

Commercial Web Site Links

Mike Thelwall

School of Computing and Information Technology

University of Wolverhampton, Wulfruna Street, Wolverhampton, WV1 1SB, UK.

Abstract

Every hyperlink pointing at a web site is a potential source of new visitors, especially one near the top of a results page from a popular search engine. The order of the links in a search results page is often decided upon by an algorithm that takes into account the number and quality of links to all matching pages. The number of standard links targeted at a site is therefore doubly important, yet little research has touched on the actual interlinkage between business web sites, which numerically dominate the web. This paper discusses business use of the web and related search engine design issues as well as research on general and academic links before reporting on a survey of the links published by a relatively random collection of business web sites. The results indicate that around 66% of web sites do carry external links, most of which are targeted at a specific purpose, but that about 17% publish general links, with implications for those designing and marketing web sites.

Keywords

Web links, Marketing, Business, Search engine, World Wide Web

INTRODUCTION

The world-wide web, although created and first used by academics, is now dominated by business sites, with 83% of web servers containing commercial content (Lawrence and Giles, 1999). The commercial importance of the web was underlined at the start of the year 2000 by the take-over of the giant media corporation Time-Warner by the Internet Service Provider, AOL. With companies now providing free unmetred access to the Internet in countries such as the USA, Australia and the UK, the commercial focus is firmly on the actions that can be engaged in by web surfers whilst online, rather than the cost of being online. Businesses can use the web to sell online but sites can also provide online information and advertising for offline purchases, provide general information and support, and build customer relationships (Matheison, 1998, McGovern, 1999; Sainoske and Durlow, 1998).

An important issue for any web site owner is how visitors will find their way to the site. This is a crucial question for business because the effectiveness of an online initiative is likely to be dependant upon the number of potential customers that access the site. A visitor can arrive at a site in two ways: by being told its address, perhaps in a magazine advert (Pardun and Lamb, 1999), an email or an online discussion forum; or by following a link from another site. This dichotomy is not exact, however, with much email and newsgroup reader software automatically turning URLs into hyperlinks and intelligent agents automating the process of recommending links to friends (Lieberman et al, 1999) and through newsgroups (Terveen *et al.*, 1998). Surveys have shown that directories and search engines are amongst the most popular sites on the web (NetRatings Inc., 2000), with Yahoo!, for example, claiming 625 million daily page views in March 2000 (Yahoo!, 2000), and so it is to be expected that for many sites they will be a significant source of new

visitors. Links from other types of sites may also attract new visitors to any given site, but, for the web as a whole, search engines and directories are by far the most significant. As discussed later, the trend in search engines is towards improving the quality of results returned from a search, rather than attempting to index the entire web. One of the measures used by some search engines for the value of the information on a site is the number and quality of the pages with links pointing to it, where quality can also be defined in terms of the number of links pointing to a page, a recursive definition (Brin and Page, 1998). This makes links into business sites doubly important: as a direct source of new visitors and an indirect source through improved search engine quality indices. This paper will analyse business web site links in the context of trends in commercial exploitation of the web and search engine design, and will present a survey of the linkage patterns of a reasonably random sample of business web sites.

Commercial exploitation of the web

The web is used in many different ways by business, with some being more dependent on links to attract new visitors than others. Moreover, usage patterns change over time, driven by technological advances and the profile of the online population (Bellman *et al.*, 1999). Changing demographics, from the early predominantly male computing specialist domination to the much wider social basis in many countries today (NetRatings Inc., 1999; Angus Reid, 2000), have an effect upon the potential customer base size and composition, and therefore upon the types of sites that are appropriate. Technological advances generally increase the scope of the types of activity that are possible. Examples of this are faster modems allowing more data to be exchanged online, version 2 of HTML allowing the forms interface to be included in web pages (Berners-Lee and Connolly, 1995), and digital certificate support in browsers allowing relatively secure online commerce (Freier *et al.*, 1996). A less visible development with a commercial impact upon the feasibility of online advertising was the ability to reliably audit page accesses (Cooper, 1996, Sumner-Smith, 1997), although this can now be broken by sophisticated new techniques (Anupam *et al.*, 1999). In parallel with online developments, the advertising of web sites has become more aggressive, from the early days of online adverts, to the inclusion of addresses in traditional media adverts and advertising directly for web sites. In the first half of 2000, for example, the number of television adverts in the UK that are exclusively for web sites has caused comment (Morris, 2000). The more expensive advertising reflects the increasing numbers online, something that also makes possible more ambitious e-commerce. Commercial web exploitation is, then an evolving complex phenomenon in which different types of web engagement have differing needs for being the target of external links. This is illustrated in the following classification of different types of web site, based upon Hoo-Im *et al.* (1998).

