The Electronic Library

A webometric analysis of major keywords and expressions in biochemistry using LexiURL Searcher

Kobra Taram Abbas Doulani

A webometric analysis of major keywords and expressions in biochemistry using LexiURL Searcher

Kobra Taram
Department of Knowledge and Information Science, Payame Noor University, Tehran, Iran, and
Abbas Doulani
Central Library, Urmia University of Medical Sciences, Urmia, Iran

Received 26 March 2014
Revised 12 April 2014
17 June 2014
20 August 2014
Accepted 27 October 2014

Abstract

Purpose – The purpose of this paper is to explore webometric analysis of keywords and expressions of the biochemistry field of study via LexiURL Searcher.

Design/methodology/approach – Interfaces for assisting users with information access have received considerable attention. Along with the extraction of data on Web sites for webometric purposes (e.g. link analysis, ranking of Web sites, etc.), LexiURL Searcher presents some information on the arrangement of links among different Web sites. Such capability enables users to identify one or more Web sites around their intended subject and, accordingly, explore all Web sites linked with their identified Web site(s). LexiURL Searcher has preceded webometric analysis by considering the main expressions and keywords derived from the MeSH database.

Findings – The worldwide survey indicated that links from countries such as England, Japan, Germany, Australia and Canada were among the Web sites that are most used in biochemistry. Alternatively, other countries such as Singapore, Thailand and Poland had the most advantageous links to the outside world, whereas South Africa, New Zealand and The Netherlands had the least link effect. Biochemistry, being a specialized domain, would benefit greatly from site linking and would provide users the most assistance in information processing.

Originality/value – Most webometric studies remain on the level of link analysis and Web site statuses; however, this paper gives information on the common thread Web sites based on a standard thesaurus.

Keywords Web sites, Webometrics, Keywords, Biochemistry, Expressions, LexiURL Searcher

Paper type Research paper

Introduction

The dramatic increase in the volume of scientific data has aggravated the problem of accessing selective, relevant and reliable information with the least time spent and has turned accessibility to required information into a major concern for Internet users. Thus, interfaces for assisting users to access required information have proliferated.

The authors acknowledge Department of knowledge and Information Science.

The Electronic Library
Vol. 33 No. 6, 2015
pp. 1163-1173

0264-0473
DOI 10.1108/EL-03-2014-0054

Webometrics is deduced from other expressions, such as bibliometrics, science metrics and so forth, and has been variously defined by scientists in a variety of manners. Herring (2010) stated that research and studies related to links and relationships on the Internet were founded in the mid-1990s. The concepts were first used by Almond and Ingversen in 1997 (as cited in Herring, 2010), referring to different manners. Hsu and Park (2012) declared that we should consider the following examples for webometrics: the longitudinal of Web links (Park *et al.*, 2011), Web basic citation analysis (Thelwall, 2008), impact factor in terms of online visibility (Woo and Yeon-ok, 2008) and e-science network mapping (Park, 2010).

Quick insight reveals that the Internet bears a structural capability that enables one to analyze data in a systematic way. In other words, capability and efficiency are gained through deep investigation on how to approach the worldwide network. Furthermore, a great deal of research and discovery has been carried out around the subject of related information retrieval over many consecutive years. In this paper, it is explored as an expanded title from different points of view by exemplifying how the necessary information retrieval in online public access catalogs is used, including the bases of some data information search and retrieval. As there is an ever-increasing number of sources on the Internet, including different Web sites, databases, Web logs and other types of sources offering access to daily information, it would provide better quality information to search information from different sites. Then, it is possible to search the sites to discover what is being sought. It has not been determined where the most comprehensive source of information could be found around a specific subject on the Internet. This issue has resulted in the emergence of a specific field in webometric research entitled "link analysis". It would be a great help for users to have quick access to their required information. There are different kinds of software which have been used to inspect links among Web sites, domains and databases. These kinds of software include World Wide Web Consortium software packages and ScoSciBot software, the former to analyze the links among different Web sites and the latter to study the links between different Web sites (Thelwall *et al.*, 2008). In addition to the separate software, there are some search engines and subject directories which contain different algorithms with statistical figures from the related Web site links. Procedures defined by using search engines and subject directories, such as Google, AltaVista and Yahoo, would be very useful to help with these issues.

