Link Analysis and Web Influence Evaluation for Enterprise Websites

Junping Qiu1, Quan'e Ren2

1 College of Information Management ,Reseach Center for Chinese
Science Evaluation , in Wuhan Uninersity
jpqiu @whu.edu.cn

2   College of Information Management ,Reseach Center for Chinese
Science Evaluation , in Wuhan Uninersity
renquane @yahoo.com.cn

Abstract. As the enterprise Informatization process goes through deep, an increasing number of enterprises deploy their web sites on the internet, making the web sites become a brand of the enterprise. Thus, it is very necessary to evaluate the quality for enterprise web sites. However, almost the total current website evaluation methods have ignored the external web influence factors of quantitative evaluation and can't work effectively. In this paper, we employ the weblink analysis method to evaluate the web influence for enterprise web sites, combing 17 link analysis indicators obtained by Google and AltaVista. We then explored the relationship between the web sites type and web influence and discussed the performance of the link indicators and link analysis tools. The whole analysis is based on our experiments, using the data accepted from 2006 Top 500 companies list of China and 2006 Fortune 500 companies list in the world.

1   Introduction

Many indicators and methods have been proposed to evaluate the enterprise web sites in China and foreign countries, however, their focus are different. Few of them referred to the evaluation system of web sites link analysis and web influence [6,13]. At present, there are four typical evaluation system concerning on the electronic business web sites, such as: 5 indicators designed by Gomez, 10 indicators designed by BizRate, indicators used by the organization of CU' consumer reports online, CNNIN evaluation indicators[4], and indicators used by China Enterprise Confederation, information exchange center which belongs to China Entrepreneurs Association has launched an appraisal campaign on the web in 2005, only using standard as "Navigation clear, unmistakable link, homepage with links to columns point inflation rate of not less than 90%. " to evaluate Internal Links[2].

From this, we divide the electronic business web sites evaluation methods into three parts: qualitative evaluation, quantitative evaluation and synthetically evaluation. The main methods of quantitative evaluation are the web traffics' indicators, often using, Alexa traffic. However, the rank is not absolutely authorities [1].

There is another quantitative evaluation method: using link analysis methods and web influence factor (WIF). Link analysis uses the webpage number of link which links to certain sites and represent its quality, this is based on these theories: if site A is linked by site B, then that means B is appreciating and using A, and the content of A and B is relevant; the more the site is linked, the bigger is the web influence. WIF measurement is based on link analysis, using WIF to represent the web influence. Therefore, we attempt to apply link analysis method and WIF measure methods to evaluate some of the typical enterprise in China and foreign countries. What's more, we explore the relationship between the type of sites and web influence.

2. Research approach

2.1 Research swatch

Because the difference exists in industry and which will influent the effect of link analysis, we define the swatch in information technology enterprises and consumable electron enterprises. The American Forrest Research company has designed a Site Need Index (SNI) to help the enterprise to decide whether they need to build a site for popular products. SNI' research results show that information technology enterprises and consumable electron enterprises hurriedly need to build their own brands and products through web sites [3, 7]. In our experiment, we selected thirty-four enterprise sites form 2006 Top 500 companies list of China and 2006 Fortune 500 companies list in the world to be our research swath. And our focus is enterprises sites in China, several foreign enterprises are used to compare with domestic web *sites*.

2.2 Indicators

In order to represent the scale and web influence of the enterprise sites comprehensively, we choose seven link analysis indicators as follows :(the website evaluated is labeled with A in this paper)

Pages Number: It represents the number of pages which are retrievable in search engines for A. It represents the scale of a website to some extent, but not necessarily represents the quality and concentration of a website.

Total Links: It is the number of pages indexed by search engines which are linking to A. This stands for the web influence of A.

Internal Links: This is the number of pages in A indexed by search engines which are linking to A itself, that is to say, it is a kind of self-link. Self-link represents the hierarchy and fullness of A's inner structure.

