

Web Presence and Impact Factors for Middle-Eastern Countries

Alireza Noruzi

Department of Library and Information Science, University of Tehran

E-mail: [nouruzi at gmail.com](mailto:nouruzi@gmail.com)

Abstract

This study investigates the *Web presence* and *Web Impact Factor* (WIF) for country code top-level domains (ccTLDs) of Middle-Eastern countries, and sub-level domains (SLDs) related to education and academic institutions in these countries. Counts of links to the web sites of Middle-Eastern countries were calculated from the output of Yahoo search engine. In this study, we compute the WIF at two levels: top-level domains, and sub-level domains.

The results show that the Middle-Eastern countries, apart from Turkey, Israel and Iran, have a low web presence. On the other hand, their web sites have a low inlink WIF. Specific features of sites may affect a country's Web Impact Factor. For linguistic reasons, Middle-Eastern web sites (Persian, Kurdish, Turkish, Arabic, and Hebrew languages) may not receive and attract the attention that they deserve from the World Wide Web community.

Introduction

Questions can arise about how global the World Wide Web really is. Particularly for Middle Eastern countries, it is important to know whether Web presence and Web Impact Factor (WIF) attract the attention they deserve from the World Wide Web community. The academic community, in particular, is ill served if an important geographical region is under-represented.

This study investigates the Web presence and Web Impact Factor (WIF) for country code top-level domains (ccTLDs) of Middle-Eastern countries, and sub-level domains (SLDs) related to education and academic institutions in these countries. The study calculated counts of links to the Web sites of Middle-Eastern countries from Yahoo! searches and computes the WIF at two levels: top-level domains and sub-level domains. The results show that the Middle-Eastern countries -apart from Turkey, Israel, and Iran- have a low Web presence.

On the other hand, these three countries' Web sites have a low inlink WIF. Specific features of sites may affect a country's WIF. For linguistic reasons, Middle-Eastern Web sites (Persian, Kurdish, Turkish, Arabic, and Hebrew languages) may not receive and attract the attention that they deserve from the World Wide Web community.

The World Wide Web is a reflection of human culture, a massive socio-cultural network of Web resources authored by millions of people and organizations around the world. As a whole, "the Web displays a striking '*rich get richer*' behavior, with a relatively small number of countries having a disproportionately large number of Web sites and pages and share of hyperlink references and traffic" (Pennock et al., 2002). Think of the scientific community on the Web, and especially online journals, as a citation network where traditional information entities and citations from them are replaced by Web pages (e-articles) and hyperlinks, respectively.

Webometric Research

The WIF is an important part of Webometric research, which studies hyperlinks and the impacts and influence of Web sites. Webometric studies display several similarities to informetric and scientometric studies and use bibliometric methods, according to Almind and Ingwersen, who first discussed these concepts in 1997. For example, simple counts and content analysis of Web resources resemble traditional publication analysis. Counts and analyses of outlinks (outgoing links from Web resources) and inlinks (backlinks pointing to Web resources)-can be seen as reference and citation analyses, respectively. Outlinks and

inlinks are similar to references and citations, respectively, in scientific e-articles. Webometric studies of the structure and content of Web sites in various countries, as well as link structures, are important to understand the international virtual highway and interconnections among countries. The WIF provides quantitative tools for ranking, evaluating, categorizing, and comparing Web sites, top-level domains, and sub-domains.

The primary objective of this study is to formulate a methodology for the calculation of WIF at two hierarchical levels: ccTLD and SLD. This involved three activities.

1. Calculate WIF for all the Middle-Eastern countries and rank them based on their inlink (backlinks coming from other countries) WIFs.
2. Calculate WIF for SLDs related to education and academic institutions and rank them based on their inlink WIFs.
3. Show the number of Web pages from these countries indexed by the Yahoo! search engine and rank them based on their Web page size.

The WIF as a useful measure of the overall influence of a Web site, using the backlinks or inlinks to the Web site, has been proposed independently by two bibliometric researchers (Rodríguez i Gairín, 1997; Ingwersen, 1998). For a detailed literature review, see Noruzi, 2005.

