

Web links as analogues of citations

[Alastair G. Smith](#)

School of Information Management
Victoria University of Wellington, New Zealand

Abstract

This exploratory study investigates the extent to which Web links are analogues to the citations in traditional print literature. A classification of Web links is developed, using the nature of the source and target pages, and the reasons for linking. Links to a sample of research oriented Websites (universities, professional institutes, research institutes, electronic journals, and individual researchers) were classified. Overall, 20% of the Web links in the study could be regarded as research links analogous to citations.

Introduction

Citations in conventional print publications have traditionally been used as indicators of links between researchers, and it is tempting to regard Web links as analogues of citations. To what extent do hypertext links between Websites indicate research links? This exploratory study examines the nature of Web links made to research oriented Websites, to determine the extent to which the Web links are made for research-related reasons and are, therefore, analogous to citations.

One obvious difference is the nature of the documents that are linked. Citations in conventional print publications are generally between research publications, while Web links may be between a wide variety of publication types: personal home pages, subject resource guides, etc.

Why are citations and Web links made? Bibliometric literature contains many studies that examine the motivation behind citing. Egghe and Rousseau (1990) review a number of studies of citation motivation. Garfield (1964) identifies reasons that include paying homage, identifying methodology, providing background, correcting, criticising, substantiating or authenticating, alerting to forthcoming or poorly disseminated work, identifying original concepts, and disclaiming or disputing previous work. Kelland and Young (1998) list several motivations, including acknowledging prior work, identifying methodology, providing background reading, correcting or criticising, substantiating claims, alerting readers to forthcoming work, authenticating data, identifying original publication of a term or concept, disclaiming work of others, and disputing priority claims. Case and Higgins (2000) found that citations might be made because the cited work (1) was a 'concept-marker', (2) promoted the authority of the citing work or (3) deserved criticism. Oppenheim and Smith (2001) studied citations in students' dissertations and found that while their motivations for citing were similar to academics, there was a trend towards citing Internet sources.

Studies of linking motivation on the Web, on the other hand, are more recent, and indicate that Web links are made for different reasons. Kim (2000) studied links from Web based scholarly publications and found a wide range of motivations for linking, some echoing citations, but also including: publicity, credit to an institution, providing an immediate access mechanism, to provide a graphical image, and an editorial policy of encouraging hyperlinks. Thelwall (2003) studied 100 Web links between university home pages, and classified them into 'ownership',

'social', 'general navigational' and 'gratuitous'; arguing that the majority of Web linking motivations were trivial compared with citing motivation. Chu (2003) analysed links to academic Websites, and found that 50% of links were to resource or directory information, while only 27% were motivated by research or teaching/learning. Because of these wider reasons for Web linking, and the more eclectic range of documents that Web links are made between, motivations are different for Web links and citations. Wilkinson *et al.* (2003) found that less than 1% of links to university department Websites were formal research citations. Vaughan and Shaw (2003) studied bibliographic and Web citations to articles in LIS journals and found that many Web citations represented intellectual impact and that for most journals there was a correlation between Web citations and the Journal Impact Factor.

The increasing use of Web links in Webmetric research means that the nature of these links, and the extent to which they serve the same function as citations, is important. The current study undertook to classify links made to a sample of research oriented Websites, to test a trial classification and to estimate the extent to which the Web links were analogous to citations.

Methodology

Fifteen research-oriented sites were studied, taking three examples of each of five types of sites. These were:

- University Websites ([Victoria University of Wellington](#), [Australian National University](#), and [MIT](#))
- Professional institute Websites ([The Royal Society of New Zealand](#), [Institute of Professional Engineers of New Zealand](#), and the [New Zealand Law Society](#).)
- Research institute Websites ([National Institute of Water and Atmospheric Research](#), [Institute of Veterinary, Animal and Biomedical Sciences](#), and the [New Zealand Institute for Economic Research](#))
- Electronic Journals ([British Medical Journal](#), [Journal of Internet Law and Technology](#), and [AtP \(AntePodium\)](#))
- Individual researchers' Web pages (an NZ [artificial intelligence researcher](#), a US [library and information management researcher](#), and a US [mathematical statistician](#))

These sites were chosen as being representative of a range of ways in which research appears on the Web, and representing both academic and professional research. This approach to selection was appropriate for an exploratory study testing the classification and methodology.

Links to each of the target sites were determined by using the AltaVista command in advanced mode (<http://altavista.com/Web/adv>):

```
link:xxx and not host:xxx
```

where xxx is the target domain. This excluded Web links made within the site itself. Site collapse was off, and the search was set to be world wide, for documents in English (to avoid having to classify sites in languages the researcher was not familiar with). Every twentieth item on the list was examined, up to a total of ten linking sites, except where fewer than 200 linking sites were found, when every tenth site was examined. If a site or link was no longer valid, the next link on the results list was chosen. 150 links were studied in total. Searches were carried out in January 2003. As this was an exploratory study, classification was carried out by the researcher without independent verification of the classifications.

