KNOWLEDGE RESOURCES HERITAGE OF INDIA: A WEBOMETRIC ANALYSIS

ABSTRACT OF
DISSERTATION
SUBMITTED FOR AWARD OF THE DEGREE OF

Master of Philosophy

IN

LIBRARY AND INFORMATION SCIENCE

UNDER THE SUPERVISION OF

Dr. S. K. Sonkar

Assistant Professor

SUBMITTED BY

Uttkarsh

BABASAHEB
BHIMRAO
AMBEDKAR
UNIVERSITY

LUCKNOW

प्रज्ञा शील करुणा

ESTABLISHED 1996

DEPARTMENT OF LIBRARY AND INFORMATION SCIENCE
(SCHOOL FOR INFORMATION SCIENCE & TECHNOLOGY)
BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
(A CENTRAL UNIVERSITY)
VIDYA VIHAR, RAEBARELI ROAD LUCKNOW, UTTAR PRADESH

ENROLLMENT NO: 398/19

2020

Abstract

Introduction

Information technology revolutionized the way of communication and sharing of information and after the invention of Internet which boost the information technology on another level which help in communication and retrieval of information. In nearly 3 decades of the invention of the Internet, it becomes one of the very important part of day-to-day life of human life. Other than communication, Internet plays vital role in influencing the education and business. In the current era of human civilisation people are very keen to know about different culture and their heritage so it is necessary for the websites of these resources put relevant and updated information.

The World Wide Web is a collection of documents linked over the Internet. WWW work as a node which connects the information to one another through hypertext. The information/ document on the web is written or layout, is in Hypertext Markup Language (HTML). From one web page to another web page which creates link between them is called as Hyperlink and these hyperlinks are called as hypermedia when images or any other multimedia is attached with the hyperlink.

Webometrics

Metric studies in Library and Information Science was started long back. In the field of metric studies Librametrics, Bibliometrics, Scientometrics and Informetrics are well known. There are vast studies has been carried out on these areas of metrics studies. Those metrics studies were all about to measure the scholarly communication, to identify research trend, to find the growth of knowledge, etc.

Since the mid-1990's increasing effort have been made to investigate the nature and properties of the World Wide Web by applying modern informetric methodologies to its space of contents, link structures and search engines (Bjorneborn and Ingwersen, 2001). In 1997, Almind and Ingwersen used the term 'webometrics' for the first time and identified that to measure the documents and information, web would be the important source (Almind and Ingwersen, 1997).

In other words, we can say that quantitative analysis of the web sources is called as webometrics. Webometricians are the person who carry out the webometric

study. The term ‘cybermetrics’ and ‘nettometrics’ used contemporary to the ‘webometrics’.

Earlier webometrics also considered as and still relevant that it is a application of bibliometric techniques to measure the different aspects of the websites. The measuring aspects of the websites as a webometrics study are such as number of web pages, keywords, hyperlinks, web search engine results, traffic on websites, pageviews, in-links, etc.

Bjorneborn and Ingwersen (2004) given webometrics accepted definition as “the study of web-based phenomena using quantitative techniques and drawing upon informetric methods”.

Thelwall (2009) defined webometrics as “the study of web content with primarily quantitative methods for social science research goals using techniques that are not specific to one field of study”.

Web Impact Factor

Web Impact Factor is an average frequency of links of a particular website. WIF was developed by Ingwersen in 1998. Web impact factor work as a quantitative indicator of the website. It is a methodology of webometric studies. Web impact factor of websites vary from one search engine to another, so it can act as an indicator for evaluation of search engine performance too.

There are four types of links, so there are four types of WIF:

I. **Simple Web Impact Factor**: It is a rate of links/ hyperlinks to the total number of web pages in the website.
II. **Self-link Web Impact Factor**: It is a rate of self-links or internal links within in the website to the total number of web pages in the website.
III. **External Link Web Impact Factor**: It is a rate of external link to the website, to the total number of web pages of the website.
IV. **Revised Web Impact Factor**: It is a rate of in-links in the website to the total number of web pages of the website.

