Online Information Review

Hyperlink analysis of the visibility of Canadian library and information science school web sites

Kwan Yi, Tao Jin,

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1468-4527.htm

Hyperlink analysis of the visibility of Canadian library and information science school web sites

Kwan Yi
Graduate School of Library and Information Science, University of Kentucky, Lexington, Kentucky, USA, and
Tao Jin
School of Library and Information Science, Louisiana State University, Baton Rouge, Louisiana, USA

Refereed article received
8 July 2007
Approved for publication
1 November 2007

Abstract

Purpose – The purpose of this paper is to probe the external visibility of the web sites of all seven ALA-accredited Canadian library and information science (LIS) schools. The number of inlinks to the schools' web sites is used as an indicator of the visibility of all or some portions of the LIS web sites.

Design/methodology/approach – Inlinks pointing to the LIS school web sites were collected using the AlltheWeb search engine. The LIS school web pages pointed to by inlinks were manually analysed to discover visible topics and contents.

Findings – Four content clusters were identified by which to group the content of all the inlinked LIS school web pages. These clusters were LIS, research, home page and resources. The most visible cluster was the LIS cluster and the least visible was the research cluster. The most visible topics were student projects/activities, LIS-related resources and course-related information, in that order. The home page of each LIS school's web site was shown to be the single web page with the most visibility.

Originality/value – This was a comparative webometric study, which collected and analysed inlinks for seven Canadian LIS school web sites at two different times, 3 years apart (2003 and 2006). In the study, the ranking of visible clusters, topics and web pages from the LIS web sites were identified.

Keywords Worldwide web, Cluster analysis, User interfaces, Canada, Inter-computer links

Paper type Research paper

Introduction

The term hyperlink was coined by Nelson (1965), who was inspired by the vision of linking or joining documents described in "As we may think", a seminal article by Bush (1945). In the early 1990s, the hyperlink notion was adopted by Tim Berners-Lee in the creation of the web, where it has been revitalised as a tool for linking, compiling and navigating web resources. A hyperlink on the web (often referred to as a "link") is a virtual link embedded in a web page that points to a portion of the same or another web page. Thus, it is viewed as a vehicle representing the connectivity of web pages and for discovering relationships between web pages.

Hyperlinks are an essential resource for organising, retrieving and accessing digital resources on the web – web pages are collected primarily by navigating the web through hyperlinks (Bar-Ilan, 2001); hyperlinks and associated text are used to index and classify

Online Information Review
Vol. 32 No. 3, 2008
pp. 325-347

1468-4527
DOI 10.1108/14684520810889655

web pages (Golub, 2006); hyperlinks have been used to indicate the authority, trust, values, evaluation and impact of web pages and web sites (Brin and Page, 1998; Gyongyi *et al.*, 2004; Ingwersen, 1998; Kleinberg, 1998). Many previous studies have focused on the application or use of hyperlinks. However, there has been much less investigation into the underlying motivations for the creation of hyperlinks, such as why they are created and what they are created for (Chu, 2005; Wilkinson *et al.*, 2003).

A web site serves as a virtual window into an organisation's programs, practices, policies and values. The evaluation of a web site may therefore prove valuable in the evaluation of the associated institution (Chu *et al.*, 2002). Moreover, academic or university web sites are often adopted and explored as role models (Li *et al.*, 2001), as the content of the web sites is often more mature and stable. Although a number of studies have been conducted on the motivation behind the hyperlinks in the web sites of academic institution's (Chu, 2005; Thelwall, 2003; Wilkinson *et al.*, 2003), there has been only limited empirical evidence to explain the hyperlinking behaviour associated with academic web sites.

The primary objective of this study was to investigate how visible the web sites of the seven Canadian ALA-accredited library and information science (LIS) schools are to the external world on the web. Visibility is understood in terms of how all or some portions of the LIS web pages or web sites are recognised and valued. The study addressed the following questions:

- How are the LIS web sites connected to the external world on the web?
- How visible is each of the LIS web sites on the web?
- Which topics and contents of the LIS web sites are more visible to the external web?

This paper proceeds in four parts. First, it will provide a review of the literature on previous research and findings related to this study. Next, it will describe the methodology used to answer the three research questions, including the rationale of the underlying approach. Then, the results will be presented and examined. In conclusion, the major findings and implications of the study will be discussed.

Literature review

In this section, a number of important terms and definitions related to hyperlinks and hyperlink analysis are presented, followed by a brief discussion of some essential techniques that have been applied in hyperlink analysis. Then, research on hyperlinks in university web sites in general and LIS school web sites in particular is reviewed.

Link terminologies

The term hyperlink refers to a clickable navigation element imbedded in a web page that leads to another web page (external linking) or to another portion within the same page (internal linking). For external linking, terms such as outlink and inlink are used often; while, for internal linking, the term selflink is common (Thelwall and Harries, 2004).

Björneborn and Ingwersen (2004, p. 1216) developed a detailed link typology that makes explicit distinctions between the link terminologies. According to their typology, an external link has two opposing meanings: "either as a link pointing out of a web site (outlink) or a link pointing into a site (inlink)" (Thelwall and Harries, 2004, pp. 86-7). In other words, an external link can be seen either as an inlink or an outlink,

depending on the web node (i.e. web page, web site or other web entity) from which it is viewed.

Figure 1 shows the basic link relationships between the web nodes of A, B and C. A has an outlink to B, while B receives an inlink from A; B has a selflink; and B has an outlink to C, while C receives an inlink from B. Thus, the same link from A to B or from B to C can be named differently.

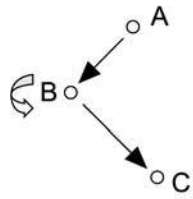
To avoid any confusion, in this study the terms are used from the perspective of the LIS school web pages. Accordingly, a hyperlink pointing to a LIS school web page is called an inlink; a web page containing such an inlink is referred to as an inlinking page; and a LIS school web page pointed to by an inlink is referred to as an inlinked page.

Webometrics

Quite often the hyperlink relations of pointing to and being pointed to are seen as analogous to those of referring to and being referred to in bibliometrics and citation analysis. The term citation was coined for web citations (Rousseau, 1997), corresponding to citation in bibliometrics. In the late 1990s, bibliometrics further inspired web studies when webometrics was coined to describe a quantitative study on the phenomena of the web.

Hyperlinks have been the primary tool in the study of webometrics, especially in the study of the usage of web sites (Almind and Ingwersen, 1997; Björneborn and Ingwersen, 2001; Ingwersen, 1998). The webometric approach has also been applied in measuring the impact, visibility and connectivity of web sites (Thelwall, 2002; Leydesdorff and Curran, 2000; Smith and Thelwall, 2001). Although webometrics does not have a long history, a good amount of literature on the topic exists. In the *Annual Review of Information Science and Technology*, Thelwall and Harries (2004) and Bar-Ilan (2004) have provided us with a relatively thorough picture of the topic.

A collection of hyperlink terms used in link analysis and webometrics is presented in Table I.



Source: Björneborn and Ingwersen (2004)

Figure 1.
Link relations

	Inlink	Outlink
Synonyms	Incoming link, inbound link, inward link, back link	Outgoing link, outbound link, outward link, forward link
Sitation	Sited	Siting
Bibliometric/citation analysis	Citation	Reference
Implication	Visibility of an entity on the web	Luminosity of an entity on the web

Sources: Thelwall and Harries (2004), Björneborn and Ingwersen (2004) and Chu *et al.* (2002)

Table I.
Distinctions between inlink and outlink

Basic techniques used in webometric studies

Link counts and content analysis are two common methods used in webometric studies. Link counting refers to the process of collecting web pages using web data collection tools (e.g. personal web crawlers, commercial search engines) and then counting the occurrence of an event from the retrieved results. It is often the first step in hyperlink analysis and webometric studies (Bar-Ilan, 2000; Vaughan and Thelwall, 2003). For example, the web impact factor (WIF) was devised as an indicator of the value of a web entity (web page, web site, etc.) (Ingwersen, 1998). WIF is defined as the ratio of "the number of links pointing to a site divided by the number of pages in the site" (Bar-Ilan, 2004, p. 258). Thus, a link count must be done to calculate a WIF.

