>
<tr>
<td>2.7.5.1 Commercial Search Engines</td>
<td>37</td>
</tr>
<tr>
<td>2.7.5.2 Personal Web Crawlers</td>
<td>38</td>
</tr>
<tr>
<td>2.7.6 Web Presence</td>
<td>38</td>
</tr>
<tr>
<td>2.7.7 Web Impact Factor (WIF)</td>
<td>39</td>
</tr>
<tr>
<td>2.7.8 Webometric Ranking</td>
<td>40</td>
</tr>
<tr>
<td>2.7.8.1 National Level Ranking</td>
<td>40</td>
</tr>
<tr>
<td>2.7.8.2 Global Level Ranking</td>
<td>41</td>
</tr>
<tr>
<td>2.7.9 Link Topology</td>
<td>42</td>
</tr>
<tr>
<td>2.7.9.1 Graph Theory</td>
<td>42</td>
</tr>
<tr>
<td>2.7.9.2 Visualization Tools</td>
<td>42</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>2.7.9.3</td>
<td>Power Laws</td>
</tr>
<tr>
<td>2.7.9.4</td>
<td>Social Network Analysis</td>
</tr>
<tr>
<td>2.8</td>
<td>Trends of Webometrics in Indian Context</td>
</tr>
<tr>
<td>2.9</td>
<td>Observations</td>
</tr>
<tr>
<td>2.10</td>
<td>Summary</td>
</tr>
<tr>
<td></td>
<td>References</td>
</tr>
</tbody>
</table>
Chapter—2

Review of Literature

2.1 Introduction

One of the ways of economizing an investigation is to review the related literature and work done by the previous researchers to set the background for the current study. The research problem "Indian Universities on the Web: Analysis of Hyperlinks through the Application of Webometric Tools and Techniques" has been taken up for investigation. In order to study in-depth and to make the conception clear, it is highly required to review literature on the focus under study. All the studies are not significant enough to enumerate here but an attempt has been made to review and highlight briefly the objectives and findings of the important studies relevant to the present study in this chapter.

A literature review surveys scholarly articles, books, Internet resources, dissertations and conference proceedings relevant to a particular issue, area of research, theory providing descriptions, summary, and critical evaluation of each work. The objective is to give an overview of significant literature published on the related aspects of the research problem. The relevant literature is a combination of two fields: statistical applications in general and different facets of LIS. Apart from these two basic facets, the third facet includes Web technology. So, this research study searches literature from wide varieties of resources like IEEE Digital Library, John Wiley & Sons, Directory of Open Access Journals (DOAJ). Some abstracting and indexing databases like Library and Information Science Abstract (LISA), Library, Information Science and Technology Abstract (LISTA), Information Science and Technology Abstract (ISTA), JCCC@UGC-Infonet, an e-
journal gateway and also citation databases like Google Scholar and Scopus have been consulted to have the wider coverage of the concerned topic. Besides, each issue of few core scholarly journals were consulted like Scientometrics, Journal of the American Society for Information Science and Technology (JASIST), Journal of Information Science, Cybermetrics, Information Research, Information Processing & Management and Webology during 2005 to 2010.

2.2 Background Information

The field Webometrics emerged from Bibliometrics, Scientometrics, Informetrics and Cybermetrics. The development of Bibliometrics and Scientometrics was noticed through the use of citation index of scientific literature developed by Garfield (1955). Online citation analysis helped the development of webometrics studies. In the mean time, a range of new terms were proposed to describe the emerging research field e.g. Netometrics (Bossy, 1995); Webometry (Abraham, 1996); Internetometrics (Almind & Ingwersen, 1996); Webometrics (Almind & Ingwersen, 1997); Cybermetrics (journal started in 1997 by Aguillo, 2002); Web Bibliometry (Chakraborty, et al 2002). The relation among Bibliometrics, Informetrics, Scientometrics, Cybermetrics and Webometrics were explained in the following diagram (Björneborn & Ingwersen, 2004).

Figure 2.1—Relation among LIS fields like bibli/infor/sciento/cyber/webo-metrics

Figure 2.1 shows that the field Cybermetrics exceeds the boundaries of bibliometrics because some activities are normally not recorded in cyberspace. Almind and Ingwersen first coined the term Webometrics in 1997. Webometrics tried to measure the World Wide Web (WWW) to get knowledge about number, type of hyperlinks and reference structure of the WWW and usage pattern. The field of webometrics entirely covers by Bibliometrics because web documents are recorded information.

According to Björneborn and Ingwersen (2004), the definition of Webometrics is “the study of quantitative aspects of the construction and use of information resources structure and technologies on the web drawing on Bibliometrics and Informetrics approach”.

During post 1990s, Internet research had gained lot of popularity due to its wide application in various fields. Therefore, the need aroused to measure the various aspects of Internet especially World Wide Web (WWW). Traditional
bibliometrics and informetric laws have been applied for measuring the Web in order to assess the performance of the websites in terms of link analysis, web content analysis and web citation analysis. One important measure may be the application of Web Impact Factor (WIF) to judge the quality of websites of the institute. WIFs are the web versions of the Impact Factors (IFs), which is published by the Institute of Scientific Information (ISI), Philadelphia for judging the quality of scientific journals. The research in the field of library and information science yielded number of theories and laws concerning the quantitative aspects of information with respect to its generation, organization, and distribution of information to the users. Some such laws are Lotka’s Law on productivity distribution among scientists, Bradford’s Law on the scattering of literature on a particular topic over different journals (Bradford, 1934), Zipf’s law of word frequencies in text. Likewise, some laws have been identified on the Web e.g. the distribution of Top Level Domains (TLDs) on a given topic (Rousseau, 1997) or inlinks per website (Albert, Jeong & Barabasi, 1999).

Rodríguez-Gairín (1997) had introduced the concept of information impact on the Internet in a Spanish documentation journal earlier. The idea of measuring average link frequencies i.e. WIF was developed by Peter Ingwersen in 1998. Generally, it is defined as the number of pages linking to a site or area in Internet divided by the number of pages in that site. A high value of WIFs indicates a site with great impact because there are relatively many pages linking to the site. Till today, researchers have to depend on the commercial search engines to collect the necessary data for calculating WIF. The popular search engines helpful in this endeavour are AltaVista, Web Crawler, Yahoo! Hotbot, etc. It is known to everybody that the search engine’s coverage and capability are limited with respect to the volume of webpages. The link analysis data are collected for universities in India using relevant commercial search engines with the provision of special keywords for webometric studies. Such data sets may be utilized to measure the Web presence of Indian universities and their impact (through WIF) on the teaching-learning-research process.