External links: It is the number of pages except A indexed by search engines which are linking to website A. This is the primary indicator, which is even better than total links for representing web influence.

Web Influence Factor (WIF=Total Links/Total pages) reflects the average level of pages for A that are linked. It also is the calculation way for WIF proposed by Ingwersen [5].

External Web Influence Factor (WIFe=External Links/Pages Number) reflects the total average level of Internal links of A. It is an objective indicator.

Internal Web Influence Factor (WIFs=Internal links/Pages Number) reflects the total average level of Internal links of A

3. Data Sources

3.1 Retrieval Methods and Tools

In the webometrics , there are two kind of link analysis tools. One is commercial search engines, the other is professional crawler. Because an proper tools for measuring WIF should Satisfy five conditions [10], there yet aren't any commercial search engines fits for them all.

A1taVista is often used in the research on webometrics. In fact, both Google and A1taVista are good tools to finish this task, what's more, the coverage and capacity of data in Google is better than A1taVista[12]. Thus, both Google and A1taVista are used in our experiment to eliminate and balance the uncertainty of link analysis, and improve the reliability and efficiency of the research results. In this paper, different indicators and retrieval expressions are used in different tools, as showed in table 1 and table 2

Table 1：Indicators and Query Expressions used in Google search engine, taking Haier Group for example

Indicators	Query Expressions
Pages Number: T	site: www.haier.com
Total Links: L	"www.www.haier.com"
Internal Links: S	"www.www.haier.com" site: www.haier.com
External Links: E	"www.haier.com" -site: www.haier.com
Web Influence Factor: WIF	L/T
External WIF: WIFe	E/T
Internal WIF: WIFs	S/T

Table 2：Indicators and Query Expressions used in A1taVista search engine, taking Haier Group for example

Indicators	Query Expressions
Pages Number: T	Host: www.haier.com
Total Links: L	Link: www.haier.com
Internal Links: S	S1:link:www.haier.com AND host: www.haier.com S2:host:www.haier.com AND link: www.haier.com
External Links: E	E1:link: www.haier.com AND NOT www.haier.com E2:link: www.haier.com AND NOT www.haier.com link:

	www.haier.com)
	E3:link: www.haier.com AND NOT www.haier.com host: www.haier.com)
Web Influence Factor: WIF	L/T
External WIF: WIFe	(E1+E2+E3)/3/T
Internal WIF: WIFs	(S1+S2)/2/T

The accuracy 34 Sino-foreign enterprises' name and their website URL in our swatch are showed in table 3. The experiment was done from May 1 to May 3, 2007, using a quick accessing way to eliminate and reduce errors caused by dynamic net information.

Table 3：Name and URL of the companies

Enterprise Name	Enterprise Website URL
Microsoft Corporation	www.microsoft.com
IBM Company	www.ibm.com
DELL Company	www.dell.com
Intel Company	www.intel.com
Oracle company	www.oracle.com
IBM China Company Limited	www.ibm.com.cn
Founder Electronics Co., Ltd	www.founder.com
Haier Co., Ltd	www.haier.com
Nokia Mainland China	www.nokia.com.cn
ZTE Corporation	www.zte.com.cn
Motorola China	www.motorola.com.cn
Philips China	www.philips.com.cn
Intel Company China	www.intel.com.cn
Hisense Corporation	www.hisense.com
Huawei Technologies Co., Ltd	www.huawei.com.cn
UFIDA Software CO.,LTD	www.ufsoft.com.cn
Shenzhen Skyworth Electronics Co., Ltd	www.skyworth.com.cn
TCL CORPORATION	www.tcl.com
Lenovo	www.legend.com.cn
Tsinghua Tongfang CO.,LTD	www.thtf.com.cn
Chunlan Group Corp	www.chunlan.com
Inspur Group	www.inspur.com
Konka Group CO.,LTD	www.konka.com
Panda Electronics Group Co., Ltd	www.chinapanda.com.cn
Gree Electric Appliances, Inc	www.gree.com.cn
Tsinghua Unisplendour Corporation Limited	www.thunis.com
changhong electric co., ltd	www.changhong.com.cn
Aucma co., ltd	www.aucma.com.cn
Media in China	www.midea.com.cn
BOE Technology Group Co., Ltd	www.boe.com.cn
Great Wall Technology Co. Ltd	www.greatwalltech.com
Datang Telecom Technology Co.,Ltd	www.datang.com
Oracle company China	www.oracle.com/global/cn
Cosun Group	www.qiaoxing.net

