Collaboration Analysis of World National Library websites via webometric methods

MOHSEN HAJI ZEINOLABEDINI0[1]
LEILA MAKTABIFARD[2]
FARIDEH OSAREH[3]

1- PhD Student, Shahid Chamran University. and Faculty member of AREO, School of Education & Psychology, Department of Library & Information Science. Ahwaz-Iran zabedini@yahoo.com

2- PhD Student, Shahid Chamran University, School of Education & Psychology, Department of Library & Information Science. Ahwaz-Iran r_maktabi@yahoo.com

3- Associate Professor of Shahid Chamran University, School of Education & Psychology, Department of Library & Information Science. Ahwaz-Iran fosareh@yahoo.com

Abstract

This article aimed to study National Library Websites (NLW) using webometric methods. The in-links and co-links to national library websites were analyzed to study: firstly, the visibility of these National libraries on the web. Secondly, the collaboration on national and international level amongst the studied national libraries websites. This study found that according to the in-link count of 38 national library websites, 3 were extremely popular and we can call them the most visible national library websites as they come below: 1. United States of America (http://www.loc.gov); 2. Australia (http://www.nla.gov.au); 3. United Kingdom (http://www.bl.uk). The results of the study also showed that, there were 5 clusters (2 cross continental and 3 international) in the studied national library websites. On the other hand, the multidimensional scaling map showed 4 major collaboration clusters: 2 cross national (both European) and 2 international (European, Asian, American, Australian). African national library websites were not seen in these clusters. It means that, African national libraries have a little collaboration with others through their websites. However, due to the problems of search engines which are used for data collection in webometric studies, this method needs to be used with caution.

1. Introduction

Nowadays, via ICT-assisted dynamic environment, NLs have very suitable facilities for playing their role. Undoubtedly, one powerful website is adequate for meeting users' needs of NLs in anywhere. In addition investing on designing and managing website, NLs need to make necessary investments for introducing and making accessible their website contents. Therefore, using webometric methods, the present study has examined visibility, in–links and co-links of NLs in all of the world. Also, it is important to mention that in performing such studies which have been derived from models of printed citation research [6] we should pay attention to some considerations. While in scientific journals, citation indexes are the most important tools for citation studies, in the web environment search engines play partly this role [15]. But we should consider that using search engines for such studies faces up to some problems. As much as bibliometric research indicates criticisms on intrinsic shortcomings of ISI products, there are several studies which demonstrate limitations of search

engines. As a result, researchers act prudently with any findings which are based on these primary tools of data collection [13].

Results of this research will reveal status of NLs website and highlight their strengths and weaknesses. It also will study collaboration rate among world NLs through their website so that managers of these sites can understand how they have met their goals and improve possible deficiencies.

2. Literature review

Webometrics has a short and relatively new background. In a survey, Smith [4] compared impact factors of website of Australian (Australian and New Zealand). After calculating links of these 2 websites , he concluded that Australian NL website is larger and has more in–links rate. Chu [10] analyzed in-links of 12 websites affiliated to librarianship schools which are approved by ALA. Methodology of his research which has examined in-links of these 12 websites via clustering method and multi-dimensional scaling is similar to ours. One of his research findings is that providing websites with various issues leads to more in–links and visibility. By comparing reasons of citation and link, in addition to past findings chu [11] offered 3 main reasons for creating hyper links in the web environment. These include: first, in linking a site, links are primarily created to sites which somehow have relation to it. Second, link are mainly created in web page or website, while in citation, reference is made to sentence, paragraph or a part of one document. And third, in citation we encounter negative references which have no compatibility with research but link is typically created to positive, relevant and valuable issues.

Using clustering method and multi-dimensional scaling, Osareh [7] analyzed structure of 95 websites of library and information science schools from 18 countries. Among them 70 sites were active which formed sample of her research. She concluded that websites examined are categorized in 7 groups including 2 national and 5 international ones. Also, two-dimensional map showed 5 relevant clusters. Among these 5 clusters, 2 cases (one from USA and the other from Canada) were national and 3 cases were international.

