

Why are Hyperlinks to Business Websites Created? A Content Analysis

Abstract: Motivations for the creation of hyperlinks to business sites were analyzed through a content analysis approach. Most links were created for business purposes confirming findings from early quantitative studies that links contain useful business data. Links to competitors were extremely rare but competitors were often co-linked, suggesting that co-link analysis is the direction to pursue for information on competitive intelligence.

Résumé : Les motivations pour la création d'hyperliens pour les sites commerciaux ont été analysées à l'aide de l'approche de l'analyse de contenu. La plupart des liens ont été créés dans l'objectif d'une utilisation commerciale, confirmant les résultats obtenus des études quantitatives préliminaires suggérant que ces liens contiennent des données commerciales utiles. Les liens vers la concurrence étaient extrêmement rares, cependant les concurrents étaient souvent co-liés, suggérant que l'analyse des co-liens est la voie à poursuivre en ce qui concerne l'information en veille concurrentielle.

1. Background and Purpose of the Study

The Web is of tremendous importance to business development, particularly e-commerce. Web hyperlinks make up a network that connects businesses, customers, suppliers etc. Earlier studies have found that Web hyperlinks contain useful business information (Reid, 2003; Tan, Foo, & Hui, 2002). Specifically, the number of hyperlinks to a company's Website correlates with the company's business performance measures such as revenue and profit (Vaughan, 2004a; Vaughan & Wu, 2004). The correlation is significant even after accounting for the size of the company (Vaughan, 2004b).

While the findings from these quantitative studies provide us with empirical evidence that links to business sites can be objects of data mining for business purposes, a deeper understanding of the nature of the hyperlinks is needed to gain full confidence for such an endeavour. Toward this end, a qualitative content analysis study was carried out to examine links that point to business Websites. Of particular interest is why links were created (i.e. motivations for linking). Also of interest is the country location of linking pages as well as the type of site that created the links. The purpose of the study is twofold: to validate findings that Web hyperlinks contain useful business information and thus can be objects of data mining; to deepen our understanding of the Web link phenomenon. It is hoped that this qualitative content analysis study will serve to triangulate results from earlier quantitative studies.

The study focuses on inlinks (also called links throughout the paper). Inlinks are links coming into (or pointing to) a Web page while outlinks are links going out from a Web page, i.e. the hyperlinks embedded in a Web page (Björneborn & Ingwersen, 2004; Vaughan, 2005). Most Webometrics research uses quantitative methods to analyze inlinks. Of the few qualitative studies that analyzed motivations of linking, all examined academic Websites (Chu, 2005; Thelwall, 2003; Wilkinson et al, 2003) or scholarly Websites (Kim, 2000; Smith, 2004). The current study examined business Websites, an unexplored area. Thus the findings from the

study will not only serve the purposes mentioned above but also complement results from earlier studies on link motivations.

2. Methodology

2.1 Websites and Links in the Study

The study focuses on a particular industry since different industries may have different Website characteristics. Our most recent study (Vaughan, 2004b), which found a correlation between business performance measures and links to company Websites, chose the information technology (IT) industry and included all Canadian and U.S. IT companies as potential candidates of the study. As the purpose of the current study is to explore link creation motivations to aid our understanding of the quantitative correlations found, it is natural that these companies be used for this study. Links to the Websites of these companies (71 Canadian companies and 209 U.S. companies) were searched using commercial search engines (details below). A random sample of x links to each company Website was taken where x is 2 and 6 for U.S. and Canadian companies respectively. In total, 418 links to U.S. companies and 390 links to Canadian companies were classified, allowing for a proper comparison between the two countries.

2.2 Search Engines and Query Syntax Used

In choosing a commercial search engine to search for links to be classified, several possibilities were explored. Earlier Webometrics studies used AltaVista and AllTheWeb for inlink search. However, both were acquired by Yahoo! in March 2004 and their databases became subsets of that of Yahoo!. The Google database is one of the largest among commercial search engines. However, it cannot filter out internal links in its link search function and for this reason it was not used as a main data collection tool in earlier Webometrics research. Therefore Yahoo! (www.yahoo.com) was chosen as the most appropriate search engine for this study. Unfortunately, around the time the study started (late May 2004), the inlink search function of Yahoo! was often unavailable. Google was thus used during the initial development of the classification scheme. The Google command “link:www.abc.com” was used to retrieve links to website www.abc.com. Internal links within the site www.abc.com were manually filtered out and only external links were used in the classification exercise.