2.3 Bibliometrics

The concept of bibliometrics is applied to various sorts of library services to study the characteristics of subjects and nature of citations of different subjects by information scientists and librarians. Dr. Ranganathan coined the term ‘Librametry’ in 1948 as the application of mathematical and statistical techniques to library problems (Sengupta, 1992). Sengupta (1992) claimed that Campbell (1896) produced the first bibliometric study using statistical methods for studying subject scattering in publications. On the other hand, Lawani (1981) and Khurshid & Sahai (1991a, b) claimed that Cole and Eales (1917) made the first recorded study on bibliometrics. They analysed publications after counting the numbers of titles, both books and journal articles in field of comparative anatomy from 1543-1860 in order to study the distribution of literature among countries. Hulme (1923) introduced the term ‘Statistical Bibliography’, which was considered as the second reported work on bibliometrics. Hulme analysed the journal articles and derived the ranking of countries by their productivities. Gross and Gross (1927) carried out the study of citation data. This was the third bibliometrics study conducted by
them. Pritchard (1969) coined the term bibliometrics (Fairthorne, 1969, Brookes, 1988, Khurshid & Sahai, 1991a, b) to replace the term ‘Statistical Bibliography’. Pritchard’s initial work was concentrated on two dimensions i.e., quantitative and qualitative study of bibliometric aspects. The quantitative analysis leads to three basic bibliometric laws:

- Lotka’s Law (1926) predicting the productivity distribution of authors.
- Zipf’s law (1933) describing the word frequency ranking.
- Bradford’s Law (1934) of scattering describes the distribution of journals in specific discipline.

2.3.1 Evaluative Bibliometrics

Bibliometrics is basically two types such as evaluative bibliometrics and relational bibliometrics. Evaluative bibliometrics uses citation techniques to assess the impact of scholarly work and compare the contribution of two authors, whereas relational bibliometrics uses citation as raw data.

2.3.2 Relational Bibliometrics

Many attempts had been taken to develop bibliometric methods to examine relations in science through ISI data. Early relational analysis tried to focus on the structure of science through simple means such as network diagrams of the flow of citations between key set of articles. One of the important applications of relational bibliometrics is the co-citation as a measure of similarity.

2.3.3 New Bibliometric Databases

The major initiatives for preparing bibliometric databases were taken by Garfield (1972), the founder of the Institute of Scientific Information (ISI), now Thomson Scientific, Philadelphia, USA. Google Scholar provides a simple way to search scholarly literature. There is a provision from one search-interface to search across many disciplines and sources: peer-reviewed papers, theses, books, abstracts and articles, from academic publishers, professional societies, preprint repositories, universities and other scholarly organizations. Another popular database is Scopus, the World’s largest abstract and citation database of peer-reviewed literature and quality web sources with smart tools to track, analyse and visualize research. In addition, there are some field specific citation databases like CiteSeer for Computer Science. CiteSeer1, which was developed in 1997 at the NEC2 Research Institute, New Jersey was the first scientific literature digital library and search engine to provide automated citation indexing and citation linking using the method of autonomous citation indexing. Citebase is another database whose coverage and capabilities are based: a) only on those citing and cited papers that their authors have already archived in the source eprint archives, b) only on those of the cited papers that can currently be successfully linked.

---

1 CiteSeer was a public search engine and digital library for scientific and academic papers. It is considered as the first automated citation indexing system. At present it is replaced by CiteSeer³

2 NEC Laboratories America, Inc. (NEC Labs) is the US-based part of NEC’s global network of research laboratories. It was created in November 2002.
2.3.4 Usage Data from Digital Libraries

The usage data from digital library in an academic environment may be a great challenge and new dimension in the area of bibliometrics. An investigation into log file analysis may add new phase in bibliometric analysis for the evaluation of documents. Digital library usage data can correlate with citation analysis because higher usage data of a particular document will able to predict to have more citation counts and vice-versa.

2.4 Informetrics

The term ‘Informetrics’ covers both sciento—and biblio—metrics impartially. It covers both documentary and electronic forms of information. Moreover, it deals with all aspects of quantitative analysis of information transfer, irrespective of media. The historical background of Informetrics started with the tables of cases cited in 1700’s and 1800’s (Shapiro, 1992). Björneborn and Ingwersen (2001) opined that the scope of Informetrics is drawn from the overlapping fields of both bibliometrics and Scientometrics. According to Tague-Sutcliffe (1992), Informetrics is “the study of the quantitative aspects of information in any form, not just records, or bibliographies, and in any social group, not just scientists”. Informetrics covers both Scientometrics and Bibliometrics (Brookes, 1990). Hood and Wilson (2001) explain the usage and history of the terms Scientometrics, bibliometrics and Informetrics. Informetric research includes studies pertaining to scattering of articles and journals, growth and obsolescence of documents, productivity and impact of research, distribution of scientific publications by country, by language and circulation studies. Wilson (1999) covered the development of Informetrics till the end of 20th century. Bar-Ilan (2008) thoroughly covered the development of Informetrics during 2000 to 2006.

2.5 Scientometrics

Scientometrics deals with the scientific measurement of the work of scientists, especially by way of analysing their publications and citations. Scientometrics is typically defined as the quantitative study of science and technology. According to Tague-Sutcliffe:

“Scientometrics is the study of the quantitative aspects of science as a discipline or economic activity. It is part of the sociology of science and has application to science policy making. It involves quantitative studies of scientific activities, including among others publication, and so overlaps bibliometrics to some extent”.

2.6 Cybermetrics

Cybermetrics deals with the quantitative aspects of Internet. Björneborn (2004) defined Cybermetrics, “the study of the quantitative aspects of the construction and use of information resources, structure and technologies on the whole Internet, drawing on bibliometric and informetric approaches.” Cybermetrics encompasses statistical studies of discussion groups, mailing lists and other computer mediated communication on the Internet (Bar-Ilan, 1997; Hernandez-Borges, Pareras &
Jimenez, 1997; Herring, 2002; Matzat, 1998) including the Web. In other words, Cybermetrics is a quantitative measure of the Internet technology, topology and traffic.

2.7 Webometrics

Webometrics is very young and attractive research field in LIS, computer science and computing & information technology. Almind and Ingwersen (1997) introduced the application of informetric methods to the WWW, so called Webometrics. They proposed a number of specific informetric parameters such as hyperlinks per webpages, link density on webpages distributed over type of documents and domain names. Björneborn and Ingwersen (2004) defined Webometrics as:

“The study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web, drawing on bibliometrics and Informetrics approach.”

A detailed link topology, Web node diagram and various terminologies were developed (Björneborn & Ingwersen, 2001). The scope of webometrics can broadly be categorized under: (a) Webpage content analysis, (b) Web technology analysis, (c) Web usage analysis, and Web link structure analysis.

In order to determine the scope of webometrics, it is required to know the relationship between Informetrics, Scientometrics, bibliometrics and webometrics. It may be noticed that webometrics is associated with bibliometrics and overlaps Scientometrics to some extent (Björneborn & Ingwersen, 2004).