The technique of link counting appears to be easy and straightforward, however, some limitations and challenges are inherent in the practice. The first challenge lies in the limitations of the web data collection tools. Most webometric studies rely entirely on experimental or commercial search engines to crawl web pages, but the coverage, update frequency, indexing rules, performance and stability of the search engines vary considerably. The second challenge comes directly from the nature of the web itself. The content of the web is inherently dynamic, so the result of a link count reflects only the particular time when the query was submitted. A third challenge may come from the researchers. Their language skills, information retrieval strategies and other human factors may lead to erroneous link counts (Björneborn and Ingwersen, 2001).

Another technique commonly used in webometrics is content analysis, which is derived from traditional communications studies (Krippendorff, 1980). In webometric studies, the following methods have been adopted: sampling pages/sites, examining the contents of the samples, clustering, classifying and topological modelling of the contents (Bar-Ilan, 2004; Thelwall and Harries, 2004; Weare and Lin, 2000; Vaughan and Thelwall, 2003). The downside of content analysis is that it is very time-consuming and the results can be inconsistent, particularly when human decisions are involved. Despite the possibility of errors, content analysis is critical in studying hidden behaviours and intentions on the web. As significant amounts of data are often needed in such studies, data sampling should be undertaken carefully.

Previous studies on academic web sites

Exploration of the hyperlinks within and between university web sites has been a popular research topic in webometrics for some years. The research has tended to address two general issues. The first is whether any correlation exists between the research performance of an academic unit (i.e. a university as a whole or a single department) and the hyperlink patterns or behaviours used in its web site. Early research findings were rather negative on such a correlation, although the unreliability of the search engines used was blamed for the negative results (Smith, 1999; Thelwall, 2000). In later studies, however, evidence was found that the number of links to a given university/departmental web site correlates with its research productivity (Thelwall, 2001; Smith and Thelwall, 2002; Thelwall and Tang, 2003; Li *et al.*, 2001; Tang and Thelwall, 2003). Thelwall and Harries (2004, p. 106) suggested that:

> [...] research produced by academics was the main reason for attracting links: Universities with better researchers attracted more links because the researchers produced more web content, rather than because the content produced was of a higher link attractiveness.

The second issue is which factors may influence the patterns and behaviours of hyperlinks between academic web sites? Several influencing elements have been identified, such as geographic, personal, political, economic, historical, cultural and linguistic factors. Two studies on the influence of geographic factors arrived at different results. Thelwall *et al.* (2005, p. 106) found that "the extent of interlinking between pairs of UK universities decreased with geographic distance" and "neighboring institutions were very much more likely to interlink than average". However, geographic factors were not found to be statistically significant in a similar study conducted using US academic web sites (Tang and Thelwall, 2003).

The influence of political, economic, historical, cultural and linguistic factors on academic web sites was examined in the following studies. Using a mathematical modelling approach, Vaughan and Thelwall (2003) analysed the hyperlink patterns of 74 Canadian university web sites. They observed that French language universities in Canada received far fewer inlinks than English language universities. Thelwall and Smith (2002) conducted a study of the interlinking of some universities in the Asia-Pacific region. Their results showed that university web sites in Australia and Japan, two industrialised countries, served as the hubs of inlinks in that region. A study on university web sites in the European Union (EU) showed that there was a great gap between the West and the East of the EU, and that the disparity may hinder their interlinking behaviour on the web (Thelwall and Harries, 2004).

Thelwall and Harries (2004) studied the effect of personal factors on the creation of hyperlinks. They investigated 2,763 personal home pages that linked to UK universities. It was found that "this source of links gave very similar quantitative results to inter-university links, even though almost a third of the links was [sic] for recreational purposes" (Thelwall and Harries, 2004, p. 106).

To summarise the research findings discussed so far:

- Analyses of hyperlinking (inlinks) to academic web sites have shown that the number of hyperlinks pointing to the research-related web contents of an academic web site is somewhat related to the research performance of the academic unit.
- To some extent, the number of inlinks can be used as an indicator of the usage of a web entity. Thus, if a departmental web site is receiving a relatively low number of inlinks, it is highly recommended that the department review its web content (Thelwall and Harries, 2004).

Previous studies on LIS school web sites

At the time this paper was being written, only three published studies focused on hyperlink analysis of LIS school web sites. Thomas and Willet (2000) conducted a study of hyperlinking to the web sites of 14 LIS departments in the UK. They observed "very limited correlation" (Thomas and Willet, 2000, p. 421) between the inlink counts and peer reviewed research performance. The authors concluded that researchers should be cautious in using hyperlinks to assess the research performance of LIS departments. They also suggested that LIS departments should host a wide variety of web materials to improve their web site visibility.

Chu *et al.* (2002) collected and analysed various link data (i.e. inlinks, outlinks, selflinks and colinks) for 53 American and Canadian LIS school web sites. They found significant differences between inlink counts and the academic rankings published in

US News & World Report (www.usnews.com) for these LIS schools. They also discovered that most inlinks to the school web sites came from Org, Edu or Net domains.

A follow-up study of 54 LIS school web sites in North America was conducted by Chu (2005) between 2003 and 2004. In the study, she investigated the link creation motivations and created a taxonomy of the visible LIS school web pages. Interestingly, almost half of all the inlinks came from resource- or directory-type web pages. According to the taxonomy, the four most visible categories of web content were teaching/learning, research, service and home page. It is worth noting that 73 per cent of the inlinks were created for the purpose of linking to a service or home page, while only 27 per cent of the inlinks were created to point to the web contents for research or teaching/learning. These findings may cast doubt on the correlation between link counts and research performance.

Despite these efforts, we still do not have sufficient and conclusive data to understand hyperlinking between, from and to LIS school web sites. Given the dynamic nature of the web, there is also an imminent need to monitor how LIS school web sites are used over a long-term span. In this study, the same experiment was conducted twice in 3 years – between 2003 and 2006 – to identify time-variant and time-invariant hyperlink patterns in LIS school web sites.

Methodology

Selection of search engine

The ideal data set for this study would be all the web pages available on the entire web that contain hyperlinks (inlinks) to the sample LIS school web sites. However, obtaining such a collection is simply infeasible due to the nature of the hidden web (a portion of the web that is not available to the public) and the limitations of web crawlers (Lawrence and Giles, 2000; Bharat and Broder, 1999). Instead, there are two alternative approaches – use of commercial search engines or the creation of a new web crawler. Similar to other webometric studies, this study relies on the use of a commercial search engine to collect data. The AlltheWeb search engine (www.alltheweb.com) was selected, primarily because this search engine was used in the previous study by Chu *et al.* (2002) and using it for the present research would allow a direct comparison of the results. In addition, when the first data set was collected in 2003, AlltheWeb was the only search engine with the ability to search for all web pages containing inlinks to a specified set of Uniform Resource Locators (URLs). A URL refers to the address of a web resource or web page. (Yahoo! search engine has since enabled this feature.)