After the first stage (described below) of the classification process was completed, the Yahoo! link search function was available so it was used from the second stage on. The Yahoo “link” and “site” commands, as documented in the Help menu, were used. This means that to search for external links to site www.abc.com, the search query “link:http://www.abc.com – site:abc.com” was entered into Yahoo’s basic search mode. The “-site:abc.com” portion of the query was used to exclude internal links within the site www.abc.com. The partial domain name “abc.com” was used instead of the complete name www.abc.com as it is conceivable that a related URL (e.g. mail.abc.com) might be used by the company besides the standard www.abc.com. At the time of the study (summer 2004), the Yahoo! help menu indicated that the “link” command would search for all links pointing to the site in question. During the classification process, it was discovered that only links pointing to the specific URL (i.e. the homepage of the site) were retrieved. However, an alternative command that would truly search for all links to the site was not available. At the time of writing this paper (spring 2005), it was found that Yahoo! has an undocumented command “linkdomain” that would

search for all links to a site rather than just the links to the specific URL. This command is recommended for future research.

2.3 Development of the Classification Scheme

For each link sampled, the linking page (the page that initiated the link) was retrieved and the content of the page as well as the context of the link were examined to record the following three variables: country, type of Websites, and the motivations for linking (i.e. why the link was made). The value of the country variable was determined by the country of the page owner rather than the country designation in the URL. Similarly, type of Website was assigned by the type of organization responsible for the content of the page, not by the domain name (e.g. .com or .org). Five types of Websites, such as educational and governmental, were used in the classification, see Appendix 1 for details.

As there was no prior research examining the motivations for linking to commercial sites, no existing classification scheme was available for ready use. The classification scheme for this study was developed through an induction process based on grounded theory (Berg, 1995, p.179-181). It used latent content analysis as opposed to manifest content analysis (Babbie, 2001, p.310; Berg, 1995, p. 176). Specifically, the category of motivations for linking was developed in a three-stage process. In the first stage, the first author examined a group of linking pages and identified eight preliminary categories with brief notes defining each category. In the second stage, the second and the third authors classified a group of linking pages using the preliminary categories and added more categories as needed. They did this exercise together to achieve a consistent understanding of the categories. They also recorded three examples for each category and refined the notes that defined categories. The process continued until they found that the existing categories were sufficient to describe all links encountered.

All three authors then discussed and finalized a classification scheme of ten categories which was used in the third stage, the final classification of 808 links of the study. In this stage, each of the 808 pages was classified by the second and the third authors independently. Inter-coder consistency was 95% for the links to U.S. companies and 94.4% for links to Canadian companies. These consistency rates are higher than that achieved in previous studies of Web page and Web link classification (Bar-Ilan, 2003; Crowston & Williams, 2000; Thelwall et al, 2003; Wilkinson et al, 2003). The very high consistency rates provided us with assurance of the validity and reliability of the data.

Initially, we considered the possibility that more than one motivation could be identified for a particular link so multiple categories could be assigned to a linking page. The experience from the first two stages showed that multiple linking motivations were rarely needed so the final classification allowed only one motivation for each linking page. Minor modification of the classification scheme took place in the third stage for fine tuning purpose. The modification was discussed and agreed upon by all involved and any affected data points were carefully revised. See Appendix 2 for the final classification scheme.

3. Data Analysis and Results

3.1 Linking Pattern by Country

Countries where the linking pages were located (where the links came from) are summarized in Table 1. “Others” include Australia and Russia etc. Clearly, the majority of the links came

from the U.S., although the number of target pages (pages to be linked to) of the U.S. and Canada are similar (418 vs. 390 respectively).

Country	Number of links	Percent
U.S.	479	59.3
Canada	200	24.8
Europe	34	4.2
Asia	6	0.7
Others	8	1.0
Unknown	81	10
Total	808	100

Table 1 – County where the links come from

A cross tabulation of linking pages and linked pages is shown in Table 2 where countries other than U.S. and Canada were merged in order to have enough data points in each cell to carry out a chi-square test. The test shows a significant relationship ($p < 0.01$) between linking country and linked country. Links targeting U.S. companies were predominantly American pages while links targeting Canadian companies tended to come from Canadian pages. In other words, there is a strong tendency towards linking to pages in the same country. Countries other than the U.S. and Canada do not seem have a tendency towards linking to either U.S. or Canadian sites.

