Hyperlinking to Academic Websites: Salient Features Examined

Ying XING
Center for Healthcare Informatics Education
Health Sciences Center Library
Stony Brook University, SUNY
Stony Brook, NY 11794, USA
Ying.Xing.1 at stonybrook.edu

Heting CHU (To receive all the correspondence)
Palmer School of Library & Information Science
Long Island University/C.W. Post Campus
720 Northern Blvd., Brookville, NY 11548, USA
hchu at liu.edu

ABSTRACT

OBJECTIVE: To explore in depth and comprehensively features of inlinks (incoming links) as opposed to that of citations so that better understanding can be achieved with regard to the limitations and implications in using links for evaluative webometric research.

DESIGN & MEASUREMENTS: A total of 446 randomly selected cases of hyperlinking to 15 medical schools' websites were analyzed and then classified into a revised version of a taxonomy created in a previous study for identifying linking motivations. The classification of the linking data was accomplished within the context of linking and linked sites as well as based on reasons for hyperlinking.

RESULTS: This research shows that only 5% and 7% of all the inlinks analyzed were made for reasons relating respectively to teaching/learning and research whereas 88% of the hyperlinks the target sites received were created for motivations germane to service and general nature. These findings demonstrate that inlinking is not the same as citing since inlinks exhibit features considerably different from that of citations in at least several aspects: 1) Inlinks mainly point to the ofness of target sites rather than their aboutness. 2) Inlinks cannot achieve the same level of quality in referencing as citations. 3) It is more difficult to determine reasons for hyperlinking given its unique nature.

CONCLUSION: Inlink counts alone cannot serve as quality indicators for scholarly and evaluation purposes. Other factors (e.g., authors and intellectual contents of linked entities) have to be considered in evaluative, link-based webometric research.

KEYWORDS

Websites, inlinks, hyperlinks, medical schools, webometrics, content analysis

I. Introduction

Hyperlinking is an integral feature of the Web to connect two sites to show either navigational relationships (e.g., "next page", "top of document") or content-based associations between the outlinking (i.e., linking) and inlinked (i.e., linked or target) websites [1]. Those content-based hyperlinks are commonly regarded as bibliographic citations or citations in short, and a new term "sitation" was also devised in time [2]. Studies using links as data or link-based research have been conducted ever since [e.g., 3].

Hyperlinks, particularly inlinks (incoming links), have then been examined and contrasted against citations from various perspectives [e.g., 4, 5, 6, 7]. The general consensus, via discussion alone [e.g., 7] or through peripheral considerations in research with other major questions [e.g., 8, 9], is that inlinks are not the same as citations although reports about how the two differ and in what ways they diverge vary in the published literature. On the other hand, links are often used as an evaluative measure for websites in link-based studies on the assumption that, like in evaluative citation analysis, there must be some substantial relationships between the linking and linked sites. According to this assumption, for instance, Cui [3] identified a group of 78 highly cited websites in health science by analyzing outlinks (outgoing links) of 19 top U.S. medical schools' library websites. Chu, He and Thelwall [6] evaluated 53 American Library Association (ALA)-accredited library and information science schools using inlinks, outlinks and colinks of those schools' websites.

The afore-mentioned assumption, however, has been extensively explored in citation analysis in the form of citation motivations or its synonymous expressions such as reasons for citing or citing behavior [e.g., 10, 11] in order to find out if citations can indeed be used as a plausible measure for evaluation purpose. Findings of such studies are far from consistent and conclusive [5]. As link analysis steadily evolves into a sub-area in the emerging field called webometrics [12], it appears necessary to investigate if the same assumption underlying citation analysis would hold true for link analysis. In other words, would inlinks exhibit similar characteristics as citations when we observe reasons for creating hyperlinks to websites? Several studies [5, 8, 9, 13] were conducted to address the question raised above, all focusing on why links are made to academic websites by classifying linked sites along with linking sites and other hyperlink data (e.g., URLs and link names).

The rationale for choosing hyperlinks to academic websites as research subjects in those why-linking studies seems obvious. First, websites for individual e-journals are still limited in number, and interlinking among them is to gain adequate momentum for link analysis. Second, besides interlinking among e-journals available on the Web, hyperlinks to academic sites perhaps are the closest cousin of citations among all other possibilities (e.g., links to and from non-academic sites) in existence. Third, links to academic sites are by far the most often used in link-based webometric studies [14].