Data collection

Two sets of web pages that linked to the LIS school web sites were collected at two separate times with almost 3 years between them (the first set on 4 November 2003 and the second set from 13 to 16 October 2006). An interval of 3 years is enough for a robust change in the web to be expected. Table II displays the list of URLs for the LIS schools and the queries submitted to the AlltheWeb search engine for the data collection.

In formulating the queries two search conditions were taken into consideration:

(1) the retrieval of web pages that linked to any web page under the LIS schools' domains rather than just the LIS home page URLs; and

(2) the exclusion of self-inlinks (or inlinks from the same school web site) (Chu *et al.*, 2002).

Abbreviation LIS school name, university	Search query (Home page URL)
Alberta	
School of Library and Information Studies, University of Alberta	link.all: slis.ualberta.ca – site: ualberta.ca (www.slis.ualberta.ca)
Dalhousie [a]	
School of Library and Information Studies, Dalhousie University	link.all: www.mgmt.dal.ca/slis – site: dal.ca (www.mgmt.dal.ca/slis) link.all: sim.management.dal.ca – site: dal.ca (http://sim.management.dal.ca)
McGill	
Graduate School of Library and Information Studies, McGill University	link.all: gslis.mcgill.ca – site: mcgill.ca (www.gslis.mcgill.ca)
Montreal	
Ecole de bibliothéconomie et des sciences de l'information, Université de Montréal	link.all: ebsi.umontreal.ca – site: umontreal.ca (www.ebsi.umontreal.ca)
Toronto	
Faculty of Information Studies, University of Toronto	link.all: fis.utoronto.ca – site: utoronto.ca (www.fis.utoronto.ca)
UBC	
School of Library, Archival and Information Studies, University of British Columbia	link.all: slais.ubc.ca – site: ubc.ca (www.slais.ubc.ca)
UWO	
Faculty of Information and Media Studies, UWO	link.all: fims.uwo.ca – site: uwo.ca (www.fims.uwo.ca)

Note: [a]The official URL of the school was changed from (www.mgmt.dal.ca/slis) to (http://sim.management.dal.ca) during the period between the 2003 and 2006 data collections

Table II.
Canadian LIS schools and corresponding web site home pages

The rationale for the first search condition was to retrieve web pages that linked to the LIS schools' sub-domains – for example, using "fis.utoronto.ca" instead of "www.fis.utoronto.ca" for the University of Toronto would ensure that inlinks such as "choo.fis.utoronto.ca" would not be missed.

As noted in Table II, the home page URL of the School of Library and Information Studies at Dalhousie University was changed sometime between the two data collection times. Therefore, the old and new URLs were both used in data collection in 2006 as a number of links to old URLs were automatically redirected to new URLs. However, web pages containing links that lead to a "page not found" message were not included in the data set.

Collection of inlinking and inlinked pages. The definitions of inlinking and inlinked pages are given in the "Link terminologies" part of the "Literature review" section. AlltheWeb responded to each search query (Table II) by returning the total number of inlinking pages (Table III). However, only a small percentage of the total inlinking pages were retrievable, as with other search engines. Of the retrieved web pages, only the web pages containing active inlinks to the LIS web sites were considered for further analysis. The following inlinks were not considered active: inlinks that no longer existed (i.e. dead links) possibly due to changes in the web structure after the web pages were last crawled by AlltheWeb; inlinks that referred to non-HTTP servers such as e-mail addresses, Gopher servers and file transfer protocol (FTP) servers; and inlinks to non-LIS content (for example, the LIS web site of the University of Western Ontario (UWO) is located under the Faculty of

School	Year	Number of inlinking pages reported[a]	Number of inlinking pages retrieved[b]	Number of distinct inlinking pages with active links[c]
Alberta	2003	874	455	194
	2006	1,240	530	46
Dalhousie	2003	214	201	184
	2006	621; 424[d]	418; 116[d]	6; 3[d]
McGill	2003	527	381	325
	2006	704	296	36
Montreal	2003	500	371	337
	2006	235	184	35
Toronto	2003	21,581	1,042	624
	2006	6,490	1,100	154
UBC	2003	7,953	1,022	816
	2006	872	696	252
UWO	2003	1,295	659	448
	2006	661	335	37
Total	2003	32,944	4,131	2,655
	2006	10,823; 10,626	3,559; 3,257	543; 540

Notes: [a]Numbers that the AlltheWeb search engine reported when the search queries in Table II were submitted; [b]AlltheWeb did not display all the web pages reported in the previous column, like other search engines; the number of web pages displayed, but inaccessible were also excluded from the figures in the column; [c]these figures were obtained by subtracting the number of web pages in which active links to the LIS schools were not found from the figures in the previous column; after that, the same web pages were counted as one; [d]the first and second figures correspond to the new school URL (sim.management.dal.ca) and the old URL (mgmt.dal.ca/slis), respectively

Table III.
Statistics of inlinking pages to the LIS web sites

Information and Media Studies which includes non-LIS programs such as journalism and mass communication). Since an inlinking page contains at least one inlink, the number of inlinks in the pages is at the least equal to or greater than the number of inlinking pages.

The inlinked pages were collected relatively easily as the URLs of the inlinked pages were equivalent to the URLs assigned to the inlinks. The link data obtained during the process are summarised in Tables III and IV.

Computer programs were written in the Perl language to analyse the inlinking pages' HTML source codes, identify the inlinks pointing to the LIS schools' web sites and collect the contents of the LIS web pages pointed to by the inlinks.

In the following section, the results from this study are compared to the results from Chu *et al.*'s (2002) study.

Results

Comparison of visibility (2001 vs 2003 vs 2006)

The number of inlinks can be used as an indicator of the visibility of the web contents receiving inlinks (Vreeland, 2000; Chu *et al.*, 2002; Wormell, 2001). The underlying assumption is that the more inlinks an organisation's web site receives, the more easily it can be seen on the web. The visibility of the seven Canadian LIS schools for the five-year-period from 2001 to 2006 is shown in Table V.

In comparing the web site visibility measured in 2001 and 2003, LIS school web sites had 2.4 times more inlinks in 2003 than in 2001. There was an increase in inlinks

School	Year	Hyperlink to HTTP servers Active	Hyperlink to HTTP servers Non-active[a]	Hyperlinks to non-HTTP servers (mail, Gopher, FTP)	ETC	Total
Alberta	2003	299	202	11	0	512
	2006	65	37	0	0	102
Dalhousie	2003	213	22	1	0	236
	2006	9	0	0	0	9
McGill	2003	393	63	1	0	457
	2006	60	5	0	0	65
Montreal	2003	459	19	2	0	480
	2006	68	8	0	0	76
Toronto	2003	876	440	103	0	1,419
	2006	211	179	46	0	436
UBC	2003	950	185	24	0	1,159
	2006	347	65	0	0	412
UWO	2003	442	280	0	75[b]	797
	2006	48	71	0	33[b]	152
Total	2003	3,632	1,211	142	75	5,060
	2006	808	365	46	33	1,252

Notes: [a]Refers to links pointing to HTTP servers that are no longer valid; it includes links producing HTTP 403 (file and directory cannot found) or HTTP 401.2 (unauthorised access) error messages and links through CGI forms or search engines; [b]count towards the web contents for the Departments of Journalism and Media Studies

Table IV.
Analysis of inlinks

	Active inlinks to the LIS school web sites								
	2001[a]			2003			2006		
School	Freq.	Per cent	Ranking	Freq.	Per cent	Ranking	Freq.	Per cent	Ranking
Alberta	70	4.6	5	299	8.2	6	65	8.0	4
Dalhousie	46	3.0	6	213	5.9	7	9	1.1	7
McGill	30	2.0	7	393	10.8	5	60	7.4	5
Montreal	202	13.3	2	459	12.6	3	68	8.4	3
Toronto	898	59	1	876	24.1	2	211	26.1	2
UBC	175	11.5	3	950	26.2	1	347	42.9	1
UWO	103	6.8	4	442	12.2	4	48	5.9	6
Total	1,524	100	N/A	3,632	100	N/A	808	100	N/A

Source: [a]Chu *et al.* (2002)

Table V.
Visibility comparison (2001 vs 2003 vs 2006)

for all the web sites, except Toronto's, which had a small decline. The amount of the increase varied from 13 times more inlinks for McGill to 2.3 times more for Montreal. Note that in the time between the two experiments, the home page URL of the Montreal LIS school changed from www.fas.umontreal.ca/ebsi, used in the previous study (Chu *et al.*, 2002), to www.ebsi.umontreal.ca, used in this study.