	Links from U.S.A	Links from Canada	Links from other countries	Total
To U.S.A Companies	346 (89.9%)	13 (3.4%)	26 (6.8%)	385
To Canadian companies	133 (38.9%)	187 (54.7%)	22 (6.4%)	342
Total	479.0	200.0	48.0	727.0

Table 2 – Relationship between linking country and linked country

3.2 Linking Pattern by Types of Websites

The types of Websites that link to U.S. and Canadian companies are summarized in Table 3. The type “organizational” includes the Websites of non-profit organizations, associations, and conferences. “Educational” includes the Websites of universities, colleges, and schools. The most common type of Website that links to the companies in the study is commercial sites (71.3%) while links from personal sites form only 6.1%. This suggests that links to company Websites are most likely business related (even some links from personal sites are business related).

Country	commercial	organizational	educational	personal	governmental	Total
U.S.	283 (67.7%)	78 (18.9%)	29 (6.9%)	23 (5.5%)	5 (1.2%)	418
Canada	293 (75.1%)	48 (12.3%)	13 (3.3%)	26 (6.7%)	10 (2.6%)	390
Total	576 (71.3%)	126 (15.6%)	42 (5.2%)	49 (6.1%)	15 (1.9%)	808

Table 3 – Types of Websites

There is a significant difference (chi-square test, $p < 0.01$) between American and Canadian companies in the types of Websites that link to them. Table 3 shows that the American

companies attracted more links from educational and organizational sites while the Canadian companies received more links from commercial sites.

3.3 Motivations for Linking

The results of the classification of motivations for linking (why the link was made) is summarized in Table 4 where the categories are sorted by the frequency of occurrences. The “others” categories include Weblog, bookmark, product review, newsletter etc. and each of these types constitutes less than 2% of the links classified. One of the categories that we originally had but did not use in the end due to the lack of occurrences was “link to competitors”. Only 2 out of the 808 links classified were links to competitors (one to an American company and one to a Canadian company). The scarcity of links to competitors is highly plausible. Links promote Web visibility and traffic to the site being linked to, something that companies do not want to provide for their competitors. From the standpoint of Web data mining, this means that analyzing links alone will not uncover much useful information on business competition.

Motivations	Number of links	Percent
Online Directory	182	22.5
List of companies	158	19.6
News articles	100	12.4
Acknowledging sponsors	61	7.5
Links to business partners	49	6.1
Links from customers	27	3.3
List of products	25	3.1
Links from job advertisements	22	2.7
List of clients	20	2.5
Others	164	20.3
Total	808	100

Table 4 – Motivations of Linking

The most common type of motivation for linking is “online directory” (22.5%, including search engine directories such as Yahoo! directories) followed by “list of companies” (19.6%). Related companies, many of which are competitors, are likely to be present together in these lists. This also applies to the category “list of products”. These three categories constitute 45.5% of links of all types. When two related companies, often competitors, both appear in a list, they are co-linked (page X and page Y are co-linked if a third page links to both). In other words, a significant percent of links are potential co-links. Thus, co-link analysis could be a fruitful direction to pursue for information on business competition. In fact, a recent study (Vaughan & You, 2005) successfully mapped business competition landscape for the telecommunication industry using co-link analysis, which confirms the above assertion. Motivations for links to Canadian and U.S. companies were compared and no significant difference was found (chi-square test, $p=0.15$). This means that motivations for linking are consistent across the two countries.

All motivations for linking in Table 4 are business related other than the few in the “other” category (e.g. resume of employees, newsletter run by the employees). This is in strong contrast to motivations for linking to university Websites where many links are not academically related (Chu, 2005; Thelwall, 2003). This suggests that links to business Websites are potentially more fruitful objects for data mining.

4. Conclusions and Discussion

Building on early findings that links to business Websites contain useful business information, the current study investigated the motivations for hyperlink creation (why hyperlinks are created) to gain a deeper understanding of the Web hyperlink phenomenon. The study takes a qualitative content analysis approach that complements the quantitative method employed in earlier studies. North American IT companies were selected for this study and a random sample of 808 links to these company Websites were manually classified. A classification scheme was developed through a three-stage induction process. The very high inter-coder consistency of 94.7% provides assurance of the reliability of the data.

Each linking page (a page that linked to a company in the study) was classified by country, type of Website, and motivations for linking. Links to U.S. companies predominantly came from U.S. sites while links to Canadian companies tended to come from Canadian sites. In other words, there is a very strong tendency towards linking to pages in the same country. The most common type of sites that linked to these companies was commercial sites (71.3%) while only 6.1% of links are from personal sites. American companies attracted more links from educational and organizational sites while the Canadian companies received more links from commercial sites.

