

Web Presence of Selected Asian Countries: A Webometric Study

By

Mr. Samir Kumar Jalal*, Dr. Subal Chandra Biswas¹ and Dr. Partha Sarathi Mukhopadhyay²

*Assistant Librarian, Central Library, Birla Institute of Technology, Mesra, Ranchi, Jharkhand. India

¹Professor, Department of Library & Information Science, University of Burdwan, W.B. India

²Sr. Lecturer, Department of Library & Information Science, University of Burdwan, W.B. India

Abstract

The paper focuses on the Web presence and visibility of Asian countries websites. The paper tries to highlight the Web presence using some webometric indicators like Internet access, webpages, number of Internet users, and link studies. The study analyzes the web presence using popular search engines like Altavista, Google, Yahoo and MSN. An attempt has also been taken to find out the Web Impact Factor (WIF) for selected Asian countries. The result shows that China (43.7%), Japan (16.7%) and India (10.4%) occupy highest web presence amongst Asian countries based on the total number of effective Internet users. China being the second highest number of Internet users having 11.8% after USA (19.7%) followed by India with 4.9% of world Internet Users and Japan is having the highest number of webpages followed by China and South Korea.

Keyword: Web Presence, Visibility, WIF and Asian Countries

1. Introduction

The Web is a reflection of human culture, a massive socio-cultural network of Web resources authored by millions of people and organization around the world. Web presence being an important part of webometric research focuses on the studies of Web presence of Asian countries with respect to various Webometric indicators like number of Internet users, number of external links, number of webpages, number of domain names etc. The present study concentrates on the Web presence and calculation of the Web Impact Factors (WIF) for top-level domain (TLD) and sub-level domain (SLD) for academic and educational institutions in these countries. It also studies the external links for the calculation of revised WIF using AltaVista search engine. The result shows that China (43.7%), Japan (16.7%) and India (10.4%) occupy highest web presence amongst Asian countries based on the total number of effective Internet users. Again, top five Asian countries are Japan, China, South Korea, Taiwan and Vietnam based on the total number of webpages.

Since mid 1990s, there have been lots of efforts to study the structure and characteristics of the Web by itself, web contents; links and web search engines using new informetric methodologies. Several studies show that web sites can be compared and ranked in different domains based on their impact factors. In 1998, Ingwersen [1] calculated the web impact factor for some Danish domains and websites. He used AltaVista for his study because he believed this search engine covers broad area of the web and provides sufficient information for webometric studies.

The easiest way of measuring Web presence of countries is with the help of large-scale search engines such as Yahoo, Google, AltaVista with its power of advanced search facilities in webometric research. The web presence of Asian countries is a helpful tool to judge the 'digital divide' between the rich and the poor countries in terms of the utilization of Information and Communication Technologies (ICTs) of the countries.

Literature Review

The World Wide Web becomes more popular especially after 1990s and millions of people and organizations around the world depend on the Web for their daily life and for information exchange. The WWW has gained popularity largely because of its ease of use and multimedia capabilities, as well as its convenient access to other types of Internet services. The Web is the fastest medium for transferring information and has universal reach crossing geographical and time boundaries. It is also easy to access information from millions of Web sites using search engines. The number of domains from 16300 in July 1992 increased to 30,000,000 in July 2001 (Gromov [2]).

Thelwall, et al [3] attempted to find out similar academic websites using links, bibliometric couplings and co-links. They did an experiment with a random sample of 500 pairs of domains from the UK academic domains to find out the similar academic websites. The result showed that using a combination of all three (links, colinks and couplings) gave surprisingly marginal improvement over links alone of identifying similar academic websites.

Aguillo, et al [4] have shown the analysis of Web presence of the universities by means of cybermetric indicators. The developing countries in Latin America are making a great effort for publishing electronically their academic and scientific result. The authors have studied the Brazilian universities regarding web presence and web visibility, domain size. Result shows that there is a tremendous increase in the commitment of the Brazilian universities to the web. The paper described the co-link maps of 167 Brazilian academic institutions.

Mukhopadhyay [5] had tried to explore the possibility of research in the field of Webometrics in the educational institutions in India using Web Impact Factor (WIF). He ranked the Indian Institute of Technology (IITs) and Indian Institute of Management (IIMs) systems based on the calculation of WIF.

Mukhopadhyay [6] in his study showed the results of Webometric investigation at different levels of domain name system. His study is based on the hyperlink analysis. He calculated Web Impact Factors for cc TLDs of SAARC countries; Sub Level Domains (SLD) related to academic and research institutes registered under Indian ccTLD and hosts under IIT and IIM educational system in India.