Therefore, hyperlinks to academic websites will again be the subjects of analysis in the present study. More specifically, this project intends to illustrate the nature of hyperlinking by analyzing and categorizing a randomly selected sample of links to over a dozen websites of the medical schools, accredited by the Liaison Committee on Medical Education (LCME), chosen

for this study. A revised version of a taxonomy created in a previous study of similar kind [5] is used for the classification. The ultimate objective of this investigation is to explore in depth and comprehensively if the inlinks academic websites receive are comparable to citations and what can be suggested for link-based webometric research.

II. Related Studies

When links, including inlinks, became one major kind of subjects in webometric research, it is natural for researchers to explore their features in comparison with that of citations. Egghe is among the first who argue that hyperlinks are different from citations commonly used in citation analysis in several aspects although his discussion of this topic is merely conceptual [7]. Egghe indicates that, first of all, hyperlinks could be bi-directional while references are always uni-directional because only later publications can cite earlier ones. Second, websites are not articles and hyperlinks are created in a context entirely different from that for making citations. Third, hyperlinks are made for a variety of reasons (e.g., direction, politics, or subject content) by webmasters who may or may not be responsible for the intellectual content of the sites. By comparison, citations are made, though also for different purposes, by authors who should be fully responsible for the documents they compose. Fourth, there is no counterpart in citations for directional links (e.g., Go to next page) at websites.

In addition, websites, unlike articles, are subject to continuous change and often do not go through the quality control mechanism (e.g., the referee process) to get published [15]. Furthermore, links could be removed [15] or disappear for reasons irrelevant to scholarly communication (e.g., malfunction of a server). Unlike citations which only exist conceptually between the citing and cited documents, links always provide physical connection to linked sites. A simple click on the link will enable one to access the linked sites [5]. While the list of disparities between inlinks and citations could go on, how valid and convincing are the points already made and published in this regard?

Thelwall [8] did a pioneer study in identifying reasons for academic hyperlink creation. Classifying a sample of 100 random inter-site links in the ac.uk domain into a self-created list of categories, Thelwall found that motivations for linking to academic websites are primarily trivial compared with reasons for citing. Drawing a random set of link data from the same source as [8], Wilkinson and his coworkers [9] again applied the classification method to the links collected and concluded that only two out of the 414 inlinks (i.e., about 0.5%) examined were equivalent to citations. This finding implies that motivations for hyperlinking on the Web are different from those for citations.

In the study by Smith [13], 150 links to 15 research-oriented sites (including two e-journal sites) from New Zealand, United Kingdom and United States were classified to find out if inlinks are analogues to citations. The finding shows that only 10-20% of such links could be regarded similar to citations in nature. But Smith particularly emphasizes that links to e-journals are not exactly equivalent to print citations based on the kinds of links e-journal received. Although Smith states that the nature of links is more varied than citations, he did not elaborate on what the variations are.

In creating a taxonomy of inlinked websites by classifying a random sample of 1,379 sets of linking data to all the ALA-accredited library and information science (LIS) schools' websites, Chu [5] reported that less than one third (27%) of all the sites analyzed in the study were linked to for reasons relating to research or teaching/learning – the two major criteria traditionally used for evaluating academic institutions. In contrast, nearly three fourths (73%) of links were made out of motivations concerning service or home page (i.e., general). According to the results obtained through the classification and analysis, hyperlinks seem to have fewer dimensions, less complexity, and little negative implications when compared with citations.

Similar inquiries were also made either about hyperlinking to others' work in their e-publications [16] or as part of some webometric studies [e.g., 17, 18]. They will not be reviewed in this section as the current study focuses on the features of hyperlinking to the websites of selected medical schools by employing a methodology applied in a previous study [5]. The present researchers hope to continue this line of research with a different set of data from the academia.

III. Methods

The methodology followed in this study is explained in detail in Chu [5], which includes content analysis of inlinks in the context of linking and linked sites, creation of a taxonomy to accommodate the link data categorized, and identification of reasons for hyperlinking based on the taxonomy constructed. However, several aspects of this study are handled differently from Chu [5] in consideration of the focus and objective of this study. First, another researcher participated in this investigation, including the classification of the link data collected. Second, the taxonomy created in Chu [5] was adapted for this study, eliminating the need of re-inventing the wheel. Third, results obtained from the investigation are discussed to show the nature of hyperlinking to academic websites as opposed to citations in the print environment.