Web Address Structure

The Web address is hierarchical in structure. This hierarchy has its origin in the Domain Name System (DNS). The DNS translates the plain-English address (ut.ac.ir) into a corresponding IP address (217.218.33.14). From right to left, the domain name structure has the following hierarchy:

- Top-level domain
- Sub-level domain
- Host-level domain (site/server domain)

In the above example, the hierarchy is as follows: .ir (top-level domain for Iran), .ac (sub-level domain of academic sites under .ir), and .ut (specific domain of the University of Tehran, operating under top-level domain .ir and sub-level domain .ac). The ccTLD is allotted for each country in accordance with two-letter codes based on ISO-3166 (.ir for Iran, .sa for Saudi Arabia, and so on). Each Middle-Eastern country has an SLD for universities and academic institutions as outlined in Table 1.

Table 1. ccTLD of Middle-Eastern countries and their SLDs for universities in these countries

Countries	TLD	SLD
Bahrain	.bh	.edu.bh
Cyprus	.cy	.ac.cy
Egypt	.eg	.edu.eg
Iran	.ir	.ac.ir
Iraq	.iq	.ac.iq
Israel	.il	.ac.il
Jordan	.jo	.edu.jo
Kuwait	.kw	.edu.kw
Lebanon	.lb	.edu.lb
Oman	.om	.edu.om
Palestine	.ps	.edu.ps
Qatar	.qa	.edu.qa
Saudi Arabia	.sa	.edu.sa
Syrian Arab Republic	.sy	.edu.sy
Turkey	.tr	.edu.tr
United Arab Emirates	.ae	.ac.ae
Yemen	.ye	.edu.ye

Methodology

The most convenient way of measuring links among countries' sites is to use the advanced search facilities of general Web search engines, several of which, notably Yahoo!, include link data in their databases. Several WIF studies have been carried out using the advanced search facilities of AltaVista, both before and after its acquisition by Yahoo! (Ingwersen, 1998; Smith & Thelwall, 2002; Thelwall, 2002; Kousha & Horri, 2004; Smith, 2004; Noruzi, 2005), and of Yahoo! itself (Noruzi, 2006). This study uses Yahoo! because it offers special commands that search for matches only in Web elements such as pages, domains, links, and so on.

Google's advanced search facility does not support the same level of Boolean querying as Yahoo! or AlltheWeb. Its advanced search can limit the source to a given domain but it cannot explicitly exclude all links from within the site itself (it cannot eliminate the self-links), a second critical gap in its functionality.

In this study, Yahoo! is used to collect data for the calculation of WIF at different levels. Yahoo! supports *linkdomain:* as a command to find pages with a backlink to a site. For example, *linkdomain:ut.ac.ir* will find all pages with at least one link to the Web site of the University of Tehran. It also supports *domain:* as a command to retrieve the number of Web pages indexed per site or domain. Using these commands, the study collected the number of Web pages and the number of link-pages, respectively, from the Yahoo! search engine. Yahoo! reports the number of Web pages retrieved against each search. For example, the following queries will retrieve the number of Web pages, the number of inlinks, and the number of self-links for the ccTLD of Iran (.ir):

[linkdomain:ir/](#)

Will report total number of Web pages in Yahoo! database that link to .ir domain (ccTLD of Iran), i.e., total number of link-pages.

[linkdomain:ir/ NOT domain:ir/](#)

Will report number of Web pages not under .ir domain but that link to .ir domain (ccTLD of Iran), i.e., inlink pages.

[linkdomain:ir/ AND domain:ir/](#)

Will report number of Web pages under .ir domain that link to .ir domain (ccTLD of Iran), i.e., self-links.

[domain:ir/](#)

Will report number of pages under .ir domain (ccTLD of Iran) indexed by Yahoo! search engine.

Data collection

Data collection took place on Sept. 29, 2005. All the domain names were searched to check whether Yahoo!'s databases include these ccTLDs. For each of the countries, a search was carried out to determine the total number of links, the number of inlinks, the number of self-links, and the total number of Web pages at the domain. Searches were carried out to determine the following:

- the total number of pages linking to the ccTLD, A, for example
[linkdomain:ir/](#)
- the number of pages at the ccTLD, D, determined by the command:
[domain:ir/](#)

- the number of inlinks (links from pages outside the ccTLD), B:
[linkdomain:ir/ NOT domain:ir/](#)
- the number of self-links (links from pages in the same ccTLD), C, measured in the following way:
[linkdomain:ir/ AND domain:ir/](#)

Table 2. Number of pages indexed by Yahoo! from each Middle-Eastern country

Country	No. of Web Pages
Turkey	29300000
Israel	27000000
Iran	5560000
Saudi Arabia	3270000
Egypt	3110000
United Arab Emirates	2210000
Lebanon	1500000
Cyprus	931000
Kuwait	854000
Jordan	538000
Palestine	287000
Bahrain	170000
Qatar	151000
Oman	70800
Yemen	53700
Syrian Arab Republic	16300
Iraq	0