Since Web had been flourished first in developed countries and its influence came gradually in developing countries. Therefore, there is a huge gap in Web presence among American, European and Asian Countries. Web presence and Web Impact Factor for middle-east countries (Noruzi [7, 8]) reflects that Turkey, Israel and Iran occupy the highest Web presence amongst middle-east countries. Will, et al [9] surveys 3738 US-based websites of colleges and universities to determine how higher education employ the Internet to communicate to key publics in general and students in particular. Mukhopadhyay had shown the ranking of SAARC countries using external- WIF.

2. Objectives of the Study

1. To know the volume of Webpages of Asian countries indexed by AltaVista search engine;
2. To compare the web presence of Asian countries and rank them based on appropriate webometric indicators;
3. To calculate revised Web Impact Factor (WIF) for ccTLD of Selected Asian Countries in order to rank them.
4. To find out the variation of webpage growth among selected Asian countries on the Web Space

3. Domain Structure

The Domain Name System (DNS) translates the language address (i.e. www.buruniv.ac.in) into a corresponding IP address (144.16.192.17). From right to left, the domain name has the following structures

- Top level domain

- Sub-level domain
- Host-level domain

In the above example, the hierarchy is: .in (Top level domain), .ac (sub-level domain working under TLD i.e. .in) and .buruniv is the host-level domain represented for University of Burdwan. The top level domain is also categorized into three i.e. a) Generic Top Level Domain (gTLD), b) International Top Level Domain (iTLD) and, c) Country Code Top Level Domain (ccTLD).

4. Methodology

5.1 Selection of Countries

Initially, all the 45 countries [10] in Asian continent had been taken into consideration and then each countries' webpages has been calculated using AltaVista search engines with the command [site:cn], where cn stands for TLD of China. All countries have been ranked based on total number of webpages and then only top 20 countries have been selected for our present studies.

5.2 Choice of Indicators

Web activity is multi-dimensional and is reflected through its web presence. So, the best way to build the ranking is combining a group of indicators that measure these different aspects. Almind & Ingwersen [11] proposed the first Web indicator, Web Impact Factor (WIF), based on link analysis that combines the number of external inlinks and the number of pages of the website, a ratio of 1:1 between visibility and size. This ratio was used for the ranking but adding two new indicators to the size component: Number of documents, measured from the number of rich files in a web domain, and number of publications being collected by Google Scholar database. Therefore, four indicators were obtained from the quantitative results provided by the main search engines as follows:

- **Size (S).** Number of pages recovered from four engines: Google, Yahoo, Live Search and Exalead. For each engine, results are log-normalised to 1 for the highest value. Then for each domain, maximum and minimum results are excluded and every institution is assigned a rank according to the combined sum.
- **Visibility (V).** The total number of unique external links received (inlinks) by a site can be only confidently obtained from Yahoo Search, Live Search and Exalead. For each engine, results are log-normalised to 1 for the highest value and then combined to generate the rank.
- **Rich Files (R).** After evaluation of their relevance to academic and publication activities and considering the volume of the different file formats, the following were selected: Adobe Acrobat (.pdf), Adobe PostScript (.ps), Microsoft Word (.doc) and Microsoft Power point (.ppt). These data were extracted using Google and merging the results for each filetype after log-normalising in the same way as described before.
- **Scholar (Sc).** Google Scholar provides the number of papers and citations for each academic domain. These results from the Scholar database represent papers, reports and other academic items.

The four ranks were combined according to a formula where each one has a different weight and the combined rank is: Webometrics Rank (position)=4*RankV+2*RankS+1*RankR+1*RankSc

5.3 Data Collection and Data Analysis

AltaVista, Yahoo and Google had been chosen to collect the data for this study. Data collection had performed during Feb 10-12, 2008 & Feb 18-19, 2009. All the domain names of Asian countries were

searched to check their validity using Yahoo! Google and AltaVista database. For each of these countries, a search was carried out to determine the total number of links, total webpages, selflinks and inlinks using the following commands:

- The total number of webpages to ccTLD, China (for example), domain:cn
- The number of total links at the ccTLD, China (for example), linkdomain:cn
- The number of inlinks can be calculated using the command, linkdomain:cn –domain:cn
- The number of self-links can be measured using the formula, linkdomain:cn domain:cn

