INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY & CREATIVE ENGINEERING (ISSN:2045-8711)
VOL.1 NO.8 AUGUST 2011

An Efficient Approach for Discovering Impact Factor of E-Books using Eigen Factor and UCINET

Ms. D. Saraswathi# Dr. A. Vijaya Kathiravan* Mr.G.Sivakumar# Ms.R.Kavitha+
#Lecturer in Computer Science, K.S.Rangasamy College of Arts & Science (Autonomous), Tiruchengode-637215, Namakkal, TN, INDIA.
** Assistant Professor in Computer Science, Government Arts College, Salem-07, TN, INDIA.*
+Assistant Professor, Dept. of MCA, AIMIT, St. Aloysius College(Autonomous), Mangalore, Karnataka, India
saraswathi.ds@gmail.com, vijayakathiravan@gmail.com,sivakumarsen@gmail.com, kavitha.rajamanii@gmail.com

***Abstract*-An Impact Factor is one measure of the relative importance of a journal, individual article or scientist to science and social science literature and research. Each index or database used to create an impact factor uses a different methodology and produces slightly different results, revealing the importance of using several sources to judge the true impact of a journal's or scientist's work. In the web environment, impact factor is measured through the number of hyperlinks counts and number of WebPages. The concept of self-citation is replaced by self-links, i.e., the links within the websites and citation is replaced by in-links, i.e., the links coming outside the websites. As we know, WIF is the logical sum of external and self-link WebPages divided by number of web pages found on that particular websites. There are number of way to find the impact of journal, paper, and Web sites etc. In this proposed system is going to find impact factor of E-books by using EigenFactor and the links of E-books represented by using UCINET software. The link of E-books can identify based on the degree, betweenness of the link. This system is used to measure the quality of E-books and to know how many of them referring the E-book. Most of the E-books are downloaded from the Web or require pages can read from the Web site itself. This was done by means of a citation analysis and a reader survey. For the citation analysis, impact factor, citing half-life, number of references per article, and the rate of self-references of a periodical were used as indicators. Webometric data have been collected through Yahoo! And Google search engines using special query syntax.**

Keywords: **Scientometrics, Webometrics, Cybermetrics, Bibliometrics, impact factor.**

I. INTRODUCTION

The World Wide Web has now become one of the main sources of information on academic and research activities, and therefore it is an excellent platform to test new methods of evaluating webometric activities. However the world scientific community has not yet accepted the Web as a full supplement or a complement to traditional scientific publishing. The science of webometrics (also cybernetics) tries to measure the World Wide Web to get knowledge about the number and types of hyperlinks, structure of the World Wide Web and usage patterns.

According to Björneborn and Ingwersen (2004) [11], the definition of webometrics is "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." Webometrics is (a) a set of quantitative techniques for tracking and evaluating the impact of web sites and online ideas and (b) the information science research field that developed these ideas. Webometric techniques include link analysis, web mention analysis, blog analysis and search engine evaluation, but from the perspective of digital library evaluation the main method is link analysis. Why can analyzing web hyperlinks help evaluate digital repositories? The reason is that the links to a web site can reveal useful information about how popular it is, which pages or resources are the most popular, why it is popular and where it is popular. Whilst all this information can also be gained from web server log file analysis, the latter can normally only be

conducted with permission of a site's webmaster. In contrast, link analysis can be applied to any web site. This means that link analysis can be used to evaluate a web site by comparing it to its competitors or to similar web sites and can also be used to identify missed audiences for a site. Links can reveal information about web sites because each link to a web site may be created to direct visitors to it. The link author believes that the target site is important or useful. For example, the course pages for an archaeology degree may contact links to the New Library of Alexandria for its images of ancient Egyptian artefacts. From the opposite perspective, discovering all the links to the New Library of Alexandria web site would give useful insights into who was using it and why. Of course, most people using a web site will not create a link to it but a link analysis can still give indicators about likely users and uses.