Links were classified on three aspects:

- Nature of source page, i.e., the page containing the link.
- Nature of target page, i.e., the page linked to.
- Reason for linking.

The classifications were derived from similar studies (for example Thelwall, 2003; Chu, 2003) but were oriented to identifying features that would indicate the research nature of the link. Table 1 shows the classifications used for the Source and Target pages.

Code	Type of page

1	General information resource, not otherwise classified
2	Research information
3	Teaching resource
4	Administrative information
5	Student assignment
6	Links list
6.1	<i>Bibliography/publication list</i>
6.2	<i>Directory/ subject guide</i>
6.3	<i>Related/useful links</i>
6.4	<i>Events list</i>
7	Discussion list archive, Blog
8	Formal publication
8.1	<i>Technical paper, report</i>
8.2	<i>E-journal article</i>
8.3	<i>Conference paper</i>
8.4	<i>News item</i>
8.5	<i>E-journal</i>
8.6	<i>Conference</i>
8.7	<i>News source</i>
9	Personal Web page
10	Main page of organisation
10.1	<i>Hosted or subsidiary organisation</i>
11	Software source
12	'About' page

Table 1: Page types

The first classes (1-5) are for general Web pages that act as information resources in research, teaching, and administration. Student Websites, created for assignments, are a significant group, and were given their own classification. A specifically Web phenomenon is the directory or links page (6) and these were subdivided into different types: a bibliography or publication list (generally referring to print publications), directory/subject guides (generally having some organised structure), related/useful links (generally an unstructured list that formed part of an individuals home page), or a list of events (usually including links to the events in question). Class 7 reflects the role of communication on the Web, with pages that are discussion lists, or Weblogs.

Class 8 was for Web pages that were formal publications analogous to print publications: e-journals, online conference papers, etc. This class distinguished between individual items (8.1-8.4) and home pages for the publications. The remaining pages were for uniquely Web documents: the personal Web page, the main page of an organisation (a subclass was created for subsidiary or hosted organisations), software sources (a common feature of pages in computer science, where software may be the 'output' of a research project), and the 'about' page, giving background information about a site – ownership, sponsorship, etc.

Reasons for linking were classified according to the schedule in Table 2:

--	--

Code	Reason for link
1	General information link
1.1	<i>Teaching/learning</i>
1.2	<i>Administrative</i>
1.3	<i>Research funding</i>
1.4	<i>Dissemination of Research</i>
1.5	<i>Employment</i>
2	Formal research citation
4	Sponsor/acknowledgement of support
5	Self link to more information about creator
6	Related pages
6.1	<i>Related individual</i>
6.2	<i>Related organisation</i>
7	Information about geographic area
8	Advertising
9	Software download

Table 2: Reasons for linking

Class 1 was for general informational reasons, subdivided into different types. Class 2 was for links where the reason was a formal research citation. Other classes were acknowledgements of various kinds: of support (4) or relationships (6). Providing contextual information was a reason: information about the creator (5) or the geographic area to which the page related (7). The advertising class (8) was used where there seemed to be a commercial motivation (including boosting an educational institution's enrolments). Dissemination of software (9) was a significant reason, particularly in the computer science area.

While there is some relationship between the classes for the reason for linking, and the classes for source and target pages, it is important that these are classified separately. In the case of a teaching resource page (source class 3) linking to a university Web page (target class 10), this could be for a variety of reasons: for teaching or learning (reason 1.1), because the university was a sponsor (reason 4) or because the creator of the page had a relationship with the university (reason 6.2).

Results

Table 3 indicates the percentages of source (i.e., the page the link was made from to the research oriented site) and target (i.e., the page on the research oriented site that was linked to) pages that were in each classification. More detailed frequency data are given in the [Appendix](#).