Table-1: Top 20 Asian countries based on webpages

Sl No	Countries	TLD	No of Webpages (Feb 2008)	No of Webpage (Feb 2009)	% Growth	Inlinks	Selflinks
1	Japan	.jp	297000	linkdomain:in AND domain:jp	Japan	327000	Linkdomain:jp AND domain:jp
2	China	.cn	30300	linkdomain:in AND domain:cn	China	104000	Linkdomain:cn AND domain:in
3	South Korea	.kr	72800	linkdomain:in AND domain:kr	South Korea	28200	Linkdomain:kr AND domain:in
4	Taiwan	.tw	43700	linkdomain:in AND domain:tw	Taiwan	31700	Linkdomain:tw AND domain:in
5	Israel	.il	7540	linkdomain:in AND domain:il	Israel	9170	Linkdomain:il AND domain:in
6	Vietnam	.vn	34400	linkdomain:in AND domain:vn	Vietnam	61900	Linkdomain:vn AND domain:in
7	Turkey	.tr	5330	linkdomain:in AND domain:tr	Turkey	29900	Linkdomain:tr AND domain:in
8	Malyasia	.my	40100	linkdomain:in AND domain:my	Malyasia	49700	Linkdomain:my AND domain:in
9	Singapore	.sg	78800	linkdomain:in AND domain:sg	Singapore	316000	Linkdomain:sg AND domain:in
10	India	.in	7100000	linkdomain:in AND domain:in	India	7100000	Linkdomain:in AND domain:in
11	Phillipines	.ph	32900	linkdomain:in AND domain:ph	Phillipines	45500	Linkdomain:ph AND domain:in
12	Indonesia	.id	14800	linkdomain:in AND domain:id	Indonesia	6720	Linkdomain:id AND domain:in
13	Iran	.ir	1680	linkdomain:in AND domain:ir	Iran	1460	Linkdomain:ir AND domain:in
14	Kazakhstan	.kz	594	linkdomain:in AND domain:kz	Kazakhstan	2940	Linkdomain:kz AND domain:in
15	Pakistan	.pk	3200	linkdomain:in AND domain:pk	Pakistan	4490	Linkdomain:pk AND domain:in
16	Soudi Arabia	.sa	497	linkdomain:in AND domain:sa	Soudi Arabia	252	Linkdomain:sa AND domain:in
17	Armenia	.am	1200	linkdomain:in AND domain:am	Armenia	1770	Linkdomain:am AND domain:in
18	UAE	.ae	2330	linkdomain:in AND domain:ae	UAE	269000	Linkdomain:ae AND domain:in
19	Uzbekistan	.uz	112	linkdomain:in AND domain:uz	Uzbekistan	351	Linkdomain:uz AND domain:in
20	Azerbaijan	.az	723	linkdomain:in AND domain:az	Azerbaijan	119	Linkdomain:az AND domain:in

Source: Yahoo! dated 28th March 2009

It has been found from above table that Japan is giving maximum links to India and simultaneously

getting highest links. Another remarkable finding is the success of South Korea and Taiwan web presence with respect to producing links to India and also received links from India. It has also been studied taking into whole 45 Asian countries (detailed result is not shown here) that more than 17 countries have incoming and outgoing links is either zero (0) or less than 100.

Following table explains generic top-level domain that links to and links from .in domain during February 2009.

Table-3: Generic TLD that links to and from Indian domains (Feb 23, 2009)

Generic TLD	# links to .in domain	Search Command	TLD	# links from .in domain	Search Command
COM	16400000	linkdomain:in AND domain:com	COM	4630000	linkdomain:com AND domain:in
NET	511000	linkdomain:in AND domain:net	NET	1210000	linkdomain:net AND domain:in
ORG	884000	linkdomain:in AND domain:org	ORG	1680000	linkdomain:org AND domain:in
INFO	121000	linkdomain:in AND domain:info	INFO	267000	linkdomain:info AND domain:in
EDU	62300	linkdomain:in AND domain:edu	EDU	73600	linkdomain:edu AND domain:in
GOV	1580	linkdomain:in AND domain:gov	GOV	49700	linkdomain:gov AND domain:in

Source: Yahoo! dated 28th March 2009

The above table shows that out of 6 generic TLDs, .com domain is producing maximum links from Indian and also getting highest links; whereas .GOV is having least links to and from India.

Following table shows the population and Internet users for Asian countries. It has been mentioned in table-7 that total number of Internet users (578538257) of Asia i.e.39.5% of World Internet users. Based on this, individual countries percentage of Internet users has been calculated and has been shown only selected countries of Asia.