In this study, LexiURL Searcher was used for the abovementioned goals for biochemistry Web sites. Along with the extraction of data on Web sites for webometric purposes (e.g. link analysis, ranking of Web sites, etc.), LexiURL Searcher presents some information on the arrangement of links among different Web sites. This capability enables users to identify one or more Web sites around their intended subject or to explore all Web sites linked with their identified Web sites.

Review of related literature

There are different programs and software for webometric studies. Web International Consortium has established and formulated a set of rules and guidelines to design and improve the quality of webometric analysis. These guidelines contain all matters related to Web sites, including objects such as their design or technical matters (Sreedhar *et al.*, 2010). One of the basic points in the efficient designing of Web sites is that users can easily connect and use them

completely and, if necessary, they can connect to other related Web sites. This process makes it necessary to have a systematic order for both inside and outside links (Barnett, 2012). Nowadays, when ranking Web pages in terms of their validity and rate of use, the link structure of a Web site is considered (Page *et al.*, 1998). As Glover *et al.* (2002) argued, for classifying sites in terms of topic, their linking network can be applied. In their research, Zuccala *et al.* (2008) used LexiURL Searcher to study and analyze digital repositories of links. The results of a survey with their users revealed that the digital repositories that lead and help users to use other linked digital repositories have been done via the same links. Hence, to examine the existing links and to create new ones in new repositories, Web site managers are required to verify these depositories every four or six months (Zuccala *et al.*, 2008). Thelwall (2008) and his research group examined different investigative Web sites for scientific cooperation. The results derived from LexiURL Searcher revealed that the links between those Web sites had an important role in attracting researchers and also workforces from farther distances. Thus, the exchange of workforces and investigators among research groups has occurred via the Web platform. In another research, Zuccala *et al.* (2007) investigated the National electronic Library for Health (NeLH) using LexiURL Searcher. Their results showed that as the library was a multifaceted digital library, it has a multitude of links among governmental, educational, commercial, organizational and other Web sites, while the library itself did not have any considerable links with any of the other similar Web sites or databases (Zuccala *et al.*, 2007). Their results emphasized that as the users of other Web sites are guided to the Web site of the NeLH through several links, which has rich sources too, they were more successful than those who directly used the NeLH Web site. The reason was that the first group who used different sites in their research could get help from several different links, databases and digital libraries, whereas the users of the NeLH alone had the least number of outside links. In another study, Huvila (2011) addressed how users get information on the Web according to the type and form of their questions. For this purpose, he used LexiURL Searcher to find some special types of questions (e.g. I asked my mom, but,) and determined the Web sites and databases which had appropriate answers to these questions (Huvila, 2011).

Hsu and Park studied the interactions among South Korean and Chinese Web sites using LexiURL Searcher. They selected 800 Web sites from among the study population. The results indicated that most of the Web sites in South Korea and China connected to each other were organizational Web sites and were mostly cooperating in educational, scholarship, research, commercial and industrial fields. By investigating the dispersion of links among the Web sites of these countries, they found that the number of links between different Web sites in South Korea was far larger than the number of links between Chinese Web sites, and these were due to the differences which obviously existed in their organizational cultures (Hsu and Park, 2012). Research by Park (2010) on the global hyperlink network was carried out on Web-based network linking country code with top-level domains representing countries, which used hyperlink connectivity by means of network analysis. The results indicated that the 2009 international hyperlink network was completely interconnected with G7 countries, and Spain was at the center of the network, and poorer countries from Africa, Asia and Latin America.

