The Web Presence of Middle-Eastern and European Countries: A Digital Divide

ALIREZA NORUZI
University of Paul Cezanne Aix-Marseille III, GERSIC/CRRM, Faculty of Science & Technology, Marseille, France Email: anouruzi@yahoo.com

Abstract

This study investigates the Web presence of country code top-level domains (ccTLDs) of European and Middle-Eastern countries. Counts of web pages from European and Middle-Eastern countries were collected from the output of the Yahoo search engine. This study shows that European and Middle-Eastern countries with a higher number of Internet users have a higher web presence. The results show that the European countries, especially Germany, the United Kingdom and Italy, have the highest web presence, while the Middle-Eastern countries, apart from Turkey, Israel and Iran, have the lowest web presence. Specific features of countries may affect a country's web presence, for example for linguistic reasons; Middle-Eastern web sites (Persian, Kurdish, Turkish, Arabic and Hebrew languages) may not attract the attention they deserve from the Web community. They are not highly linked and thus they are not indexed well by search engines. Consequently, they have a low visibility on the Web. The results show that there is even a digital divide between European countries. A further gap between the developed and developing world in the uptake of information and communication technologies is obvious within the global community, and may be of even greater significance.

Keywords: Webometrics, Web presence, Internet access, Middle-East, Europe

1. Introduction

In the global digital information age, those who are unable to access the Internet and the World Wide Web through the application of information and communication technologies (ICTs) are increasingly disadvantaged in their access to information [1]. The Web is an unprecedented medium where everybody can publish his or her ideas. The 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. Overall, "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" [2].

Web presence evaluation is an important part of webometric research, which studies the Web presence of different countries. 'Webometric' studies display several similarities to informetric and scientometric studies and use bibliometric methods [3]. For example, simple counts and content analysis of web pages are like 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. Webometric studies of the structure and content of web sites in various countries, as well as link structures, are important for understanding the international virtual highway and interconnections between countries. Web presence studies provide quantitative tools for ranking, evaluating, categorizing and comparing web sites, top-level domains and second-level domains.

The most convenient way of measuring web presence of countries' is to use the advanced search facilities of large-scale search engines, such as Yahoo and Google. Previous studies have been carried out using the advanced search facilities of Yahoo [4] [5]. Web presence evaluation can be a useful measure of the overall presence of a country, using the number of web pages published by the given country. Web presence studies can be done to determine "*digital divide*" between rich and poor countries.

The *digital divide* is a generic term used to describe the lack of access to information and communication technologies (such as computer, Internet and Web) due to linguistic, socio-cultural, political, educational, economic and geographic factors. From an Internet user's point of view, the *digital divide* is a gap between those who have ready access to the Internet and make effective use of the Internet as a communications and information medium and those without such access or skills.

The term "*digital divide*" was coined by James D. Wolfensohn [6], president of the World Bank, to describe the perceived growing gap between countries that have access and skills to use ICT and those that, for socio-economic, political, geographical, educational, attitudinal and generational factors, have limited or no access. There was a particular concern that ICT would exacerbate existing inequalities. Wolfensohn has stressed the need to bridge the technological gap between rich and poor nations. According to Wolfensohn, "the digital divide is one of the greatest impediments to development and it is growing exponentially" [7].

As an example, a study in the United States presented overall differences between whites and African Americans on computer access and general Web use. This study found that individuals who own a home computer and have access to a computer at work are much more likely than any other group to have used the Web. This study has also shown that whites are more likely to use the Web than African Americans. While income explains race differences in home computer ownership, whites are still more likely to own a home computer than African Americans at each and every education level, despite controlling for differences in education [8]. The digital divide is more marked for Internet access. Various statistics show that North America and Europe have the highest number of Internet hosts and Internet users.

Following the enlargement of the European Union (EU) from 15 to 25 member states, the digital divide in the EU widened substantially (national Internet connectivity varying from less than 10 to more than 60 per cent). Inequality in Internet-connectivity in the EU will increase dramatically, with all consequences for communication, dissemination of information, economy (e-commerce), caused by regional lack of technological infrastructure as well as cultural and psychological factors [9].