Thelwall, Vaughan and Björneborn (2005) contributed a classical article on webometrics to demonstrate basic concept, origin, scope and coverage of webometrics and related reviews. The issues on data collection methods and measurement techniques of Web related activities were addressed. Almind and Ingwersen (1997) took an initiative to introduce and argue that it is possible to apply informetric methods to the Web. Webometrics covers all network-based communication using informetric or other quantitative measures.

Thomas and Willett (2000) described a webometric analysis of the linkages to websites associated with individual departments of Library and Information Science (LIS) in United Kingdom (UK) universities. The findings of the study revealed that it was not possible to identify any significant correlation between the citation data and peer evaluations of research excellence embodied in the Research Assessment Exercise (RAE) rankings.

2.7.1 Webpage Content Analysis

Content analysis is the systematic study of the content of website. Webpage content analysis yields the structure of the websites and the creation of hyperlinks between webpages. Being an important area of website analysis for its performance enhancement and improvement, it helps to know the taxonomical structure of website, use of appropriate keywords, which may be helpful for the
index of search engine and assigns meaning and contextual keywords to represent the content of website. Rosenbaum (1998) made a study of contents of websites of 24 web-based community networks in Indiana universities to know the structure of the websites. Haas and Grams (2000) identified seven different page types using a content analysis of 331 randomly selected webpages. Dumais and Chen (2000) explored the use of hierarchical structure for classifying a large heterogeneous collection of web content to support classification of search results.

2.7.1.1 Webpage Classification

Since the Web is uncontrolled by nature, extra challenges to webpage classification as compared to traditional text classification is required. The categorization of webpage helps to understand the relative importance of different types of pages within the structure of website. Following are the basic categories of webpages:

- a) Homepage
- b) Index Pages
- c) Resources

Webpage classification is more difficult than text classification due to the involvement of more noise information and lack of permanency. Shen, et al (2004) proposed new web page classification algorithm based on Web summarization for improving the accuracy. The result reflected that the improvement of results as compared to text-based classification was 8.8%. Qi and Davidson (2009) reviewed the web page classification and described the state of the art practices. Asirvatham and Ravi (n.d.) proposed a method for automatic categorization of webpages into a few broad categories based on the structure of web documents and images in it.

2.7.1.2 General and Commercial Link Creation

Link creation is a natural phenomenon in websites. Kim (2000) found out the motivations for creating links in electronic publications in order to find out the relationship between citations and scholarly e-journals. Three factors have been identified such as scholarly, social and technological reasons. Harrison (2002) identified a few principles of link creation and proposed an approach for classification of links. Park (2002) conducted a survey of 64 Korean webmasters of commercial websites to assess their motivations for linking to other websites and found that webmasters were more likely to hyperlink to websites possessing practical content, information or services.

2.7.1.3 Academic Web Link Creation

Chu (2005) pointed out reasons for creation of links by analysing sample of links and found that only 27% of links were made for research or teaching motivations. Kousha and Horri (2004) made a survey of Iranian university and found that 63% of hyperlinks were made for navigational purpose. Wilkinson, et al (2003) surveyed 414 links between UK university websites and found that less than 1% of hyperlinks created for formal scholarly articles in journals or conferences; 90% of targeted materials were some way or rather related to scholarly activity. Bar-Ilan (2004a) made academic link studies of inter-university links in Israel and found that 20% of links related to research category. Thelwall (2003) surveyed a sample of 100 random inter-site links to UK university homepage and found four types of
motivations: ownership, social, general and navigation reasons. Thelwall (2001a) made an attempt to distinguish links between research related and non-research oriented.

2.7.2 Web Technology Analysis

Web technology analysis is an important part of webometric research though most concentration lies on hyperlink analysis. Cronin (1984) argued that most important articles tend to be cited more. Google applies this assumption to the web with its PageRank algorithms to find most important page on the web. Therefore, the application of citation analysis techniques to the web is not always a one-way relationship: Google’s success has led to PageRank’s adoption for bibliometrics (Thelwall, 2002c; Thelwall & Vaughan, 2004). Almind and Ingwersen (1997) used AltaVista extensively in webometric research as it supports the advanced Boolean features in its search box. Yahoo! search engine had also extended the facility for webometric research through its site explorer.

Bar-Ilan (1999, 2002) tested the performance of search engines over time and found that search engines did not yield accurate results. In fact, commercial search engines are having some drawbacks in extracting link data. Besides, search algorithms of various search engines are not known by scientific researchers. Lawrence and Gill (1998, 1999) estimated the coverage of commercial search engines on the web by investigating several queries and retrieving results. It was found that the largest search engines Northernlight can able to index only 16% of the webpages. Mayr and Tosques (2005) mentioned that Google introduced Web Application Programming Interface (API), an instrument for webometric research. Google’s Web API allow Internet researcher to get data for webometric analysis like: (a) analysis of times series, (b) Journal web coverage, (c) TLD analysis, (d) distribution of file format on the web.

2.7.3 Web Usage Analysis

The web usage analysis is basically focused on log files of user’s searching and browsing behaviour. The objective of such study is to see which pages are most frequently viewed and to identify patterns of surfing with a view to improving site navigation (Huntington, Nicholas & Jamali, 2007). The main problem of such webometric study is that log files are covered by the same server and access to these log files are often restricted. Pirolli and Pitkow (1999) analysed website server logs over a ten days period to compare different paths reconstructions and to investigate how fast surfing behaviour predicts future surfing choice.

2.7.4 Web Link Structure Analysis

Web link structure analysis is the core part of webometrics and it is the focus of our study. The importance of web links is duly acknowledged by Berners Lee (1997), the founder of WWW. Link structure includes the studies of inlinks, outlinks, self-links, total links and co-links. Larson (1996) made an investigation of link structures in academic web spaces using Altavista. The co-citation analysis for earth science related websites produced a cluster of websites that had topical similarities. Henzinger (2001) reviewed link structures analyses from a computer
science perspective, showing how links could be used in search engine ranking algorithms. Bjørneborn (2004) has contributed the idea of link structure in his PhD thesis on ‘Small world link structures across an academic web Space: a library and information science approach’. It described the link structure of the academic institutions of UK. Harries, et al (2004) expressed that hyperlinks between academic websites can be used to map disciplinary structures and identify evidence of connections between disciplines.

Early university link studies investigated the relationship between interlinking counts of selected set of universities on a national level and the research activity indicators of those universities. The findings were negative (Smith, 1999; Thomas & Willett, 2000) and the main reasons were the failure in the part of search engines to obtain the link counts and the number of links created for reasons unrelated to research. Hemerks and Van Den Besselaar (2006) analysed and compared hyperlink networks using a variety of linking units on the different levels of aggregation and specificity. It has been shown how the linking units like countries, universities, departments, and individuals produced heterogeneous linking patterns that might provide information about knowledge production.