Code	Type of page	Source %	Target %
1	General information resource	9	5
2	Research information	4	2
3	Teaching resource	3	2
4	Administrative information	1	0
5	Student assignment	0	0
6	Links list	0	0
6.1	<i>Bibliography/publication list</i>	9	1

6.2	<i>Directory/subject guide</i>	48	11
6.3	<i>Related/useful links</i>	3	0
6.4	<i>Events list</i>	1	0
7	Discussion list archive, Weblog	2	0
8	Formal publication	0	0
8.1	<i>Report</i>	2	4
8.2	<i>E-journal article</i>	1	5
8.3	<i>Conference paper</i>	0	0
8.4	<i>News item</i>	1	1
8.5	<i>E-journal</i>	1	13
8.6	<i>Conference</i>	1	2
8.7	<i>News source</i>	4	3
9	Personal Web page	5	10
10	Main page	2	36
10.1	<i>Host/subsidiary</i>	1	3
11	Software source	0	1
12	'About' page	3	1

Table 3: Frequency of page types

The most common source page overall was the directory or subject guide (page type 6.2) with 72 links (48% of all links) made from these. This reinforces the role of directory and subject guide in facilitating surfing on the Web. However, bibliography and publication lists (page type 6.1, which might be regarded as the more traditional analogues of Web directories), as well as general information sources (page type 1), were significant sources of links, with 13 (9% of all links) each. Formal publications (page types 8.1-8.7, combining technical papers, e-journals, conferences, and news sources) were also significant, with 13 links (9% of all links).

The most common target page type is the main page of an organisation (page type 10, with 36% of all links). This type, along with subsidiary/hosted organisation, page type 10.1, personal Web pages, page type 9, and e-journal home pages, page type 8.5) could be regarded as a wider class of 'Entry points'. These were the targets of 67% of total links. Directory/Subject guides (page type 6.2), together with publication lists (page type 6.1), were the target of 12% of total links, indicating that pages that point to further information play an important role in the research oriented Web, and in attracting links. On the other hand, specific formal publications (page types 8.1-8.4) were the targets of only 10% of total links. Target pages that could be regarded as 'information content' (page types 1-4, 8.1-8.4, and 11) as opposed to entry or directional, accounted for 20% of total links.

Table 4 shows the percentage of links for each 'reason for linking'.

code	reason for link	%
1	General information	39
1.1	<i>teaching</i>	7
1.2	<i>administrative</i>	0
1.3	<i>Research funding</i>	1
1.4	<i>Dissemination of Research</i>	7
1.5	<i>Employment</i>	2
2	citation	10
4	sponsor	7
5	information about creator	1
6	Related pages	1

6.1	<i>Related individual</i>	6
6.2	<i>Related organisation</i>	17
7	geographic area	1
8	Advertising	0
9	Software download	1

Table 4: Frequency of reasons for linking

The most common reason for linking was the provision of further information to amplify the content of the source page (link types 1-1.5 and 5, 57% of total links). Relationship links, for example to related individuals or organisations, comprised 23% of total links. There were 15 links for which the reason was formally citing research (10% of total citations). Of those links which might be considered analogous to print citations, 9 were to e-journals, 4 to individual researchers sites, and 2 to other kinds of university sites.

The targets of the links to related organisations (link type 6.2) were largely professional organisations and research institutes (88% of this link type). Interestingly, there were no related organisation links made to general university sites – this may reinforce the view of universities as 'ivory towers'. Not surprisingly the bulk of 'teaching and learning' links were made to universities (7 out of 11 links of link type 1.1).

A significant number of target pages for the researchers Websites were directory subject guides (10 out of 30 were page type 6.2) indicating that researchers may perform a valuable function in providing subject guides on the Internet. Most of the target pages at e-journals were to the e-journal as a whole, rather than specific articles (7 were page type 8.2 while 19 were page type 8.5). This reinforces the view that Web links to e-journals are not exactly equivalent to print citations.

Discussion

The classification proved to be useful, although it has not been tested for consistency across different individuals acting as classifiers. The class 6.2 in the page types (directory and subject guides) proved to be broad, and could be broken down into overall subject guides (Yahoo! etc), specialised subject guides (e.g., SOSIG) and informal subject guides (researchers' personal lists of interesting sites). In the current study, any site listings that were organised in some way were classified as directory and subject guides, and this may have been too inclusive. Software sources (in this sample, only two links) were included as a classification since in computer science, software is often seen as a research output. However it would be useful to distinguish between software originating from the organisation, and software being redistributed.

To what extent were the Web links in this study analogous to citations, i.e., that they are citing a research source? A Web link might be regarded as being equivalent to a citation if the target page was a research source, or if the reason for linking was made for research type reasons.

Some classes of target pages that could be regarded as indicating substantive research links are:

- Research information (page type 2)
- Technical papers, reports, e-journal articles, and conference papers (page types 8.1-8.3)
- Software sources (page type 11)

13% of total Web links fell into these classes. Not surprisingly, e-journals were the target of almost half of these links: 5% of total links.

Classes of reasons for Web linking that that could be regarded as indicating research links are:

- Dissemination of research (link type 1.4)
- Formal research citation (link type 2)
- Software download (link type 9)

17% of total links were made for these reasons. While most of these were formal research citations, a significant

number were general hypertext links that had the object of disseminating research: indicating that perhaps new forms of citation are evolving on the Web.