A. Data Collection

A systematic random sample of 15 medical schools was drawn from a total of all 142 that are accredited by LCME. This task was accomplished by compiling information available from the directory of accredited medical education programs (http://www.lcme.org/directry.htm), and the listing of medical schools in the U.S. and Canada from Association of American Medical Colleges (http://www.aamc.org/members/listings/msalphaae.htm[*]). The reason for choosing medical schools as the subjects for this study is two-fold: 1) One of the authors works at a medical school library and her knowledge about the field would help in the research process. 2) Medical schools, like the LIS schools analyzed in Chu [5], belong to the category of professional schools, which could enhance the comparability of the current study with its predecessor. Table 1 lists the 15 medical schools chosen for this study.

[*] The URL has been changed to http://services.aamc.org/memberlistings/index.cfm?fuseaction=home.search&search_type=MS.

Table 1 List of 15 Selected Medical Schools

School Name	Website URL
University of Alabama School of Medicine	main.uab.edu/uasom/
Stanford University School of Medicine	med-www.stanford.edu
George Washington University School of Medicine and Health Sciences	www.gwumc.edu
Rush Medical College of Rush University	www.rushu.rush.edu/medcol/
University of Kansas School of Medicine	www.kumc.edu/som/som.html
Boston University School of Medicine	www.bumc.bu.edu
University of Nebraska College of Medicine	www.unmc.edu/UNCOM/
Albany Medical College	www.amc.edu
Stony Brook University Health Sciences Center School of Medicine	www.informatics.sunysb.edu/som/
Oregon Health & Science University School of Medicine	www.ohsu.edu
University of Pittsburgh School of Medicine	www.medschool.pitt.edu
Meharry Medical College School of Medicine	www.mmc.edu
University of Washington School of Medicine	www.washington.edu/medical/som/
University of Calgary Faculty of Medicine	www.med.ucalgary.ca
University of Toronto Faculty of Medicine	www.library.utoronto.ca/medicine/

AlltheWeb (www.alltheweb.com) was used to collect data because it was chosen for a similar study early [5] and remains to be one of the leading search engines for collecting webometric data. The query for each selected medical school typed in AlltheWeb's Boolean Search Box is:

link:XXX NOT site:YYY (e.g., link:main.uab.edu/uasom/ NOT site:uab.edu)

Where XXX is the URL of a medical school's website (e.g., main.uab.edu/uasom/) and YYY is the root network address of the school's parent institution (e.g., uab.edu). As shown, self-links (i.e., internal links from a chosen medical school's own site or from its parent institution) are excluded from the data collected by having "NOT site:YYY" in the search query. The reason for excluding self-links is because such links are mainly for publicity, showing organizational structure (e.g., an academic unit within an institution) or other technical purposes [19].

The data collection completed within one week of time at the end of 2004. A total of 28,689 hits containing links to the 15 medical schools were retrieved, of which 5,764 (20%) could be actually displayed on screen due to AlltheWeb's own policy for result presentation. This, however, seems to be the prevailing practice among publicly accessible Web search engines and consistent with what being experienced in Chu [5]. All of the 5,764 hits were saved for data stability. A stratified, systematic sample of 507 hits was extracted from the original data set saved, forming the pool of link data for content analysis. This data group was then saved separately so that the two researchers would work with the exactly same link data during the process of content analysis.

B. Content Analysis & Taxonomy Revision

Each of the 507 hits was accessed on the Web independently by both authors to locate an inlink to a particular medical school. The linked site was then classified into the taxonomy created in a previous study [5] according to the connotations of each category (see Appendix) and reasons for hyperlinking. The link data (i.e., link name and URL), the context in which the link pointing to a target site, and the content of the linked site were also taken into consideration during the process of content analysis and taxonomy revision. The entire process was carried out in three rounds, each dealing with linking data to five medical schools. After the first round, the initial taxonomy was revised based on the classification results both authors obtained. Further but minor revisions of the taxonomy were also made after the second around. In the third around, the same procedure of content analysis and result comparison was followed.

The agreement rate between the two classifiers is on the average close to 50% (48.8%) for all the three rounds, which is lower than the 92.7% of inter-coding reliability Cronin, Snyder, Rosenbaum, Martinson and Callahan [20] achieved but higher than the 30% upper limit of inter-indexing consistency Cleverdon [21] reported. One possible reason for this mediocre consistency rate is perhaps due to the fact that the taxonomy used in this study contains more categories than those, for example, in Cronin, et al [20] and Wilkinson, et al [9]. Specifically, there are 22 categories in the taxonomy used in this study while the number of categories in Cronin, et al [20] and Wilkinson, et al [9] are 11 and 10 respectively. Although Cooper [22] stated that interindexer consistency is not necessarily an indicator of accuracy, the present authors consider the cross-classification a worthwhile practice which also meets the expectation of the research community [10].