2.7.4.1 University Links

Middleton, McConnell and Davidson (1999) opined that university websites act as tool for communication to internal and external users. Using hyperlinks to investigate the online informal scholarly communication is now established as a practical and useful approach (Wilkinson, Thelwall & Li, 2003). Thelwall and Smith (2002) described the pattern of international interlinking between 13 Asia-Pacific university websites. Alta vista’s advanced search facility was used for data collection and network diagrams were used to portray the results. The study found that Japan was placed as top due to source or target of more links than any other country.

Onyancha and Ochalla (2007) in their study used link analysis to compare Kenyan and South African universities based on several web based indicators such as number of webpages, number of inlinks and outlinks. The external outlinks were examined to determine the institutions targeted by Kenyan and South African universities and also investigated the networks links between universities. Results were indicated that Kenyan universities like most African universities have embraced the Internet and its web sites are at initial stages of construction. On the other hand, South African universities have made tremendous progress in their web presence. Fereshteh and Marzieh (2010) analysed website links to know the impact of top universities of Islamic world on the Web through statistical data.

Many link patterns have been identified for the universities both at national and international level, but departmental interlinking has been relatively ignored initially. Thomas and Willett (2000) described webometric analysis of linkages to websites associated with department of LIS. Sitation data are not well suited to the quantitative evaluation of the research status of LIS departments. Li, et al (2005) had taken up interest to study national and international university’s departmental website interlinking. Departmental interlinking patterns were illustrated through the study of physics, chemistry and biology in Australia, Canada and UK to study
link pattern differences and compared them. Tang and Thelwall (2003) investigated into departmental interlinking in USA for history, chemistry and psychology. These three departments were used the web differently in a range of different ways (Tang & Thelwall, 2004) showing that there are clear web linking differences between different disciplines i.e. hard sciences from social sciences.

2.7.4.2 Commercial Links

Vaughan (2004) made a study of top ranking information technology (IT) companies in US & China and found that there is a significant correlation between business performance measures and inlinks to company websites. Vaughan and Wu (2004) examined the websites of China’s top 100 IT companies. Link counts to company’s websites were found to correlate with the companies revenue, profit and research & development. On the other hand, it was also found that hyperlinks to commercial sites could be a business performance indicator and thus a source of business information.

2.7.4.3 Government Website’s Links

Danesh, Soheili and Shafiei (2008) studied 20 websites of Iranian government ministries concerning staffs, visibility, WIF and collaboration rate of the websites of Iranian government ministries through the process of link analysis using Yahoo! for a period of (Jan 10–Feb 22, 2008). The research results revealed that Ministry of Cooperatives (MOC) websites with 282 inlinks, Ministry of Science, Research & Technology (MSRT) with 126 and Ministry of Industry & Mines (MIM) with 109 inlinks were the most visited sites. Li and Fu (2009) had chosen 15 provincial government websites to search their link-based data. The result obtained through link analysis using factors with government efficiency scores and found that most obviously correlated indexes were total link and external link.

2.7.4.4 Political Party’s Website Links

Park and Thelwall (2007) made an investigation into political communication expressed by links connecting politician’s websites in South Korea. The study examined the linking patterns between sites and found that the pattern did not change between 2003 and 2004. Mukherjee (2009) made a comparative study of Indian political party’s websites to judge the popularity of Internet and to know whether Internet can be used as a tool for Indian politics through link analysis using search engines like Google, Yahoo! and Altavista. Esteban and Liwen (2010) explored the feasibility of using web hyperlink data to study the political websites of 96 European Union (EU). The result showed that Web hyperlink data reflected the political pattern in the EU.

2.7.4.5 Web Blog Links

Web blogs are special kind of websites. The analysis of these blogs may come under webometric analysis. Kim Holmberg is popular webometrician maintaining his personal blog. Kim’s website is having 265 webpages and inlinks is 87 using AltaVista. Torres-Salinas, et al (2011) made an attempt to undertake a webometric analysis of 1108 personal and corporate blogs on LIS between November 2006 and
June 2009 indexed on the Lib-orm search engine characterizes the community’s behaviour quantitatively. Over the study period, there was 52% of decrease in the number of active blogs. Despite the drop in production over this period, the average number of posts per blog remained constant (14 per month). The most representative blogs in the discipline are identified.

### 2.7.4.6 Classification of Links

Wilkinson, et al (2003) conducted a survey on (.ac.uk) domain to find out the motivations for creating links between university websites. For this purpose, a classification scheme was created and applied to classify 414 of such links. It was found that over 90% of links were created for broadly scholarly reasons. Chu (2005) applied two part process while developing classification of hyperlinks for inlinked pages with four categories i.e. teaching-learning (14%), research (13%), service (46%) and homepage (27%).

### 2.7.4.7 Motivation of Links

Kim (2000) found through interview method that there are 19 motivational factors, which may be grouped under scholarly, social and technological issues. Park (2002) surveyed webmasters in Korea to know the reasons of creating hyperlinks and found that credibility and usefulness were prime important factors. Even, the correlation between research performance and links to websites had been noticed (e.g. Smith & Thelwall. 2002; Thelwall, 2002a). Chu (2005) found that less than 27% of the links were made out of research or teaching-learning motivation while Thelwall (2003) found the majority of linking motivations were trivial compared with citation motivations. Vaughan, Gao and Kipp (2006) discussed some motivational factors for the creation of hyperlinks to business sites using random sample of 808 links, which were manually examined to determine why the links was created. The study found that most links were created for business purpose and links to competitors were less but competitors were often co-linked. The study also took qualitative content analysis approach that complements the quantitative approach. Each linking page was classified by country, type of websites and motivations for linking.

### 2.7.4.8 Link Analysis Algorithm

When users are looking for something from the Web with the expectation of getting the relevant result, the search engine takes the query and searches for the relevant information and retrieves the result. The rate of success to retrieve the relevant and authoritative information depends on the power of algorithms. Therefore, link analysis algorithms have a strong influence to retrieve the relevant information from the Web. Different search engines use different link analysis algorithms for their research. The popular hyperlink analysis algorithms are HITS and PageRank.

#### 2.7.4.8.1 HITS

Hypertext Induced Topic Search, popularly known as HITS was developed by eminent information scientist, John Kleinberg in 1999 (Kleinberg, 1999). HITS
involves two steps: (a) Building a neighbourhood graph \( N \) related to the query terms, and (b) Computing authority and hub scores for each document in \( N \), and present the two ranked list of the most authoritative and most “hubby” documents. The HITS principle is based on relevance and query-dependence.