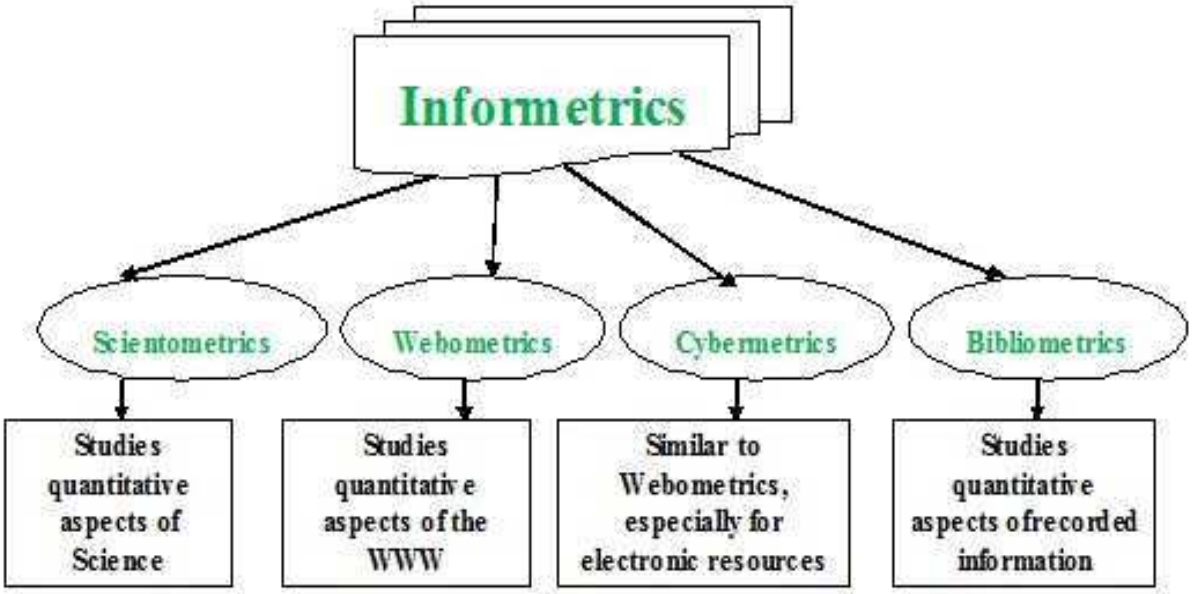

Fig 1: Taxonomy of Informetircs

II. TAXONOMY OF INFORMETIRCS

Informetrics is the study of quantitative aspects of information. This includes the production, dissemination and use of all forms of information, regardless of its form or origin. As such, informetrics encompasses the fields of

- Scientometrics, which studies quantitative aspects of science.
- Webometrics, which studies quantitative aspects of the World Wide Web.
- Cybermetrics, which is similar to webometrics, but broadens it's definition to include electronic resources.
- Bibliometrics, which studies quantitative aspects of recorded information.

Informetrics can be classified based on the specialization as given below in Fig 1.

I. Bibliometrics Analysis

Bibliometrics is the application of mathematical and statistical methods to publications (from biblos: book and metron: measurement). Bibliometrics is often used to assess scientific research through quantitative studies on research publications [7]. Bibliometric assessments are based on the assumption that most scientific discoveries and research results eventually are published in international scientific journals where they can be read and cited by other researchers. The number of citations to a journal article can be considered to reflect the article's impact on the scientific community. Applied bibliometrics, as it is used today, analyses the number of scientific articles published by a selected number of authors, citations to these articles and connections between articles, authors and subjects.

II.Scientometrics Analysis

Scientometrics is the science of measuring and analyzing science. In practice, scientometrics is often done using bibliometrics which is a measurement of the impact of (scientific) publications. It is part of sociology of science and has application to science policy-making. It involves quantitative studies of scientific activities, among others, publication, and so overlaps bibliometrics to some extent [10]. Bibliometrics and scientometrics are two closely related approaches to measuring scientific publications and science in general, respectively. In practice, much of the work that falls under this header involves various types of citation analysis, which looks at how scholars cite one another in publications. This data can show quite a bit about networks of scholars and scholarly communication, links between scholars, and the development of areas of knowledge over time.