The hypothesis of this study is that countries with higher number of Internet users publish more web pages than countries with limited access to the Internet.

Objective

The primary aim of this study is to evaluate the Web presence of different countries at country code top-level domain (ccTLD) levels. More specifically the objectives of the study are:

1. To compare the Web presence of all the Middle-Eastern countries and rank them based on their web page size;
2. To compare the Web presence of all the European countries and rank them based on their web page size;
3. To show the number of web pages from these countries indexed by the Yahoo search engine; and
4. To show the digital divide between countries.

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 (e.g. ut.ac.ir) into a corresponding IP address (e.g. 217.218.33.14). From the right hand side, the domain name structure has the following hierarchy:

- Top-level domain
    - Second-level domain
        - Host level domain (site domain)

In the above example, the hierarchy is: *.ir* (Top-level domain for Iran), *.ac* (Second-level domain for academic sites under *.ir*), and *.ut* (Specific domain of the University of Tehran, operating under top-level domain *.ir* and second-level domain *.ac*). The ccTLD is allotted for each country in accordance with two letter codes based on ISO-3166 (e.g. *.uk* for United Kingdom, *.fr* for France, *.ir* for Iran). Each country has a Top-Level Domain (TLD) as outlined below (see tables 1 and 2).

Methodology

It is important to understand how to collect data about the number of web pages under any given top-level domain. Large-scale search engines are used for collecting such data. Most of the well-known search engines (such as Yahoo and Google) offer special commands to search for matches in web elements such as pages, domains and links. Yahoo and Google are currently the most widely-used search engines.

In this study, Yahoo is used to collect data for purposes of comparison at different levels. Yahoo supports "*domain*:" as a command to retrieve the total number of web pages indexed per domain or site. This data collection method extensively uses the special command "*domain*:" to collect the number of web pages from Yahoo. This search engine reports the number of web pages retrieved against each search. For example, *domain:uk/* will find the number of pages under .uk domain (ccTLD of the United Kingdom) indexed by the Yahoo search engine. Yahoo provides a quantitative tool for ranking and

evaluating top-level domains and second-level domains. For example, this search engine can be used to compare the web presence of second-level domains under .uk domain (co.uk/, ac.uk/, org.uk/, net.uk/, gov.uk/, etc). For instance,

domain:ac.uk/

Data Collection

The number of web pages under ccTLDs of Middle-Eastern and European countries was counted using the Yahoo search engine. Data collection took place on March 6, 2006. All the domain names were searched to check whether Yahoo includes these ccTLDs. For each of the countries, a search was carried out to determine the total number of web pages at the domain. Searches were carried out to determine the total number of web pages at each ccTLD, for example:

domain:uk/

The data shown in column 3 of tables 1 and 2 relating to the Number of Web Pages were collected from the output of the Yahoo search engine. The data shown in columns 4 and 5 of tables 1 and 2 relating to Internet Users and Total Domains (org, com, net, info, biz) derived from the following web sites respectively: InternetWorldStats.com [10] and WebHosting.info [11]. The data shown in columns 6 and 7 of tables 1 and 2 relating to Population and Literacy derived from the CIA's World Factbook [12]. Tables 1 and 2 include all Internet users, who have accessed the Internet. Although this data is not the most up-to-date available for all countries listed, for example the data for 2006 is not available.

Table 1: Number of pages indexed by Yahoo from each Middle-Eastern country

Middle-Eastern Countries	**ccTLD**	**No. of Web Pages**	**Internet Users** [10]	**Total Domains** (org, com, net, info, biz) [11]	**Population** [12]	**Literacy (total population)** [12]
Turkey	.tr	38,300,000	10,220,000	320,463	69,660,559	86.5%
Israel	.il	35,500,000	3,200,000	57,173	6,276,883	95.4%
Iran	.ir	6,460,000	7,500,000	54,413	68,017,860	79.4%
Saudi Arabia	.sa	3,790,000	2,540,000	27,283	26,417,599	78.8%
Egypt	.eg	3,770,000	4,200,000	25,335	77,505,756	57.7%
United Arab Emirates	.ae	2,660,000	1,384.800	16,765	2,563,212	77.9%
Lebanon	.lb	1,760,000	600,000	10,821	3,826,018	87.4%
Kuwait	.kw	884,000	600,000	8,779	2,335,648	83.5%
Jordan	.jo	628,000	600,000	3,532	5,759,732	91.3%
Palestine	.ps	420,000	160,000	1,005	3,259,363 [10]	NA
Bahrain	.bh	252,000	152,700	2,278	688,345	89.1%
Qatar	.qa	247,000	165,000	918	863,051	89%
Oman	.om	102,000	245,000	3,469	3,001,583	75.8%
Yemen	.ye	63,800	220,000	1,099	20,727,063	50.2%