Using AltaVista, Noruzi [3] studied how many links have been created to Iranian university websites. He found that the websites of Iranian universities have a low in-link. This indicates that for linguistic reasons, Iranian (Persian) sites may not receive and attract the attention that they deserve for the world wide web.

3. Research objectives

This research aims at examining visibility of world NLs websites as well as their impact factor and determining collaboration rate among these websites.

4. Research questions

1- How is visibility of world NLs websites?

2- Which websites have the highest rate of self-links?

3- How is impact factor of world NLs websites?

4- Which websites have the highest rate of co-links?

5- Using clustering, how many clusters will be categorized in world NLs?

6- Using multi-dimensional scaling and drawing map of co-links of world NLs websites, how many groups (clusters) will be identified?

5. *Methodology*

Methodology is Webometrics which will be done by means of links analysis. Also, using clustering method and multi-dimensional scaling, we will analyze links.

6. *Method of data collection*

Data relating to world NLs have been collected from NLs list available in IFLA website http://www.ifla.org/VI/2/p2/national-libraries.htm. So, this reliable list was selected as our research basis. In mentioned list, there were details of 160 NLs which among them 69 libraries had internet address (URL). From these 69 NLs, 12 countries had more than one internet address because these countries have some libraries playing roles similar to NLs.

However, we only chose NLs websites of each country as our research sample. It is worth saying that 11 websites were omitted because they were not active when doing our research (Nov. 2005). In order to collect data, we used AltaVista, because of its potentials for webometric research. Researchers such as Smith [5], Noruzi [7] and Kousha and Hori [14] also recommend AltaVista because in their opinion, in addition to counting the number of pages in the website studied and the number of pages linking to the web site it has a large database, covering as much of the web as possible, provides Boolean operators and retrieves more consistent results. Then, via AltaVista in–links and the number of pages indexed for each website were calculated and registered. After data collection, we inserted earned results in Excel software. In regard to structure and features of statistical software used namely SPSS as well as because of detailed data available in webometric studies that create problems for calculations, we refined colleted data (as Ingwersen). In this regard, Ingwersen [19] believed that "search engines can't index all of the web, their overlapping is not considerable and their retrieval factures are so simple that can not be used for online broad webometric analyses. Thus, sampling is so important, critical and difficult that requires refinement".

As a result, 69 websites mentioned earlier reduced to 38 ones because we omitted 31 cases from countries including Andorra, South Africa, Chile, Cuba, Faroe Islands, Papua New Guinea, Georgia, India, Indonesia, Israel, Kenya, Kazakhstan, Macedonia, Latvia, Namibia, Pakistan, Panama, Philippines, Sri Lanka, Singapore, Saudi Arabia, Romania, Turkey, Venezuela, Uganda, Ukraine, Uruguay, Vatican City, Taiwan, Serbia and Quebec.

Afterwards, using SPSS software we analyzed data entered and in 2 steps through clustering and drawing multi–dimensional scaling, which will be fully described in research findings, determined clustering as well as linkage map of websites.

7. *Data collection tools*

We utilized AltaVista, as mentioned previously, in order to collect data and broadsheet software (Microsoft Excel) as well as SPSS (version 11.5) for data analysis.

Research results

We found findings as following:

In response to question 1, we showed that in-link counts of a website in the web indicate its visibility status [10]. The more one site has in-links indicates that the more users are interested in its information. In the web environment, this interest or need is showed by link i.e., the more links count, the more site reliability. So, the most determinant element of importance, reliability, use as well as websites visibility is in-links count of websites. In order to determine websites visibility, in-links count was extracted via AltaVista. For collecting in–links data we used the following formula:

(link:http://www.nlai.ir/ OR link:nlai.ir/) NOT (host:http://www.nlai.ir/ OR host:nlai.ir/)

Results of this section are provided in column A of table 1. As you can see, LC had the highest in-links count (rate), 249000, and Slovenian NL site had the least in-links count, 249.