Figure 1. Middle-Eastern countries with the highest web presence

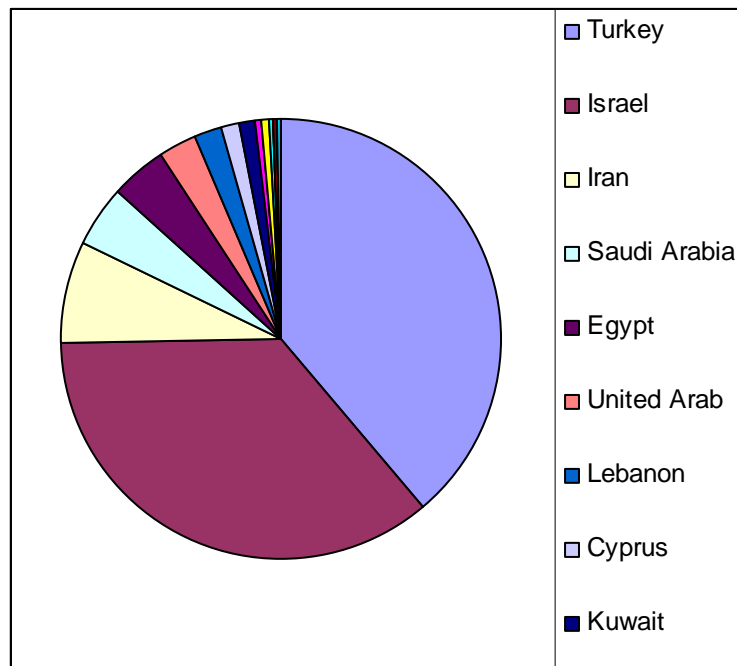


Table 3. WIF for ccTLDs of Middle-Eastern countries

Country	WIF (A/D) Web Impact Factor (overall WIF)	A Links to ccTLD (total)	B Inlinks to ccTLD, without self-links	C Self-links to ccTLD	D Web pages indexed by Yahoo	R-WIF (B/D) Revised Web Impact Factor
Oman	2.56	181000	154000	15700	70800	2.18
Syrian Arab Republic	1.31	21300	16600	6640	16300	1.02
Yemen	1.28	68500	41800	10700	53700	0.78
Palestine	1.03	297000	218000	68400	287000	0.76
Qatar	0.75	114000	82300	25700	151000	0.55
Bahrain	0.69	118000	80000	38000	170000	0.47
Saudi Arabia	0.71	2330000	1520000	762000	3270000	0.46
Lebanon	0.60	902000	567000	304000	1500000	0.38
Jordan	0.51	274000	154000	110000	538000	0.29
Cyprus	0.48	447000	233000	202000	931000	0.25
United Arab Emirates	0.47	1030000	506000	546000	2210000	0.23
Israel	0.42	11300000	5210000	6390000	27100000	0.19
Kuwait	0.42	355000	142000	202000	854000	0.17
Iran	0.37	2050000	758000	1280000	5560000	0.14
Turkey	0.35	10400000	3690000	6430000	29300000	0.13
Egypt	0.37	1150000	384000	777000	3110000	0.12
Iraq	0.00	0	0	0	0	0.00

For universities and academic institutions in these countries, the following searches were carried out to determine:

- the total number of pages linking to the SLD, A, for example
[linkdomain:ac.ir/](#)
- the number of pages at the SLD, D, determined by the command:
[domain:ac.ir/](#)
- the number of inlinks (links from pages outside the SLD), B:
[linkdomain:ac.ir/ NOT domain:ir/](#)
- the number of self-links (links from pages in the same SLD), C, measured in the following way:
[linkdomain:ac.ir/ AND domain:ir/](#)