3.2 Web link data

In our experiment, we used 17 link analysis indicators to analyze the data, 7 of which is through Google, and 10 of which is through Altavista, as showed in Table 4, 5 (limited space, only display the part business' link data) :

Table 4：Data from Google search results. This was done from 19:10 to 24:10, on May 1,2007. * represents the abnormal data, which will be discussed in part five

Oracle company	117000	1020000	116000	993000	8.718	8.487	0.991
IBM China Company Limited	21	36100	9	36100	1719	1719	4.5
Founder Electronics Co., Lt	1260	24800	1300	16400	19.683	13.016	1.032
Haier Co., Ltd	15100	66000	12300	43900	4.3709	2.907	0.815
Nokia Mainland China	3890	44100	3160	32700	11.337	8.406	0.812
ZTE Corporation	11500	45100	9420	18400	3.922	1.6	0.819
Motorola China	10900	45600	8490	22000	4.183	2.018	0.779
Philips China	384	12100	369	9170	31.51	23.88	0.961
Intel Company China *	1	10400	1	10900	10900	10900	1
Hisense Corporation	2270	14500	2120	9810	6.388	4.322	0.934
Huawei Technologies Co., Ltd	148	33100	9	33100	223.649	223.649	0.0608
UFIDA Software CO.,LTD	8850	30600	7110	9350	3.458	1.056	0.803
hen Skyworth Electronics Co.	127	4550	127	4530	35.827	35.669	1
TCL CORPORATION	3830	13100	3190	11100	3.42	2.898	0.833
Lenovo	95	6030	82	5380	63.474	56.632	0.863
Tsinghua Tongfang CO.,LTD	313	10500	306	959	33.546	3.064	0.978

Table 5：Data from Altavistasearch results. This was done from 19:10 to 24:10, on   May 1,2007. * represents the abnormal data, which will be discussed in part five

Enterprise Name	T	L	S1	S2	E1	E2	E3	WIF	WIFe	WIFs
Microsoft Corporation	27900000	92000000	243000	244000	1250000	1250000	3380000	3.287	0.07	0.009
Intel Company	8890000	31500000	26000	26000	253000	259000	884000	3.543	0.052	0.003
Founder China *	8650000	20200000	133	140	20200	20600	22000	2.335	0.0024	0
IBM	9210000	25200000	166000	167000	190000	190000	1290000	2.736	0.06	0.018
Konka Group	2910	82000	444	445	16000	16000	24800	28.179	6.506	0.153
uawei Technologies Co.,Lt	1400	63400	692	824	32500	32700	26300	45.286	21.786	0.541
Oracle China *	182000	314000	0	0	754	757	756	1.725	0.004	0
DELL Company	5900000	33800000	6600	6630	139000	140000	377000	5.729	0.037	0.001
Intel Company China	175000	752000	32800	35400	23800	23700	19200	4.297	0.172	0.195
ng Telecom Technology Co.	59400	398000	29	29	1820	1820	3630	6.7	0.041	4.882
Philips China *	57900	128000	2	2	18800	18700	27500	2.211	0.374	
IBM China	216000	612000	108000	117000	21000	21000	30500	2.833	0.112	0.521
Haier Group	114,000	834000	55	180	36800	36800	47800	7.316	31.947	0.004
Hisense Group	1,640	160000	119	120	22500	22400	33000	97.561	142.500	0.291
ZTE Corporation	770	32400	149	149	17700	17600	17400	42.078	22.814	0.194