Table-4: Asia's Internet Usage and Population in 2008

Sl No	ASIA	Population	Internet Users	Internet Users (%)	Users	User Growth
Asia	3776181949	56.60%	114,304,000	578538257	39.50%	406.14%
Rest of the World	2899938339	43.40%	246,681,745	885094104	60.50%	258.80%
World Total	6676120288	100.00%	360985745	1463632361	100.00%	305.50%

Source: www.internetworldstats.com

The above table shows there is a significant percentage (39.5) of Internet usage in the world. With respect to population percentage (56.6), Internet users are little less compared to rest of the world. The reason may be most of the countries in Asia are economically in developing nature.

5. Findings of the Study

There are 45 Asian Countries having their web presence in Web space but top 20 selected countries based on their webpages have been taken for this study. Japan is having highest number of webpages among Asian countries, followed by China and South Korea; India got sixth position based on number of webpages. South Korea is able to achieve remarkable increase (i.e. 367 times) in terms of web pages during 2008-2009. Turkmenistan also witnessed highest inlink counts 554.3 per thousand webpages but the country has not been shown in the table due its low webpages and also it has witnessed negative growth (54.15%) from Feb 2008 to Feb 2009. The credit of achieving highest overall WIF and revised WIF goes to Armenia, whose webpage is only 1730. Therefore, volume of webpage is an important indicator for

influencing WIF of any country or institutions. The growth of highest Internet users' growth is Uzbekistan. Taiwan attained highest number of academic webpage (68, 80,000) beating China (45, 20,000) and even Japan (53, 80,000). India becomes highest growth of Internet users during last one year i.e. April 2007 to April 2008.

6. Conclusion

It is a fact that nearly, one billion people around the world are on Online. Seven Asia-Pacific countries are in the top 20 countries for numbers of Internet users among them three are China, Japan, and India in the top five. Internet Penetration in Asia is 17.4% whereas world average is 23.8% [13] on March 2009. Asia is having 41.2% [13] Internet users in the world. The latest internetworldstats.com statistics show that worldwide Internet penetration has increased to 16.0%, due to increased contribution from Asian countries. India has more than 39 million [13] Internet users i.e. 13% of the World's Web population. Japan is the world's third largest Internet market with an estimated 68 million subscribers. China being the second highest number [14] of Internet users having 11.8% after USA (19.7%) followed by India with 4.9% of world Internet Users. Therefore, there is a huge importance of web presence study in the Asian countries. Besides hyperlink analysis are also required to know the interconnection of one country to another with the power of search engines. The present study of webometric analysis of selected Asian countries is an attempt to bring out the importance of webometric research and tried to throw some lights on few aspects specifically hyperlink studies so as to reflect the present status of Asian countries and their relative position among themselves.

References

1. P. Ingwersen. The calculation of Web Impact Factors. *Journal of Documentation*, 54(2), 236-43; 1998.
2. G.R Gromov. History of Internet and WWW: the roads and crossroads of Internet history, 2002. Retrieved Dec 5, 2004 from <http://www.netvally.com/intvalstat.html>
3. M. Thelwall and D. Wilkinson. Finding similar academic websites with links, bibliometric couplings and co-links. *Information Processing and Management*, 43(3), 515-526; 2004
4. I.F. Aguillo, J.L. Ortega, and B. Granadino. Brazil academic webuniverse revisited: a cybermetric analysis, 2006. Retrieved dated 4th Feb 2009 from: <http://digital.csic.es/bitstream/10261/4200/1/R-12.pdf>
5. P.S. Mukhopadhyay. Measuring Web Impact Factors: a webometric study based on the analysis of hyperlinks. In. *National seminar on information support for rural development, India. IASLIC*, Dec 2004.
6. P.S. Mukhopadhyay. The calculation of Web Impact Factors for educational institutes of India: A Webometric analysis. In. *Information Management in e-Libraries*, 26-27 February 2002.
7. A. Noruzi. The Web Impact Factor: A Critical Review. *The Electronic Library*, 24(4), 490-500; 2006
8. A. Noruzi. Web presence and impact factors for middle-eastern countries. *Online*, 30(2), 22-28; 2006
9. E.M. Will and C. Callison. Web presence of universities: Is higher education sending the right message online. *Public Relations Review*. 32(2), 180-183; 2006
10. Countries of Asia. http://www.timberhunt.com/timber_trade/asia.html
11. T.C. Almind and P. Ingwersen. Informetric analyses on the World Wide Web: methodological approaches to Webometrics, *Journal of Documentation*, 53(4), 404-26; 1997
12. S. Wasserman and K. Faust. Social network analysis: methods and applications. Cambridge: Cambridge University Press, 1994
13. Internet World Stats (2009). Internet World Stats: Usage and population statistics, (April 7, 2009). <http://www.internetworldstats.com>
14. World Internet Penetration. Accessed dated 19th July 2009 from (http://www.readwriteweb.com/archives/world_internet_penetration_sept06.php)