Barnett (2012) assumed that "decomposing" or "cracking" *.com sites should lead poorer countries on the periphery to a more accurate description of the international hyperlink network. The research proposed a method of decomposing *.com which was applied to the proposed procedure, and the investigation revealed the differences between two international hyperlink networks: the one including *.com and the other excluding *.com data. The adjusted international hyperlink network showed substantial differences in the centrality of several countries making extensive use of *.com. For example, in contrast to the hyperlink network excluding *.com, the network including *.com showed a 30-fold increase in the out-degree centrality of the USA and a change in the country's eigenvector centrality that was greater than that of any other country (Barnett *et al.*, 2011). In addition, the centrality of China, Japan and India was increased in another study by Barnett (2012). The author reviewed recent developments in the global telecommunication network and found that peripheral nations are clustered into regional groups of Latin America, Asia and countries that made up the former Soviet Union. Over time, the integration of the global community and the collapse of the Eastern Block, the reintegration of Hong Kong into China and the dotcom bubble have resulted in a lot of changes in the pattern of global telephone flows (Barnett, 2012).

Using different means, each of the studies referred to in this section addresses the status of the Web sites in some vein in the network. Given the large volume of information on Web sites, portals, blogs and so forth, applying webometric tools including LexiURL Searcher software to identify and determine the status of topic-related Web sites can, to a large extent, help users to access targeted information, as well as to maximally decrease false drops.

Research methodology

An introduction to LexiURL Searcher

LexiURL Searcher is a free program for collecting data from search engines through the application of programming interfaces for webometric purposes. In addition to loading data from search engines and saving them as plain text files, LexiURL Searcher facilitates the possibility of generating different types of reports and graphs. LexiURL Searcher contains a classic interface that is used in complex works for the production of webometric reports. The interface is free for the use of researchers in research-only (non-commercial) works. Analysis of data from the search engines is performed in three phases. The first phase involves preparing a text file (in WordPad or NotePad) that includes a list of Internet addresses or keywords to be searched. In the second phase, LexiURL Searcher loads these queries into a search engine selected by the researcher. Subsequently, the results of the search are presented in a plain WordPad text file after a rather long period and then saved. Finally, in the third phase, the Searcher can process the results of the files for the representation of more detailed and formatted results and to produce outputs as tables and graphs. It is worth mentioning that LexiURL Searcher can analyze 1,000 queries per search (McCown and Nelson, 2007).

LexiURL Searcher is considered a powerful webometric utility whose significance and usefulness is recognized through the analysis of inbound and outbound links of Web sites and their connections. Given that LexiURL Searcher could search and recognize the Web sites without having the URLs and only by accepted keywords of

every field of science (extracted from subject headings), it can measure and present the status of information published around every specific area in the Web network through calculating links and visualizing their connections. On the other hand, regarding the mission of the library and information science field that is to locate suitable information in a suitable amount of time and in the right place for users, it could be stated that these kinds of software packages are highly valuable, as they can supply a selected set of information portals for users through Internet addresses and particularly through effective keywords in every field of science. Considering the limitation of this information to a specific field, LexiURL Searcher could prevent false drops to a large extent.

All of the research data were gathered through LexiURL Searcher with different choices, such as:

- reporting Web impact by keywords and phrases;
- reporting link impact on the bases of link numbers in a Web site or Web page; and
- reporting the Web impact and link impact through classic interface.

LexiURL Searcher has also an advanced choice for each of those functions. In a recent research study, all options of LexiURL Searcher were used to their greatest extent. Thus, for the first biochemistry field of study, the main and accepted keywords were extracted out of the MeSH database. Then the field experts not only studied these keywords and expressions to add other new cases but also suggested examining the keywords and expressions. Finally, after re-examining them in four phases, the group accepted 26 keywords and expressions as the main cases. Next, after installing LexiURL Searcher and selecting the Web impact report about Web mentions of more words or phrases, the keywords and expressions were saved in a new WordPad file (each word written in a separate line and each phrase used in a quotation mark). After searching the Web and analyzing the results for a few hours, the final report became a collection of Web pages' interlinks.