III.Webometrics Analysis

In the wake of Internet/Web developments, some bibliometricians drew analogies between Webbased and research documents and came up with the idea that the scientific content of the Web could be analysed in the same way as the science journal system. Webometrics is the quantitative analysis of web phenomena, drawing upon informetric methods, and typically addressing problems related to bibliometrics. Webometrics [12] was triggered by the realisation that the web is an enormous document repository with many of these documents being academic-related [12]. Moreover, the web has its own citation indexes in the form of commercial search engines, and so it is ready for researchers to exploit. In fact, several major search engines can also deliver their results automatically to investigators' computer programs, allowing large-scale investigations. One of the most visible outputs of webometrics is the ranking of world universities based upon their web sites and online impact [6]. Webometrics includes link analysis, web citation analysis, search engine evaluation and purely descriptive studies of the web. Webometrics, a modern, fast-growing offshoot of bibliometrics is reviewed in detail [9].

III.THE PROPOSED ARCHITECTURE FOR DISCOVERING IMPACT FACTOR OF E-BOOKS

In order to properly to know, we first have to define the impact factor. The ISI (the Institute for Scientific Information) impact factor was designed by Eugene Garfield around 1960 as a means to measure the impact of a specific journal, and it gives an average value on how many times an article in the journal has been cited. It is defined as the average number of citations given in a specific year to documents