In response to question 2, we found that if one site has more self–links, it indicates that information and pages within it are well–connected. It is worth saying that high rate of self-links indicates that related resources available in one site are well-linked and users are properly conducted to optimal documents [2]. Search engines track self–links and therefore, create precise indexes from a website. Although, self-links are not calculated in webometric analyses, the more self-links are in one site, the more and the better its information as well as web pages can be indexed by search engines and therefore, site contents will be better retrieved.

To estimate self-links of these websites, we utilized the following formula:

(link:http://www.nlai.ir/ OR link:nlai.ir/) AND (host:http://www.nlai.ir/ OR host:nlai.ir/)

Research findings can be considered in column B of table 1. As can be seen, LC website had the highest rate of self-links, 89600, and Russian as well as Swiss NLs websites as the least rate of self-link had no self-link.

In response to question 3, research findings can be summed up as follows:

In order to determine impact factor of studied websites, revised impact factor was calculated by the following formula:

The number of in-links divided by total number of web pages indexed by used search engine

Revised impact factor is indicator of real ranking of a website. If revised impact factor is high, of course, website has highly ranking and vice versa [8].

Table 1: In-links, self-links, total number of indexed web pages and revised impact factor of world NLs websites

Row	NLs	URLs	A	B	C	D
			In-links	Self-links	Web pages indexed by search engines	Revised impact factor
1	Argentina	http://www.bibnal.edu.ar/	7620	129	288	26.45
2	Australia	http://www.nla.gov.au/	229900	43400	198000	1.16
3	Austria	http://www.onb.ac.at/	6420	7850	6840	0.93
4	Belarus	http://natlib.org.by/	1150	96	30	38.33
5	Belgium	http://www.kbr.be/	3130	271	11700	0.26
6	Brazil	http://www.bn.br/	15900	16100	70300	0.22
7	Canada	http://www.collectionscanada.ca/	6220	17	80500	0.07
8	China	http://www.nlc.gov.cn/	24400	1	22900	1.06
9	Croatia	http://www.nsk.hr/	2740	2	1550	1.76
10	Czech	http://www.nkp.cz/	4350	1310	5000	0.87
11	Denmark	http://www.kb.dk/	13500	9990	16700	0.80
12	Finland	http://www.lib.helsinki.fi	4130	1720	12300	0.33
13	France	http://www.bnf.fr/	36700	24700	5950	6.16
14	Germany	http://www.ddb.de/	5400	2140	1960	2.75
15	Hungary	http://www.oszk.hu	851	716	9650	0.08
16	Iceland	http://www.bok.hi.is/	2590	35	1230	2.10
17	Iran	http://www.nlai.ir/	2800	22	694	4.03
18	Ireland	http://www.nli.ie/	5230	79	11700	0.44
19	Italy	http://www.bncrm.librari.beniculturali.it/	1950	101	57	34.21
20	Jamaica	http://www.nlj.org.jm/	8440	21	83	101.68
21	Japan	http://www.ndl.go.jp/en/index.html	977	1	4200	0.23
22	Korea north	http://www.nl.go.kr/	80900	41	1950	41.48
23	Lithuania	http://www.lnb.lt/	2580	407	569	4.53
24	Malaysia	http://www.pnm.my/	5730	128	147	38097
25	Netherland	http://www.kb.nl/	12100	36300	23200	0.52
26	Newzealand	http://www.natlib.govt.nz/	12500	699	5150	2.42
27	Norway	http://www.nb.no/	7520	2450	2750	2.73
28	Peru	http://www.binape.gob.pe/	2010	2	11	182.72
29	Poland	http://www.bn.org.pl/	10400	849	31	335.48
30	Portugal	http://www.bn.pt/	7480	1370	42	178.09
31	Russia	http://www.nlr.ru/eng/	1040	0	8230	0.12
32	Slovakia	http://www.snk.sk/	845	650	6490	0.13
33	Slovenia	http://www.nuk.uni-lj.si/vstop.cgi	249	1	2240	0.11
34	Spain	http://www.bne.es/	86800	2130	4020	21.59
35	Sweden	http://www.kb.se/	4490	3200	17100	0.26
36	Switzerland	http://www.snl.admin.ch/slb/	3410	0	1920	1.77
37	United Kingdom	http://www.bl.uk/	97400	284	99900	0.97
38	United States of America	http://www.loc.gov/	249000	89600	452000	0.55