Countries |TLD |Webpage (A) |Inlinks (B) |pdf |ps |doc |ppt |Rich Files [R] |Google Scholar (Sc) |logA
|logB |LogR |logSc |Index Value |WISER Rank |Japan |jp |535000000 |276000000 |25100000 |41600
|845000 |141000 |26127600 |13900000 |9.73 |8.44 |7.42 |7.14 |67.78 |1 | |China |.cn |2360000000
|253000000 |21300000 |15100 |1300000 |181000 |22796100 |13500000 |9.37 |8.40 |7.36 |7.13 |66.85 |2 |
|South Korea |.kr |2200000000 |123000000 |2620000 |1100 |54800 |12500 |2688400 |1070000 |9.34 |8.09
|6.43 |6.03 |63.5 |3 | |Taiwan |.tw |1240000000 |58900000 |12100000 |6870 |255000 |42500 |12404370
|1910000 |9.09 |7.77 |7.09 |6.28 |62.64 |4 | |India |.in |124000000 |32700000 |16400000 |12500 |895000
|116000 |17423500 |6290000 |8.09 |7.51 |7.24 |6.8 |60.28 |5 | |Turkey |.tr |112000000 |18900000 |13900000
|6470 |362000 |37100 |14305570 |1750000 |8.05 |7.28 |7.16 |6.24 |58.6 |6 | |Vietnam |.vn |280000000
|11400000 |12000000 |1340 |288000 |30800 |12320140 |1170000 |8.45 |7.06 |7.09 |6.07 |58.28 |7 |
|Singapore |.sg |83600000 |11800000 |17300000 |16700 |252000 |41600 |17610300 |1740000 |7.92 |7.07
|7.25 |6.24 |57.62 |8 | |Israel |.il |116000000 |8600000 |14600000 |28000 |562000 |49000 |15239000
|3380000 |8.06 |6.93 |7.18 |6.53 |57.58 |9 | |Malaysia |.my |95000000 |8390000 |12000000 |930 |223000
|29500 |12253430 |1020000 |7.98 |6.92 |7.09 |6.01 |56.75 |10 | |Indonesia |.id |55800000 |5300000
|20000000 |1770 |267000 |36800 |20305570 |1290000 |7.75 |6.72 |7.31 |6.11 |55.81 |11 | |Philippines |.ph
|40200000 |5830000 |11300000 |735 |178000 |25400 |11504135 |1080000 |7.60 |6.77 |7.06 |6.03 |55.37 |12 |
|Iran |.ir |30000000 |2830000 |8290000 |1600 |210000 |82300 |8583900 |1190000 |7.48 |6.45 |6.93 |6.08
|53.77 |13 | |Saudi Arabia |.sa |11500000 |2670000 |3550000 |597 |73400 |11600 |3635597 |391000 |7.06
|6.43 |6.56 |5.59 |51.98 |14 | |Pakistan |.pk |10400000 |1390000 |9570000 |743 |212000 |22200 |9804943
|1070000 |7.02 |6.14 |6.99 |6.03 |51.63 |15 | |Kazakhstan |.kz |10700000 |2240000 |2320000 |636 |49000
|7360 |2376996 |192000 |7.03 |6.35 |6.38 |5.28 |51.12 |16 | |UAE |.ae |9020000 |2850000 |1140000 |307
|41200 |6940 |1188447 |198000 |6.96 |6.45 |6.07 |5.3 |51.1 |17 | |Armenia |.am |7720000 |1720000 |2560000
|948 |61000 |6410 |2628358 |194000 |6.89 |6.24 |6.42 |5.29 |50.42 |18 | |Azerbaijan |.az |4440000 |459000
|1930000 |216 |39400 |5110 |1974726 |122000 |6.65 |5.66 |6.30 |5.09 |47.32 |19 | |Uzbekistan |.uz |3960000
|339000 |2020000 |351 |35000 |4610 |2059961 |133000 |6.60 |5.53 |6.31 |5.12 |46.75 |20 | | Source: AltaVista,
Feb 18-19, 2009 & Google March 20-21, 2009

* Address for Communication

Mr. Samir Kumar Jalal
Central Library, Birla Institute of Technology, Mesra
Ranchi, Jharkhand, Email: jalal_gtz@yahoo.co.in