Table 5 shows the number and proportion of Web links that satisfied one or other of the definitions for being analagous to citations, i.e., were research-oriented links (target page type 2,8.1-8.3,11; or link type 1.4, 2, or 9). Percentages are of the possible links, e.g., 8 of the 30 links to universities (27%) were analagous. Over all the types of sites, 30 of the 150 links (20%) were analagous. University sites and electronic journals had relatively high proportions of links made to them that were analagous to citations; while the professional institute and research institute sites had relatively low proportions. However the sample in this exploratory study is too small to put much weight on these differences.

	No.	%
Universities	8	27
Professional institutes	3	10
Research institutes	4	13
Electronic Journals	9	30
Researchers	6	20
All sites	30	20

Table 5: Research-oriented links

Conclusion

This exploratory study addresses the question of the extent to which Web links are analagous to conventional citations. It proposes a classification for Web links, based on classifying source and target pages, as well as the reason for linking. The classification method was applied to links made to a small sample of research-oriented sites. Not surprisingly, a high proportion of links were made from directory or subject guides. Formal publications (technical reports, e-journal articles, conference papers) were noticeable as targets of links, but less so as sources. Formal research citations only amounted to 10% of all citations. However if a broader definition of research oriented Web links is used, 20% of the links in this sample could be regarded as analagous to citations.

This study reinforces the view that the nature of Web links are more varied than print citations. However in the case of links made to these research-oriented sites, a sizeable minority were analagous to citations.

Note:

This article is based on a paper presented at ISSI 2003, 25-29 August 2003, Beijing. The author thanks the reviewers and participants of this conference for their useful comments.

References:

- Case, D.O., & Higgins, G.M. (2000). How can we investigate citation behaviour? A study of reasons for citing literature in communication. *Journal of the American Society for Information Science*, **51**(7), 635-645.
- Chu, H. (2003). *Reasons for sitation (hyperlinking): what do they imply for Webometric research?* Paper presented at the International Conference on Scientometrics and Informetrics, 9th. 25-29 August 2003, Beijing.
- Egghe, L.L., & Rousseau, R. (1990). Citations and citers' motivations. In, Leo Egghe & R. Rousseau, (Eds.) *Introduction to informetrics: quantitative methods in library, documentation, and information science* (pp. 211-227). Amsterdam: Elsevier Science Publishers.
- Garfield, E. (1965). Can citation indexing be automated? In, M.E. Stevens, V.E. Giuliano, & L.B. Heilprin (Eds.), *Statistical association methods for mechanized documentation, symposium proceedings, Washington, DC, 1964* (pp. 189-192). Washington, DC: National Bureau of Standards. (NBS Miscellaneous Publication 269).

10.1	Hosted or subsidiary organisation	0	0	1	3	0	0	0	0	1	1	2	4
11	Software source	0	2	0	0	0	0	0	0	0	0	0	2
12	'About' page	0	0	4	0	1	0	0	1	0	0	5	1

Frequency of page types

Code	Reason for link	Uni	Proff	RI	Ej	Reschr	Total
1	General information link	12	7	13	18	9	59
1.1	<i>Teaching/learning</i>	7	0	1	1	2	11
1.2	<i>Administrative</i>	0	0	0	0	0	0
1.3	<i>Research funding</i>	0	1	0	0	0	1
1.4	<i>Dissemination of Research</i>	4	2	2	0	2	10
1.5	<i>Employment</i>	0	2	1	0	0	3
2	Formal research citation	2	0	0	9	4	15
4	Sponsor/acknowledgement of support	3	3	4	0	1	11
5	Self link to more information about creator	0	0	0	0	2	2
6	Related pages	1	0	0	0	0	1
6.1	<i>Related individual</i>	0	0	0	0	9	9
6.2	<i>Related organisation</i>	0	14	9	2	1	26
7	Information about geographic area	0	1	0	0	0	1
8	Advertising	0	0	0	0	0	0
9	Software download	1	0	0	0	0	1

Frequency of link types

Find other papers on this subject.

How to cite this paper:

Smith, A.G. (2004) "Web links as analogues of citations." *Information Research*, 9(4) paper 188 [Available at <http://InformationR.net/ir/9-4/paper188.html>]

Articles citing this paper, [according to Google Scholar](#)

Check for citations, [using Google Scholar](#)

5808
Web Counter

© the author, 2004.
Last updated: 1 May, 2004



- [Contents](#) |
 - [Author index](#) |
 - [Subject index](#) |
 - [Search](#) |
 - [Home](#)
-