The most common type of links came from online directories (22.5%). This parallels findings from Smith (2004) study where links from directories or subject guides were the most common, although that study examined links to research oriented Websites. The vast majority of links are business related which confirms findings from early quantitative studies that links contain useful business information. This is in strong contrast to links to university Websites where a significant percentage is not related to research. Despite the finding that there is a correlation between research performance and the count of links to the university Websites (e.g. Smith & Thelwall, 2002; Thelwall, 2001), Chu (2005) found less than 27% of the links were made out of research or teaching/learning motivations while Thelwall (2003) found the majority of linking motivations were trivial compared with citation motivations. This is perhaps the result of the difference between university Websites and business Websites. Company employees are generally not allowed to have personal pages on company Websites while university professors or students can have personal pages where personal interests such as hobbies can attract links that are unrelated to academic matters. The finding that the vast majority of links to business sites are business related suggests that data mining on business Websites could be more promising than that on academic sites.

Few links were created to point to business competitors (only 2 out of the 808 classified). However, competitors are frequently co-linked by a third party. This shows that simple link analysis would not provide much information on business competition but co-link analysis could be a fruitful direction to pursue. In summary, results from the study not only help in our understanding of hyperlink phenomenon but also point to directions for future research. However, findings from the study are based on links to North American IT companies. Whether they can be generalized to other industries needs to be determined by future studies.

Acknowledgement: This study is part of a larger project funded by the Initiative on the New Economy (INE) Research Grants program of the Social Sciences and Humanities Research Council of Canada (SSHRC).

References

- Babbie, E. (2001). *The practice of social research*, 9th ed. Belmont, CA: Wadsworth/Thomson Learning.
- Bar-Ilan, J. (2004). A microscopic link analysis of academic institutions within a country – the case of Israel, *Scientometrics*, 59(3), 391-403.
- Berg, B.L. (1995). *Qualitative research methods for social science*, 2nd ed. Boston: Allyn and Bacon.
- Björneborn, L. & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 55(14), 1216-1227.
- Chu, H. (2005). Taxonomy of inlinked Web entities: What does it imply for Webometric research? *Library & Information Science Research*, 27(1), 8-27.
- Crowston, K. & Williams, M. (2000). Reproduced and emergent genres of communication on the world wide web, *The Information Society*, 16, 201-215.
- Kim, H. J. (2000). "Motivations for hyperlinking in scholarly electronic articles: a qualitative study." *Journal of the American Society for Information Science*, 51(10), 887-899.
- Reid, E. (2003). Using Web link analysis to detect and analyze hidden Web communities. In: D. Vriens, (ed). *Information and Communications Technology for Competitive Intelligence* (pp. 57-84). Hilliard, Ohio: Ideal Group Inc.
- Smith, A.G. (2004). Web links as analogues of citations. *Information Research*, 9(4) paper 188 [Available at <http://informationr.net/ir/9-4/paper188.html>]
- Smith, A. & Thelwall, M. (2002). Web impact factors for Australasian universities, *Scientometrics*, 54(3), 363-380.
- Tan, B., Foo, S., & Hui, S. C. (2002). Web information monitoring for competitive intelligence. *Cybernetics and Systems*, 33(3), 225-251.
- Thelwall, M. (2003). What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation. *Information Research*, 8(3) paper no. 151. [Available at <http://informationr.net/ir/8-3/paper151.html>]
- Thelwall, M., Vaughan, L., Cothey, V., Li, X. and Smith, A. G. (2003). Which academic subjects have most online impact? A pilot study and a new classification process. *Online Information Review*, 27(5), 333-343.
- Thelwall, M. (2001). Extracting macroscopic information from Web links. *Journal of the American Society for Information Science and Technology*, 52 (13), 1157-1168.
- Vaughan, L. (2005). Web Hyperlink Analysis. In Kempf-Leonard, K. (ed.), *Encyclopedia of Social Measurement*, San Diego, CA: Academic Press, 949-954.

Vaughan, L. (2004a). Web hyperlinks reflect business performance: A study of U.S. and Chinese IT companies. *Canadian Journal of Information and Library Science*, 28(1), 17-31.

Vaughan, L. (2004b). Exploring Website features for business information. *Scientometrics*, 61(3), 467-477.