Cyberstores

Cyberstores accept payment online for goods or services whether delivered online (if electronic) or offline. An online shop needs to be found to be used but strategies have evolved to facilitate this, such as the creation of cybermalls to help to attract customers through the creation of a common customer base. As an example of this, one company offers the service of linking any cyberstore to 1263 cybermalls (Majon International, 1999). Smaller shops would seem to be more dependent on links to generate visitors than larger ones, some of which have deployed large advertising

budgets (Wilder and Dalton, 1999). Cybermalls would presumably also be more likely to have an advertising budget, but should still be concerned to attract new visitors through links or any other means. Although electronic commerce continues to grow rapidly, some types of product are more suitable than others (Phau and Poon, 2000), and it may not be appropriate for certain types of business for the foreseeable future (Wysocki, 1999). One trend does, however, reduce the importance of links: the use of programs known as intelligent agents to automatically collate data from shops and give the owner a choice of the best ones. These agents can be small programs owned by the user such as BuyWiz (buywiz.com) which facilitates purchasing from a range of cyberstores, or the agents could be based in a website, for example shopsmart.com. They potentially downplay the importance of links because it will be impractical to design them to crawl the web for shop sites, rather they would be pre-programmed to visit known locations, where the data format is also recognised. These may become very widely used because they can be programmed to identify the cheapest location to buy a particular item. This is then likely to affect initially the more common items and larger stores that the product identification algorithms can identify. It is a future technical possibility for intelligent agents to dominate all online shopping, for example if an XML standard (Connolly, 2000) is adopted for all sites for the storage of product information, in which case links could again be important to help agents find a site. Alternatively, if a central database or portal was set up for cybercommerce sites that intelligent agents could use, in a similar vein to current search directories, then other links could become less relevant if ranking was not a problem.

Online Promotion

Online promotion appears to be the most common reason for creating a web site (Hoo-Im *et al.* 1998), where this is classified as promoting or advertising products or services with the intention that the customer purchases them offline. Since this is essentially an advertising function, links pointing to such a site will be important, but perhaps more so for the less well-known or smaller company without an advertising budget for the web. Although online promotion sites seem to dominate the web numerically, they may not do so financially in terms of total revenue generated. An example of an online promotion site is that of a photographer explaining and illustrating his services at www.ericstrandstrom.com.

Product or service enhancement

Product or service enhancement web sites are designed to give added value to the customer. This may take the form of instant access to information, such as the location of a parcel sent, the latest service release for a piece of software, or online help files. It can also take various other forms, such as building online user groups or common interest forums. This type of site may well not benefit greatly from links because it is designed to be used after the product has been bought, and so the site address can be promoted on its instructions or packaging. The Federal Express site www.fedex.com is a cyberstore and, with its shipment tracking service, also a service enhancement web site.

Advert display

Advert-displaying web sites earn their revenue by providing a useful service for visitors and displaying banner or other adverts. Examples include most search engine and directory sites, free online news sites and information sources (Business

Information Review, 1996). These depend directly on the number of users and so would find links important. It is easy to arrange advertising through an online broker (Gibbel, 1996) so that even small sites can use this model, but again the better financed sites and more well-known brands are likely to be more self-sufficient in their ability to attract new visitors.

Internal

Internal sites are business sites that are publicly viewable on the Internet but are not intended to attract new visitors. Examples include holding pages for domain names; company Intranets that do not contain sensitive information and have been left unprotected, and experimental and temporary, under construction sites. These would, in general, not benefit from links since new visitors are not needed. Although this type of site forms a numerically large proportion of the web (Lawrence and Giles, 1999; Thelwall, 2000a) this does not weaken in any way the argument for the importance of links because of the number of sites in the other categories for which links are still beneficial.

In summary, there are some types of areas of the web for which links are important and others where they are not, both accounting for a significant proportion of the web. The numbers and proportions are likely to continue to change and, although there is a technical possibility of their diminished importance in the arguably key area of online shopping, their role is unchallenged in the numerically large area of product or service promotion web sites, particularly for smaller and less well-known companies.

Web Portals

The necessity of having portals for the inherently unorganised collection of sites that is the web was recognised early on in its history, with a manually created, category based index site developed early on by Yahoo! and the World Wide Web Consortium. The first automatic searching and indexing robot, the World Wide Web Worm, was released in February 1994. This program crawled the web by automatically downloading web pages and extracting their links for subsequent downloading, compiling a very simple index of the web pages (Chun, 1999). Since then, many more search engines have been created, with increasingly sophisticated indexing algorithms, although Yahoo! and other imitators of its manually created directory structures have continued to survive and flourish. Over 200 robots were referenced in April, 2000 on a page that attempts to keep track of them (bots.internet.com/search/). Search engines face two problems: to index as much of the web as possible, and to identify the quality of the information on a site and then its degree of pertinence for any particular search. The relevance of a page to a search for a set of keywords could be judged by simply comparing the keywords entered by the searcher to the words in the document. In many cases, however, there will be hundreds or thousands of matches, and so algorithms have been developed to rank pages in response to searches. These algorithms are normally kept secret by the developing companies, but are known to usually give higher weightings to certain parts of a web page, for example the page title, headings and official HTML keywords tag (Pringle *et al.*, 1998). Despite these developments, a common complaint of search engine users has been that the information returned is not relevant to them (Clarke, and Willett, 1997; Gordon and Pathak, 1999; Kirsch, 1998; Pollock and Hockley, 1997). In response, many search engines now incorporate popularity ratings for pages in their ranking algorithms (Brin and Page, 1998; Kirsch, 1998). The exact description of the iterative