C. Methodological Difficulties Encountered

The use of publicly accessible search engines like AlltheWeb for collecting webometric data poses difficulties, and this study is not exempted from them. Besides limitations in coverage [23, 24], in locating scientific contents [25], and regarding inadequate functionality [4], AlltheWeb stops supporting truncation, a feature that is necessary for getting linked pages and sites located below home pages in structure. For example, the URL for research programs at Oregon Health and Science University is www.ohsu.edu/research/. As truncation is no longer supported at AlltheWeb, any links to the research programs will not be retrieved. All the links AlltheWeb produced using the query “link:www.ohsu.edu NOT site:ohsu.edu” would only point to the home page and nothing beyond. The lack of truncation at AlltheWeb virtually eliminates all links to pages or sites that structurally position below the chosen medical schools’ home pages. This results in fewer numbers and less variety of linked sites to be placed into certain categories of the taxonomy, as shown in the next section.

Both Yahoo and AltaVista were also tried in an attempt to overcome the difficulty described above. However, the linkdomain option Yahoo offers would not work if a URL contains any subdirectory information (e.g., /uasom/ in linkdomain:main.uab.edu/uasom/ NOT site:uab.edu). Yet, there are seven sites out of the 15 included in this study whose URLs have subdirectories. AltaVista, like AlltheWeb, can only retrieve inlinks to a site's home page. This

limitation of search engines in locating inlinks to a site has an impact of the current study. In other words, no search engines that can be accessed in the public domain now support truncation the way they used to be. Meanwhile, it must be pointed out that inlinks are only one type of data used in classifying linked sites selected for this research. Outlinks (both the URL and link name), link names for the inlinks, and the contexts surrounding inlinks are also collected and analyzed in the study being reported.

Out of the 507 hits selected for this study, 13 are inaccessible for various reasons. Although such sites were tried in Internet Archive (www.archive.org), which is not of much help because what could be located there is usually the home page while linking sites most of the time are not. In addition, 48 results from the selected data set turned out to be self-links due to the unique composition of two medical schools' URLs. Take George Washington University School of Medicine and Health Science (www.gwumc.edu) as an example, there are two possible kinds of self-links. One would be from gwumc.edu (the medical school itself) and the other from gwu.edu (the school's parent institution George Washington University). Yet, AlltheWeb only allows one of the two root network addresses to be included in the query rather than the ideal expression as in link:XXX NOT (site:YYY OR site:ZZZ). Consequently, 446 hits out of the total were actually analyzed in this study.

Content analysis is the method chosen for this study over the two other available options (i.e., interview and survey). Its pros and cons are closely examined in previous studies [e.g., 5, 10]. The qualitative nature of this method, however, becomes outstanding especially when the authors tried to determine if a particular medical school site was linked to as directory or resource information. A decision was made, after serious discussion between the authors, that a medical school's website is linked to as directory information only if the linking site is a directory service. Most of such linking sites are in the .com domain. Other measures (e.g., classifying the data in three batches, comparing notes between the authors) are also taken to ensure the quality of the content analysis.

IV. Results and Discussion

A. Revised Taxonomy of Inlinked Sites

As indicated before, the taxonomy displayed in Table 2 is a revised version of the one created in an early study [5]. From Table 2, it can be seen that, with one exception of name change, the same four categories (i.e., teaching/learning, research, service and general) are maintained in the taxonomy. The first three of the four categories reflect the traditional components of academic institutions while the fourth one, being relabeled from home page to general, represents the general dimension of one medical school that does not fit specifically into the three other categories.

Table 2 Taxonomy of Inlinked Sites

Category	Frequency	% in Category	% in Taxonomy
Teaching/Learning	***22***	***100***	***5***
Continuing Education	3	13.6	
Course Offering	3	13.6	
Specific Program/Degree	16	72.8	
Research	***30***	***100***	***7***
Research News	21	70.0	
Research Project/Center/Forum	4	13.3	
Research Resource	5	16.7	
Service	***149***	***100***	***33***
Announcement/Description	1	0.7	
Application/Admissions	17	11.4	
Financial Aid/Scholarship	5	3.3	
Job/Co-op/Residency/Fellowship	8	5.4	
News	3	2.0	
Professional Organization	1	0.7	
Resource	111	74.5	(24.9[*])
Student Organization	3	2.0	
General	***245***	***100***	***55***
Affiliation/Relation	26	10.6	
Casual Reference/Client	17	6.9	
Degree Granting Institution	7	2.9	
Directory	92	37.6	(20.6)
Host/Organizer/Sponsor	4	1.6	
Parent Institution	40	16.3	
Partner/Affiliated Institution	34	13.9	
Specialization	25	10.2	
Total:	**446**		**100**