### 2.7.4.8.2 PageRank

PageRank, a popular link analysis algorithm was developed by Larry Page based on the principle of ‘use of importance rather than relevance’. PageRank algorithm was used by Google to rank the website. The strength of PageRank is that it is query independence and retrieval is faster. Google PageRank is measured from 0 to 10. The PageRank is being calculated on the basis of backlinks. The more and more quality backlinks results in higher PageRank.

### 2.7.4.9 Web Citation Analysis

Web citation analysis deals with the analysis of citations for the scholarly scientific documents published on the Web. Gross and Gross (1927) made pioneering research work to evaluate the importance of scientific work using citation counts. Cronin (2001) pointed out that the significance of Web from bibliometric perspective. Web-based retrieval system allows us to go beyond traditional citation. Zhang (2006) pointed out that open access journals have higher percentages of citations than other type of documents. Yang and Meho (2006) presented a case study comparing citations found in Scopus and Google Scholar with Web of Science (WoS). CiteSearch analysed combined data from multiple citation databases to produce citation-based quality evaluation measures.

#### 2.7.4.9.1 Citations per Publication

Yi and Jin (2008) conducted a study to perform bibliometric analysis of all pentachlorophenol related publications in the Science Citation Index (SCI) during 1994-2005. The analysis includes parameters like document type, language of publication, page count, publication output, authorship, publication pattern, citation and country of publication. The indicator citation per publication was successfully applied to evaluate the impact of number of authors, countries and journals. The result found that the mean value of citations per publication of collaborative papers was higher than that of single country publication. The same indicator may be applied to web environment.

\( h \)-index is an indicator to measure the research productivity through number of citations and total number of publications. \( h \)-index may be defined as “A scientist has index \( h \) if \( h \) of his or her \( N_p \) papers have at least \( h \) citations each and the other \( (N_p - h) \) papers have \(< h \) citations each” (Hirsch, 2005).

#### 2.7.4.9.2 Web Co-Citation Analysis

Co-citation analysis is a unique method for studying the cognitive structure of science. It is used to map the topical relatedness of clusters of authors, journals, articles or webpages on a particular field. Web co-citation analysis (WCA) is relatively new technique based on pairing web colinks instead of bibliographic co-
Website co-citation is nothing but URLs of two websites cited in a paper. Yang and Xiong (2009) made an exploring study to apply the co-citation in the analysis of academic websites, where Chinese top 100 universities' websites were taken up as an example in the empirical analysis. The relationship among websites visualized through Pajek. The results showed that co-citation approach is efficient tool to reveal the relatedness among websites.

2.7.4.9.3 Web Coupling

Kessler (1963) proposed a technique, known as bibliographic coupling to measure the similarity between two scientific documents in terms of number of citations they make in common. Small (1973) proposed another technique, co-citation to measure the similarity between two documents as the number of common documents that cites both documents.

These two concepts may be extended in the web environment, where citations may be replaced by hyperlinks. Therefore, the concept of bibliographic coupling may be analogous to Web coupling. Thelwall and Wilkinson (2004) attempted to find out similar academic websites using links, bibliometric couplings and co-links techniques. An experiment with a random sample of 500 pairs of domains from the UK academic domains revealed the similar academic websites using a combination of all three (links, colinks and couplings).

2.7.4.10 Alternative Document Model (ADM)

Alternative document models (ADMs) were created with the purpose of reducing the extent to which anomalies occur in counts of web links at the page level, and have been used extensively within webometrics as an alternative to using the web page as the basic unit of analysis. All previous web link studies used webpage as primary source document for counting purpose until Thelwall (2002b) argued that this was not necessarily ideal. It is due to the fact that individual webpages are often the only choice if search engines are used for collecting raw data. The original ADM exploited simple URL-based heuristic to automatically merge webpages for counting purposes (Thelwall, 2002b). Payne and Thelwall (2009) experimented through webometric analysis of universities of UK, Australia and New Zealand to ascertain which ADM model gives the most relevant result. The findings from the study showed that domain ADM gives the most consistent results than using page ADM as the standard units of measurement.

2.7.5 Webometric Tools and Techniques

Webometric research is based on web-based data, which may be collected through either commercial search engines or personal web crawlers. These are as follows:

2.7.5.1 Commercial Search Engines

Commercial search engines are extensively used in Webometric research, although there is inconsistency of search engine results (Bar-Ilan, 1999, 2004b; Rousseau, 1999) and also are biased in coverage (Vaughan & Thelwall, 2004; Vaughan & Zhang, 2007). Vaughan and Zhang (2007) made a study to examine search engine
coverage of websites across countries and domains. Websites of commercial, educational, governmental, and organizational were selected from four countries i.e. U.S., China, Singapore, and Taiwan. Total 1664 sampled sites were examined using four major search engines like Google, Yahoo!, MSN, and Yahoo! China. The study found that US websites got higher coverage rates than other countries. The language of website did not affect much. Yahoo! China was reflected to have better coverage of websites from China than global counterpart. Webometric research used hit count estimates as raw data for many studies of web information (Ingwersen, 1998). Uyar (2009) made an investigation to know the accuracy of search engine hit-counts for search queries using single and multiple terms.

2.7.5.2 Personal Web Crawlers

Two popular academic web crawlers i.e. SocSciBot3 and LexiURL4 are freely available on the Web. These softwares are developed by Mike Thelwall, University of Wolverhampton, UK in order to find out alternative link analysis strategy. SocSciBot is a crawler designed for webometric research. Together with supporting programs like SocSciBot Tools and Cyclist, SocSciBot3 can be used to conduct link analysis research on websites to produce standard statistics about the interlinking and network diagrams.

LexiURL Searcher has been replaced by Webometric Analyst, which automatically analyses webpages and creates network diagrams of collections of websites. It automatically submits queries to search engines and process the results. The functional difference between LexiURL and Webometric Analyst is that LexiURL does the hyperlink search whereas Webometric Analyst does the citation searches. In social science, informetric data about WWW collected by Web crawling (Cothey, 2004). Blinker, another web crawler was used to extract data to find the number of webpages and domains for large number of universities. Data were analysed using Ucinet 6.109 to build network graphs using NetDraw 2.28.

2.7.6 Web Presence

Web is the best showcase for universities. Web presence can be measured through various web indicators like webpages, inlinks, self-links and total links, rich files, and scholarly publications. Besides, age of websites, space occupied by a particular domain in the whole web space, PageRank etc. were also useful indicators for measuring web presence.

Aguillo, Ortega, and Granadino (2006) made an analysis of Web presence of the universities through cybermetric indicators. The developing countries in Latin America are making a great effort for publishing electronically their academic and scientific result. The websites of Brazilian universities regarding their web presence were analysed through visibility and domain size. The result showed that there was tremendous increase in the commitment of the Brazilian universities to the Web.