algorithm used by Google was published (Brin and Page, 1998), showing ratings based upon the number of links to a page from other pages and the popularity of the linking pages, measured in the same way. The reasoning behind this strategy was that a page linked to by many others is more likely to contain high quality information than one with few incoming links. The algorithm will thus give a particularly high weighting when a very popular site, such as Yahoo! is the source of a link, effectively regarding it as a high quality recommendation. Links into a site, therefore, are vitally important for obtaining new visitors through search engines.

As an example of the potential importance of links, an examination of the server logs from the Wolverhampton University Computer Based Assessment Project revealed users arriving following searches for assessment but also that a large number of visitors had been diverted to it from unrelated search engine searches. On the same day in April, 2000, links had been followed to it from a search on AltaVista for 'Wolverhampton University' (ranked 1 out of 1,998), a search on Excite for 'Wolverhampton Grammar School' (ranked 14 out of an unspecified number of matches), and a search on Yahoo! UK & Ireland for 'Wolverhampton University' (mentioned twice in a page of 14 unranked matches). Based upon information that is known about search engine ranking algorithms (Pringle *et al.*, 1998), it is believed that this site ranks well because of the number of links to it. It is in fact included in one of the human-reviewed Yahoo! categories, and it may therefore be seen as valuable information by being linked to from such a popular site. Clearly the ability to both attract users who search for appropriate topics and even to divert others to a site would be extremely advantageous for many commercial web sites.

Web link research

The web was originally developed by the Physicist Tim Berners-Lee from an idea for a medium by which research and other documents could be seamlessly linked together electronically and be accessible through a single interface (Berners-Lee *et al.*, 1992). Its name-giving purpose was to allow a large degree of document interlinking, creating a web of information. There is some evidence of this happening in university web sites, there having been several studies of academic and academic-related site links (Beall, 1997; Ingwersen, 1998; Lawrence *et al.*, 1999b; Smith, 1999), mostly focussing on links for research purposes. There is extensive interlinking, with each page in a number of sites surveyed being the target of an average between 0.44 and 2.71 links from external sites (Smith, 1999). As mentioned by Smith, such results should be treated with caution due to being based upon search engine indexes, which are known to be incomplete. The extent to which university sites attract or publish links for research purposes is, however, very variable with at least one site publishing no research documents on the web (Thelwall, 2000b) and another having prohibited external access (Snyder and Rosenbaum, 1998). The existence of meaningful links in other areas of the web is conjectured and Amento *et al.* (1999) demonstrate the potential to automatically improve non-academic information gathering through a crawler that follows external links to find new related sites. There is more academic research on web site links in general, but much of it ignores external links, focussing on different issues such as internal site structure (Bauer and Scharl, 2000) and navigation (Wan and Chung, 1998; Schneider and Lederbogen, 1999; Kim and Yoo, 2000), support for navigation provided in browsers (Cockburn and Jones, 1996), and the speed of downloading linked pages (Campbell, and Maglio, 1999).

Link-related research has also been used to measure search engine effectiveness with Lawrence and Giles showing that pages which are the target of

many links are more likely to appear in search engine indexes (Lawrence and Giles, 1999). Henzinger *et al.* (1999) used a method of creating random walks between web pages to assess the quality of search engine indexes. This study was of the entire web, rather than just the commercial areas, and was limited in its ability to find pages not already indexed by search engines. A by-product of the research was a list of the most commonly targeted sites and pages, which was dominated by pages to download resources, information pages and search engine home pages, completed by some major site home pages and a banner exchange site. The most common sites were those hosting these pages plus a page counter site, some large computing companies, an online bookstore, and the members' home pages site of an Internet Service Provider. This does not, however, give any information about the extent to which the less well-known sites are interlinked or the distribution of less common links, and this appears to be a genuine gap in the literature.

In the absence of concrete research, indications of the occurrence of external links may be gained by studying further the context in which they may be used. One of the main motivations for academic links, to allow researchers and educators to reference and share each other's work, does not necessarily apply in the more competitive business world. There are certainly those who recommend that commercial sites should link to complimentary sites in order to give a service to visitors (Matheison, 1998) and build a trust-based customer relationship. In support of this idea, there are indications that building customer relationships is a top priority for at least the largest companies (Dutta & Segev, 1998). Many sites such as BannerExchange.com also advocate link swapping to help to attract new customers, with some success as shown in by the study of Henginger *et al.* (1999). There is, nevertheless, caution expressed in some quarters about publishing external links. Any link to an external site is an extra exit route for visitors (Siegel, 1997) and so a natural reaction may be to avoid these altogether, creating an isolated self-contained structure (Grossman, 1997). Such an entity could attract visitors through external means such as traditional advertising. Gehrke and Turban (1999) surveyed web site design articles and reported the general recommendation that links should be exchanged, but with the proviso that "if outside links are necessary, they should not be placed on the home page".