Syria	.sy	39,800	800,000	6,604	18,448,752	76.9%
Iraq	.iq	0	36,000	435	26,074,906	40.4%

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

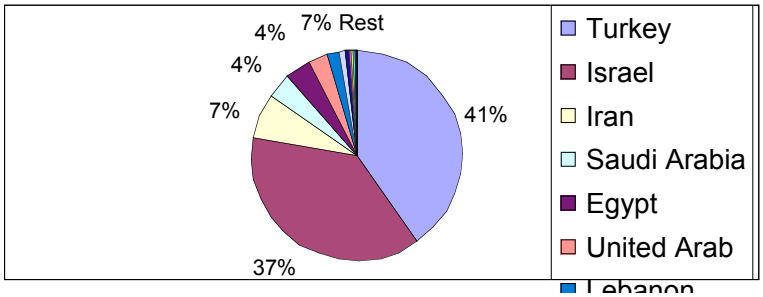


Figure 2: Middle-Eastern countries with the highest number of domains

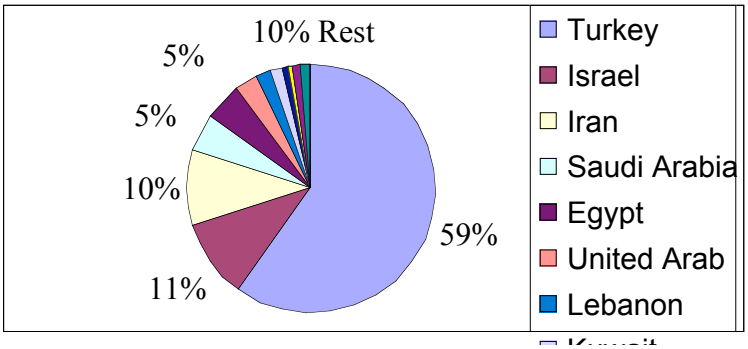


Table 2: Number of pages indexed by Yahoo from each European country

European Countries	**ccTLD**	**No. of Web Pages**	**Internet Users** [10]	**Total Domains** (org, com, net, info, biz) [11]	**Population** [12]	**Literacy (total population)** [12]
Germany	.de	1,390,000,000	48,722,055	3,329,943	82,431,390	99%
United Kingdom	.uk	735,000,000	37,800,000	2,666,108	60,441,457	99%
Italy	.it	356,000,000	28,870,000	659,064	58,103,033	98.6%
Netherlands	.nl	321,000,000	10,806,328	485,157	16,407,491	99%
France	.fr	239,000,000	26,214,174	1,443,767	60,656,178	99%
Poland	.pl	202,000,000	10,600,000	105,565	38,635,144	99.8%
Sweden	.se	171,000,000	6,800,000	189,176	9,001,774	99%
Switzerland	.ch	150,000,000	4,944,438	163,308	7,489,370	99%
Denmark	.dk	145,000,000	3,762,500	249,150	5,432,335	100%
Czech Republic	.cz	141,000,000	4,800,000	80,365	10,241,138	99.9%
Norway	.no	129,500,000	3,140,000	155,065	4,593,041	100%
Spain	.es	124,000,000	17,142,198	784,261	40,341,462	97.9%
Austria	.at	123,000,000	4,650,000	150,103	8,184,691	98%
Finland	.fi	105,000,000	3,286,000	115,800	5,223,442	100%
Belgium	.be	104,000,000	5,100,000	93,210	10,364,388	98%