In the next stage, they entered the derived site dimensions in a new WordPad file and saved them, such that each of them was set in a separate line. Then, reinstallation of LexiURL Searcher was begun, and they selected the link impact report on links for one or more Web sites or Web pages. LexiURL Searcher returned the link impact report after a few hours.

In the third stage, the authors used the site domains derived from the first report to make a network diagram. After several hours of analysis, the diagram showed a network of links of analyzed Web sites. It should be mentioned that data collection was done after the software had been received from http://lexiurl.wlv.ac.uk in November 2012.

Findings

In this study, keywords and expressions in biochemistry were analyzed via webometrics through LexiURL Searcher in three phases. First, after consulting with the experts and using the Delphi method, keywords and phrases were selected in several stages. Then, 26 words and expressions were selected as the main ones and were saved in a WordPad file using LexiURL Searcher. Then, the report on the webometric analysis of the keywords and expressions was prepared. The released Web impact report about these keywords and expressions indicated that all the keywords and expressions were

Base query	URLs	Domains	STLDs	TLDs
Chemistry	985	756	65	48
Biochemistry	951	745	61	48
Water	902	714	59	38
Protein	888	627	41	34
Inorganic	713	601	53	41
Neurochemistry	625	550	44	35
Metabolomics	647	524	43	31
Chemistry Bioinorganic	595	494	59	50
Amino acid	595	494	59	50
Enzyme	526	426	32	24
Molecular epidemiology	481	405	51	37
Glycomics	488	394	39	26
Hormones	488	394	39	26
Immunohistochemistry	464	373	30	26
Lipids	434	373	33	29
Nucleoid acid	464	373	30	26
Acid-base	464	373	30	26
Immunochemistry	420	357	39	30
Coenzyme	420	357	39	30
Proteomics	434	353	53	40
Pathology, Molecular	415	344	33	27
Molecular biology	394	323	31	25
Prussian Blue reaction	372	299	42	33
Periodic Acid-Schiff reaction	361	288	46	40
Carbohydrate biochemistry	342	286	47	37
Histocytochemistry	155	131	34	29

Table I.
Frequency of uniform resource locators (URLs), domains, STLDs and TLDs

used in 14,023 URL addresses and 11,354 domains. It was also reported that nearly 1,132 domains belonged to second top-level domains (STLDs) and 886 belonged to top-level domains (TLDs). Thus, the keywords and expressions of Table I have the highest Web impact in terms of the addresses, TLDs and STLDs.

In the second phase, to show the participation of world countries in the production of Web sites involved in the biochemistry field, the report derived from LexiURL Searcher was arranged based on the domains of countries. To determine the least participation by the countries, the Web site of countries that yielded five unique Web sites in that report was considered in the statistics. Hence, 27 countries that had produced a distinctive number of Web sites in biochemistry were identified. The report revealed 724 Web sites regarding the domain of different countries (Table II).

Furthermore, the third phase was dedicated to reporting on the impacts of linking the Web sites according to the names of domains extracted from the second report. At this stage, LexiURL Searcher reported the number of every site domain in outside links. As mentioned above, there were 724 Web sites reported in this way. By entering the Web site addresses into a WordPad file, LexiURL Searcher reported the impacts of the links as a table. According to the results, there were 77,085 links which were exceeding those Web sites. Nevertheless, the number of these links was different in each country, as they are separately given in Table III. Accordingly and without considering the numbers of Web sites in terms of the links, England, Japan and Canada had the highest numbers of

Country	Domain name	No. of sites (domains)	(%)
England	uk	145	20.0
Japan	jp	81	11.1
Germany	de	73	10.0
Australia	au	70	9.6
Canada	ca	63	8.7
The Netherlands	nl	34	4.6
New Zealand	nz	28	3.8
Brazil	br	19	3.1
France	fr	20	3.0
India	in	17	2.3
Spain	es	17	2.3
Sweden	se	17	2.3
Switzerland soes	ch	17	2.3
China	cn	15	2.0
Greece	gr	11	1.5
Russian Fed	ru	11	1.5
Czech Republic	cz	9	1.3
Norway	no	9	1.3
South Africa	za	9	1.3
Poland	pl	10	1.3
Finland	fi	8	1.1
Portugal	pt	8	1.1
Taiwan	tw	8	1.1
South Korea	kr	7	1.0
USA	us	7	1.0
Ireland	ie	6	0.8
Singapore	sg	5	0.6