A surprising phenomenon shown in the results was a sharp plummet in the visibility of all seven school web sites between 2003 and 2006 (Table V). The total number of inlinks received by all the seven web sites in 2006 was equal to only 22 per cent of that received in 2003 and 53 per cent of that in 2001. The dramatic plunge seems to be due primarily to the much smaller ratio of the number of web pages

containing active inlinks to the number of total pages retrieved by the search engine in 2006 than in 2003 (Table III). In 2003, at least one active inlink was found and investigated in 2,655 of the 4,131 pages retrieved, which is approximately 64 per cent (2,655/4,131 × 100), whereas in 2006, the ratio fell to 15.3 per cent (543/3,559 × 100) using the old Dalhousie URL, or 16.6 per cent (540/3,257 × 100) using the new Dalhousie URL[1]. The decrease in the ratio from 2003 to 2006 may have been caused by a variation in the precision of the AlltheWeb search engine in the time between the two query trials. Precision – a popular method for measuring the performance of an information retrieval system – is the number of relevant documents retrieved divided by the total number of documents retrieved. During this experiment, it was often observed that a considerable number of the web pages returned from the query did not contain any links to the LIS schools used, which is an example of the low precision of the search engine. In summary, the total number of inlinks obtained appeared to be significantly related to the precision of the AlltheWeb search engine, specifically of the advanced query command (link and site). Therefore, a direct comparison of visibility only based on the frequency of active inlinks would not accurately reveal the change in visibility as the frequency is greatly influenced by the precision of the search tool.

Indirect comparison based on relative ranking of visibility is more useful. Comparing the 2001, 2003 and 2006 results, the school web site visibility rankings were approximately constant and there was no radical drop or rise found (Table V). The largest drop in ranking occurred for UWO, which fell from the fourth most visible in 2003 to the sixth in 2006, and the largest rises were in Alberta from 2003 to 2006, McGill from 2001 to 2003 and UBC from 2001 to 2003. The Alberta web site continuously increased in visibility over the years and the UBC web site had increased visibility in recent years.

An important finding in visibility is that the frequency of inlinks was unevenly distributed between the seven schools' web sites. Specifically, more than 50 per cent of all the inlinks to the LIS school web sites were concentrated on only one site (Toronto) in 2001, and two (Toronto and UBC) in 2003 and 2006. In percentage, the Toronto school web site was the most visible in 2001 and the UBC site became the most visible in 2003 and 2006. Over the years, the Dalhousie site was consistently the least visible.

It is possible that the change of a LIS school's URL may lead to a decline in visibility. For instance, the proportion of the inlinks to Dalhousie dropped from 5.9 per cent in 2003 to 1.1 per cent in 2006, with a URL change between those years. Although it is not clear that the URL change led directly or indirectly to the drop, it is certain that a sufficiently long duration of automatic redirection from the old URL to the new URL would help to maintain visibility as it was.

Inlinking pages and active inlinks

The distribution of the number of active inlinks present in inlinking pages was further explored as the ratios of active inlinks were found to be significantly different between 2003 and 2006. In Figure 2, the relationship between the number of active inlinks per inlinking page (x-axis) and the frequency of distinct inlinking pages (y-axis) is illustrated. The graphs in the figure correspond to the 2003 and 2006 data and the frequency values appear near markers on the graphs. Each marker represents a cumulative value for all seven LIS schools. Frequency values for individual schools can be found in Appendix 1.

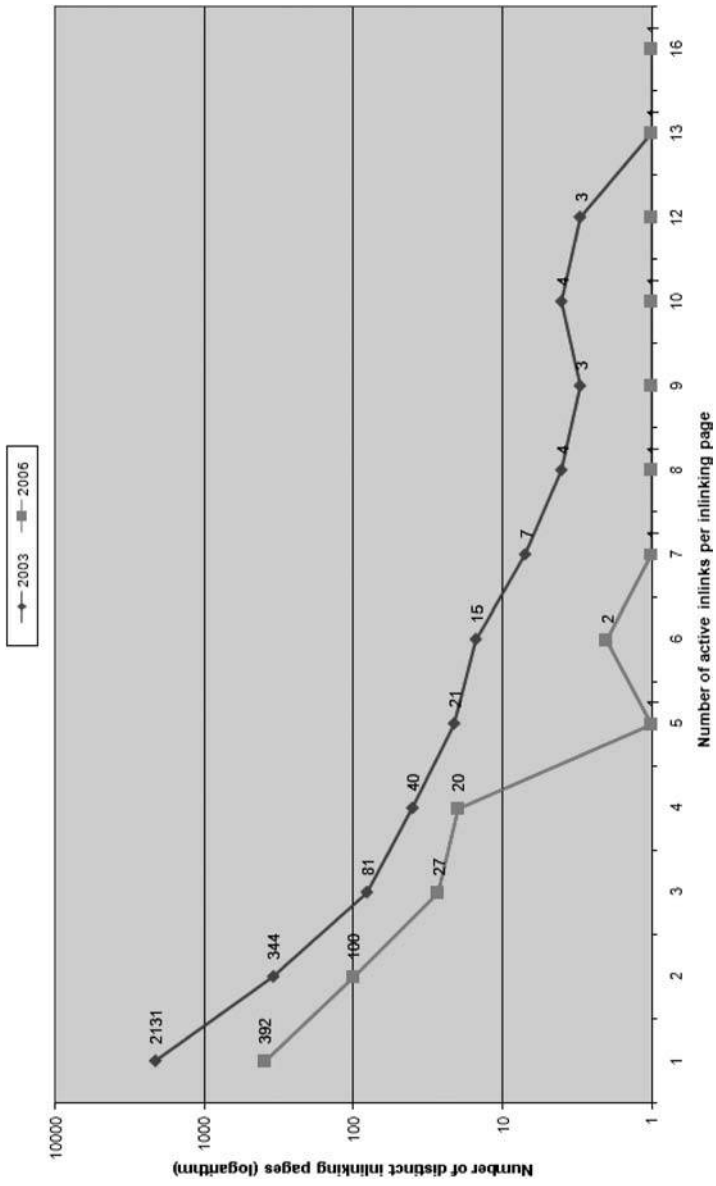


Figure 2.
Distribution of the frequency of inlinking web pages over the number of active inlinks to LIS web sites

The total number of distinct inlinking pages was 2,655 for 2003 and 543 or 540 for 2006, respectively, as shown in the last row of Table III. The proportion of inlinking pages containing only one inlink to a LIS school web page was 80 per cent for 2003 and 72 per cent for 2006. An additional 13 and 18 per cent contained two inlinks for the same years. Both of the graphs are somewhat similar to the shape of the Zipf law (Figure 5), except for a sharp drop in the number of web pages containing five active inlinks for the year 2006. Ingwersen, 1998s data, the inlinking web page with the largest number of inlinks (13) was an Association for Library and Information Science Education 2004 report (http://ils.unc.edu/ALISE/2004/Curriculum/Tableper cent20III-5a.htm) which contained inlinks to all seven LIS schools. In the 2003 data, the inlinking page with the largest number of inlinks (16) was a private company web page (www.ccrpyramid.com/resources/5883.htm – no longer exists) where all the links were to the Toronto web site.