Hungary	.hu	58,800,000	3,050,000	25,302	10,006,835	99.4%
Romania	.ro	32,200,000	4,940,000	32,605	22,329,977	98.4%
Ireland	.ie	36,900,000	2,060,000	65,504	4,015,676	98%
Greece	.gr	32,600,000	3,800,000	23,542	10,668,354	97.5%
Ukraine	.ua	32,200,000	5,278,100	37,188	47,425,336	99.7%
Slovakia	.sk	28,900,000	2,276,000	8,480	5,431,363	99.6%
Portugal	.pt	22,700,000	6,090,000	50,319	10,566,212	93.3%
Estonia	.ee	19,000,000	670,000	11,050	1,332,893	99.8%
Lithuania	.lt	18,600,000	968,000	7,882	3,596,617	99.6%
Iceland	.is	11,800,000	225,600	3,312	296,737	99.9%
Croatia/Hrvatska	.hr	16,500,000	1,303,000	19,596	4,495,904	98.5%
Latvia	.lv	12,200,000	810,000	7,979	2,290,237	99.8%
Slovenia	.si	11,300,000	950,000	18,422	2,011,070	99.7%
Serbia and Montenegro	.yu	8,900,000	1,200,000	12,409	10,829,175	96.4%
Bulgaria	.bg	7,290,000	2,200,000	68,653	7,450,349	98.6%
Luxembourg	.lu	5,580,000	270,800	16,770	468,571	100%
Belarus	.by	4,580,000	1,600,000	4,319	10,300,483	99.6%
Moldova	.md	2, 600,000	406,000	3,500	4,455,421	99.1%
Macedonia	.mk	1,770,000	392,671	321	2,045,262	96.1%
Liechtenstein	.li	1,640,000	20,000	4,678	33,717	100%
Bosnia - Herzegovina	.ba	1,240,000	225,000	2,749	4,025,476	94.6%
Cyprus	.cy	1,220,000	298,000	13,963	780,133	210,000
Malta	.mt	1,150,000	301,000	4,971	398,534	92.8%
Faroe Islands	.fo	853,000	31,000	204	46,962	NA
Man, Isle of	.im	350,000	NA	6,337	75,049	NA
Andorra	.ad	309,000	24,500	3,725	70,549	100%
Gibraltar	.gi	270,000	6,200	24,572	27,884	80%
Albania	.al	248,000	75,000	1,993	3,563,112	86.5%
San Marino	.sm	224,000	14,300	1,107	28,880	96%
Guernsey	.gg	201,000	36,000	397	65,228	NA
Jersey	.je	138,000	27,000	506	90,812	NA
Monaco	.mc	132,000	16,000	28,702	32,409	99%
Vatican City State	.va	145,000	93	3,831	921	100%

Figure 3: European countries with the highest web presence

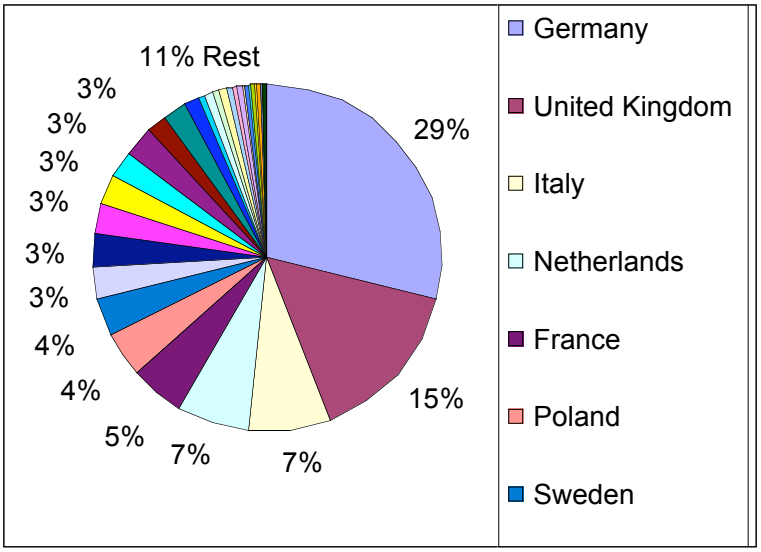
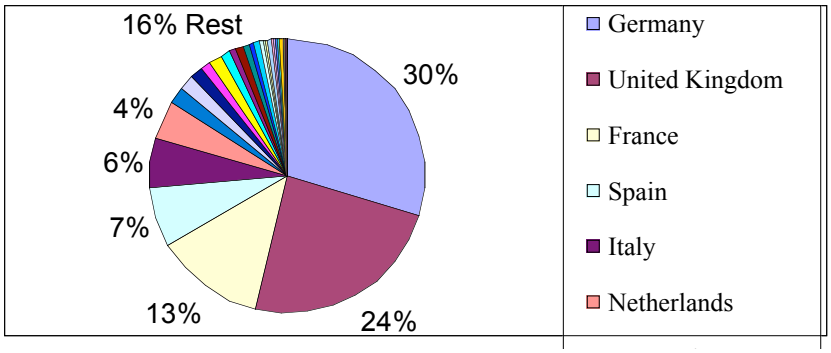