Table II. Frequency of domain name, number of sites (domains)

links, while New Zealand, Greece and South Africa had the lowest numbers of the links to Web sites. In terms of the number of Web sites in every country and their outside links, Singapore, Taiwan and Poland had the most links to other Web sites, while South Africa, New Zealand and The Netherlands had the lowest numbers of outside links, respectively (Table III).

In the third phase, to study mutual cooperation among Web sites in different countries, the authors drew a network link diagram with LexiURL Searcher. Based on the previous sample (724 Web site domains from different countries) and using stratified random sampling, 254 addresses were selected according to the Morgan table (to facilitate making a network diagram). Then the addresses were saved in a WordPad file. In addition, by selecting the network link diagram in LexiURL Searcher, the amount of connections and links between the Web sites was assigned. The derived diagram (Figure 1) from LexiURL Searcher indicates that Web sites which were utilized in the biochemistry field with related matters in different countries would not be able to build any special link and connection with each other. On carefully studying this diagram, we realized that among all these 254 Web sites, only eight of them were unilaterally linked to each other; while others lacked such links and connections (Figure 2).

Series	Country	Domain name	No. of sample sites (domains)	Link impact report	Average no. of link per page
1	England	uk	145	13,292	91.6
2	Japan	jp	81	7,648	94.4
3	Canada	ca	63	7,082	112.4
4	Australia	au	70	4,882	69.7
5	Germany	de	73	4,574	62.6
6	Spain	es	17	2,947	174.9
7	Switzerland soes	ch	17	2,673	157.2
8	Poland	pl	10	2,663	266.3
9	Brazil	br	19	2,662	140.1
10	China	cn	15	2,533	168.8
11	France	fr	20	2,509	125.4
12	Taiwan	tw	8	2,219	277.3
13	Norway	no	9	2,147	238.5
14	India	in	17	2,068	121.6
15	Singapore	sg	5	2,033	406.6
16	Finland	fi	8	1,928	241
17	Russian Fed	ru	11	1,799	163.5
18	The Netherlands	nl	34	1,524	44.82
19	Sweden	se	17	1,425	83.8
20	Portugal	pt	8	1,418	177.2
21	South Korea	kr	7	1,237	176.7
22	USA	us	7	1,236	176.5
23	Ireland	ie	6	1,227	204.5
24	Czech Republic	cz	9	1,138	126.4
25	New Zealand	nz	28	1,094	39
26	Greece	gr	11	966	87.8
27	South Africa	za	9	161	17.8

Table III.
Frequency of sample sites (domains), link impact report and average number of link per page

Conclusion

The main objective of the present study was to conduct a webometric analysis of keywords and expressions of the biochemistry field of study using LexiURL Searcher to see if the results pointed toward related Web sites to narrow down how many sites users need to examine when seeking biochemistry information. Also, to identify the countries involved in the biochemistry field of study and how they cooperate with each other, the authors restricted the primary results to the domains of those countries; consequently, countries active in the biochemistry field and the way they connect with each other were determined. In general, there is no connection or relation among the biochemical Web sites of the countries under investigation and no interaction is observed among them (Figure 2); however, with respect to connections with each other, the Web sites have some links to very important databases, such as with the National Center of Biotechnology Institution.