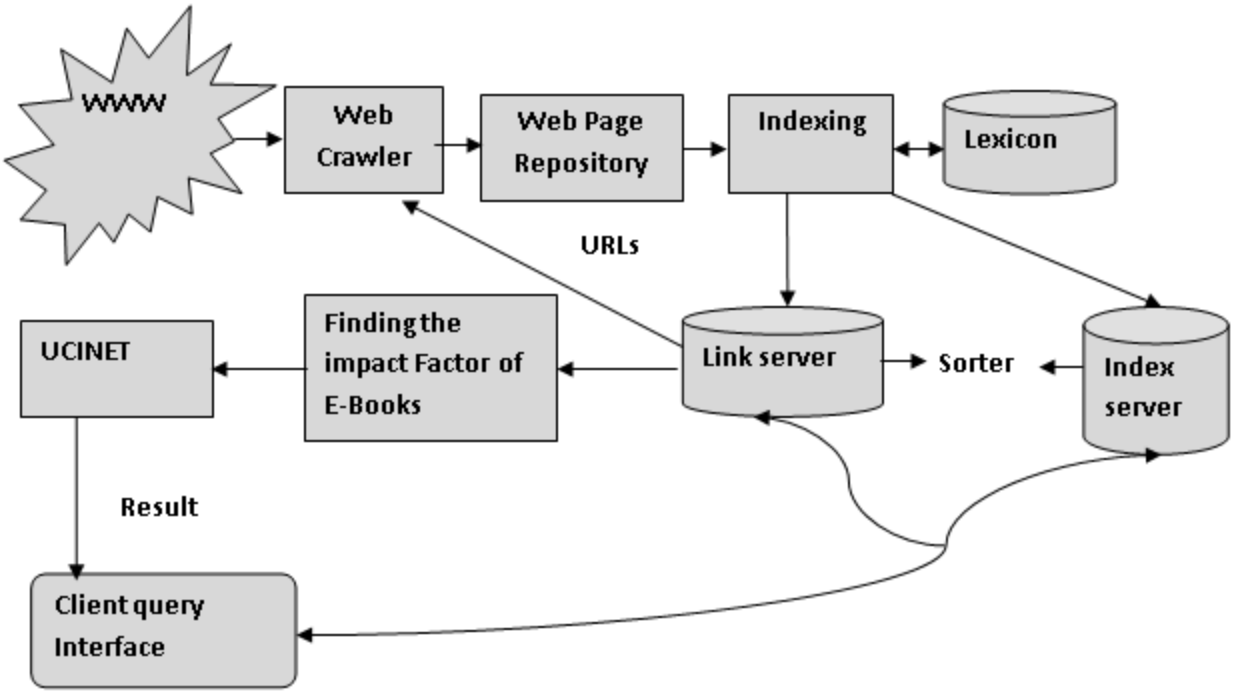

Fig 2: Proposed Architecture for Discovering Impact Factor of E-Books

published in that journal in the two preceding years, divided by the number of documents published in that journal in those two years. In this proposed system is used to find out the impact factor of E-Books and determine the quality of their EBooks. The E-books available in various web server of the web are downloaded by the crawler using either depth first crawler or breadth first crawler and the document corpus collected are stored in the web page repository i.e., a web warehouse. The E-books URL addresses in the web repository are indexed using indexing methods. The URL addresses and the structural connectivity of the E-books URL addresses are stored in the link server and the indices with respect to their E-book ID are stored in the index server The E-Books of impact factor is calculated based some parameter such as E-Books count, Citations and impact factor, and Co-citation by using EigenFactor and related links of that E-books represented in diagrammatically by using UCINET. This is done by using UCINET Software [13], measure centrality of degree and betweenness in network. It is based on the set of the users most cited E-books and the number of citations that they have received in other users. The impact factor of E-books is returned to the client and it is possible to measure the quality of their E-books. Fig 2 gives an overview of this architecture.

IV. RELATED WORKS

Jalal, Biswas and Mukhopadhyay [1] had shown in their hyperlink study for the state universities of West Bengal that IIT Kharagpur occupied the first rank among the universities based on WISER indicators and Uttar Banga Krishi Vishwavidyalaya got the last position from the point of view of webometric ranking Smith and Thelwall [5] calculated Web Impact Factors (WIFs) for Australian universities using a specially designed crawler and the AltaVista search engine. Links between UK, Australian and New Zealand Universities had been reflected. Both the number of pages at the site, and the number of academic staff members, were used as measures of the size of the universities.

The WIFs were compared with conventional measures of research output: rankings by Asiaweek magazine, the number of publications per staff member, and the number of citations per staff member. There is a good correlation between the crawler and AltaVista in estimating the link counts. The WIFs do not appear to correlate well with conventional measures of research output.

They also discussed some of the methodological issues in the calculation of WIFs. Mukhopadhyay [8] tried to explore the possibility of research in the field of webometrics in the educational institutions in India using Web Impact Factor (WIF). Li [3] studied hyperlinks extensively by applying existing bibliographic methods and made an exhaustive review the development of WIF. Li pointed out the origin of WIF and techniques for data collection using commercial search engines. The study also highlights the development of WIF - origin, traditional measures and its improvements.

Jayshankar and Babu [4] in a webometric study examined the websites of 45 universities in Tamil Nadu to analyse the number of WebPages, links, calculate various types of WIFs. The result found that although some universities of Tamil Nadu have quite large number of WebPages but very low number of inlinks and hence low WIF.

Ravikumar [10] investigated the link pattern of selected academic libraries in India using UCINET computer software to visualize the network pattern that existed among peer group libraries.

Elgohary [2] made a study in order to investigate the WIF of Arab universities. The study included 99 universities representing 20 Arab countries. The advanced search facility of AltaVista was used for data collection. Two rounds of data collection were conducted to retrieve the links as well as the web presence of the included universities. The findings revealed that Jordanian universities represent 40 percent of the top ten universities with the revised web impact factor. However, this was not the case in terms of the universities' web presence. Results indicated a strong correlation between external links and web presence.

V. RESULTS AND DISCUSSION

Step 1: The social network is visualized as a graph having each user as nodes and their friend's links as edges as shown in fig 2. The first target is to identify nodes which have maximum indegree. The indegree of each node is calculated by counting the number of links pointing towards a node. For this aspect we consider two factors,

1. Number of messages arrived to an user per a day (MPD)
2. Number of unique dispatchers(UD)
3. Number of identical dispatchers (ID).

We find out a nodes indegree by using below given formula ((MPD/UD) – ID).We eliminate the inflowing links from identical users by subtracting the number of identical dispatchers from the computed result.

Step 2: We explore the relationship between each user by finding out the relationship such as friend, Classmate, colleague etc.We find the whether there is a path from each node to every other node in the graph by finding the connectivity of the graph. We also find the neighbors of each user by considering the adjacency (those nodes connected with a path of length

1. We also explore how reach a node at an optimum speed and duration by finding out the shortest path using Dijkstra's algorithm.