Figure 4: European countries with the highest number of domains



Data Analysis

The data obtained from various search statements by following the above-mentioned methodology are in Tables 1 and 2. The Web presence of each country has been shown at ccTLD level by considering all web pages of each country. The ranking is based on the total number of web pages, as this is the reflection of the degree of presence of a country on the Web.

Pearson correlation coefficients and two-tail Ps were calculated to determine whether there was any significant correlation between the number of Internet users and the number of web pages.

Pearson correlation coefficients and their levels of significance indicated that there is a significant correlation at the 0.01 level between the number of Internet users and the number of web pages of Middle-Eastern countries. Certainty the Web presence of each country is positively related to the number of Internet users. In other words, countries with a higher number of Internet users tended to have a higher visibility on the Web.

Pearson correlation coefficients and their levels of significance indicated that there is a significant correlation at the 0.01 level between the number of Internet users and the number of web pages of European countries. Thus, the Web presence of each European country is positively related to the number of Internet users.

Results

Tables 1 and 2 show that Middle-Eastern and European countries with a higher number of Internet users have a higher web presence. The more access users have to the Internet, the more web presence a given country will have.

Table 1 and Figure 1 show that Turkey, Israel and Iran respectively have the highest web page size among Middle-Eastern countries. Table 2 and Figure 3 show that Germany, the United Kingdom and Italy respectively have the greatest web page size among European countries. Table 2 also shows that between the first European country's web presence, Germany, and the last, Vatican City State, there is a divide of over 99.99 per cent. Moreover, the divide between the top five (e.g. Germany and France) is significant – 17.19 per cent.

The greater the web page size, the greater the Web presence for the country. Therefore, countries that publish more web resources have a greater web presence. Additionally, results suggest that web sites from Middle-Eastern countries are somewhat limited and thus there is a digital divide between Europe and the Middle-East. There is even a digital divide between European countries: for example, the number of web pages of Germany is 3 times higher than Italy, 48 times higher than Greece, and 5 times higher than France. As another example, the total number of domains in Germany is 5 times higher than Italy, 141 times higher than Greece, and twice as high as France.

Discussion

Search engines as primary data gathering instruments may create problems in the conclusions of Web presence studies. We are using a tool for Web presence analysis that is not specifically meant for the task. Search engines are designed for contents retrieval and not analysis of web presence. These problems are technical and could be resolved if the search engine programmers had incentives to work on them. As it is relatively easy to gather data by using commercial search engines, webometrics has all the potential to evolve as a tool for performance evaluation of any country.

One of the main limitations of the current study is that although several thousands of European and Middle-Eastern web sites have generic top-level domains such as ".org", ".com", ".net", given the current features of the search engines which serve as the basic data mechanism, it is not possible to determine how many web sites have generic TLDs. So the current research has considered only top-level domains from these countries. Despite the recognition that web presence evaluation, based on search engines, is an imperfect measuring tool, there is no obvious alternative. Thus, those forced to use search engines for direct web presence comparison should be encouraged to remain open-minded and cautious aware of the inherent limitations of their use.