The Survey

A survey was conducted to ascertain the extent of the use of external links in business web sites. The first problem was to find a random selection of commercial web sites to investigate, but first some details of the working of the Internet will be needed to explain the methodology used. Web pages on the Internet will all be hosted by a web server, and requests for a web page using HTTP are transported across the Internet using TCP/IP, referenced by a set of four numbers constituting the address of the host. These hosts can be selected at random by sending requests to randomly generated addresses. Most valid addresses do not host an active web server, they are instead either unused or used for a different function. In these cases there will either be no response or an error page will be sent. This method of randomly selecting servers, similar to that used by Lawrence and Giles (1999) is not perfect because the same web space can be referenced by multiple IP addresses. This can happen, for example when an Internet Service Provider is holding a number of addresses for later use by companies, but redirects them to its own home page. Another problem is that multiple domain names can be hosted by a single IP number in a mechanism known as a virtual server, which was introduced in HTTP 1.1 (Fielding *et al.*, 1999). This is,

however, a better method than to use search engines since their coverage is known to be biased by the amount of links to a site (Lawrence and Giles, 1998). It is also believed to be better than random selections based upon domain names for a world-wide survey, e.g. Thelwall (2000d), because of the uneven spread of domain name usage across the globe. Such a survey could take a relatively random sample of domain names from each national domain, but in order to decide upon the proportion to take from each country to give a representative sample, reliable information would also be needed on relative national domain usage. The lack of such data makes this type of survey impractical.

The search for IP addresses was conducted from January to March 2000. Each web server found was automatically put through a series of tests designed to filter out the irrelevant ones. The first test was for domain: only IP addresses with registered domain names that were from the designated commercial area, .com or national variations such as .co.il and .com.eg were kept, along with those from the European multi-purpose domains such as .de and .nl. The second test was of the home page returned for the address. These were checked and those containing the error messages such as 'File Not Found' or 'Permission denied' were excluded. After these tests the home pages were checked manually to filter out non-commercial sites and other non-public access services as well as all temporary holding pages and sites that were still under construction. The final count of sites after all the checking was 232.

Once the site list had been compiled, the sites were crawled by an automatic crawler program that started at the home page and attempted to index all pages on the site. Web site crawling is known to be sensitive to the specific technical parameters under which the program is operating (Thelwall, 2000d) and so, although not essential to the understanding of this paper, these need to be declared in order that the method may be reproduced by others. These parameters are detailed in the next paragraph.

The crawler operated in March 2000, sending requests identical to the Netscape 4.7 web browser. It indexed and crawled standard HTML anchor links, hidden automatic links to other pages (server redirection META tags), and browser based clickable image links (client side image maps). Web downloads generating transfer errors were repeated, as were those returning web pages without a </HTML> end of document tag, and a 60 second timeout was enforced for non-responding servers. The only web page errors that were corrected were those in links without matching quotes, with link URLs being automatically terminated at the end of tag or end of line character, whichever was first. Only HTML documents were indexed and two pages with different names were counted as one if they were in the same folder and contained identical HTML. Links containing URL-encoded data were ignored, as were any using the Front Page client side image map backup mechanism. Web crawlers need to have the ability to ignore problem pages since, for example, some sites including www.1freespace.com (accessed 31 March, 2000) contain anti-spam links which go to large collections of pages designed with the sole purpose of stopping web crawlers from indexing the site to find embedded email addresses for spam lists. In the sample chosen, however, no pages needed to be ignored.

The output of the web crawler was a list of the web pages found on each site, together with a list of all the external links found in them. Links were counted as internal if the domain name was the same as the home page, or differed only in the initial part. This allowed sites to use virtual hosting, for example with www.a.com, mail.a.com, ftp.a.com, and www2.a.com all counting as part of a single site. It would also have been possible to attempt to count the links into these same sites by using a search engine such as AltaVista that allows appropriate advanced queries. However,

this method was not used because the results from search engines have been shown not to be reliable enough for academic purposes. This is due to their uneven coverage of the web (Lawrence and Giles, 1999; Smith, 1999; Snyder, H. and Rosenbaum, 1999) and their variability over time (Bar-Illan, J, 1999; Rousseau, 1999).

Results

The database of links produced by the web crawler was used to manually check and classify the links in all of the sites, with the results shown in Table 1. The categories that were initially chosen had to be modified several times to accommodate links that were difficult to place under the original schema, but there were no changes after the first twenty sites had been surveyed.