Content analysis of the LIS inlinked web pages

The content analysis of the LIS inlinked web pages aimed to discover the most visible content topics of the LIS school web sites. We identified 3,632 active inlinks from 2,655 distinct inlinking pages (averaging 1.38 active inlinks per inlinking page) in 2003 and 808 active inlinks from 546 distinct inlinking pages (averaging 1.48 active inlinks per inlinking page) in 2006 (Table V and Appendix 1). Note that the number of active inlinks is equal to the number of inlinked web pages.

The content analysis enabled us to identify a set of content clusters into which all the LIS inlinked web pages could be classified. The analysis was conducted manually in a repeated manner using the site maps of the LIS school web sites. A site map is a visually and/or textually organised (usually hierarchically) model representing the full content of a web site, allowing for easy perception of the entire contents and the relationship between parts. First the inlinked LIS school web pages were mapped onto the seven site maps. As a result of the mapping, each category of the site maps had a specific number of inlinked web pages. The seven LIS site maps were all different in structure and terminology, so direct comparison or merging of the different site map categories was not always appropriate. Thus, in this analysis, we sought a set of clusters to define categories of inlinks through the repeated manual process of merging or splitting the site map categories and the linked web pages assigned to these by comparing the contents of the site map categories. From this, four broad content clusters were identified:

(1) *LIS cluster*. A collection of inlinked web pages about the LIS program itself, including:
   - departments, educational programmes and courses-related information; and
   - human resources information about staff and students, including students' products but excluding research work or projects.

(2) *Home page cluster*. The home pages of the LIS school web sites.

(3) *Research cluster*. A collection of inlinked web pages about research projects and publications from faculty members and research fellows. Web pages about conferences and workshops held by the unit were included as well.

(4) *Resources cluster*. A collection of inlinked web pages about online information resources of any kind, but especially career and LIS-related web resources and associations or groups.

Figures 3 and 4 show the distribution of visibility between the four clusters in 2003 and 2006, respectively. In each figure, the proportion of inlinked web pages in each of the four clusters is charted for each school and for all seven schools in total. For brevity, the abbreviation of each school name is given along the *x*-axis. Visibility is shown by percentage (along the *y*-axis), rather than by pure frequency, to normalise the difference in frequency between the schools to allow a better comparison of the distribution between schools.

In the 2003 data, there were three cases in which a cluster received more than 50 per cent of the total visibility of a school web site:

(1) the resources cluster of the Montreal LIS web site (the school visibility ranking was four as shown in Table V);

(2) the LIS cluster at McGill (ranked 5th); and

(3) the home page cluster at Dalhousie (ranked 7th).

A similar skewed pattern where one cluster received more than 50 per cent visibility occurred in the same three schools in the 2006 data as well:

(1) the LIS cluster at Montreal (school visibility ranking of 3);

(2) the LIS cluster at McGill (ranked 5th); and

(3) the home page cluster at Dalhousie (ranked 7th).

It is noteworthy that there were some single web pages that received significantly more inlinks than the home pages of the schools' web sites. Such examples included pages titled "Recruiting for diversity" (Toronto, 3.8 per cent of 876 in 2003), "Ontario's archival information network" (Toronto, 3.8 per cent of 211 in 2006), "Planning and building libraries" (UBC, 12.8 per cent of 950 in 2003 and 25.9 per cent of 347 in 2006), an article published in the proceedings of a conference held in Alberta (Alberta, 7.7 per cent of 65 in 2006), *Marginal Magazine* (McGill, 31 per cent of 393 in 2003 and 26.7 per cent of 60 in 2006). Surprisingly, a number of LIS web pages on students' projects and products received a considerable amount of inlinks.

In the 2003 data, the LIS cluster received the most visibility in four schools (McGill, Toronto, UBC and UWO, the home page cluster received the most in two schools (Alberta and Dalhousie), and the resources cluster received the most visibility in one school (Montreal). The situation was similar in the 2006 data in that the LIS cluster received the most visibility in four schools (Alberta, McGill, Montreal and Toronto) and then each of the remaining clusters received the most visibility in one school.

When all of the seven LIS schools were considered together, the same ranking order of the clusters occurred in 2003 and 2006 (see "All" in Figures 3 and 4). In descending order, this was: LIS, resources, home page and research. The summative results are based on the accumulation of all the inlink frequencies, not the proportions. Therefore, the ranking is greatly affected by the summation of the top three schools because their combined frequency was equivalent to 62 per cent of the total in 2003 and 77 per cent of the total in 2006 (Figure 5, Table V).

According to the results, LIS research-related web pages were the least visible to the public. This is not consistent with the previous finding of a positive relationship between hyperlinks and research productivity.

More detailed intermediate data from the process of mapping web pages into the site map clusters can be found in Appendixes 2 and 3.

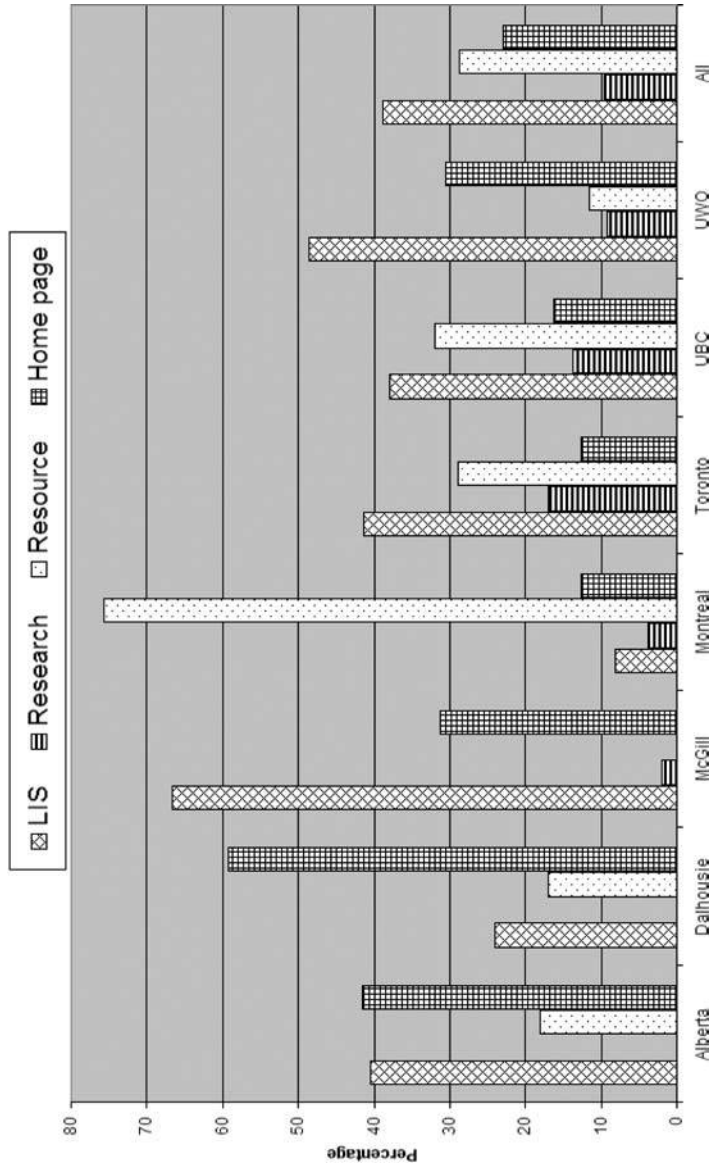


Figure 3.
Comparison of the visibility of LIS web sites by cluster (2003)