Moreover, a web presence study is a "*snapshot*" of a search engine database at a specific time. Comparisons should be performed with caution, and preferably be carried out within the same snapshot, because the Web presence varies over different snapshots taken at short intervals. A retrospective web presence is not reproducible, because it depends on dynamics (expansion or contraction) of web sites [5] and the *Web is a growing organism* [13].
The challenge of the digital divide on the Web goes to the heart of the Internet's mission to provide equitable access to information for all users, regardless of language, ethnic group, religion, social class, sex, age, or any other factor. Every cybercitizen has a right to information [13].

Conclusion

The results of the present study confirm the hypothesis that countries with higher number of Internet users publish more web pages than countries with a limited number of Internet users. The results show that the Middle-Eastern countries, apart from Turkey, Israel and Iran, have a limited web presence. A comparison of Middle-Eastern countries' domains raises interesting questions about the place of different countries, cultures and languages on the Web. These countries are outside the main Web area, dominated by the USA, the UK, Canada, Europe, Australia, Japan, China, etc. It appears that Middle-Eastern countries may have a more limited presence on the Web because they are outside the current mainstream of the Internet, which is dominated by developed countries. The current digital divide is a warning to cybercitizens.
Overall, it can be concluded that due to the relatively limited web presence of various Middle-Eastern countries, there is a digital divide, and the level of web presence of European countries varies. The ready access to the Internet in European countries is one of the main reasons for high web presence. Therefore, easier access to ICT in Middle-Eastern countries may increase their web presence.
Further research is needed to gain a better understanding of the nature of search engines, and further research may be necessary to find reasons for the limited number of Middle-Eastern web pages, considering other factors for each country, such as language, ICT facilities, search engine problems with sites in languages other than English, etc. It could also be interesting to investigate the Web presence of African countries.

Acknowledgments

The author wishes to thank Mrs. Marjorie Sweetko and Mr. Kayvan Salehi for their helpful comments.

References

1. R. Cullen, Addressing the digital divide. *Online Information Review*, Vol. 25 No. 5, (2001): 311-320.
2. D.M. Pennock, G.W. Flake, S. Lawrence, E.J. Glover, and C.L. Giles, Winners don't take all: Characterizing the competition for links on the Web. *PNAS* 99, No. 8 (2002, April 16): 5207-5211.

3. T.C. Almind and P. Ingwersen, Informetric analyses on the World Wide Web: methodological approaches to Webometrics. *Journal of Documentation*, Vol. 53 No. 4, (1997): 404-426.

4. A. Noruzi, Web Presence and Impact Factors for Middle-Eastern Countries. *Online Magazine*, Vol. 30, No.2, (2006): 22-28.

5. A. Noruzi, The Web Impact Factor: A critical review. *The Electronic Library*, Vol. 24, No. 4, (2006), forthcoming.

6. J. Wolfensohn, Interactive debate. In: *United Nations Conference on Trade and Development*, Tenth session, Bangkok, 12-19 February, (2000). <http://www.unctad.org/en/docs/ux_tdxm11.en.pdf>.

7. P. Norris, *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide*. Cambridge University Press, (2003): 40.

8. D.L. Hoffman and T.P. Novak, Bridging the racial divide on the Internet. *Science*, Vol. 280 (April 17, 1998): 390-391.

9. S. Hubregtse, The digital divide within the European Union. *New Library World*, Vol. 106 No. 3/4, (2005): 164-172.

10. Internet World Stats. Internet World Stats: Usage and Population Statistics. (2005, December 31). <http://www.internetworldstats.com>.

11. WebHosting.info. Country-wise Total Domains Data as of 02/27/2006. <http://webhosting.info/domains/country_stats/>. <http://www.webhosting.info/registries/country_stats/>.

12. Central Intelligence Agency, *The World Factbook*. (2005), <http://www.cia.gov/cia/publications/factbook/geos/we.html>.

13. A. Noruzi, Application of Ranganathan's Laws to the Web. *Webology*, Vol. 1 N. 2, (2004), Article 8, <http://www.webology.ir/2004/v1n2/a8.html>.