Each of the four top categories is further broken down into more divisions in Table 2. Connotations for every category are provided in the Appendix. Revisions done at this level could be grouped into three types. The first type of revisions is the removal of subcategories that contain no entry. They consist of “course material”, “student project/writing/notes” under the teaching/learning top category, “conference/meeting”, “journal/proceedings”, “research document” under the research category, “report” under the service category, and “personal page”, “school’s home page” as the intermediary plus its subordinate component “log/statistical

[*] Numbers in parentheses are not counted in computing totals.

report" under the general category. The withdrawal of the truncation feature at AlltheWeb attributes to most of the deletions enumerated above.

The second type of revisions involves the change of subcategory names to better describe what they imply. This group of subcategories includes "continuing education" (from "continuing education/workshop"), "specific program/degree" (from "specific program/degree/requirement") under teaching/learning, "research project/center/forum" (from "research project/center") under research, "job/co-op/residency/fellowship" (from "job/career"), "news" (from news/newsletter/e-zine") under service, and "casual reference/client" (from "casual reference"), "host/organizer/ sponsor" (from "exhibitor/host/organizer/sponsor") under home page. Modifications made onto this group in part reveal the differences between medical schools and LIS schools.

The third kind of alterations in the taxonomy adds in new subcategories to encompass what is present in hyperlinking to medical schools but absent in the case of LIS schools. This list comprises "research news" under research, "application/admissions" under service, and "specialization" under the general category. Medical research generates a lot of news, which requires a subcategory by itself. Unlike LIS schools, there are agencies (e.g., American Medical College Application Service) established specifically for processing applications for medical schools. This fact justifies the need of having a subcategory for that purpose. As there are many specializations (e.g., asthma or transplantation) in both medical teaching and practices, it seems inappropriate to just put the link site into the "specific program/degree" subcategory under teaching/learning. A new subcategory is therefore added into the taxonomy under general.

The taxonomy, as indicated early, is revised in several rounds according to the contents of linked sites, link data, as well as the context linking sites provide. The frequency and percentage figures listed in Table 2 indicate the distribution of linked sites in the taxonomy. This taxonomy subsequently becomes the basis for identifying reasons for hyperlinking to be discussed below.

B. Reasons for Hyperlinking

The reasons identified in Table 3 for hyperlinking correspond to the categories given in the taxonomy (see Table 2). The rationale behind this approach is rooted in the way how the taxonomy is developed as well as the target sites are classified. For example, if a linking site about continuing education makes a hyperlink to a target school, the possible reason for the linking site doing so is because the linked site contains information on the same topic – continuing education. In other words, the reason for hyperlinking in this case is to provide information about continuing education.

Table 3 Reasons for Hyperlinking

Category of linked Sites (Frequency)	**Reason for Hyperlinking**
Teaching/Learning (Total: 22)	
Continuing Education (3)	1. To provide information about continuing education
Course Offering (3)	2. To point to a curriculum or course catalog
Specific Program/Degree (16)	3. To provide information about programs for specific ethnic groups (e.g., African Americans) and physically impaired people (e.g., hearing loss), about pre-medicine or para-medicine programs, and distance learning programs
Research (Total: 30)	
Research News (21)	4. To indicate the source institution of research news
Research Project/Center/Forum (4)	5. To provide information about a relevant research project, center or forum of the school
Research Resource (5)	6. To point to or publicize a research resource such as funding agencies or a list of sites related to the linking site
Service (Total: 149)	
Announcement/Description (1)	7. To point to an announcement or provide descriptions about the target school
Applications/Admissions (17)	8. To provide information about admissions requirements or application procedures for a medical school
Financial Aid/Scholarship (5)	9. To provide information about financial aid and scholarship to targeted people
Job/Co-op/Residency/Fellowship (8)	10. To point to opportunities for jobs, co-op, intern, residency, and fellowship
News (3)	11. To provide news information about a target school
Professional Organization (1)	12. To point to a professional organization's local chapter at the target school
Resource (111)	13. To point to a relevant resource such as medical dictionaries, glossaries, information about a target school, ranking of medical schools
Student Organization (3)	14. To point to a student association's regional chapter at the target school
General (Total: 245)	
Affiliation/Relation (26)	15. To indicate one's affiliation or relation with the school
Casual Reference/Client (17)	16. To refer to the school in weblogs or as a client for services (e.g., Web design)