4 Data analysis

Table 4 and Table 5 shows, use Google search and Altavista search for the enterprise links indicators specific data gap greatly. For example : the value of WIF, WIFe, WIFs of Microsoft website were 26.655, 4.286,1.484 with Google, while with Altavista were 3.287,0.07. 0.009. Clearly, database capacity and search mechanism of the two tools are different. If only using the original data to evaluate and rank, the conclusion wouldn't be objective. Therefore, we score each of the indicators, and then add in all indicators to gain final scores., and produce a total ranking, as showed in Table 6 ,7, 8 (Due to space limitations, only display the total ranking of the top 20 enterprise-data) :

Table 6 : rankings of enterprises site searched by Google

Enterprise Name	google Search						
	T Rank	L Rank	S Rank	B Rank	WIF Rank	WIFe Rank	WIFs Rank
Microsoft	1	1	1	1	8	19	2
Intel	3	3	3	3	14	10	16
IBM	2	4	2	4	17	14	3
DELL	5	2	5	2	9	7	17
Founder	21	13	20	12	10	9	5
IBM China	33	10	32	7	2	2	1
Haier	6	6	6	6	23	21	24
Philips China	25	20	24	18	7	6	12
Intel	34	24	34	14	1	1	6
Hua Wei	29	11	33	8	3	3	34
Oracle	4	5	4	5	16	12	8
Konka	22	21	27	15	12	11	33
ZTE	7	8	7	11	25	24	23
Hisense	17	15	16	16	20	18	15
Nokia	13	9	14	9	11	13	25
DaTang	15	30	15	25	34	29	27
Oracle China	19	25	21	33	21	33	31
TCL	14	19	13	13	27	22	22
UFIDA Software	9	12	9	17	26	26	26

Table 7 : rankings of enterprises site searched by Altavista

Enterprise Name	altavista Search									
	T Rank	L Rank	S1 Rank	S2 Rank	B1 Rank	B2 Rank	B3 Rank	WIF Rank	WIFe Rank	WIFs Rank
Microsoft	1	1	1	1	1	1	1	19	27	23
Intel	3	3	5	5	2	2	3	18	29	26
IBM	2	2	8	8	4	4	5	12	32	29
DELL	6	6	6	6	5	5	4	23	30	28
Founder	4	4	2	2	3	3	2	27	28	21
IBM China	8	8	23	19	9	9	9	9	11	24
Haier	12	12	10	10	8	8	7	22	23	22
Philips China	15	15	20	22	12	12	10	1	3	13
Intel	10	10	3	3	13	13	11	24	26	9
Hua Wei	19	19	12	11	10	10	13	2	13	8
Oracle	5	5	19	21	14	14	16	28	34	34
Konka	17	17	9	9	6	6	6	17	15	12
ZTE	20	20	18	20	16	16	18	4	12	16
Hisense	18	18	14	15	17	17	15	5	18	17
Nokia	11	11	24	24	26	26	27	10	31	3
DaTang	14	14	7	7	7	7	8	20	21	18
Oracle China	9	9	4	4	11	11	17	13	25	15
TCL	7	7	22	13	18	18	14	32	24	27
UFIDA Software	28	28	15	16	21	21	20	31	7	5

Table 8 : total rankings of enterprises site searched by Google and Altavista

Enterprise Name	Google Search		Altavista Search	
	Total Score	Rank 1	Total Score	Rank 2
Microsoft	212	1	274	1
Intel	193	4	254	2
IBM	199	2	244	4
DELL	198	3	231	8
Founder	155	7	254	3
IBM China	158	6	221	12
Haier	153	8	216	13
Philips China	133	12	227	11
Intel	131	13	228	9
Hua Wei	124	15	233	6
Oracle	191	5	160	19
Konka	104	23	236	5
ZTE	140	10	190	15
Hisense	128	14	196	14
Nokia	151	9	157	21
DaTang	70	32	227	10
Oracle China	62	33	232	7
TCL	115	18	168	17
UFIDA Software	120	16	158	20