Figure 2. Middle-Eastern countries with the highest academic web presence

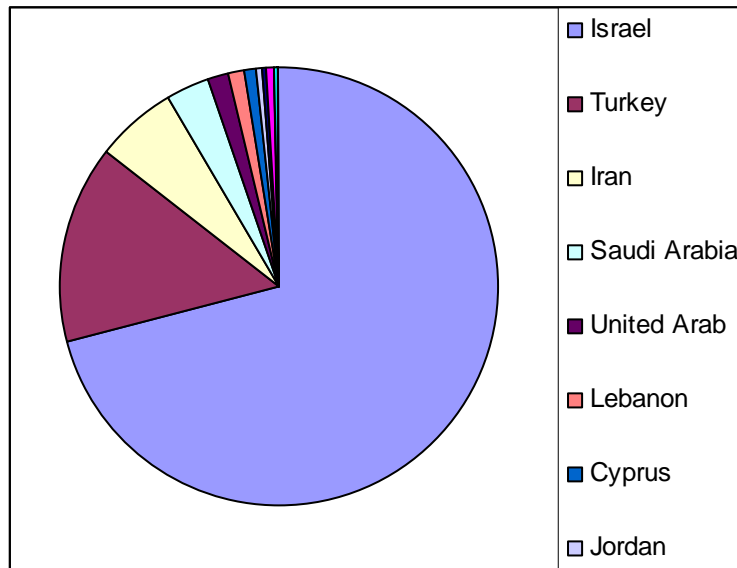


Table 4. WIF for academic SLDs

Country	WIF (A/D) Web Impact Factor (overall WIF)	A Links to SLD (total)	B Inlinks to SLD, without self-country links	C Self-country links to SLD	D Web pages indexed by Yahoo	R-WIF (B/D) Revised Web Impact Factor
Oman	8.04	80200	85500	4210	9980	8.57
Yemen	2.13	2720	1920	903	1280	1.50
Syrian Arab Republic	1.83	241	154	109	132	1.17
Bahrain	1.04	4430	3190	1630	4250	0.75
Kuwait	1.11	34000	19700	16800	30700	0.64
Palestine	1.35	778	364	585	577	0.63
Saudi Arabia	1.13	259000	119000	139000	230000	0.52
Turkey	1.39	1570000	458000	1130000	1130000	0.41
Jordan	1.01	37800	13700	14200	37400	0.37
Egypt	0.73	19500	9050	6300	26700	0.34
Cyprus	0.87	56300	20900	19400	65000	0.32
Qatar	0.91	9470	3040	7470	10400	0.29
Lebanon	0.96	90500	25800	64200	94600	0.27
United Arab Emirates	0.72	85500	29700	55600	118000	0.25
Iran	1.55	725000	90900	634000	468000	0.19
Israel	0.37	2020000	711000	1300000	5420000	0.13
Iraq	0.00	0	0	0	0	0.00

Results

The data obtained from various search statements by following the above-mentioned methodology are in Tables 2 to 4. The WIF for each ccTLD and SLD has been calculated at two levels -overall WIF by considering all the link-pages, and inlink WIF by considering only inlink pages without self-links. The ranking is based on revised WIF (inlink WIF), as this is the true reflection of the degree of impact of the domain spaces on the Web.

Table 2 and Figure 1, show that Turkey, Israel, and Iran respectively have the highest Web page size among Middle-Eastern countries, while Table 3 shows that they have a low WIF. This unexpected result occurred because the higher number of Web pages generated a comparatively lower number of link-pages. The higher the Web page size, the lower the WIF for the country. Therefore, countries that publish many Web resources may not have as high an impact as countries with few Web pages, because the high Web resources rate counteracts the high inlinks rate. Additionally, results overall suggest that Web sites from these countries are somewhat insular: nationally well-interconnected, but less-well-known internationally.

Limitations of using the WIF

Webometrics is the extension of the theory and practice of bibliometric techniques in the Web. Bibliometric research has been criticized for certain inherent limitations of ISI (now part of Thomson Scientific) products (Moed, 2002; Seglen, 1997). Similarly, search engines as primary data-gathering instruments may create problems in drawing conclusions for WIF studies. The tool being used for WIF analysis is not specifically meant for the task. Search engines are designed for content retrieval not link analysis. These problems are technical and could be resolved if the search engine programmers had incentives to work on them. As the data-gathering mechanism is quite easy to follow by using commercial search engines, Webometrics has all the potential to evolve as a tool for performance evaluation of any Web site instantly (Mukhopadhyay, 2004).

A primary limitation of the current study is that, although several thousand Middle-Eastern Web sites have generic top-level domains such as dot-org, dot-com, dot-net, given the current features of the search engines that serve as the basic data mechanism, it is not possible to determine how many Web sites have generic TLDs. Thus, the current research has considered only top-level domains from these countries. For example, the number of pages of Persianblog, one of the well-known Iranian blogs, is two times greater than Iranian university Web pages:

domain:persianblog.com/	955,000
domain:ac.ir/	468,000

In the case of Iraqi-Kurdistan Universities, which have not used SLDs such as '*edu.iq*' or '*ac.iq*', we calculated their WIFs separately (see Table 5).