Step 3: We use node-by-node matrix which has as many rows and columns as there are nodes in a network. The data entry is done by placing O's or 1's into a spreadsheet. The O's represent no connection and 1's signify a link which represents a binary matrix. An input values are given manually in Table 1, the centrality of the each node by performing Degree Centrality and Betweenness Centrality in Table 2 and corresponding Web Graph in Fig 3.

	A	B	C	D	E	F	G	H
A	1	0	0	1	1	1	1	1
B	1	1	1	1	1	1	1	0
C	0	0	0	1	0	0	1	1
D	1	1	0	1	1	1	1	0
E	1	0	0	1	1	1	1	1
F	0	0	0	1	1	0	0	1
G	1	1	1	0	0	1	1	0
H	1	0	1	0	1	1	1	0
I	0	1	1	1	1	1	0	0
J	1	0	1	0	1	1	1	1

Table 1: Sample Input values for UCINET

	Betweenness	nBetweenness		OutDegree	InDegree
I	5.317	7.384	B	8.000	3.000
G	3.683	5.116	A	7.000	6.000
J	3.633	5.046	J	7.000	7.000
D	3.350	4.653	D	7.000	6.000
F	2.150	2.986	H	6.000	5.000
H	2.117	2.940	E	6.000	7.000
E	2.050	2.847	G	6.000	7.000
A	1.900	2.639	I	6.000	9.000
C	1.700	2.361	F	5.000	8.000
B	1.100	1.528	C	5.000	5.000

Table 2: Betweenness & Degree Centrality Measures

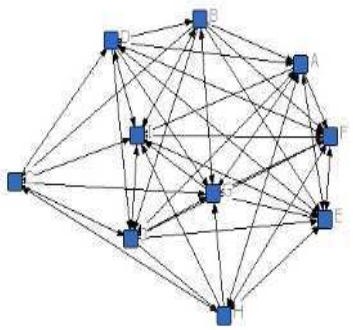

Fig 3: Web Graph for Link Structure

VI. CONCLUSION

This study has been exploratory and there is scope for future Webometrics research in this area. It would be useful to carry out a more comprehensive study comparing more E-Books and identify the quality of E-Books. There was a lot research in measuring the quality of Web sites for few Universities and institutions. Webometrics is concerned with measuring aspects of the web: web sites, web pages, parts of web pages, words in web pages, hyperlinks, web search engine results. The importance of the web itself as a communication medium and for hosting an increasingly wide array of documents, from journal articles to holiday brochures, etc.

REFERENCES

1. Jalal, Biswas, Mukhopathyay, Web Impact factor and link analysis of selected Indian Universities. 57(2010),109-121
2. Elgohary A, Arab Universities on the Web: A webometric Study, The Electronic Library, 26(3)(2008) 374-386.
3. Li X, A review of the development and application of the Web Impact Factor, Online Information Review, 27(6)(2003) 407-417.
4. Jayshankar R and Ramesh Babu B, Websites in universities in Tamil Nadu: a webometric study, Annals of library and information studies, 56(2)(2009), 69-79
5. Smith A and Thelwall M, Web Impact Factors for Australasian universities, Scientometrics, 54(1/2) (2002) 363-380.
6. WISER, Web Indicators for Scientific Technological and Innovation Research: A Survey of Practice, December 2005 43. Ingwersen, op. cit., 236-43
7. Jahaysnkar R and Ramesh Babu B,Narendra Kumar.A.M, Siruseri, Dr.Nageswara Rao.P, Web Credibility of Engineering College Websites Affiliated to Anna University, Tirunelveli (Tamil Nadu, India), October 19-22, 2010.
8. Mukhopadhyay P S, The calculation of Web Impact Factors for educational institutes of India: A Webometric analysis In. Information Management in e-Libraries, 26-27 February 2002.
9. Vishal Kumar Saxena, Webometrics, Informetrics, Scienctomerics, October 19-22, 2010.
10. Ravikumar S, A webometric study of selected academic libraries in India using link analysis, Proceedings of Fifth International Conference on Webometrics, Informetrics and Scientometrics. Dalian, September 13-16, 2009.
11. http://en.wikipedia.org/wiki/Webometrics
12. http://webometrics.wlv.ac.uk/
13. *www.analytictech.com/ucinet/.*