5 Experiment results and discussion

5.1 discussion on link analysis tools

In Webometrics study, the Altavista is the preferred search engine for researchers [8].Can we take it with other tools in the integrated use to achieve optimal results? We can answer it with Table 8 : In Ranking 1, 2 and total ranking order, the first one are both Microsoft and are consistent with the total order; while the second one is slightly different, that is IBM Corporation with Google search results, but with Altavista search results and overall rankings are both Intel Corporation; Hisense Group, in an order, are all ranked 14th in the Sort 1,Sort 2 and the total ranking ;TCL Group, Sort 1 Sort 2 and the total ranking is in turn No. 18, No. 17 and No. 18, a difference of weak. The resulting gains rough conclusions: Google web links can be as analytical tools used in conjunction with the Altavista. Although both the search results with slight deviation, they can achieve balance and comprehensive abatement through synthesizing indicators. This from another angle test our hypothesis : Altavista can make integrated use with other web link tools to achieve optimal results.

5.2 indicators and enterprises web influence

In Table 4 and Table 5, the Enterprise website with * sign can be seen, and abnormal data mainly exist in the indicators S and WIFs of all these indicators, while for indicators E and WIFe, with either Google or Altavist retrieval, the results are relative normal. Therefore, we can deduce: indicators E and WIFe is the most effective indicators of many indicators for website link analysis and its influence force evaluation.

As for the underlying factors for fluctuation of value in Table 6 and Table 7, which caused the inconsistencies of two search engines link analysis results, come from such objective reasons as the dynamic network information variability characteristics and the different search mechanisms of the search engine, simultaneously; this also reflects the availability of some enterprises websites being in question. In many literature of evaluating websites, availability is a major indicators which include the appearance ratio of webpages in search engine, the structure level of website resources and the effectiveness of internal link [14]. However, from Table 4, 5, we can see some abnormal data that produced using Google and Altavista search, such as the value is 0 or 1 of T, S. For example, the indicators T, indicators S and indicators WIFs of Intel(China) website are 1, while for IBM (China) they were respectively 21,9 and 4.5, and Huawei' S value is 9,using Google. With Altavista search, for Oracle Corporation (China), Shenzhen Skyworth Electronics Co., Ltd. and Panda Electronics Group Limited, are all 0 of S1, S2; for Philips (China), S1, S2 are 2; as for Chunlan Group Limited and Changhong Electronic Co. Ltd. , S1, S2 are 1, Great Wall Technology Co.Ltd., T is 1. These shown: without regard to search engines factors, the appearance ratio of webpages in search engine, the structure level of website resources and the effectiveness of internal link of these enterprise are not optimistic concept.

Finally, based on the assumptions of relatively availability of our retrieval and analysis tools, we can further explore the difference of some website and how such factors influence the correlation between website and its web influence , in terms of Chinese enterprises and

other country enterprises, information display websites and online trading sites, and different industries type. So we can provide reference information for Chinese enterprises to optimize website construction and upgrade website influence force.

（1）Disparity exists between Sino-foreign enterprises web influence .Results from the Comprehensive rankings can be seen in this paper, selected 34 Chinese and foreign enterprises, several of the top foreign enterprises are websites, such as Microsoft, Intel, IBM, DELL, only the second one is the Founder Group, IBM Corporation (China), Haier Group, Philips (China), Intel (China), Huawei. This shows that, there is still a big gap between Chinese enterprises on the web site construction and promotion network in comparison with foreign countries. The existing causes include our late started Enterprise Informatization, infrastructure lags and other objective reasons. Furthermore, Chinese enterprises' decision-making should build awareness of the importance of websites on. From the enterprise marketing strategy, enterprises website is a integrated network marketing tools. So with an influential corporate website will enable enterprises to operate its entities into unexpected virtuous circle.