3 Web crawler and link analyser for the social sciences. <http://socscibot.wlv.ac.uk/>
4 Webometric analyst web analysis software. <http://lexiurl.wlv.ac.uk/>
Noruzi (2006) investigated the web presence and WIFs of Middle-Eastern countries through ccTLDs and SLDs for academic institutions. The result showed that as a whole the web presence of Middle-Eastern countries is low except for Turkey, Israel and Iran.

Jalal, Biswas and Mukhopadhyay (2010b) focused on the web presence and visibility of websites of Asian countries using some webometric indicators like Internet access, webpages, number of Internet users, and link counts. The study analysed the web presence using popular search engines like Altavista, Google, Yahoo! and MSN. An attempt was made to find out the WIFs for selected Asian countries. The result showed that China (43.7%), Japan (16.2%) and India (10.4%) were in forefront amongst Asian countries based on the total number of effective Internet users. Japan is having the highest number of webpages followed by China and South Korea.

### 2.7.7 Web Impact Factor (WIF)

Ingwersen (1998) introduced the concept of WIF, which made a remarkable advancement in webometric research. WIF is defined as the ratio between the number of total links and total webpages of a specific domain. Total number of links consists of internal links plus external links. For a large website, there are good numbers of internal links which are actually used to organize the internal pages and did not deserve any credit for the reputation of a website. Links from external pages give more useful information and thus enhance the reputation of the institute. Ingwersen’s study proposed three types of WIF: internal, external and overall. The early development of the calculation of WIFs found to be crude and the results was not satisfactory (Smith, 1999; Thelwall, 2000; Thomas & Willett, 2000; Björneborn & Ingwersen, 2001).

Thelwall (2000) extended the concept of impact factor for web-based resources and used the power of search engines to cover other domains on the Internet. He conducted a survey in order to test the coverage of search engines before calculating WIF, which was found sometimes extremely uneven and leads to misleading calculation. Thelwall (2001c) found in a study through search engine that results are problematic because of the variable impact factors; sometimes, it does not correlate with the research rating. Li (2003) made an exhaustive review of the development of WIF and studied hyperlinks extensively by applying existing bibliometric methods. Li pointed out the origin of WIF and techniques for data collection.

Smith and Thelwall (2002) calculated WIFs for Australian universities using both web crawlers and search engine through links between UK, Australian and New Zealand universities. Number of webpages at the site and academic staff members were used as measure of the size of the universities. Some of the methodological issues were discussed for the calculation of WIFs. Thelwall (2002a) made comparison of sources of links for academic web impact factor calculations. It had been demonstrated that several versions of the metric could produce results that correlate with research rating of British universities.
Noruzi (2005, 2006) investigated the WIFs for Iranian universities and Middle-Eastern countries respectively and introduced a new system of measurement. Counting links were made from the output of AltaVista search engine: WIF is calculated dividing link counts by pages for each university. These WIFs were then compared to study the impact factor, visibility and influence of websites of Iranian universities.

Amipour and Payam (2007) studied the websites of forty universities of Iranian Ministry of Health on the web and calculated their WIFs. These WIFs were used to rank these universities and found that there is significance different in the result obtained through WIFs. Elgohary (2008) made a study in order to investigate the WIF of Arab universities using 99 universities representing 20 Arab countries through the advanced search facility of AltaVista. Data were retrieved for links as well as the web presence of the universities. The findings revealed that Jordanian universities represent 40% of top ten universities based on revised WIF, whereas the same result was not found from the point of view of web presence. A strong correlation between external links and web presence was found from the study.

2.7.8 Webometric Ranking

Webometric ranking implies the ranking of institutes or organizations through web indicators by analysing their websites. There are basically two types of ranking: academic ranking and webometric ranking. Ranking may be at the national level or global level.

2.7.8.1 National Level Ranking

The first public ranking system was brought out for universities in US by “US News and World Report” in 1983. At present, more than 50 countries have their own ranking system and the number is still growing day-by-day. US News and World Reports. USA: Ranking by Sunday Times. Ireland and UK: Asahi Shimbun. Japan: NAAC. India: Centre for Higher Education Development (CHE) Ranking. Germany are some examples.

The methodology and selection of indicators are the main reasons for the successfulness of any ranking system. Aguillo, Ortega and Fernandez (2008) have shown that there is huge academic digital divide between higher education institutions in United States, European Union, and of course Asian countries. Hence, the ranking system using web indicators may not reflect the true picture about the performance of the universities.

Jalal, Biswas and Mukhopadhyay (2009b) focused on the importance of webometric indicators to judge the performance of the websites of Indian universities. Also, it was discussed various webometric indicators by different methodologies of ranking in terms of their advantages and disadvantages. The study reported that University of Hyderabad occupied the top position in Southern Region through RWWU, NAAC and 4ICU ranking systems. The study reviewed the relevance of Web indicators to obtain realistic ranking for all universities. Jalal, Biswas and Mukhopadhyay (2010a) made an attempt to rank central universities in India using appropriate webometric indicators.
Chapter 2: Review of Literature

2.7.9 Link Topology

Topology is the physical layout of a network. Web link topology is the linked structure of the World Wide Web. Similarly, link topology of websites is the linked structure of the particular websites. Web link topology is a complex phenomenon. To understand this concept clearly, visualization techniques are used. Gibson, Kleinberg and Raghavan (n.d) developed the concept of hyperlinked communities on the WWW through an analysis of the link topology.

2.7.9.1 Graph Theory

To explain the topological structure of a network, graph theory may be applied. A graph is a mathematical representation of a network in which nodes are connected by its edges (Björneborne, 2004). Graph may be directed or undirected one. In a directed graph, the edges represent directional relation between the nodes, whereas in an undirected graph, the direction of links is ignored. Graph theoretical approaches are used to analyse the citation networks in bibliometrics and Scientometrics (Shepherd, Watters & Cai, 1990; Fang & Rousseau, 2001; Egghe & Rousseau, 2002).

2.7.9.2 Visualization Tools

Graphical representation is one of the best techniques to represent large data in a manageable and understandable form. Many software are available to form the graph based on the data. One is ViAGraph, a tool for graph visualization and analysis. Another software is Pajek, which worked with SocSciBot to generate the graphs based on the data from SocSciBot tools (Batageli, 2010). NetDraw2.2 is another visualization tool, which may be used to build network diagram. Truong and Dkaki (2006) explained the various graphical techniques including ViaGraph and also proposed a new approach of node replacement based on geographic constraints.

2.7.9.3 Power Laws

There are few established bibliometric laws and techniques such as Lotka, Zipf and Pareto power laws. Lotka (1926) showed that productivity is highly concentrated within a small number of researchers. Pareto’s law is given in terms of cumulative distribution function (cdf). Rousseau (1997) identified some behavioural pattern on the web using AltaVista search engine and found that Lotka’s law is fitted very well. Rousseau (1997) has shown distribution pattern of websites, site links and site self links. Rousseau’s study used 343 websites as retrieved through AltaVista using query string: Informetrics or Bibliometrics or Scientometrics. The study showed that the distribution of TLDs and inlinks for the investigated sites followed the ubiquitous power-laws like Lotka’s distribution.