As can be seen in column D of table 1, Polish NL website had the highest rate of revised impact factor, 335.48, and Canadian NL website had the least rate of revised impact factor, 0.07. Here, one of the most important problems in calculating impact factor is clear. As previous mentioned, LC website has the highest rate of links, in-links, self–links as well as number of web pages indexed by search engine and therefore, is supposed as the most powerful site in all of world NLs websites. But, calculating web impact factor (WIF) indicates that LC website is weak while Polish NL website has quality status. As scientometric studies in which total number of citations given to published articles in one journal at a certain point in time (usually biennial) divided by the total number of published articles (citable items) in the same journal and time is a criterion for estimating journals impact factor (JIF) [8], in webometric studies impact factor also should be calculated in terms of total number of a website pages indexed by a search engine. Therefore, total number of a website pages indexed needs to be determined. We used the following formula to determine total number of pages:

domain:nlai.ir OR domain:www.nlai.ir

As shown in column C of table 1, LC website had the highest rate of pages indexed in search engine, 452000, while Peruvian NL website had the least rate of page indexed, 11.

In regard to question 4, co-link means that link of two sites is seen in third one. This position is similar to co-citation in printed resources. Osareh [9] believed that co-link is important because demonstrates a relation between two websites. In other words, these two sites have some similarities in their working field. So, in webometric analyses, studying co-links is important because it can help to identify core sites of each field.

In order to determine co-link status of world NLs websites, below command was used in AltaVista:

Co-link rate = URL of first site+ space +URL of second site

http://www.bn.pt/ http://www.nlia.ir/

For example, according to this Formula, co-links rate between Iranian and Portuguese NLs websites equal with 1360 items. This means that in links of 1360 websites there were both Iranian and Portuguese NLs websites.

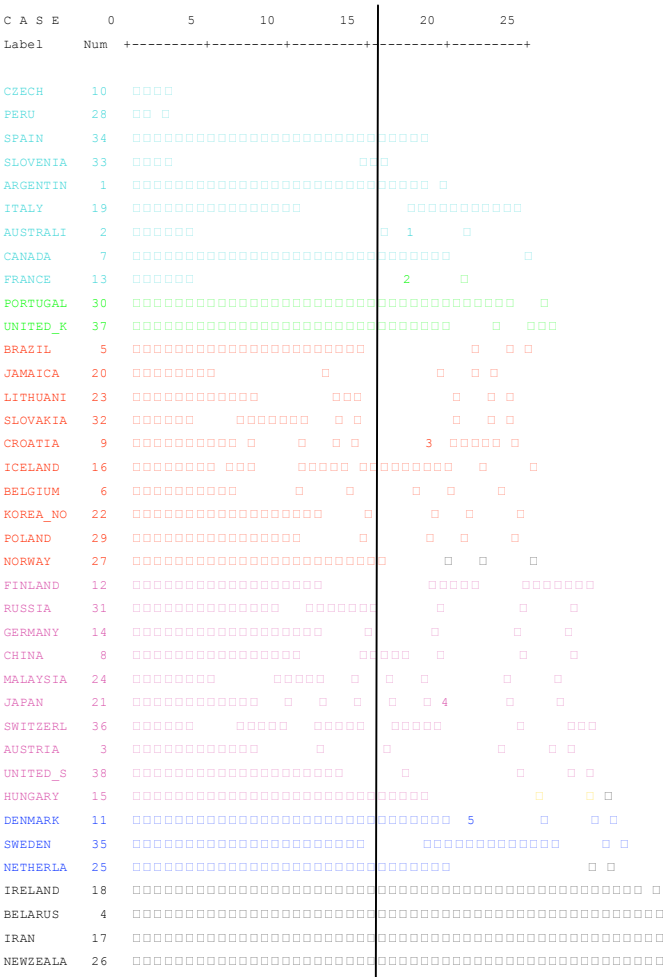
Results of co-links derived from calculation in SPSS, have been provided in the form of a clustering diagram (figure 1) and the connection map of websites (figure 2).