（2）Web influence depend on industries type .While the selection of samples have been taken into account the relationship between nature of the trade and its website construction ,and only selected information technology businesses and consumer electronic products enterprises, we study still found that these two field industries site also exist differences. Overall, the web influence of information technology and software products enterprises website is bigger than that of the consumer electronics business. From Table 6, 7, 8 in the rankings, we can see that the information technology or information products corporations like Microsoft, Intel, IBM, DELL, Founder, Huawei website ranks forward, but consumer electronics companies such as Haier, Konka, etc. were ranked then. So, the enterprise website building and the diffusion of different industries can take different strategies according to the characteristics of enterprises. A reasonable position of sites function is the premise to achieve perfect networks influence force for the enterprise.

（3）Web influence is relational with its function type. According to the scale and function, websites can be divided into three broad categories: information display only, online direct marketing websites, and e-commerce website[11].Information display-only website is an information carrier, the main function of which is to publish and display information (e.g. Konka Co.). If website add online orders and payment functions, it has the online marketing conditions (e.g. DELL Co.), who is on the initiative of the "linear ordering patterns." If website includes the entire business process integration of information processing systems, it is an e-commerce website (e.g. Haier Co.).

In this paper, the enterprise website link analysis and web influence evaluation showed that DELL website is better than Haier Group, and Haier Group is better than Konka Group. We presumably think that web influence and enterprises' functional types are correlated, that is, the online direct marketing enterprise website influence is better than integrated e-commerce websites and an integrated e-commerce website is better than the display of information for enterprises.

But, as for the accuracy of the results, Ronald Rousseau found that the flaw of AltaVista's arithmetic lead significant fluctuations of the search result [9]. At the same time, we may have missed some related businesses in selecting samples. What we have done is just selecting parts of business websites as representative to analyze so as to describe the problem existing. Therefore, these shortcomings and deficiencies may lead to the research errors. It suggests that it is important to reasonably understand the data results and analysis conclusions when referring to this evaluation ranking, which is also the direction we should improve.

6 Acknowledgement

References

1. Alexasir,com (January 10, 2007);http://www.alexasir.com/tech.html.
2. Cec-ceda,org (January 7, 2007); http://www.cec-ceda.org.cn/yxqywz/pxbz.htm
3. Forrester (January 12, 2007);http://www.forrester.com.
4. A.B. Huang, D.M. Zhao, "A research on the evaluation indicator system for B2C website", *Commercial Research*, 2006(1): 192-194.
5. Ingwersen. "The calculation of WEB attractiveness factor",. *Journal of Documentation*, 1998, 54,   236-243 www.//bacterio.cict.fr/backlinks.html.
6. Y.H. Li, S.J. Tang, "An evaluation on electric business website", *Library Theory and Practice*, 2006(1): 64-65.
7. Marcomconsulting (January 6, 2006);http://www.marcomconsulting.net
8. M. Thelwell, "A Comparation of Source of Links for Academic Web Attractiveness Factor Calculations", *Journal of Documentation,*2002, 58(1), 60-72.
9. R. Rousseau, "Daily time series of common single worsearches in AltaVista and Northern Light", [2007-3-2]http://cybermetrics.cindoc.csic.es/pruebas/v2i1p2.htm.
10. T. Yang, "The Webometrics Study of Twenty China Mainland Universities", *Library and Information Service*,2003(9)：61-66.
11. Z.P. Wang , *web marketing*, Wuhan: Central China normal university press, 2003.9 :P53-58.
12. J.K. Xu, "Studying by Comparison the Four Searching Engines in Common Use in the Research of Network Information Measurement", *New Technology of Library and Information Service*, 2004(11)：46-48.
13. M.Z. Yang, Y.L. Zhuang, "A summary on the theory and practice for e-business website evaluation", *Market Modernization*, 2006(2): 140-141.
14. Y. Zhang, Q.H. Zhu, "Application of Fuzzy Analytic Hierarchy Process for Enterprises Websites Evaluation", *New Technology of Library and Information Service*, 2006(6):68-71.