Huberman and Adamic (1999) identified power laws in the distribution of webpages. A power law is identified as the frequencies [n] of variable [x], which is proportional to 1/x^b. The power law is well fitted in the distribution of webpage. In Web, there is no boundary for connectivity. Scale free distributions of web inlinks
and outlinks supports long power law tails, which indicates that only few webpages have many links whereas majority has less.

2.7.9.4 Social Network Analysis

Social network analysis views social relationships in terms of network theory through nodes and ties. Nodes represent individuals or organizations whereas ties indicate the relationships between two nodes.

Ortega and Aguillo (2009) tried to visually display the most important universities in the world through the study, which described the Web relationship among universities of different countries and continents using the data set of 1000 universities from the list of ‘Ranking Web of World Universities’. It was also possible to build the network graphs and maps from the search engine data. Social network analysis techniques were used to analyse the structural properties of the network. An interesting result found that US dominates the world network, and within Europe the British and the German sub-networks stand out.

Krackplot 3.0 is graph layout software for social network, which allows the users to display the node’s attribute using colour and shape. It has menu driven interface with large set of hot keys to load data files, save files, create printer files in an easy manner. Krackhardt, Blythe and McGrath (2010) made a detailed study on features and facilities and utilization aspects of the Krackplot 3.0.

Small world is part of social network analysis. Small world topologies are deeply concerned with information science issues such as navigability and accessibility of information. Watts and Strogatz (1998) introduced a small world network model characterized by highly clustered node.

2.8 Trends of Webometrics in Indian Context

Though much contribution on Bibliometrics, Scientometrics and Informetrics were recorded by our Indian scientists but Indian contribution on Webometrics is not encouraging during last 12 years of the development of the subject. The pioneering work was done by Mukhopadhyay (2002), who tried to explore the possibility of research in the field of webometrics specifically in educational institutions in India. Indian Institute of Technology (IITs) and Indian Institute of Management (IIMs) were ranked using WIFs. Mukhopadhyay (2004) analysed hyperlinks at different levels of domain. The calculation of WIFs for ccTLDs of South Asian Association for Regional Cooperation (SAARC) countries; SLDs related to academic and research institutes registered under Indian ccTLD and hosts under IITs and IIMs were made.

Jalal, Biswas and Mukhopadhyay (2009a) made comprehensive review on webometrics in their paper ‘Bibliometrics to webometrics’. WIF is being calculated for the IITs and IIMs systems and the results were compared with the previous study conducted by Mukhopadhyay (2004).

websites of 27 universities of two prominent and largest states of India e.g. Uttar Pradesh and Rajasthan to find out facilities provided by them. They analysed authority detail, contact details, currency details, navigation links, user support links and status of universities. Rank of each university is obtained based on scores and graded from excellent to very poor. The study recommended that a committee might be formed to evaluate websites for future improvement.

Sutradhar (2008) made a webometric study on ‘NAAC scores and webometrics accredited ranks of Indian universities: An analytical study’, where it has been shown that there is a strong correlation between the NAAC score and RWWU system. Walia and Kaur (2008) reported the results of webometric investigation done on websites of selected library association’s websites of India. The author reported that the library associations could boost their web visibility by hosting a wide range of materials as much as possible and increase the number of inlinks from other similar websites.

Jalal, Biswas and Mukhopadhyay (2008) made a pilot study on the hyperlink analysis of universities in West Bengal. It concentrates on the critical analysis of the calculation of WIF with existing and new formulas in order to draw the reliable ranking based on the value of WIF. The main findings of the study show that based on the ranking of total links, Indian Statistical Institute (ISI) Kolkata possess the first rank followed by IIT Kharagpur and Calcutta University among 20 universities in West Bengal. On the basis of WISER ranking, IIT Kharagpur got the first place followed by ISI Kolkata and Bengal Engineering Science & University whereas Calcutta University got fifth place.

Rameshbabu, Jeyshankar and Rao (2010) made a webometric study on central universities in India. The study developed a network diagram showing link structures between web nodes in webometric analysis. Walia and Kaur (2010) made a study on government of India websites to review whether it is fulfilling its objectives using some webometric indicators and found that different ministries excelled from different point of view.

Pahari (2011) argued that Indian universities even elite institutions are not getting any significant position in the reputed internal ranking systems. The validity and reliability of grading the universities and colleges by NAAC and NBA was raised.

Two international conferences on webometrics were organized in India: (i) Third International conference on Webometrics, Informetrics and Scientometrics (WIS), during 6-9 March 2007 at New Delhi. (ii) Sixth International conference on Webometrics, Informetrics and Scientometrics (WIS) during October 19-22, 2010 at Mysore. Many issues on webometrics were discussed at the peripheral level.

2.9 Observations

Following section provides facet-wise summary of important works and observations.
Table 2.1 — Major areas with salient works in literature review

<table>
<thead>
<tr>
<th>Major Topics</th>
<th>Researchers</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Webometrics</td>
<td>• Almind &amp; Ingwersen (1996, 1997)</td>
<td>• New discipline ‘Webometrics’ has been recognized.</td>
</tr>
<tr>
<td></td>
<td>• Björneborn (2001, 2004)</td>
<td>• Various terminologies related to webometrics have been defined.</td>
</tr>
<tr>
<td></td>
<td>• Björneborn &amp; Ingwersen (2001)</td>
<td></td>
</tr>
<tr>
<td>Search Engines</td>
<td>• Bar-Ilan (1999)</td>
<td>• Search engine reports a fraction of web documents;</td>
</tr>
<tr>
<td></td>
<td>• Thelwall (2000)</td>
<td>• Search engine results are sometimes not reliable.</td>
</tr>
<tr>
<td>WIF</td>
<td>• Ingwersen (1998)</td>
<td>• Formula for the calculation of WIF has been introduced.</td>
</tr>
<tr>
<td></td>
<td>• Li (2003)</td>
<td>• Thelwall encouraged to use ‘total number of faculty members’ instead of webpage for the calculation of WIF.</td>
</tr>
<tr>
<td></td>
<td>• Noruzi (2005, 2006)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Thelwall (2001-2004)</td>
<td></td>
</tr>
<tr>
<td>Hyperlink Analysis</td>
<td>• Henzinger (2001)</td>
<td>• Alternative Document Model (ADM) have been introduced</td>
</tr>
<tr>
<td></td>
<td>• Thelwall (2001-2008)</td>
<td>• No. of inlinks has been given due weightage rather than self-links</td>
</tr>
<tr>
<td>Webometric Ranking</td>
<td>• Aguillo, Ortega &amp; Fernandez (2008); Jalal, Biswas &amp; Mukhopadhyay (2009, 2010)</td>
<td>• Size, visibility, Rich Files and Google Scholar are the basic webometric indicators for ranking:</td>
</tr>
<tr>
<td></td>
<td>• Yi &amp; Jin (2008), RWWU (2007)</td>
<td></td>
</tr>
<tr>
<td>Reasons for Hyperlink</td>
<td>• Kim (2000)</td>
<td>• Mainly, there are three factors— scholarly, social and technological reasons.</td>
</tr>
<tr>
<td></td>
<td>• Harrison (2002)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Wilkinson, et al (2003)</td>
<td>• Webmasters were more likely to create hyperlink to websites possessing practical content, information or services.</td>
</tr>
<tr>
<td></td>
<td>• Bar-Ilan (2004a)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Vaughan (2004)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Kousha &amp; Hori (2004)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Chu (2005)</td>
<td></td>
</tr>
<tr>
<td>Link Topology</td>
<td>• Björneborn &amp; Ingwersen (2001)</td>
<td>• Link topology among universities helps to identify the link pattern:</td>
</tr>
<tr>
<td></td>
<td>• Björneborn (2004)</td>
<td></td>
</tr>
</tbody>
</table>