Statement of the Research Problem

Objectives

Following are the objectives of the study:

1. To classify the domain of the websites.
2. To calculate the number of domain authority and page authority of the websites.
3. To calculate the external and internal links of the websites of the Knowledge Resources Heritage.
4. To find out the number of discovered and lost linking domains.
5. To evaluate the simple Web Impact Factor, self-link Web Impact Factor, external web
   impact factor and revised web impact factor.
6. To reveal the world traffic rank, pageviews and in-links through Alexa Web Ranking
   Tool.
7. To find out the Google PageRank of the Websites.

Hypotheses

Hypotheses for the study are as follows, on the basis of objectives:

1. Websites have different types of web domain name.
2. There are less lost linking domains of the websites.
3. There is significant difference between internal and external links among the websites.
4. There is similarity between self-link web impact factor and external web impact factor among the websites.
5. Global traffic rank of among the websites is significantly different using Alexa Web Ranking Tool.

Scope of the Study

The scope of the study includes the 16 websites of Knowledge Resources Heritage under Ministry of Culture, Government of India. Following are the lists of the Knowledge resources Heritage-

Table 1.1

Knowledge Resources Heritage of India

S. No.	Knowledge Resources Heritage	Website Links	Est. Year	Location
1.	Rampur Raza Library	http://razalibrary.gov.in/	1774	Rampur, Uttar Pradesh
2.	Asiatic Society	https://www.asiaticsocietykolkata.org/	1784	Kolkata, West Bengal
3.	National Archives of India	http://nationalarchives.nic.in/	1891	Kolkata, West Bengal
4.	Khuda Baksh Oriental Public Library	http://kblibrary.bih.nic.in/	1891	Patna, Bihar
5.	Thanjavur Maharaja Serfoji's Sarasvati Mahal Library (TMSSML)	http://www.tmssmlibrary.com/	1918	Thanjavur, Tamil Nadu
6.	Anthropological Survey of India (ANSI)	https://ansi.gov.in/	1945	Kolkata, West Bengal
7.	Delhi Public Library (DPL)	https://dpl.gov.in/	1951	Delhi
8.	National Library	https://www.nationallibrary.gov.in/	1953	Alipore, West Bengal
9.	Central Reference Library	http://crlindia.gov.in/	1955	Kolkata, West Bengal
10.	Nehru Memorial Museum and Library (NMML)	http://nehrumemorial.nic.in/en/	1964	New Delhi
11.	Central Institute of Higher Tibetan Studies (CIHTS)	https://cihts.ac.in/webpage/index.aspx	1967	Sarnath, Uttar Pradesh
12.	Raja Rammohun Roy Library Foundation (RRRLF)	http://rrrlf.nic.in/	1972	Kolkata, West Bengal

13.	Indira Gandhi Rashtriya Manav Sangrahalaya (IGRMS)	https://igrms.gov.in	1978	Bhopal, Madhya Pradesh
14.	Gandhi Smriti and Darshan Samiti (GSDS)	https://www.gandhismriti.gov.in/	1984	New Delhi
15.	Maulana Abul Kalam Azad Institute of Asian Studies (MAKAIAS)	http://makaias.gov.in/	1993	Kolkata, West Bengal
16.	Central Institute of Himalayan Cultural Studies (CIHCS)	http://www.cihcs.edu.in/	2003	Dahung, Arunachal Pradesh

Methodologies

The study is aimed to analyse the 16 websites of knowledge resources heritage of India webometrically. Few Search Engine Optimization (SEO) tools were used to collect requird data such as Open Site Explorer (www.opensiteexplorer.org), Alexa Internet (www.alexa.com/siteinfo/) and Check PageRank (https://checkpagerank.net/). Google search engine (www.google.com) was also used to collect required data using Boolean formula.

Significance of the Study

In recent year tourism in India increased exponentially and people visits the country to understand the culture. It might be possible that people before visiting the place they search on Internet about the places to visit, also researcher and information seeker access websites to get the desired information about the heritages. So, in this context it is necessary to evaluate the effectiveness of the website functioning, its structure and its content.

Chapter 2- Review of Literature

The literatures related to the subject was reviewed deeply. The literature review were categorised in 4 category, namely- Development of Webometrics- which deals with the evolution of webometric concept; Web Link Analysis- which deals with the analysis of different types of links like internal links, external links, out-links, in-links,

and so on; Web Impact factor- which deals with the literatures related to all four WIFs; and Website Evaluation- which deals with the literature about evaluation of websites which includes the study of traffic rank. The literature of all 4 categories are reviewed from the beginning of the conception to till date and arranged chronologically.