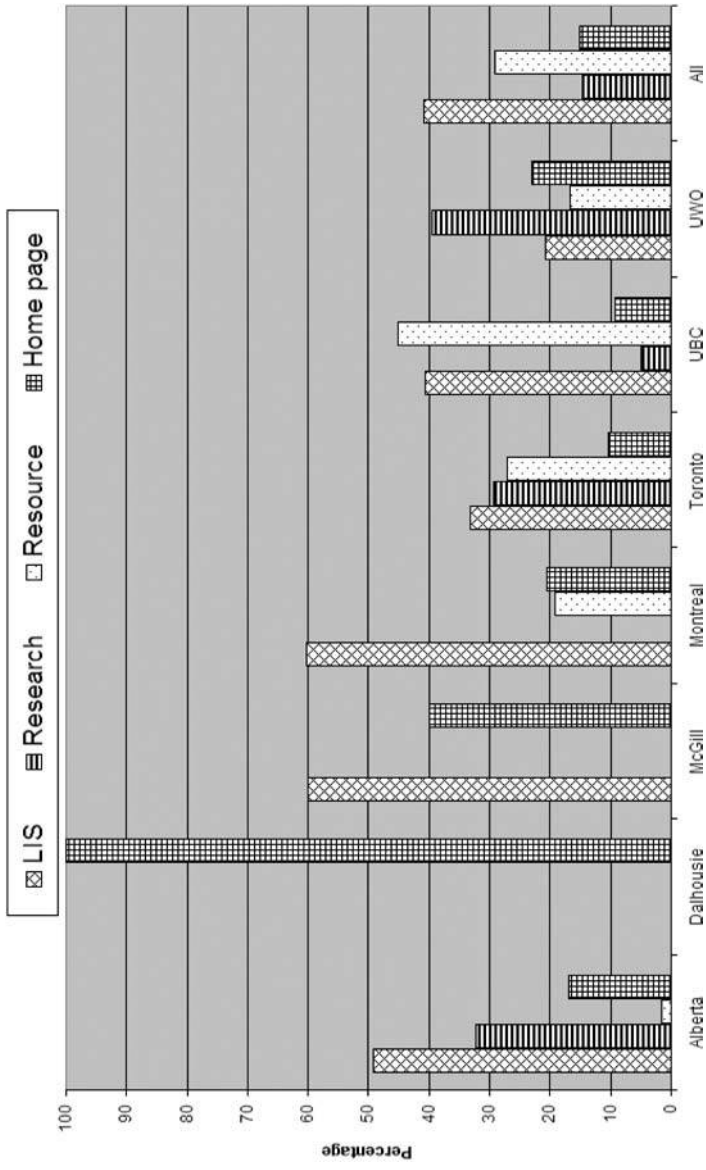


Figure 4.
Comparison of the visibility of LIS web sites by cluster (2006)

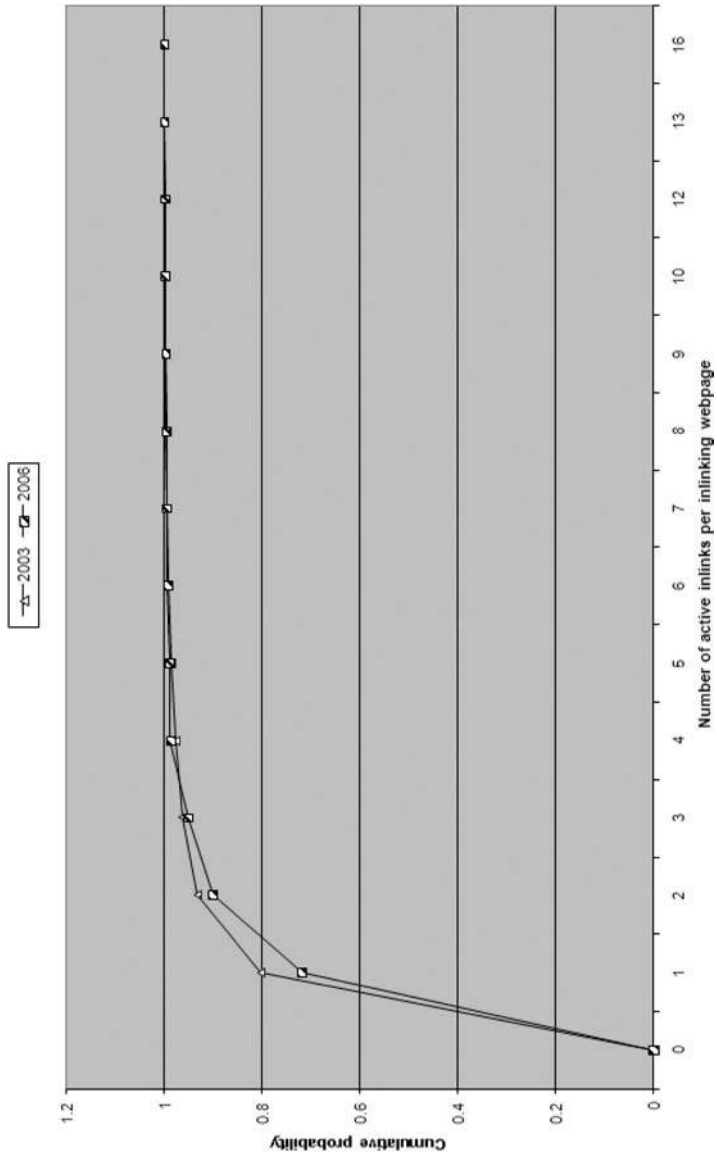


Figure 5.
A cumulative distribution function on the number of active inlinks to LIS web sites

Conclusion and discussion

This study provides experimental results from a hyperlink study of the visibility of the web sites of the seven ALA-accredited Canadian LIS schools over a 3 year period. In this study, a set of inlinks pointing to the seven LIS school web sites was collected and the contents of the corresponding inlinked web pages were manually analysed using the LIS school web site maps. The summative findings of this study on visibility (based on the combined data of 2003 and 2006) are:

- Four clusters were identified to classify the contents of all the inlinked web pages: LIS, research, home page and resources.
- The visibility of the clusters in descending order was LIS, resources, home page and research.
- The most visible topics over all LIS web sites were student projects or activities, LIS-related resources, and course-related information, in that order.
- The single most visible web page of each LIS school's web site was the home page.

The LIS school web sites received a wide range of visibility, in terms of active inlinks to the web sites. The LIS school web site with the most visibility received more than four times that of the least visible in 2003, and more than 40 times that of the least in 2006. Our analysis shows that the LIS web sites that were rich in inlinks (or visibility) grew richer, while the poor became poorer over the three years. This phenomenon is not surprising – major commercial search engines have adopted hyperlink-based ranking algorithms, such as PageRank™, which give pages with a high number of inlinks a higher ranking in query results. Consequently, it is highly likely that schools that begin with more attention and recognition continue to receive additional inlinks. Such a pattern supports the theory that the rich get richer and the poor get poorer.