Vaughan, L. & Wu, G. (2004). Links to Commercial Web Sites as a Source of Business Information. *Scientometrics*, 60(3), 487-496.

Vaughan, L. and You J. (2005). Mapping Business Competitive Positions Using Web Co-link Analysis. Accepted by 10th International Conference of the International Society for Scientometrics and Informetrics, to be held in Stockholm, Sweden, July 24-28, 2005.

Wilkinson, D., Harries, G., Thelwall, M. & Price, E. (2003). Motivations for academic Web site interlinking: Evidence for the Web as a novel source of information on informal scholarly communication, *Journal of Information Science*, 29(1), 29(1), 59-66.

Appendix 1 – Types of Websites

- Commercial
- Organizational (e.g. non-profit organizations, associations, conferences)
- Educational (including universities, colleges, and schools. Professors' homepages are included here)
- Personal
- Governmental

Appendix 2 – Motivations of Linking

Three examples are listed below each category as illustrations. Notes that further define categories are shown within brackets.

Acknowledging cooperate sponsors

A link from www.sportpei.pe.ca to www.aliant.ca. The former presents a big logo (image link) of the latter which is its sponsor.

A link from www.shpe.org to www.accenture.com. The former indicates that the latter sponsored its national technical and career conference 2004.

A link from www.outsourcinginstitute.com to www.acs-inc.com. The latter is the sponsor of the former.

Link from customers

A link from www.furmanagers.com to www.acenetx.com. The latter is the former's ISP. The former has a logo of the latter on its homepage.

A link from www.redcloudinc.com to www.atroad.com. The former is the latter's customer. The former places a logo (link) of the latter on its webpage.

A link from www.nestegg.com to www.above.net. The former is the latter's customer. The former places a link to the latter at the beginning of its introduction page.

Online directory (including search engine directory)

A link from Yahoo! directory dir.yahoo.com to www.abelsoft.com.

A link from Yahoo!'s small business directory dir.yahoo.com to www.activcard.com.

A link from an open directory project site www.krislyn.com to www.aliant.ca.

Link from job advertisements

A link from career.berkeley.edu to www.amd.com. The former is Berkeley's job listing for its students.

A link from www.computerwork.com to www.aetea.com. This is an ad for a job at Aetea.

A link from www.torjobs.com to www.bevertec.com. This is an ad for a job at bevertec.com.

Link to business partners

A link from www.enmax.com to www.accenture.ca. This is an announcement of a partnership between the two companies.

A link from takeataxi.ca.phtemp.com to www.aliant.ca. These two companies are partners.

A link from www.m-cassociates.com to www.algx.com. These two companies are partners.

News Articles about products or companies (news articles mention products or companies. Not an in depth review. Not a simple list of companies nor a simple list of products)

A link from www.bladelogic.com to www.atstake.com. This is a news article about a product produced by www.atstake.com.

A link from thewhir.com to www.atroad.com. This is a news article about application service providers and it mentioned www.atroad.com.

A link from contact-center-today.newsfactor.com to www.accenture.com. This is an article about outsourcing and it mentioned accenture.

List of products (Simple list of products. Not product review)

A link from www.mioti.com to www.abelsoft.com. This is a list of products made by the latter.

A link from yolinux.com to www.ad-opt.com. This is a list of products by various companies including the latter.

A link from www.rfcellutions.com/pages/3/ to www.wavesat.com. This is a list of products made by various manufactures including the latter.

List of Clients

A link from www.nocturne.ca/clients.htm to www.q1labs.com. The latter is one of the clients of nocturne.ca.

A link from www.absolutehrsolutions.com/clientlist.html to www.unislumin.com. The latter is one of the clients of absolutehrsolutions.com.

A link from www.jeffkoke.com/clients_xtria.html to www.xtria.com. The later is one of jeffkoke.com's clients.

List of companies (Simple list of companies. Not a list of sponsors, nor a list of partners. In other words, if the link can be classified into a category listed above, then do not put it here).

A link from www.it-careers.ca to www.ad-opt.com. The former is a career site linking to IT companies. (This is different from the job ad category. Here it is just a simple list of company names without a specific job ad.)

A link from www.eagleeyeinc.com to www.atsva.com. The former lists top 100 companies that have the highest total number of contracts in 2001.

A link from www.nvo.com to www.cmigroup.com. The former lists Systems Integration companies with brief profiles.

Others (e.g. Weblog, bookmark, product review, newsletter etc.)