Table 5. WIF for Iraqi-Kurdistan Universities

University	URL	WIF (A/D) Web Impact Factor (overall WIF)	A Links to the site (total)	B Inlinks to the site, without self-links	C Self-links	D Web pages indexed by Yahoo	R-WIF (B/D) Revised Web Impact Factor
University of Sualimani	univsul.com	9.37	553	701	14	59	11.88
University of Dohuk	uni-dohuk.net	4.35	648	722	109	149	4.85
Salahaddin University	usalah.org	3.49	136	110	26	39	2.82

Therefore, the WIF is not a perfect tool to measure the quality, or even the quantity, of Web sites from a country. Despite recognizing that the WIF is an imperfect measure with 10 years of criticism, there is no obvious alternative. Those forced to use this tool for direct Web site comparison should be encouraged to remain open-minded and cautious, with an awareness of

the inherent limitations of its use. Although the WIF is arguably useful for quantitative intra-country comparison, application beyond this (to inter-country assessment) has little value.

Conclusion

A comparison of Middle-Eastern countries' sites raises interesting questions about the place of different countries, cultures, and languages on the Web. These countries are outside the main Web area, which is dominated by the U.S., Canada, Europe, Australia, India, Japan, and China. It appears that Middle-Eastern Web sites may achieve a lower visibility on the Web because their language and culture are different from the current mainstream of the Web, dominated by English-speaking countries. This should be a warning to cybercitizens.

Further research is needed to gain a better understanding of the nature of Web links, and further research may be necessary to find reasons for the limited number of Middle-Eastern Web pages. It could also be interesting to investigate the Web presence of African countries to see if their situation is similar.

Acknowledgements

The author wishes to thank Mrs. Marjorie Sweetko for her helpful comments.

References

- Almind, T.C., & Ingwersen, P. "Informetric analyses on the World Wide Web: methodological approaches to Webometrics." *Journal of Documentation* 53, No. 4 (1997): pp. 404-426.
- Ingwersen, P. "The calculation of Web Impact Factors." *Journal of Documentation* 54, No. 2 (1998): pp. 236-243.
- Kousha, K., & Horri, A. "The relationship between scholarly publishing and the counts of academic inlinks to Iranian university web sites: exploring academic link creation motivations", *Journal of Information Management and Scientometrics*, Vol. 1, No. 2 (2004): pp. 13-22.
- Moed, H.F. "The impact factors debate: the ISI's uses and limits." *Nature* 415, (2002): pp. 731-2.
- Mukhopadhyay, P. "Measuring Web Impact Factors: A webometric study based on the analysis of hyperlinks." In: *Proceedings of National Seminar on Information Support for Rural Development*, India, IASLIC, December (2004).
- Noruzi, A. "The Web Impact Factor: A critical review." *The Electronic Library* 24, No. 4 (2006): pp. 490-500. Available at: <http://eprints.rclis.org/archive/00005543/>
- Noruzi, A. "Web Impact Factor for Iranian Universities." *Webology* 2, No. 1 (April 2005), Article 11. Available at: <http://www.webology.ir/2005/v2n1/a11.html>
- Pennock, D.M., Flake, G.W., Lawrence, S., Glover, E.J., & Giles, C.L. "Winners don't take all: Characterizing the competition for links on the Web." *PNAS* 99, No. 8 (April 16, 2002): pp. 5207-5211.
- Rodríguez i Gairín, J.M "Valorando el impacto de la información en Internet: AltaVista, el "Citation Index" de la red." [Impact assessment of information on the Internet: AltaVista, the Citation Index of the Web]. *Revista Española De Documentación Científica* 20, No. 2 (1997): pp. 175-181.
- Seglen, P.O. "Why the impact factor of journals should not be used for evaluating research." *British Medical Journal* 314, (1997): pp. 498-502.
- Smith, A.G. "Citations and links as a measure of effectiveness of online LIS journals." *World Library and Information Congress: 70th IFLA General Conference and Council, 22-27 August 2004*, Buenos Aires, Argentina. Retrieved September 29, 2005, from <http://www.ifla.org/IV/ifla70/papers/049e-Smith.pdf>
- Smith, A.G., & Thelwall, M. "Web Impact Factors for Australasian Universities." *Scientometrics* 54 No. 3 (2002): pp. 363-380.
- Thelwall, M. "A comparison of sources of links for academic Web Impact Factor calculations." *Journal of Documentation* 58, No. 1 (2002): pp. 66-78.