Degree Granting Institution (7)	17. To point to the school from which a degree is obtained
Directory (92)	18. To list the school in a directory of various types
Host/Organizer/Sponsor (4)	19. To refer or give credit to the school as a host, organizer or sponsor
Parent Institution (40)	20. To indicate the school as the parent institution of the linking site
Partner/Affiliated Institution (34)	21. To indicate the school as a partner or affiliated institution of the linking site
Specialization (25)	22. To point to specialized practices or programs in asthma, transplantation, etc.

As shown in both Table 2 and Table 3, the majority of reasons for hyperlinking are related to service and general while the other two categories, teaching/learning and research, contain just three reasons each. More specifically, the percentage distribution of all the target sites analyzed in the study for linking motivations in the four top categories is as follows: general – 55, service – 33, research – 7, and teaching/learning – 5. Compared with the findings reported in an early study of similar nature [5], the above results are not shocking. However, the low percentage both research and teaching/learning receive in reasons for hyperlinking seems unanticipated. On one hand, this finding confirms that hyperlinking to academic sites is generally created for reasons unrelated to research and teaching/learning, particularly if inlinks come from domains outside of the academia. On the other hand, the absence of the truncation feature at AlltheWeb causes the failure in retrieving more links to sites belonging to these two categories but located below the home page level in structure. A micro-analysis of the reasons listed under each top-level category would help further explore linking motivations.

There are three discrete reasons in the teaching/learning category, of which "specific program/degree" accounts for over two thirds (72.8%) of linked sites classified in this group. Furthermore, affirmative action seems to be the major factor that facilitates the formation of this large subcategory of hyperlinking motivation.

In the research category, "research news" subsumes most of the inlinks the target sites received. Exactly 70% of links are created in that line as medical schools often function as sources of research news. The other two reasons in this category do not appear as prominent in comparison.

Strictly speaking, most of the eight divisions clustered under the service category could also be labeled as "resource". But they are named separately so that the "resource" subcategory would not be gigantic in the taxonomy. Even so, almost three thirds (74.5%) of the sites in the category and nearly one fourth (24.9%) of the sites in the taxonomy received links for reasons relating to "resource". Various information available from the selected medical schools can be valuable resources from the viewpoints of linking sites, which should by and large explain why "resource" is the top reason for hyperlinking among all the ones identified in this study. In addition, both "announcement/description" and "professional organization" contain only one

entry each. They are kept in nevertheless because they not only existed in the original version of the taxonomy, but also could become larger if AlltheWeb had supported truncation.

The general category consists of eight reasons, among which "directory" seems to be a principal reason for hyperlinking. In fact, one fifth (20.6%) of the sites examined in this research are linked to for motivations concerning "directory". Several medical schools received inlinks basically for being listed in directories. As mentioned in Chu [5], Web is a good platform for presenting directories, and directories of all kinds (e.g., by state, by region, by country) flourish on the Web as a result. The other two reasons for which the chosen medical schools obtained considerable inlinks are regarding "parent institution" and "partner/affiliation". Many, if not all, medical schools have subsidiaries (e.g., research clinics) with different URLs, to which the medical schools are parent institutions. Meanwhile, a medical school often organizationally belongs to a health center that includes a teaching hospital and other divisions. To them, a medical school is regarded as a partner or affiliated institution. Those institutions normally have their own URLs so their links to the medical school would not be treated as self-links.

In a nutshell, what has been presented in this section echoes some of the findings reported in other studies on the same topic [5, 8, 9, 13]. Reasons for hyperlinking are different from that for citing. Likewise, inlinks are not the same as citations.

C. Features of Inlinking

Several features of inlinking have been observed during the course of this research in contrast with citing. The most noticeable one is that inlinking can mainly point to the ofness of the linked sites rather than their aboutness. For example, a faculty member works at medical school X (i.e., ofness) and specializes in subject Y (i.e., aboutness). When a site on subject Y attempts to link to the same subject that faculty member specializes in, what receives the linking would be the medical school instead of the subject which rarely has a distinctive URL for hyperlinking purpose. In the case of citation, the physical presence that represents the cited item's aboutness is not required because the connection between citing and cited items exists conceptually. If the subject is neurology, a URL for it is needed before it could be linked to while it can always be cited as long as it is about neurology.