It is obvious that to link Web sites from a specialized domain could help their users to better process information. In his study conducted on a research group, Thelwall (2008) found that all connections among the users and exchange of information among the experts took place through links and connections on the Web. Also, according to the enclosed research, it is obvious that exploring the Web sites and Web pages in

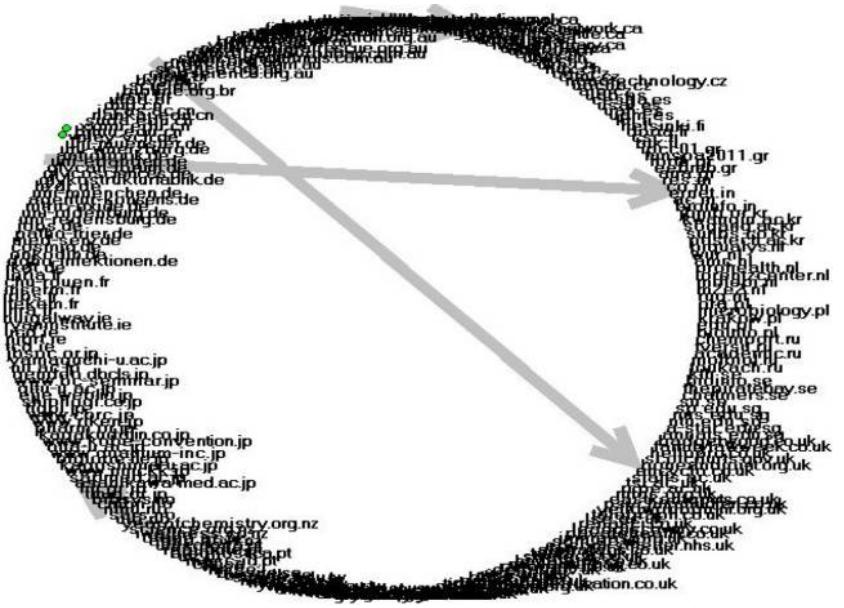

Figure 1.
Network diagram of the top biochemistry countries' sites

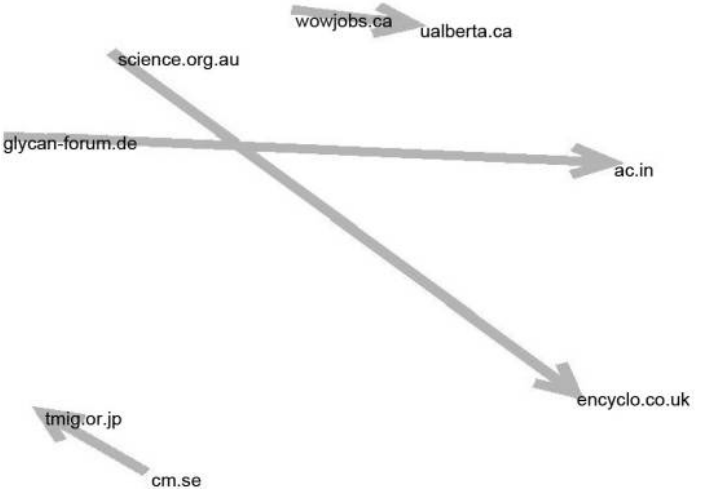

Figure 2.
Something or other (no title)

the biochemistry field has formed the basis of keywords derived from the MeSH database that have been confirmed and approved by the experts. Finally, the authors believe that the related Web sites carry effective links which enable users to access tremendous amounts of information that are loaded with required levels of quantity and quality. In a study on the NeLH, the results showed that users who use Web sites which are connected to rich sources and databases have a better chance to obtain more relevant information (Zuccala *et al.*, 2007). Another point in this regard pertains to the issues of

webometrics and the visibility of Web sites. Under the rubric of search results in ranking the Web sites, search engines take into account citations to available contents, the existence of in links and out links, along with other indices, such as being upgraded and so forth (Glover *et al.*, 2002). Also, the existence of proper links – especially out links – to topic-related Web sites leads users to a network of related Web sites and contents that in turn increases the percentage and rate of clicking by the users. This also improves the ranking of the Web site in the ranking of search engines, thereby increasing their visibility.

References

Barnett, G.A. (2012), "Recent developments in the global telecommunication network", *Proceedings of the 2012 45th Hawaii International Conference on System Science (HICSS)*, pp. 4435-4444.