Take in Table I

The results indicate that approximately two-thirds of business sites do, in fact, carry links to external sites. The most common type of link found, on 31% of sites, was to other businesses that there was a relationship with. The closest connection represented here was a product site linking to that of its manufacturer but there were also sites linking to other businesses in a larger group. A number of sites linked to their clients, often companies in the service sector using existing clients as a mark of their achievement. Some resellers also contained links to the source manufacturing companies of their products. Links listed as complementary in this category were to businesses that offered some service that enhanced the functionality of the originator's, for example one site offered surveillance equipment and linked to the web site of a private investigator. A small number of sites had Internet-based relationships with other sites through being a member of a link or banner exchange. These are organised method of attempting to gain extra visitors from other sites in exchange for a reciprocal service. In a banner exchange, each member site submits a graphical advert and these are automatically rotated around all the members in turn. A selection of links of these types were checked to count the number of registered links to the same target with the search engine HotBot. This search engine was chosen for its ability to process reliably searches for specific sites, in addition to the normal keyword searches. Its figures should be regarded as considerable underestimates because it does not index the entire web. Most of the sister company link sites registered no links or one link in HotBot, although one site did show 117. Of the complementary companies, there was a greater variety of links, with fewer zeros and one site with over 1000, and there was a similar picture for client and customer sites. The logical explanation for the difference between the two types is that the sister site links are in some sense disguised internal links, whereas if a genuinely external company is linked to then that company must have publicised their address in some way, and would therefore be more likely to attract wider links. The banner exchange target sites could not be assessed using a search engine test because in each case the banner was an automatically generated rotating one, from which a web crawler would not extract a target site link. One of the sites does, however, claim to be 'the Web's largest network' with 'over 450,000 sites' (adnetwork.bcentral.com), a claim that is given some credibility by an earlier study of links (Henzinger *et al.* 1999). There were three sites with links exchanges, two of which seemed to be functioning correctly and each site in the ring having at least three registered links. One of the three links exchanges included links to sites that were not reported to be the target of any links in HotBot. This may have been related to the fact that some of the links were to sites

with adult content, and the links page descriptions contained swear words, known to be a factor taken into account by search engines (Kirsch, 1998).

Twenty-six percent of the sites included links designed to enhance the functionality of the site. Most of these had links to other sites that provided additional information, for example one site selling laser discs linked to www.dolby.com for an explanation of one aspect of their technology. HotBot registered at least one link for each of the information sites, and several had thousands. The other type of link in this category was to sites allowing the downloading of a resource. In the survey, five sites were linked to: Microsoft for browsers and other software; Netscape for browsers; Adobe for the Acrobat portable document format reader; Apple for the QuickTime viewer; and Macromedia for the Shockwave or Flash multimedia plug-ins. A few sites also linked to cyberstores or cybermerchants to allow their products to be bought online. The targets of the resource links were sites that could be expected to attract a large number of hits, which was verified with a HotBot check. It reported more than 50,000 pages linking to the acrobat reader page, more than 10,000 linking to the Netscape browser download page, and more than 100,000 for the Internet Explorer home page. Both of the shop sites linked to also registered thousands of links.

Many sites contained links crediting an agency with a service or with having conferred an award or membership. These links were typically manifested by the company logo implemented as a link. HotBot was again used to get approximate lower bounds for the number of links to the target sites, giving results between 2 and 692, the largest being for an Information Technology testing organisation, www.nstl.com. On a few sites a page counter was used that doubled as a link to the providing site. One registered "more than 100,000 links" in HotBot and all except one (with one registered link, and not from the site in the survey) over 264. A more common type of link was to the web site design company, but there were also links to regional and national membership organisations and to companies that had given some kind of award in recognition of quality in web site design or business achievements. The numbers of links reported here were checked and adjusted, when necessary, by jumping to the end of the results list. This was necessary because in some cases the search engine's own estimates were inaccurate.

Seventeen percent of the sites carried links as a service to the customer, termed here philanthropic links. Dieberger (1997) has remarked upon the presence of such links in commercial web sites as an example of a social navigation aid, positing laying "emphasis on companies' involvement in the Internet" as an explanation for two sites studied. These tended to come from whole pages of links, in most cases revolving around a theme, such as other online businesses in the local area or online companies selling a related product. There were also some sites with links pages without any apparent pattern, being presumably to sites that the designer or owner liked. Although the percentage of sites with this type of link was relatively low, each site with links contained more than one and most contained more than five, and so the total number of external links for all commercial web sites may well be of a similar order to the number of sites. This is unlikely to mean that all or most web sites are the target of such a link, however, since factors such as site quality, company profile and search engine registration would seem to be likely to impact on the likelihood of a site being picked. It is difficult to give an accurate prediction of the proportion of sites linked to since it is impossible to verify that any given site is not the target of any links without indexing the entire web, a feat not claimed by any search engine. The directory sites linked to ranged from relatively small ones, with 14 registered links, to

large search sites with over 100,000. Most of the other sites had at least 15 links, although a small number did not have any.

