Motivations for URL Citations to Open Access Library and Information Science Articles

Kayvan Kousha[*] and Mike Thelwall[**]

[*]*kkoosha@ut.ac.ir*
Department of Library and Information Science, University of Tehran,
Jalal-Al-e-Ahmed, Tehran, Iran.

[**]*m.thelwall@wlv.ac.uk*
School of Computing and Information Technology, University of Wolverhampton, 35/49 Lichfield Street
Wolverhampton WV1 1EQ, UK

Abstract
We define URL citations as mentions of an URL in the text of a Web page, whether hyperlinked or not. The proportions of formal and informal scholarly motivations for creating URL citations to the Library and Information Science open access journal articles were identified. Five characteristics for each source of URL citations equivalent to formal citations were manually extracted and the relationship between Web and conventional citation counts at the e-journal level was examined. Results showed that 282 research articles published in the year 2000 in 15 peer-reviewed LIS open access journals were invoked by 3045 URL citations. Of these URL citations, 43% were created for formal scholarly reasons equivalent to traditional citation and 18% for informal scholarly reasons. Of the sources of URL citations, 82% were in English, 88% were full text papers and 58% were non-HTML documents. Of the URL citations, 60% were text URLs only and 40% were hyperlinked. About 50% of URL citations were created within one year after the publication of the cited e-article. A slight correlation was found between average numbers of URL citations and average numbers of ISI citations for the journals in 2000. Separating out the citing HTML and non-HTML documents showed that formal scholarly communication trends on the Web were mainly influenced by text URL citations from non-HTML documents.

Introduction

Open Access (OA) journals have rapidly become a global environment for scholarly communication and one of the platforms for publishing the scientific literature. We can now find a significant portion of the scientific literature appearing only in the peer-reviewed OA journals, although e-journal use varies by discipline. At the end of 1995, a survey of full-text, peer-reviewed journals in the areas of science, technology and medicine discovered over 100 online titles (Hitchcock, Carr, & Hall, 1996). The Web of Science, with approximately 8,700 of the highest impact research journals, covers nearly 200 OA journals (ISI press release, 2004), showing their gradual acceptance into the mainstream of research. In 2004 a study reported that there were 24,000 peer-reviewed research journals worldwide, but that only 5% (1,200 titles) were open access (Harnad et al., 2004). Currently (March 2005), the Directory of Open Access Journals indexes more than 1,500 full text and quality controlled scholarly journals, covering various subject areas (DOAJ, 2005).

The increase in open access journals indicates a new rapidly evolving publishing model. Maguire (2003) found that almost 90% of LIS professionals were willing to publish in peer-reviewed, open-access LIS journals and nearly 60% were eager to participate in building and maintaining such a journal. Today, in several science disciplines such as physics and computer science, the Web is often the first choice for authors to publish the results of current research, even before appearing in the non-OA journals. There is also evidence that the number of OA articles in established journals is increasing. Hawkins (2001) found that the number of articles in 28 LIS journals had risen from 26 in 1995 to 250 articles per year in 2001.

From the early 1990s, the importance and potential of OA publishing in scholarly communication has been widely discussed (e.g., Harnad, 1990; Harnad, 1991; Harter, 1996; Harnad, S., 1999), but only recently has strong evidence been found that OA journals and non-OA journals have similar citation impacts (ISI press release, 2004). Whilst research in this area continues to investigate the citation impact of OA journals in different disciplines (Brody et al., 2004), results of

Proceedings of ISSI 2005, pages 67–77.
Edited by P. Ingwersen and B. Larsen.

previous studies show that in some disciplines like computer science placing an article online can increase its citation impact (Lawrence, 2001).

In most related studies measuring the impact of OA journals, bibliometric techniques have been used (Borgman & Furner, 2002), for example to compare citation counts for OA articles with pay-to-access articles.