Compared with citing, it is hard to achieve the same level of quality in inlinking. This feature primarily results from the nature and quality of information presented at academic websites. Websites for academic institutions could contain scholarly information but also include information related to service, administration and other general themes. The latter, however, seems to be the main target for hyperlinking as discussed earlier. By contrast, cited information usually is put out in scholarly publications, whose quality is controlled by, although not controversy-free, the long-lasting review mechanism. Other factors could also attribute to the issue of inlinking quality. For example, any individual who knows how to create a link on the Web can make as many links to any site as s/he wishes while citations are typically made by scholars.

It is known that citation is a private and complex process [11], which defines the difficulty in identifying reasons for citing. Nevertheless, the way hyperlinks are created makes it

even harder to determine inlinking motivations. First, neither linking nor linked entities are finite and stable. What can be obtained and examined in link-based studies is basically a snapshot of the ever-changing Web. Second, links to academic websites are not necessarily created for scholarly purposes. For instance, some medical schools are linked to because they are the clients of commercial services, as found in this research. Third, it is normally impossible for link creators to recall reasons for hyperlinking to hundreds, if not thousands, of Web entities over a long period of time. Studies of this kind hence could only use the indirect method of content analysis for identifying linking motivations, which leaves much to be desired for in terms of validity [5].

All the features of inlinking examined above, together with what has been discussed in other parts of this study, indicate that inlinks have their own distinct characteristics and should not be regarded the same as citations.

V. Conclusion

Hyperlinks to 15 selected medical schools have been analyzed in conjunction with associated data (e.g., linking sites) using a methodology developed in an early study [5] to further explore the features of inlinking as opposed to that of citing. The findings of this research demonstrate that inlinks differentiate themselves considerably from citations in at least several respects.

As pointed out early, medical schools are chosen for this study because LIS schools were the targets of analysis in the previous research [5]. Although certain methodological procedures (e.g., the taxonomy revision) should be altered to address the differences between these two kinds of professional schools, the fundamental framework of the methodology as well as the findings of both studies remain unchanged. This fact proves the reliability of the research methodology as well as the study results.

Nevertheless, one should be fully aware of its qualitative nature and subsequently its validity in carrying out research of this kind. The present authors thus employed different techniques such as cross-checking and multiple revisions throughout the study process, trying to minimize the adverse effects this qualitative investigation may bring. Another point that needs to be made relates to the use of publicly accessible search engines for data collection in this type of research. The limitations of search engines (e.g., the lack of truncation at AlltheWeb) could have negative impacts on the research as well.

In conclusion, inlinks exhibit features dramatically different from that of citations based on the findings of this study. Inlink counts alone thus cannot serve as quality indicators for scholarly and evaluation purposes. Other factors (e.g., the author and intellectual content of the webpage) have to be considered in evaluative, link-based research.

Appendix: Taxonomy Categories and Their Connotations

Teaching/Learning

- Continuing Education – Information about continuing education
- Course Offering – Curricula or course catalogs
- Specific Program/Degree – Programs for specific ethnic groups (e.g., African Americans) and physically challenged people (e.g., hearing loss), pre-medicine or para-medicine programs, and distance learning programs

Research

- Research News – News about research conducted at a target medical school
- Research Project/Center/Forum – Descriptions of research projects, centers or forums at a target medical school
- Research Resource – Funding agencies, listings of sites related to the linking site.

Service

- Announcement/Description – Announcements about or descriptions of a target school
- Application/Admissions – Information about admissions requirements and application procedures
- Financial Aid/Scholarship – Information about financial aid and scholarship for students
- Job/Co-op/Residency/Fellowship – Opportunities for jobs, co-op, intern, residency, and fellowship
- News – News about a target school
- Professional Organization – Local chapters of professional organizations at target schools
- Resources – Medical dictionaries, glossaries, information about a target school, ranking of medical schools, etc.
- Student Organization – Regional chapters of student associations at target schools

General

- Affiliation/Relation – A faculty member, a student, an adjunct, etc.
- Casual Reference/Client – As mentioned in weblogs or referred to as a client for service (e.g., Web design)
- Degree Granting Institution – Degrees conferred to graduates
- Directory – Lists of medical programs and teaching hospitals in a state, a region, in a country, etc.
- Host/Organizer/Sponsor – Host/organizer of an activity, sponsor of an event, etc.
- Parent Institution – Parent institutions for subordinate units such as research centers
- Partner/Affiliated Institution – Collaboration with others on medical projects or affiliate institution of the linking site
- Specialization – Specialized practices and programs in asthma, transplantation, etc.