Chapter 3- Knowledge Resources Heritage: Profile

In this chapter, there is brief information about the Ministry of Culture under Government of India, its formation since beginning and the mission and objectives. The Ministry of Culture is mandated to preserve and conserve the cultural heritage of India and it has divided all cultural heritage into 3 groups namely- Tangible Cultural Heritage, Intangible Cultural Heritage and Knowledge Resources Heritage. This research is about Knowledge Resources Heritage which is responsible to make policy for the development of libraries. The National Archives of India is governing body of this group and this group is about those cultural heritage which have archival records. There are 16 Knowledge Resources Heritage which was briefly discussed in the Chapter.

Chapter 4 – Data analysis and Interpretation

This chapter gives the insights of the tools used in data collection and deals with the analysis of the results and interpretations of the results. There were various tools used to collect the required data like Link Explorer, Google Search Engine, Alexa Web Analytics Tools and Google PageRank to collect the data and the collected data is analysed quantitatively to prove the hypotheses framed for the study. Webometric indicators are used to rank the website as per their popularity and ranked according to their global traffic ranking.

Findings and Discussions

The analyses, interpretation and evaluation of the websites of Knowledge Resources Heritage of India was according to objectives formulated in the beginning of the study. The major findings drawn from the study are as follows:

1. While going through the list of Knowledge Resources Heritage on website of Ministry of Culture, Government of India, the name of the knowledge resources heritage were not correct such as Central Institute of Higher Tibetan Studies was named as Central University of Higher Tibetan Studies while Thanjavur Maharaja Serfoji Sarasvati Mahal Library was named as Thanjavur Maharaja Serfoji Mahal Library. It is necessary for the authority to provide correct information to the users to maintain the credibility of information given on ministry website.

2. Out of 16 websites, 14 websites of knowledge resources heritage were using restricted gTLDs such as .gov, .nic, .edu and .ac and all these websites had .in ccTLD. While website of Asiatic Society and Thanjavur Maharaja Serfoji Sarasvati Mahal Library were using gTLD of .org and .com respectively. Websites having ccTLD is of more benefit for them in getting higher SEO value on country-based search result and also SERP give search result related to the website which have same local ccTLD. Another benefit is, the local user visits the website with confidence of getting authentic information and it also increases the credibility of websites in local people.

3. Domain Authority of Khuda Baksh Oriental Public Library was highest among the 16 websites i.e. 62. Google identifies those websites having high domain authority as a trustworthy websites and Google take it as a factor which helps in determining the rank of websites on SERP. High domain authority helps the website in getting good position in search engine result pages.

4. Again, out of 16 websites Khuda Baksh Oriental Public Library had scored highest page authority of 50 and secured the rank 1st. It has same benefits as of domain authority. To increase page authority of any web page of a particular website, it is necessary to create new pages with high quality content with detailed and expanded information. Also, other web pages should be updated regularly.

5. (i) Khuda Baksh Oriental Public Library had highest number of internal links which was about 3,10,86,062 and secured the rank 1 among the 16 websites.

(ii) In case of external links, Khuda Baksh Oriental Public Library secured the rank 1 in the list of external links of the 16 websites of knowledge resources heritage, with total external links of 21,12,980.

(iii) The website of Khuda Baksh Oriental Public Library possesses highest number of total links that is about 3,31,99,042 and secured rank 1.

High number of internal and external links improve the search engine optimization of website (SEO). Internal links helps in engaging the user to the website through relevant and quality information which improves the page views so as, improves the PageRank which directly boost the position of website on search engine result pages.

Website do have good amount of external links but it doesn't mean that add any external links to the pages. The external links must have of good quality, informative and trustworthy, which helps in improving the credibility of the website. Also, valuable external links improve the authority of the website which directly improve the position of website on SERP.

6. (i) Simple web impact factor of all 16 websites of knowledge resources heritage was calculated and it was found that Anthropological Survey of India had highest simple WIF of 5,437.88 and secured rank 1st .

(ii)Asiatic society had scored the highest self-link web impact factor of 0.094 and secured 1st rank among the 16 websites.