Barnett, G.A., Chung, C.J. and Park, H.W. (2011), "Uncovering transnational hyperlink patterns and web-mediated contents: a new approach based on cracking.com domain", *Social Science Computer Review*, Vol. 29 No. 3, pp. 369-384.

Glover, E.J., Tsioutsiouliklis, K., Lawrence, S. and Pennock, D.M. (2002), "Using web structure for classifying and describing web pages", *Proceedings of the 11th International Conference on World Wide Web*, *ACM, New York, NY*, pp. 562-569.

Herring, S.C. (2010), "Web content analysis: expanding the paradigm", *The International Handbook of Internet Research*, available at: http://link.springer.com/chapter/10.1007%2F978-1-4020-9789-8_14 (accessed 18 May 2013).

Hsu, C.-L. and Park, H.W. (2012), "Korean and Chinese webpage content: who are talking about what and how?", *Journal of Computer-Mediated Communication*, Vol. 17 No. 2, pp. 202-215.

Huvila, I. (2011), "I asked my Mum, but and other cases of unsuccessful information seeking by asking Isto", *Information Research*, Vol. 16 No. 1, pp. 1-15, available at: http://InformationR.net/ir/16-1/paper460.html (accessed 10 December 2013).

McCown, F. and Nelson, M.L. (2007), "Agreeing to disagree: search engines and their public interfaces", *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*, *ACM, New York, NY*, pp. 309-318.

Page, L., Brin, S., Motwani, R. and Winograd, T. (1998), "The PageRank citation ranking: bringing order to the web", Stanford InfoLab Technical Report, available at: http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf (accessed 15 December 2013).

Park, H.W. (2010), "Mapping the e-science landscape in South Korea using the webometrics method", *Journal of Computer-Mediated Communication*, Vol. 15 No. 2, pp. 211-229.

Park, H.W., Barnett, G.A. and Chung, C.J. (2011), "Structural changes in the 2003-2009 global hyperlink network", *Global Networks: A Journal of Transnational Affairs*, Vol. 11 No. 4, pp. 522-542.

Sreedhar, G., Chari, A.A. and Ramana, V. (2010), "A qualitative and quantitative frame work for effective web site design", *International Journal of Computer Applications*, Vol. 2 No. 1, pp. 71-79.

Thelwall, M. (2008), "Quantitative comparisons of search engine results", *Journal of the American Society for Information Science and Technology*, Vol. 59 No. 11, pp. 1702-1710.

Thelwall, M., Li, X., Barjak, F. and Robinson, S. (2008), "Assessing the international web connectivity of research groups", *Aslib Proceedings*, Vol. 60 No. 1, pp. 18-31.

Woo, P.H. and Yeon-ok, L. (2008), "The Korean presidential election of 2007: five years on from the 'Internet Election'", *Journal of Contemporary Eastern Asia*, Vol. 7 No. 1, pp. 1-4.

Zuccala, A., Thelwall, M., Oppenheim, C. and Dhiensa, R. (2007), "Web intelligence analyses of digital libraries: a case study of the National electronic Library for Health (NeLH)", *Journal of Documentation*, Vol. 63 No. 4, pp. 558-589.

Zuccala, A., Oppenheim, C. and Dhiensa, R. (2008), "Managing and evaluating digital repositories", *Information Research*, Vol. 13 No. 1, pp. 1-13, available at: www.informationr.net/ir/13-1/paper333.html (accessed 15 December 2013).

Further reading

Yen, B., Hu, P.J.-H. and Wang, M. (2007), "Toward an analytical approach for effective web site design: a frame work for modeling, evaluation and enhancement", *Electronic Commerce Research and Applications*, Vol. 6 No. 2, pp. 159-170.

About the authors

Kobra Taram has an MS in Library and Information Science, and his main fields are information retrieval, Web sites and software.

Abbas Doulani is a PhD student in Library and Information Science, and his main fields are information retrieval, Web sites, computer networks and software. Abbas Doulani is the corresponding author and can be contacted at: a.doulani@gmail.com