Findings relating question 5:

As can be seen in figure 1, via clustering method we have identified five main clusters including 3 international and 2 continental (European) clusters. Interestingly, four websites have not formed any cluster and have remained independent.

In the first cluster (international), there are NLs websites of Czech, Peru, Spain, Slovenia, Argentina, Italy, Australia, Canada and France. The second cluster [continental (European)] relates to NLs websites of Portugal and UK. In the third cluster (international), there are NLs websites of Brazil, Jamaica, Lithuania, Slovakia, Croatia, Iceland, Belgium, North Korea, Poland and Norway. The fourth cluster (international) involves countries including Finland, Russia, Germany, China, Malaysia, Japan, Switzerland, Austria, USA and Hungary. Countries such as Denmark, Sweden and Netherlands are in the fifth cluster which is continental (European). Also, Iran, New Zealand, Ireland and Belarus have not formed any cluster and can be seen independent in the figure 1. Among studied websites, there were websites from all continents except Africa. This does not mean that African NLs websites have not designed any website but their websites (Kenya, Namibia, South Africa and Uganda) have been omitted for reasons mentioned earlier. According to Figure 1, some sites such as Czech and Peru have proactively and fast formed independent relation. Some sites including Portugal and UK have formed independent connection but from a distance. Some libraries as what can be seen in the third and fourth clusters have formed more as well as broad connections from different distances. At the end of figure 1, NLs websites are considered which have formed few co-links with other libraries. In other words, these websites could not absorb co-links for reason such as linguistic problems, political factors and especially inability in providing other countries with necessary and valuable information.

Fig 1: Clustering of co-links rate between world NLs websites



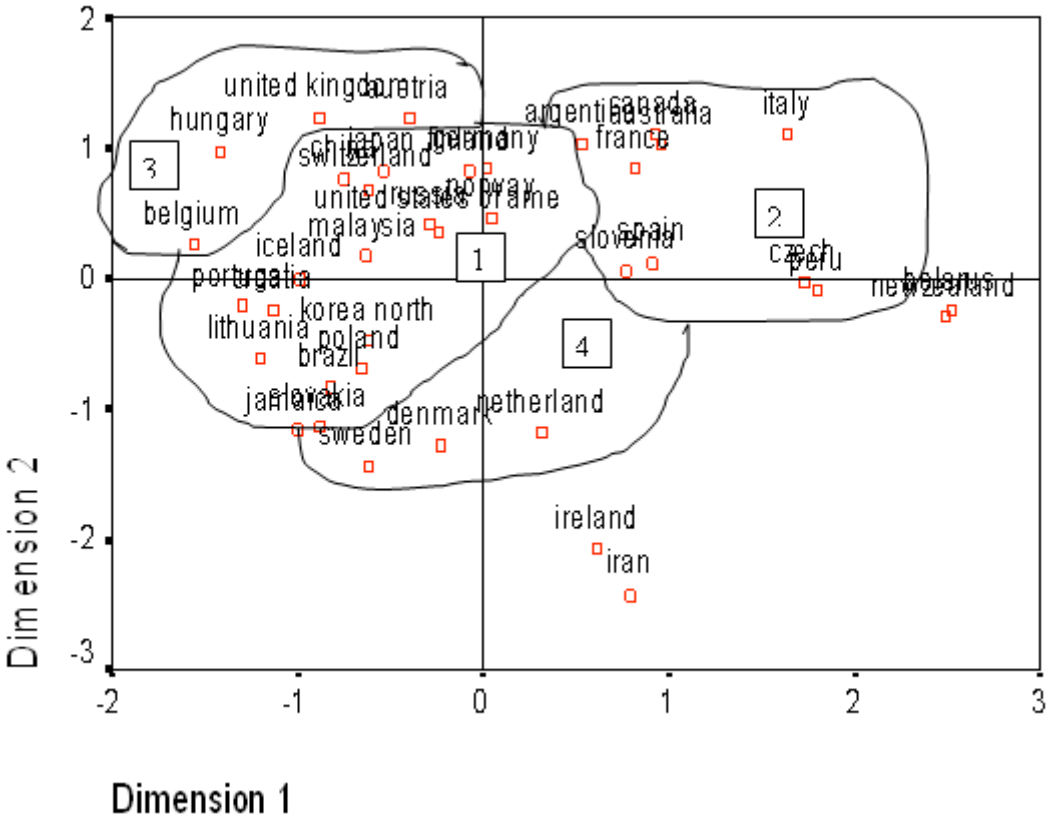
For replying to question 6, we have attained the following results:

To better understand co-links status of world NLs websites, we have drawn map of linkage connections of websites by using multi-dimensional scaling (figure 2).

As shown in the map, we can consider 4 clusters. Among analyzed data, we have identified 2 international clusters from various continents and countries as well as 2 European clusters. International cluster (first cluster) includes 17 countries: US, Russia, Japan, China, Germany, Switzerland, Finland, Malaysia, Iceland, Portugal, North Korea, Poland, Lithuania, Slovakia, Brazil and Jamaica. The other international cluster (second cluster) includes 7 countries: Canada, Argentina, Australia, France, Italy, Spain and Slovenia.

Identified continental clusters relating to European are the third cluster which has four countries (UK, Austria, Hungary and Belgium) as well as the fourth cluster with 3 countries (Denmark, Sweden and Netherlands). Also, countries New Zealand, Belarus, Ireland and Iran which have no co-links with other countries can be independently seen on the map.

Fig 2: Map of linkage connections of world NLs through multi-dimensional scaling



By making a comparison between figures 1 and 2, we can see changes which have been created in connection status among websites. Some websites that in clustering map were distant and independent are more adjacent in the linkage map. For instance, we can name countries such as New Zealand, Belarus as well as Iran and Ireland. In addition, some countries including Croatia, Portugal, US and Russia that in Figure 1 were in the same cluster but distant from each other, in the Figure 2 are fully connected and linked.

8. Conclusions

This research finding provides a picture of various NLs status in terms of their website quality and performance. Results show that based on this research criteria, LC website is the most powerful one among world NLs websites. Factors affecting on LC website power are high rate of pages, various and valuable information elements in all areas of library and information science, English language, timely updating, being user-oriented, universal coverage and so forth. Also, LC on-line catalog (http://www.loc.gov/catalog) as the most reliable source on cataloging and classification of printed and electronic resource is another reason for LC website power as well as quality.

Results relating to revised impact factor indicate that Polish and Canadian NLs websites were the best (335.48) and the worst (0.07) respectively. This result highlights one of the basic problems of webometric studies. In relation to this challenge, Noruzi [2] believes that “suppose we have two websites, A and B. website A has 100 in–links and also has published 100 web pages, while website B has 1000 in-links and has published 1000 web pages. Based on determined formula, both of them have the same impact factor namely 1. But can we really say that both of sites have the same impact on their scientific areas? Maybe they have equally been successful in absorbing links, but website B has more publication rate and therefore, its impact is 10 times more than website A”.

Using clustering method, we have identified 5 main clusters namely 3 international and 2 continental (European) clusters. Also, there were 4 independent websites.

Using multidimensional scaling method, we have identified 4 main clusters. 2 international clusters from various continents and countries as well as 2 European clusters.