Conclusions

Business web sites are designed for different purposes, from online sales to company promotion. Most types need or would benefit from being the target of links from external sites, and the consequent potential for increased traffic. The benefit is greatly enhanced by the way in which the popular search engines use links to find and rank pages. Larger companies may have the advertising budget to reduce the significance of links, and intelligent agents may bypass the task of the user finding individual web sites for online purchases, but for promotional sites and those that provide a service to gain revenue from online advertising links seem set to remain important.

It is possible to design a commercial web site that does not link to rest of the Internet, and, indeed, a third of sites in the survey had taken this option. In fact it is not intrinsically essential for any type of business site, from a simple advert for a traditional business to a cyberstore, to link to other sites, but the majority do. Some of these links reflect real-world business relationships, claim an achievement or credit aspects of the web site functionality. Others use the interlinked nature of the web to enhance site functionality though linking to downloadable resources or to additional information pages. The reasons for these links show that the commercial web does have a real need to be interconnected, paralleling the need for research papers to reference each other, as in the original conception of the web. Business web-site interlinking is, therefore, a natural and widespread phenomenon. On their own, however, these types of links may mainly group together closely related sites. Many users start at portal sites, such as search engines, and this is likely to continue for economic reasons the evolution of large links sites that serve as portals to the rest of the web is almost inevitable (Dewan *et al.* 1999). A few portal sites, then, such as directories and search engines, link to many sites but the many link to a relatively small number for information and resources. Intertwined are links between geographically and industrially related organisations, some of which are reciprocal, effectively creating clustering. The reality is more disorganised than this because of the philanthropic links potentially linking more disparate groups of sites. Whilst there is disagreement over whether a site should publish external links as a service to customers, it is certainly true that many search engines are commercially successful despite being premised upon providing external links to all visitors. An important implication for web site designers and marketers of the widespread use of philanthropic links is that it is feasible to attempt to get a site linked to in this way by others. This can be achieved by identifying sites that contain such links and sending a polite request to also link to the new site. Such sites could be discovered by searching for industrially or geographically related sites, or, more directly, by using the advance search features of search engines like HotBot or AltaVista to find all sites that link to similar sites to the one designed. It seems, then, that business web sites are likely to continue to be extensively interconnected and that the current commercial domination of the web has not undermined its original conception as a globally interlinked network.

References

- Aberg, J. and Shahmehri, N. (2000), "The role of human Web assistants in e-commerce: an analysis and a usability study", *Internet Research*, Vol. 10 No. 2, pp. 114-125.

- Amento, B., Hil, W., Terveen, L., Hix, D. and Ju, P. (1999), 'An empirical evaluation of User Interfaces for Topic Management of Web Sites', *CHI 99 Conference Proceedings*, pp. 552-559, Addison Wesley, New York.
- Angus Reid, (2000), "Second Digital Gold Rush to be led by women", http://www.angusreid.com/MEDIA/CONTENT/displaypr.cfm?id_to_view=1006, Accessed 11 April, 2000.
- Anupam, V., Mayer, A., Nissim, K., Pinkas, B. and Reiter, M. K. (1999), "On the security of pay-per-click and other Web advertising schemes", *Computer Networks*, Vol. 31 No.11-16, pp.1091-1100.
- Bar-Illan, J. (1999), "Search Engine Results over Time - A Case Study on Search Engine Stability", *Cybermetrics*, Vol. 2/3 No. 1, <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.html>, Accessed 11 April, 2000.
- Bauer, C. and Scharl, A. (2000), "Quantitative evaluation of Web site content and structure", *Internet Research*, Vol.10 No.1, pp.31-43
- Beall, J. (1997), "Cataloging World Wide Web sites consisting mainly of links", *Journal of Internet Cataloging*, Vol. 1 No. 1, pp. 83-92.
- Bellman, S., Lohse G.L. and Johnson, E. J. (1999), "Predictors of online buying behavior", *Communications of the ACM*, Vol. 42 No.12, pp.32-38.
- Berners-Lee, T., Cailliau, R., Groff, J.F. and Pollermann, B., (1992), "World-Wide Web: the information universe", *Internet Research*, Vol. 2 No.1, pp.52-8.
- Berners-Lee, T. and Connolly, D. (1995), "Hypertext Markup Language - 2.0", <http://www.ietf.org/rfc/rfc1866.txt>, Accessed 11 April, 2000.
- Brin, S. and Page, L. (1998), "The Anatomy of a large scale hypertextual web search engine", *Computer Networks and ISDN Systems*, Vol. 30 No. 1-7, pp. 107-117.
- Business Information Review (1996), "First Web advertising placement study debuts", *Business Information Review*, Vol.13 No.2, pp.109-111.
- Campbell, C. and Maglio, P. (1999), "Facilitating Navigation in Information Spaces: Road-Signs on the World Wide Web", *International Journal of Human-Computer Studies*, Vol. 50 No. 4, pp. 309-327.
- Chun, T. Y. (1999), "World Wide Web Robots: An Overview", *Online & CD-ROM Review*, Vol 23 No. 3, pp. 135-142.
- Clarke, S. J. and Willett, P. (1997), "Estimating the recall performance of Web search engines", *Aslib Proceedings*, Vol. 49 No. 7, pp. 184-189.
- Cockburn, A. and Jones, S. (1996), "Which way now? Analysing and easing inadequacies in WWW navigation", *International Journal of Human-Computer Studies*, Vol. 45, pp. 105-129.
- Connolly, D. (2000), "Extensible Markup Language (XML)", <http://www.w3.org/XML/>, Accessed 11 April, 2000.
- Cooper, L. F. (1996), "More than just hits", *Information WEEK*, No. 608, pp. 63, 68, 72.
- Dewan, R., Freimer, M. & Seidmann, A. (1999), 'Portal Kombat: The battle between web pages to become the point of entry to the World Wide Web.', *Proceedings of the 32nd Hawaii International Conference on System Sciences*, (cd-rom).
- Dieberger, A., (1997), "Supporting Social Navigation on the World Wide Web", *International Journal of Human-Computer Studies*, Vol. 46, pp. 805-825.