After reviewing more than one hundred twenty research papers related to the concerned research topic, following observations may be mentioned here:

a. Bossy (1995) proposed new ways of measuring the impact of scientific contribution of authors;
b. Almind and Ingwersen (1997) introduced the application of informetric methods to the WWW. The term ‘Webometrics’ was coined by them in 1997;
c. Ingwersen (1998) developed the concept of Web Impact Factors (WIFs) in line with Garfield’s Impact Factor (IF);
d. Molyneux and Williams (1999) reviewed "Measuring the Internet" and highlighted some of the issues: which insists the development in Webometrics;

e. The scope of Webometrics broadly be categorized as: (a) Webpage content analysis, (b) Web technology analysis, (c) Web usage analysis, (d) Web link structure analysis;

f. Aguillo (2002) pointed out, "Webometrics is still in its infancy as a scientific domain with its own different, methods to be developed and problems to be solved."

g. Huberman and Adamic (1999) identified power laws in the distribution of webpage;

h. Rousseau (1997) identified some behavioural pattern on the Web using AltaVista search engine and found that Lotka’s Law is fitted very well;

i. Björnborn (2004) in his PhD thesis had focused on small-world properties in academic web spaces and laid s strong theoretical foundation on various aspects of webometrics, link terminologies and diagrams showing the relationship among Bibliometrics, Informetrics, Scientometrics and Webometrics;

j. Rowlands (2000) reviewed cybernetic approaches to quantitative aspects of documents production and use on the Internet within the bibliometric framework;

k. Bar-Ilan and Peritz (2002) give an excellent theories and methods for exploring the Internet with focus on general informetric techniques to be applied in the both web studies and non-web Internet research;

l. Webometric tools and techniques include both search engines and personal web crawlers as data collection tools; statistical techniques as data analysis; techniques related to citation analysis are equally applicable in the webometrics;

m. Björnborn and Ingwersen (2001) have studied the search engine coverage from the point of view of content analyses and also pointed out problems in measuring WIF;

n. After ten years of research, Ingwersen (2006) realized that alternative data collection methods and fusion techniques for performing reliable and sophisticated webometric analyses are highly required;

o. Another important webometric research area in academic web space is scientific e-journals available on the web (Harter & Ford, 2000). Since much of webometrics has been motivated by citation analysis, it is interesting to see whether such techniques can be applied to e-journals;


q. Noruzi (2006) investigated the web presence and WIFs of Middle-Eastern countries through ccTLDs and SLDs for academic institutions. The result showed that as a whole the web presence of Middle-Eastern countries is low except for Turkey, Israel and Iran;

r. Jalal, Biswas and Mukhopadhyay (2010b) focused on the web presence and visibility of websites of Asian countries using some webometric indicators like Internet access, webpages, number of Internet users, and link counts;
Chapter 2: Review of Literature

s. The first public ranking was brought out for US universities by “US News and World Report” in 1983:

r. Ranking Web of World Universities (2007) is an initiative taken by the Cybermetrics Lab, Spain for the academic and research institutions:

u. Academic Ranking of World Universities (ARWU), 2003:

v. Thomas and Willett (2000) studied the websites of LIS departments in UK and found that there is no significant correlation between inlink counts and research ratings:

w. Thelwall (2001b) showed that the counts of inlink of 25 UK universities correlated significantly with their average research productivity using five-yearly UK Government Research Assessment Exercise (RAE):

x. Chu (2005) applied two part process while developing classification of hyperlinks for inlinked pages with four categories i.e. teaching-learning (14%), research (13%), service (46%) and homepage (27%):

y. Wilkinson, et al (2003) conducted a survey on (.ac.uk) domain to find out the motivations for creating links between university websites. It was found that over 90% of links were created for broadly scholarly reasons; and

z. Chu (2005) found that less than 27% of the links were made out of research or teaching-learning motivation while Thelwall (2003) found the majority of linking motivations were trivial compared with citation motivations.

2.10 Summary

The literature review helped to make the remark that many researchers (Almind & Ingwersen, 1997, 1998; Björneborn, 2001, 2004; Bar-Ilan, 1999, 2002, 2004; Egghe, 2000; Rousseau, 1997) have conducted studies based on the belief that bibliometric and informetric laws can be applied to the Web. Web data are basically the instrument in webometric research. An important area in webometric research is hyperlink analysis, which is analogous to citation analysis in print environment (Smith, 2004). Webometrics gained its strength from other disciplines like computer science, library science, information science, theoretical physics and sociology. Many motivational studies (Wilkinson et al, 2003; Chu, 2005) were conducted to reveal the factors behind link creation. Several WIFs studies were conducted to calculate various WIFs for TLD, SLDs at institutional level to help their respective stakeholders. Webometric research outputs in global as well as Indian context were analysed specifically to find out the gap of webometric research.

Throughout the discussion in literature review, it was found that there were no comprehensive webometric studies available on Indian universities so far. No initiatives at the individual as well as organization level were taken up for ranking Indian universities. Even, there is no established sound methodology for ranking the universities at the regional, national and international level.

The next chapter deals with webometric tools and techniques for hyperlink analysis. Various hyperlink analysis techniques were also explained.
References


Björneborn, L. (2004). Small world link structures across an academic web space: a library and information science approach. (PhD thesis from Department of Information studies, Royal School of Library and Information Science, Denmark)


Chapter 2: Review of Literature


Chapter 2: Review of Literature


Matzat, U. (1998). Informal academic communication and scientific usage of
Chapter 2: Review of Literature


doi:10.1145/1459352.1459357


Information Science, University of Burdwan, Burdwan, W.B)