In regard to co-link and its factors, we can not suggest a single consensus. Thelwall [18] also believes that "there are some theoretical reasons for co-link in different situation but we have little knowledge about models and motivations of linking. As a result, we can not evaluate issues on co-link as well as linking". The most important co-link reasons in studied NLs website include guiding lists provided in websites, on-line national bibliographies, important information resources, news, working programs, and electronic full-text resources. Thus, we can not consider a couple of limited factors as only reasons of a website success [10]. Therefore, webometric research must be conducted with caution and researchers need to consider all factors comprehensively.

References

1. A. Noruzi, The Web Impact Factor : a critical review, The Electronic Library. No. 24, 2006. Also available online: http://eprints.rclis.org/archive/00005543/
2. A. Noruzi, The web impact factor: A survey of some Iranian university websites, Studies in Education and Psychology affiliated to Ferdowsi University of Mashhad, Iran, 5(2), p.105-119, 2005.
3. A. Noruzi, Web Impact Factors for Iranian Universities, Webology, V. 2, Number 1, April, 2005. Available at: http://www.webology.ir/2005/v2n1/a11.html
4. A. Smith, ANZAC webometrics: exploring Australasian Web structures, 1999. [Online] Available at: http://www.csu.edu.au/special/online99/proceedings99/203b.htm
5. A. Smith,Citations and links as a measure of effectiveness of online LIS journals, IFLA Journal: Official journal of the International Federation of Library Associations and Institutions, v. 31, no. 1, P. 75-84, 2005.
6. D. Tunger and C. Plott, Bibliometric analysis as part of a trend recognition system in science, Iranian Journal of Information Science and Technology. 3(2), July/December, P. 1-17, 2005.
7. F. Osareh, Mapping the structure of library and information schools (LIS) websites using cluster and multidimensional, Paper presented at The International Conference on Scientmetrics and Informetrics, 9th.25-29 August 2003, Beijing.
8. F. Osareh, Methods and applications of infometrics, Approach, 25(3), p.94-100, 2002.
9. F. Osareh, Scientometrics: Aspects, methods and applications. Proceedings of conferences held by Iranian Library Association edited by Mohsen Haji Zeinolabedini, vol 2, p.271-287, Tehran: Iranian national library, Iranian library association, 2005.
10. H. Chu, A webometric analysis of ALA acdredited LIS school websites, In Proceedings of the 8th International Conference on Scientometrics and Informetrics, 16-20 July. Edited by Mari Davis and C. S. Wilson, Sydney: BIRG, UNSW, 2001.
11. H. Chu, Taxonomy of inlinked Web entities: What does it imply for webometric research?, Library & Information Science Research. 27(1), Winter 2005, P 8-27. Also available online via Science Direct, 2005.
12. H. Park and M. Thelwall, Hyperlink analyses of the World Wide Web: A review, Journal of Computer-Mediated Communication, 8(4), 2003. Available: http://www.ascusc.org/jcmc/vol8/issue4/park.html
13. H. Snyder and H. Rosenbaum, Can search engines be used for web-link analysis? A critical review, Journal of Documentation. 55 (4): 375-384, 1999. http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html
14. K. Kousha and A. Hori, The Relationship between Scholarly Publishing and the Counts of Academic Inlinks to Iranian University Web Sites: Exploring Academic Link Creation Motivations, In: Kretschmer, H. et al. (ed.), *Proceedings of International Workshop on Webometrics, Informetrics and Scientometrics*, March 2-5. Indian Institute of Technology, India, 136-149, 2004. Also Available at: http://www.koosha.tripod.com
15. K. Kousha, A comparative study on Iranian newspapers websites via web impact factor, Informology, 1(2), p.87 -114, 2002.

16. K. Kousha, Extracting Macroscopic Information from sources of URL Citaion To scholarly Open Access LIS Journals: A Webometric approach, 2004. Avaiable at: http://www.koosha.tripod.com/articles/ifla2.doc

17. L. Bjorneborne and P. Ingwesen, Perspectives of webometrics, Scientometrics, 50 (1), p.65-82, 2001.

18. M. Thelwall, Web use and peer interconnectivity metrics or academic web sites, Journal of Information Science, 29 (1). pp. 1–10, 2003.