In this study, four clusters were identified for the taxonomy of the LIS web site content (the topics in each cluster are listed in Appendix 2). A number of different taxonomies of web pages or web resources have been used for the classification of academic web pages, such as university web pages in UK (Thelwall, 2003) and LIS schools web sites (Wilkinson *et al.*, 2003). The four clusters identified in this study (LIS, resources, research and home page) almost exactly correspond to the four categories from Chu (2005) (teaching/learning, service, research and home page), except for the classification of the following four topics: personal home page, student organisation, news/newsletter/e-zine and announcement/description. In Chu's (2005) study, 54 ALA-accredited LIS school web sites, including the seven Canadian schools, (collected in the year of 2003) were categorised. Aligning the clusters outlined in Appendices 2 and 3 with Chu's categories yields almost identical taxonomies. As a consequence, of the alignment, the percentages for the teaching/learning, service, research and home page categories are changed to 23, 42, 13 and 22 per cent, which are comparable to 39, 29, 10 and 23 per cent for the corresponding clusters in this study.

The difference in the results of the Canadian web sites studied here and the 54 North American LIS school sites studied by Chu is due primarily to the inlinks falling under the topic of student-related activities (projects and associations). Our results reveal that the topic of student-related activities received 15 per cent of all the inlinks to the Canadian web sites, whereas it received only 4.4 per cent of all the inlinks to the North American school web sites. That is, each of the four clusters, excluding the portion

of student activities, received a similar percentage of attention (visibility or inlinks) from the outside world in the Canadian LIS school web sites as in the North America LIS school sites. However, student-related activities were found to be more visible in the Canadian school sites than in the 54 North American school web sites.

Based on the findings in this study, we make the following recommendations to improve the visibility of LIS school web sites:

- develop unique and valuable web contents;
- host a wide range of topics; and
- maintain a consistent home page URL.

For the improvement of the future research in webometrics, we suggest the following. First, the development of enhanced research tools is needed. The biggest challenge in webometrics (including this study) may be the inconsistent and imprecise performance of search engines. As Chu *et al.* (2002) pointed out, the accuracy and consistency of web search engines need to be improved for conducting webometric research. The development of a variety of effective search tools for webometric study, such as customised web crawlers that are limited to a certain scope or domain, or the refinement of search engine results through meta-search processes, would also benefit further research.

Second, studies on real life dynamic online activities are vital. Hyperlinks can be seen as static paths to navigate the web. So far, most studies in this field are focused on and limited to the study of static hyperlinks, rather than the study of the real use (dynamic online activities) of hyperlinks in online activities. Investigating dynamic online activities of hyperlinks would add value to the field of webometrics. Further research could explore how web sites are navigated by real users.

Note

1. During the period between 2003 and 2006, the Dalhousie LIS school web site URL was changed.

References

Almind, T.C. and Ingwersen, P. (1997), "Information analyses on the world wide web: methodological approaches to webometrics", *Journal of Documentation*, Vol. 53 No. 4, pp. 404-26.

Bar-Ilan, J. (2000), "The web as an information source on informetrics? A content analysis", *Journal of the American Society for Information Science*, Vol. 51 No. 5, pp. 432-43.

Bar-Ilan, J. (2001), "Data collection methods on the web for informetric purposes – a review and analysis", *Scientometrics*, Vol. 50 No. 1, pp. 7-32.

Bar-Ilan, J. (2004), "The use of web search engines in information science research", in Cronin, B. (Ed.), *Annual Review of Information Science and Technology*, Information Today, Inc, Medford, NJ, pp. 231-88.

Bharat, K. and Broder, A. (1999), "Mirror, mirror on the web: a study of host pairs with replicated content", *Computer Networks: The International Journal of Computer and Telecommunications Networking*, Vol. 31 Nos 11/16, pp. 1579-90.

Björneborn, L. and Ingwersen, P. (2001), "Perspectives of webometrics", *Scientometrics*, Vol. 50 No. 1, pp. 65-82.

Björneborn, L. and Ingwersen, P. (2004), "Towards a basic framework of webometrics", *Journal of the American Society for Information Science*, Vol. 55 No. 4, pp. 873-85.

Brin, S. and Page, L. (1998), "The anatomy of a large-scale hyper-textual web search engine", *Proceedings of the 7th International World Wide Web Conference, Brisbane, Australia, 14-18 April*, pp. 107-17.

Bush, V. (1945), "As we may think", *The Atlantic Monthly*, Vol. 176 No. 1, pp. 101-8, available at: www.cs.sfu.ca/CC/365/mark/material/notes/Chap1/VBushArticle/vbush.txt (accessed 5 October 2007).

Chu, H. (2005), "Taxonomy of inlinked web entities: what does it imply for webometric research?", *Library & Information Science Research*, Vol. 27, pp. 8-27.

Chu, H., He, S. and Thelwall, M. (2002), "Library and information science schools in Canada and USA: a webometric perspective", *Journal of Education for Library and Information Science*, Vol. 43 No. 2, pp. 110-25.

Golub, K. (2006), "Automated subject classification of textual web documents", *Journal of Documentation*, Vol. 62 No. 3, pp. 350-71.

Gyongyi, Z., Garcia-Monlina, H. and Pedersen, J. (2004), "Combating web spam with trustrank", *Proceedings of the 30th International Conference on Very Large Data Bases, Toronto, Canada, 30 August-3 September*, pp. 576-87.

Ingwersen, P. (1998), "The calculation of web impact factors", *Journal of Documentation*, Vol. 54 No. 2, pp. 236-43.

Kleinberg, J. (1998), "Authoritative sources in a hyperlinked environment", *Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), San Francisco, CA, 25-27 January*, pp. 668-77.

Krippendorff, K. (1980), *Content Analysis: An Introduction to Its Methodology*, Sage, Beverly Hills, CA.

Lawrence, S. and Giles, C.L. (2000), "Accessibility of information on the web", *Intelligence*, Vol. 11, pp. 32-9.

Leydesdorff, L. and Curran, M. (2000), "Mapping university-industry-government relations on the internet: the construction of indicators for a knowledge-based economy", *Cybermetrics*, Vol. 4 No. 1, available at: www.cindoc.csic.es/cybermetrics/articles/v4ilp2.html (accessed 14 January 2007).

Li, X., Thelwall, M., Musgrove, P. and Wilkinson, D. (2001), "The relationship between the WIFs or inlinks of computer science departments in UK and their RAE ratings or research productivities in 2001", *Scientometrics*, Vol. 57 No. 2, pp. 239-55.

Nelson, T.H. (1965), "Complex information processing: a file structure for the complex, the changing and the indeterminate", *Proceedings of the 20th National Conference, New York, NY*, pp. 84-100.

Rousseau, R. (1997), "Sitations: an exploratory study", *Cybermetrics*, Vol. 1, available at: www.cindoc.csic.es/cybermetrics/articles/v1i1p1.pdf (accessed 14 January 2007).

Smith, A. (1999), "A tale of two web spaces: comparing sites using web impact factors", *Journal of Documentation*, Vol. 55 No. 5, pp. 577-92.

Smith, A. and Thelwall, M. (2001), "Web impact factors and university research links", *Proceedings of the 8th International Conference on Scientometrics & Informetrics, Vol. 2, Sydney, 16-20 July*, pp. 657-64.

Smith, A. and Thelwall, M. (2002), "Web impact factors for Australian universities", *Scientometrics*, Vol. 54 No. 3, pp. 363-80.

Tang, R. and Thelwall, M. (2003), "Disciplinary differences in US academic departmental web site interlinking", *Library & Information Science Research*, Vol. 25 No. 4, pp. 437-58.

Thelwall, M. (2000), "Web impact factors and search engine coverage", *Journal of Documentation*, Vol. 56 No. 2, pp. 185-9.