- Dutta, S & Segev, A. (1999), 'Transforming business in the Marketplace', *Proceedings of the 32nd Hawaii International Conference on System Sciences*, (cd-rom).
- Fielding, R., Irvine, U. C., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P. and Berners-Lee, T., (1999), "Hypertext Transfer Protocol -- HTTP/1.1", <ftp://ftp.isi.edu/in-notes/rfc2616.txt>, Accessed 12 December 1999.
- Freier, A. O., Karlton, P. and Kocher, P. C. (1996), "The SSL Protocol Version 3.0", <http://home.netscape.com/eng/ssl3/draft302.txt>, Accessed 11 April, 2000.
- Gehrke, D. & Turban, E.. (1999), 'Determinants of successful Website Design: Relative Importance and Recommendations for Effectiveness', *Proceedings of the 32nd Hawaii International Conference on System Sciences*, (cd-rom).
- Gibbel, S. (1996), "Web ad networks give marketers a new option", *Business Marketing*, Vol. 81 No.7, pp. M4.
- Gordon, M. and Pathak, P. (1999), "Finding information on the World Wide Web: the retrieval effectiveness of search engines", *Information Processing and Management*, Vol. 35, pp. 141-180.
- Grossman, W. (1997), "Get linked", *Information Age*, Vol.1 No.20, pp.21-22.
- Henzinger, M. R., Heydon, A., Mitzenmacher, M. and Najork, M. (1999), "Measuring index quality using random walks on the Web", *Computer Networks and ISDN Systems*, Vol. 31 No. 11-16, pp. 1291-1303.
- Hooi-Im, N., Pan, Y. J. and Wilson, T. D. (1998), "Business Use of The World Wide Web: A Report on Further Investigations", *International Journal of Information Management*, Vol. 18, pp. 291-314.
- Ingwersen, P. (1998), "Web Impact Factors". *Journal of Documentation*, Vol. 54 No. 2, pp. 236-243.
- Kawamoto, D. (2000), "Behind the scenes of the world's largest merger", *CNET.com*, <http://news.cnet.com/news/0-1005-200-1521535.html>, Accessed 20 April, 2000.
- Kim, J. and Yoo, B. (2000), "Toward the optimal link structure of the cyber shopping mall", *International Journal of Human-Computer Studies*, Vol. 52, pp. 531-551.
- Kirsch, S. (1998), "Infoseek's experiences searching the Internet", *SIGIR Forum*, Vol. 32 No. 2, pp. 3-7.
- Lawrence, S., Giles, C. L. (1998), "Searching the World Wide Web", *Science*, Vol. 280, pp. 98-100
- Lawrence, S. and Giles, C. L. (1999), "Accessibility of information on the web", *Nature*, Vol. 400, pp. 107-109.
- Lawrence, S. Giles, C. L. and Bollacker, K. (1999), "Digital Libraries and Autonomous Citation Indexing", *Computer*, Vol. 32 No. 6, pp. 61-71.
- Leiberman, H., van Dyke, N. and Vivacqua, A. (1999), "Let's browse: a collaborative browsing agent", *Knowledge-Based Systems*, Vol. 12, pp. 427-431.
- Majon International (1999), "Mall Link", www.majon.com/malllink.html, Accessed 20 April, 2000.
- Matheison, K. (1998), "Building effective marketing web sites", *EDI Forum: The Journal of Electronic Commerce*, Vol. 11 No. 2, pp. 14-21.
- McGovern, G. (1999), *The Caring Economy*, Blackhall Publishing, Dublin
- Morris, J. (2000), "Dot comedy", *PC Pro*, Vol. 68, pp. 39.
- NetRatings Inc., (1999), "Women On The Net", http://www.nua.ie/surveys/index.cgi?f=VS&art_id=863081585&rel=true, Accessed 11 April, 2000.