Although it is possible to use the "Cited Reference Search" facility in the ISI Web of Science to retrieve citations to an OA journal in the references of other journals indexed by the ISI, in the context of the Web, this method will not reveal the links equivalent to citations to OA articles (Web citation) that are not in ISI-indexed articles. In other words, the traditional citation analysis techniques are not necessarily the best measures to explore the impact of OA journals. In fact, there may be a significant portion of formal citations on the Web to the OA journals from other Web documents (such as preprints, e-archives, online dissertations, and research reports) which will never appear in ISI indexes. Moreover, other Web pages may target OA articles for informal scholarly reasons which will never be recorded in conventional citation databases. Academic staff, for instance, can link from their homepages to OA articles for class reading lists. Thus, it is interesting to use Web citation analysis techniques to investigate creation motivations for links to OA journals and trends for using them in formal/informal scholarly communication. The development of electronic publishing on the Web has therefore created the possibility for new measures, spawning the field of Webometrics (Thelwall, Vaughan and Björneborn, 2005).

The current study identifies and classifies apparent creation motivations for URL citations (i.e. hyperlinks or non-linked URLs in the text of a page) to 15 peer-viewed library and Information Science (LIS) OA journal articles published in the year 2000. It also determines the characteristics of sources of URL citations and investigates the relationship between Web and conventional citation counts at the individual journal level.

Related studies

Conventional Citation Impact of E-journals

Although the problems and possible meanings of citations have been debated (e.g., MacRoberts & MacRoberts, 1989), citation analysis is still a well-known and frequently used technique. Using citation analysis techniques, a recent study conducted by the Institute of Scientific Information (ISI) showed that there were no impact differences between the 191 OA journals and the 8,509 non-OA journals indexed by the ISI, (ISI press release, 2004). There is another ongoing study across all disciplines, using a 10-year sample of 14 million articles from the ISI database to present a more general view of citation impact of open access journals in different disciplines (Brody et al., 2004). Lawrence showed that free online availability substantially increases a paper's impact and that more highly cited articles and more recent articles in computer science are significantly more likely to be online. He found that in computer science citations were three times higher for open access articles than for papers only available for payment in print or online. Kurtz (2004) reached almost the same conclusion in the field of astrophysics. Shin (2003) found that the impact factor of journals in the field of psychology (over two periods, 1994–1995 and 2000–2001) increased when they became available in electronic form, indicating that the greater availability of the electronic format leads to more citations. Maguire (2003) found that approximately 50% of LIS professionals have cited an on-line, peer-reviewed, open access journal in the past.

Since a significant portion of scholarly OA journals were not indexed by the ISI in 2004 (only 200 of 1,500 refereed OA journals), an important question is whether their impact can be better measured based not only upon the conventional citation databases, but also on the Web environment.

Web Citation Impact of Online Journals

Whilst conventional citation analysis techniques can only reveal formal communication patterns, one interesting nature of Webometrics is its potential for applying the same theories of traditional bibliometrics analysis for exploring both formal and informal scholarly communication models on the Web. From this basis, since 1996, many articles have been written on Web links and their interesting nature for exploring a kind of scholarly communication (e.g., Almind & Ingwersen, 1997; Rousseau, 1997; Ingwersen, 1998; Borgman & Furner, 2002). Some of the above researchers have drawn an

analogy between citation and Web links. For instance, Rousseau (1997) applied the term "Sitation" to refer to a cited site, Ingwersen (1998) proposed "Web Impact Factor" as a Web counterpart of the ISI's Impact Factor; and Borgman & Furner (2002) claimed a strong analogy between "linking and citing".

Smith (1999) used Web citation analysis techniques for 22 Australasian refereed e-journals from a range of disciplines, finding no significant relationship between inlinks and ISI Impact Factors. He concluded that links to e-journals are different to citations because the former target the whole journal whereas the latter target individual articles. Smith did not use article inlink counts (using web site inlink counts instead) or qualitative methods (creation motivations for links to journals).

Harter & Ford (2000) studied 39 scholarly e-journals, also not related to a specific discipline. Links to journals and articles were compared with ISI citations and no significant correlation was found between link and ISI impact factors. The authors classified the link creation motivations for about 300 sampled inlinks to "e-articles" into 13 categories. Nevertheless, the selected journals were relatively small and not related to a specific discipline, which is problematic because of disciplinary differences in both citations and web use.

Vaughan & Hysen (2002) analyzed journals of Library and Information Science indexed by the ISI. The journals in their study were not open access journals but were traditional journals with independent Web sites. The study found a significant correlation between the number of external links and the journal impact factor for LIS journals.

Vaughan & Thelwall (2003) studied 88 Law and 38 Library and Information Science (LIS) ISI indexed journals. They found that journals with more online content tended to attract more links as did older journal Web sites.