References

1. Baron L, Tague-Sutcliffe J, Kinnucan MT. Labeled, typed links as cue when reading hypertext documents. J Am Soc Inf Sci 1996; 47(12):896-908.
2. Rousseau R. Sitations: An exploratory study. Cybermetrics [serial online] 1997 [cited 2005 March 30]; 1(1):[6 screens]. URL: http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html.
3. Cui, L. Rating health Web sites using the principles of citation analysis: A bibliometric approach. J Med Internet Res [serial online] 1999 Jul-Sep [cited 2005 March 30]; 1(1):[7 screens]. URL: http://www.jmir.org/1999/1/e4/index.htm/.
4. Björneborn L, Ingwersen P. Perspectives of webometrics. Scientometrics 2001; 50(1):65-82.
5. Chu H. Taxonomy of inlinked Web entities: What does it imply for webometric research? Lib Inf Sci Res 2005; 27(1):8-27.
6. Chu H, He S, Thelwall M. Library and information science schools in Canada and USA: A Webometric perspective. J Edu Lib Inf Sci 2002; 43(2):110-125.
7. Egghe L. New informetric aspects of the Internet: Some reflections – many problems. J Inf Sci 2000; 26:329-335.
8. Thelwall M. What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation. Inf res [serial online] 2003 Apr [cited 2005 March 30]; 8(3):paper no. 151. URL: http://informationr.net/ir/8-3/paper151.html.
9. Wilkinson D, Harries G, Thelwall M, Prize L. Motivations for academic Web site interlinking: Evidence for the Web as a novel source of information on informal scholarly communication. J Inf Sci 2003; 29(1):49-56.
10. Case DO, Higgins GM. How can we investigate citation behavior? A study of reasons for citing literature in communication. J Am Soc Inf Sci 2000; 51(7): 635-645.
11. Cronin B. The citation process: The role and significance of citations in scientific communication. London: Taylor Graham; 1984.
12. Almind TC, Ingwersen P. Informetric analyses on the World Wide Web: Methodological approaches to "Webometrics". J Doc 1997; 53(4):404-426.
13. Smith AG. Web links as analogues of citations. Inf res [serial online] 2004 Jul [cited 2005 March 30]; 9(4):paper no. 188. URL: http://informationr.net/ir/9-4/paper188.html.
14. Thelwall M, Vaughan L, Björneborn L. Webometrics. An Rev Inf Sci Tech 2005: 39:81-135.
15. Glänzel W. On some principle differences between citations and sitation links: A methodological and mathematical approach. (PowerPoint slides of Nerdi Lecture delivered at NIWI, KNAW, Amsterdam on 2003 Feb 13).
16. Kim HJ. Motivations for hyperlinking in scholarly articles: A qualitative study. J Am Soc Inf Sci 2000; 51(10):887-899.
17. Thelwall M. The top 100 linked pages on UK university Web sites: high inlink counts are not usually directly associated with quality scholarly content. J Inf Sci 2002; 28(6):485-493.
18. Thelwall M, Harries G. The Connection between the research of a university and counts of links to its Web pages: An investigation based upon a classification of the relationships of pages to the research of the host university. J Am Soc Inf Sci 2003; 54(7):594-602.
19. Bar-Ilan J. Self-linking and self-linked rates of academic institutions on the Web. Scientometrics 2004; 59(1):29-41.
20. Cronin B, Snyder HW, Rosenbaum H, Martinson A, Callahan E. Invoked on the Web. J Am Soc Inf Sci 1998; 49(14): 1319-1328.

21. Cleverdon CW. Optimizing convenient online access to bibliographic databases. Inf Ser Use 1984; 4:37-47.
22. Cooper WS. Is interindexer consistency a hobgoblin? Am Doc 1969; 20:268-278.
23. Cothey V. Web-crawling reliability. J Am Soc Inf Sci 2004; 55(14):1228-1238.
24. Lawrence S, Giles CL. Accessibility of information on the Web. Nature 1999; 400:107-109.
25. Jepsen ET, Seiden P, Ingwersen P, Björneborn L, Borlund P. Characteristics of scientific Web publications: Preliminary data gathering and analysis. Web-crawling reliability. J Am Soc Inf Sci 2004; 55(14):1239-1249.