- NetRatings Inc., (2000), "Top 10 Web Properties: Month of February 01, 2000", <http://209.249.142.27/nnpm/owa/NRpublicreports.toppropertiesmonthly>, Accessed 11 April, 2000.
- Pardun, C.J. and Lamb, L. (1999), "Corporate Web sites in traditional advertisements", *Internet Research*, Vol. 9. No. 2, pp. 93-99.
- Phau, I. And Poon, S. M. (2000), "Factors influencing the types of products and services purchased over the Internet", *Internet Research*, Vol. 10 No. 2, pp. 102-113.
- Pollock, A. and Hockley, A. (1997), "What's Wrong with Internet Searching", *D-Lib Magazine*, March 1997, <http://www.dlib.org/dlib/march97/bt/03pollock.html>.
- Pringle, G., Allison, L. and Dowe, D. L. (1998), "What is a tall poppy among Web pages?", *Computer Networks and ISDN Systems*, Vol. 30 No. 1-7, pp. 369-377.
- Rousseau, R. (1999), "Daily time series of common single word searches in AltaVista and NorthernLight", *Cybermetrics*, Vol. 2/3 No. 1, <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html>, Accessed 11 April, 2000.
- Sainoske, K. and Durlow, S. G. (1998), "The customer-driven supply chain", *Electronic Commerce World*, Vol. 8 No. 12, pp. 20-24.
- Schneider, B. and Lederbogen, K. (1999), "Navigation concepts for Internet applications", *IM Information Management*, Vol.14 No.1, pp.103-109.
- Siegel, D. (1997), *Creating killer web sites*, Hayden Books, Indianapolis.
- Singh, S. N. & Dalas, N. P. (1999), "Web Home Pages as Advertisements", *Communications of the ACM*, Vol.42 No. 8, pp. 91 - 98.
- Smith, A. G. (1999), "A tale of two web spaces: comparing sites using Web Impact Factors", *Journal of Documentation*, Vol. 55, pp. 577-592.
- Snyder, H. and Rosenbaum, H. (1998), "How Public is the Web?: Robots, Access and Scholarly Communication", *Proceedings of the ASIS 98 Annual Meeting*, pp. 453-462.
- Snyder, H. and Rosenbaum, H. (1999), "Can search engines be used for web-link analysis? A critical review", *Journal of Documentation*, Vol. 55 No. 4, pp. 375-384.
- Sumner-Smith, D. (1997), "A global understanding", *Marketing*, 27 March 1997, pp. 27-9.
- Terveen, L., Hill, W. and Amento, B. (1998), "Collaborative Filtering To Locate, Comprehend, and Organize Collections of Web Sites", *SIGART Bulletin*, Vol. 9 No. 3-4, pp. 10-17.
- Thelwall, M. (2000a), "Commercial Web sites: Lost in Cyberspace?", *Internet Research*, Vol. 10 No. 2, pp. 150-159.
- Thelwall, M. (2000b), "Results from a Web Impact Factor crawler", *Journal of Documentation*, to appear.
- Thelwall, M. (2000c), "Effective Web Sites for Small to Medium Sized Enterprises", *Journal of Small Business and Enterprise Development*, Vol. 7 No. 2, pp. 149-159.
- Thelwall, M. (2000d), "Quality Issues for Web Crawlers", University of Wolverhampton.
- Wan, H. A. and Chung, C. (1998), "Web page design and network analysis", *Internet Research*, Vol. 8 No. 2, pp. 115-122.

- Wilder, C. and Dalton, G. (1999), "E-Commerce Dividends - Companies Are Making Bigger Investments In - And Reaping Rewards From - Online Sales", *Information Week*, May 3, pp. 18.
- Wysocki, B. (1999), "What's in store for online shopping", *The Wall Street Journal*, April 26, pp. A1(W).
- Yahoo! (2000), "Yahoo shares fall after earnings report", *Yahoo.com*, <http://docs.yahoo.com/docs/pr/1q00pr.html>, Accessed 20 April, 2000.

Table I A summary of the number of sites out of 232 surveyed containing different types of external link and 95% confidence intervals for these percentages

Type of Link	Number of sites	Percentage	Lower bound	Upper bound
Sister sites	36	16%	11%	21%
Clients/Suppliers	18	8%	4.7%	12%
Complementary businesses	16	7%	4.0%	11%
Links exchanges, banners and adverts	8	3%	1.5%	6.7%
Any affiliated business link	72	31%	25%	37%
Information sites or Bulletin boards	37	16%	11%	21%
Resource sites	30	13%	8.9%	18%
Shop sites	8	3%	1.5%	6.7%
Any extra functionality link	61	26%	21%	32%
Credits / Awards / Memberships	41	18%	13%	23%
Counter sites	9	4%	1.8%	7.2%
Any credit link	47	20%	15%	26%
Business sites	35	15%	11%	20%
Directory sites	10	4%	2.1%	7.8%
Any philanthropic link	40	17%	13%	23%
Any external link	152	66%	59%	72%