Thelwall, M. (2001), "Extracting macroscopic information from web links", *The Journal of the American Society for Information Science and Technology*, Vol. 52 No. 13, pp. 1157-68.

Thelwall, M. (2002), "An initial exploration of the link relationship between UK university web sites", *Aslib Proceedings*, Vol. 54 No. 2, pp. 118-26.

Thelwall, M. (2003), "What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation", *Information research*, Vol. 8 No. 3, available at: http://informationr.net/ir/8-3/paper151.html

Thelwall, M. and Harries, G. (2004), "Can personal webpages that link to universities yield information about the wider dissemination of research?", *Journal of Information Science*, Vol. 30 No. 3, pp. 243-56.

Thelwall, M. and Smith, A. (2002), "A study of the interlinking between Asia-Pacific university web sites", *Scientometrics*, Vol. 55 No. 3, pp. 363-76.

Thelwall, M. and Tang, R. (2003), "Disciplinary and linguistic considerations for academic web linking: an exploratory hyperlink mediated study with Mainland China and Taiwan", *Scientometrics*, Vol. 58 No. 1, pp. 155-81.

Thelwall, M., Vaughan, L. and Björneborn, L. (2005), "Webometrics", in Cronin, B. (Ed.), *Annual Review of Information Science and Technology*, Information Today, Inc, Medford, NJ, pp. 39, 81-135.

Thomas, O. and Willet, P. (2000), "Webometric analysis of departments of librarianship and information science", *Journal of Information Science*, Vol. 26 No. 6, pp. 421-8.

Vaughan, L. and Thelwall, M. (2003), "Scholarly use of the web: what are the key inducers of links to journal web sites?", *Journal of the American Society for Information Science and Technology*, Vol. 54 No. 1, pp. 29-38.

Vreeland, R.C. (2000), "Law libraries in hyperspace: a citation analysis of world wide web sites", *Law Library Journal*, Vol. 92, pp. 9-25.

Weare, C. and Lin, W.Y. (2000), "Content analysis of the world wide web: opportunities and challenges", *Social Science Computer Review*, Vol. 18 No. 3, pp. 272-92.

Wilkinson, D., Harries, G., Thelwall, M. and Prize, L. (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*, Vol. 29, pp. 49-56.

Wormell, I. (2001), "Informetrics and webometrics for measuring impact, visibility and connectivity in science, politics and business", *Competitive Intelligence Review*, Vol. 12 No. 1, pp. 12-23.

Further reading

Vaughan, L. and Thelwall, M. (2005), "A modeling approach to uncover hyperlink patterns: the case of Canadian universities", *Information Processing & Management*, Vol. 41 No. 2, pp. 347-59.

Appendix 1

School	Number of active inlinks to LIS	Number of distinct inlinking pages 2003	2006
Alberta	1	173	35
	2	16	8
	3	3	N/A
	4	2	2
	6	N/A	1
Dalhousie	1	164	9
	2	15	N/A
	3	4	N/A
	7	1	N/A
McGill	1	270	20
	2	47	11
	3	4	2
	4	3	3
	5	1	N/A
Montreal	1	270	16
	2	50	12
	3	8	2
	4	2	4
	6	1	1
	7	2	N/A
	8	1	N/A
	9	1	N/A
	10	2	N/A
Toronto	1	488	115
	2	100	26
	3	18	8
	4	3	5
	5	5	N/A
	6	3	N/A
	8	2	N/A
	10	1	N/A
	12	2	N/A
	13	1	N/A
	16	1	N/A
UBC	1	728	192
	2	67	38
	3	7	15
	4	8	6
	5	1	N/A
	6	5	N/A
	10	N/A	1
UWO	1	389	27
	2	53	9
	3	4	1
	4	1	N/A
	6	1	N/A
Total		2,655	546

Table AI.
Frequency of inlinking pages and active inlinks

Appendix 2

Table AII.
Visibility of LIS school web contents by category, topic and school (2003)

Cluster (frequency)	Percentage in cluster	LIS school (frequency)
LIS (1,410)	*100*	Alberta (121)
Student projects/association (550)	39	Dalhousie (51)
Course material (413)	29	McGill (262)
Program admission/scholarships/continuing education (169)	12	Montreal (37)
Faculty-related information including home pages/biographical information/areas (157)	11	Toronto (363) UBC (361)
LIS in general – news/about/search (78)	6	UWO (215)
People directory (43)	3	
Research (346)	*100*	Alberta (0)
Research projects (261)	75	Dalhousie (0)
Conferences and workshops/seminars (41)	12	McGill (8)
Research in general (36)	10	Montreal (17)
Research products (articles/reports/publications/slides) (8)	2	Toronto (149)
Resources (1,044)	*100*	Alberta (54)
LIS-related web resources and directories (488)	47	Dalhousie (36)
Resources in general (336)	32	McGill (0)
LIS-related associations/centres/groups (156)	15	Montreal (347)
Job-related information – site/search/position (64)	6	Toronto (253) UBC (303) UWO (51)
Home page (832)	*100*	Alberta (124)
www.slais.ubc.ca (155)	19	Dalhousie (126)
www.fims.uwo.ca (135)	16	McGill (123)
www.mgmt.dal.ca/slis (126)	15	Montreal (58)
www.slis.ualberta.ca (124)	15	Toronto (111)
www.gslis.mcgill.ca (123)	15	UBC (155)
www.fis.utoronto.ca (111)	13	UWO (135)
www.ebsi.umontreal.ca (58)	7	
Total (3,632)		Alberta (299) Dalhousie (213) McGill (393) Montreal (459) Toronto (876) UBC (950) UWO (442)

Appendix 3

Cluster (frequency)	Percentage in cluster	LIS school (frequency)
LIS (330)	*100*	Alberta (32)
Courses – syllabus and materials (169)	51	Dalhousie (0)
Student products/groups/association (81)	25	McGill (36)
Program admission/scholarships/continuing education (40)	12	Montreal (41)
Faculty-related information including home pages/biographical information/areas (15)	5	Toronto (70)
LIS in general – news/about/search/listserv (21)	6	UBC (141)
People directory (4)	1	UWO (10)
Research (119)	*100*	Alberta (21)
Research products (articles/reports/publications/slides) (39)	33	Dalhousie (0)
Conferences and workshops/seminars (37)	31	McGill (0)
Research projects (30)	25	Montreal (0)
Research in general (13)	11	Toronto (62)
		UBC (17)
		UWO (19)
Resources (236)	*100*	Alberta (1)
LIS-related web resources and directories (209)	89	Dalhousie (0)
Job-related information – site/search/position (15)	6	McGill (0)
LIS-related associations/centres/groups (12)	5	Montreal (13)
Resources in general (0)	0	Toronto (57)
		UBC (157)
		UWO (8)
Home page (123)	*100*	Alberta (11)
www.slais.ubc.ca (32)	26	Dalhousie (9)
www.gslis.mcgill.ca (24)	20	McGill (24)
www.fis.utoronto.ca (22)	18	Montreal (14)
www.ebsi.umontreal.ca (14)	11	Toronto (22)
www.slis.ualberta.ca (11)	9	UBC (32)
www.fims.uwo.ca (11)	9	UWO (11)
www.mgmt.dal.ca/slis (6)	5	
http://sim.management.dal.ca (3)	2	
Total (808)		Alberta (65)
		Dalhousie (9)
		McGill (60)
		Montreal (68)
		Toronto (211)
		UBC (347)
		UWO (48)

Table AIII.
Visibility of LIS school web contents by category, topic and school (2006)

Corresponding author
Kwan Yi can be contacted at: kwan.yi@uky.edu