(iii) Central Institute of Higher Tibetan Studies scored the highest external link web impact factor of 0.686 and ranked 1st .

(iv) Website of Thanjavur Maharaja Serfoji's Sarasvati Mahal Library has scored highest revised web impact factor of 0.568 between the 16 websites and occupied the 1st rank.

The website having high web impact factor show the 'international visibility' of the website i.e the popularity of website. So, the website with high WIF their visibility will be high and vice versa. It helps in comparing the website of same field to improve the performance of underperforming website.

7. Using Alexa Internet, Global Traffic Ranking of the 16 websites were evaluated and Khuda Baksh Oriental Public had the highest traffic rank of 2,469 among the 16 websites also Khuda Baksh Oriental Public Library had Indian Traffic Ranking of 207 and ranked 1st among the 16 websites. The website having high Alexa rank shows the popularity of the website and it also helps users to trust the website. High popularity also means that the website contains good quality content so other website place links of the website in their content which helps the website in getting external link which effects the position of website in SERP.

8. In case of PageViews, Nehru Memorial Museum and Library had highest PageViews of 4 and secured 1st rank among the 16 websites of knowledge resources heritage. High pageviews denotes the website has good quality information. Through pageviews, website come to know which page is seen most so other website can provide information similar to that to increase its pageviews.

9. Khuda Baksh Oriental Public Library had highest number of In-links that is about 3646 which was calculated using Alexa Internet and ranked 1st among the websites of other knowledge resources heritage. In-links also known as internal links. High number of in-links helps website in getting high rank in SERP. When we add more internal links to the website, Google index those web pages or internal links after crawling which improves the indexing of the website in Google database which helps the internal links getting good position in SERP.

10. The 16 websites of knowledge resources heritage were also evaluated on the basis of Google PageRank and it was found that Khuda Baksh Oriental Public Library had Google PageRank of 6 out of 10 which is highest among the selected websites. The websites need to improve their PageRank by publishing content regularly. High PageRank of website helps Google to index the content super-fast i.e. within minutes of publishing. It also increases the credibility of websites.

5.3 Suggestions

1. Ministry of Culture must cross-check the information on website before making it public.
2. Websites must provide ‘last date of update’ to show the currency.
3. Websites should provide multilingual accessibility for people who have limited English proficiency.
4. The websites which have small number of internal links must add more contents to its website, so increased number of internal links helps the websites to get topped in the search engine result pages.
5. Websites must provide detailed and relevant information on the webpages which attracts more external links which influence the websites ranking in SERPS, domain authority and page authority.
6. To increase the pageviews, websites have to update the information in real time and it also help in getting more traffic which improves the websites Global Traffic Ranking.
7. Websites should use a smaller number of ‘nofollow’ tag HTML code, which helps the crawler to index pages in their search engine directory which increase the visibility of the webpages on the search engine result pages at top.
8. Websites must improve their follow internal links and follow external links which help in getting better web impact factor.
9. Websites should interlink themselves for better flow of information to the visitors and it will increase the traffic on the website.
10. Webmaster of the websites use Search Engine Optimization (SEO) tools to check the performance of the website over the search engine and make necessary improvement in the website for better performance.
11. The websites of knowledge resources heritage should interlink to each other in order to share information and enhance websites performance.
12. Websites should provide feedback facility for users to give suggestion to enhance the website better than before.
13. Important announcements, changes in policy, starting of new services, setting up of a new facility are to be quickly brought to the notice.

Scope for Further Research

The present study includes only web link analysis and web impact factor. Following are the areas for further research:

1. Web content analysis of websites of Knowledge Resources Heritage of India.
2. Web usage analysis of websites of Knowledge Resources Heritage of India.
3. Impact of Blogs, Twitter, Facebook Pages, Reviews and Testimonials on Knowledge Resources Heritage of India: A Webometric Study

Conclusion

There is number of webometric studies regarding library and organisations websites but this study is the only webometric study of the websites of the Knowledge Resources Heritage of India. All the knowledge resources heritage has their web presence. The study is done in very precise manner. There are many websites need necessary improvement because they performed very poorly in the study which hampering their visibility and accessibility over the search engine. The findings and suggestions will be useful for the webmaster of the websites to improve their performance.