Vaughan & Shaw (2003) compared citations to journal articles from the ISI index with citations to them in the Web. They found significant correlations, suggesting that online and offline citation impacts are in some way similar phenomena. A classification of 854 Web citations indicated that many "represented intellectual impact, coming from other papers posted on the Web (30%) or from class readings lists (12%)".

Vaughan & Shaw (in press) studied the number and type of Web citations to journal articles in four areas of science. Most of the journals in their study were not open access but were traditional ISI indexed journals with independent Web sites. On the individual paper level, a significant correlation was found between ISI and Web citations as well as a significant relationship between the Journal Impact Factor and the average number of Web citations a journal receives. They suggested that Web and ISI citation counts are measuring the same things in assessing the impact of journals or their papers. Thus, Web citation counts might potentially supplement or replace ISI citation counts as an impact measure.

Although most Webometrics studies have applied quantitative methods (correlation studies) and relatively little research directly explores link creation motivations, one exception is Kim's (2000) small study of motivations for hyperlinking in scholarly electronic articles. He found that in scholarly electronic environments scholars use hyperlinks for a variety of scholarly and non-scholarly purposes, and that hyperlinking is a multidimensional behaviour involving different levels of motivations.

Using CiteSeer, Goodrum, et al. (2001) analyzed citation patterns in online PostScript and PDF computer science papers. She found that conference papers were more frequently cited online in computer science.

In summary, whilst in 1999-2000 no significant correlations were claimed between links to journal Web sites or e-articles, more recent studies have found significant relationships between Web links and citations and between Web citations and traditional citations.

Research questions

We define URL citations as mentions of an URL in a Web page, whether hyperlinked or not. Five questions were addressed to investigate creation motivations and general characteristics of URL citations to LIS open access journal articles, and to examine the relationship between Web and conventional citation patterns at the individual journal level. The inclusion of non-hyperlinked URLs in our study, through our URL citation definition, is a novel approach compared to previous research, which either investigated links or traditional (i.e. non-URL) citations.

1. What proportions of motivations for URL citations to open access LIS journal articles are related to formal scholarly communication (equivalent to formal citation), informal scholarly activities and navigational?
2. What type of documents host URL citations on the Web?
3. What are the characteristics of the sources of the URL citations in terms of language (English/other languages), publication year (2000-2004), content level (full text/bibliographic), file format (PDF, HTML, DOC, etc.), and type (text URL / hyperlink citations)?
4. How are the overall results influenced by separating out HTML and non-HTML documents in terms of content level and Web citation type?
5. At the journal level, is there a correlation between ISI and URL citation counts or the average number of ISI and URL citations?

Methods

Journal and Article Selection

The Library and Information Science discipline was selected as a pilot study for much more comprehensive doctoral research on several science and social science disciplines (Kousha, 2004). For the purpose of this study, OA journals are free accessible English journals only available on the Web with articles that have undergone some kind of peer review or editorial process. An initial study based upon the Directory of Open Access Journals (www.doaj.org) and other directories showed that there were 25 open access electronic only LIS journals. Of these, 10 were excluded for the following reasons.

- Storage in a database, because commercial search engines would have technical problems finding links to them (see Thelwall, Vaughan and Björneborn, 2005).
- Ceasing publication before or publication beginning after the year 2000.
- The existence of a non-electronic version.
- The absence of refereed articles.

URL citations to the individual journals' articles in each issue from 2000 were examined. Note that journals which didn't have an independent Web site were included in this study, because links to whole journal Web sites were not needed. The study only covered the official Web sites of OA journals (the journal publisher's Web site) and mirror sites were not examined. For the 15 OA journal included in the study, all full-text research articles (omitting reports, editorials, book reviews, etc.) published in the year 2000 were selected, a total of 282. The year 2000 was chosen as the sample year to allow about 4 years for articles to be cited on the Web and in ISI journals. The titles and URLs (either to HTML or non-HTML versions of articles) of all 282 articles were recorded.

URL Citation and ISI Citation Counts

Using Google searches, all the URL citations to the 282 articles were retrieved within the same week during September 2004. URL citations to OA articles were examined and classified based on a pre-defined classification scheme. Google was chosen because results of previous studies showed that it provides the most comprehensive (Bar-Ilan, 2004) and the most stable search results over the time (Vaughan, 2004; Vaughan & Shaw, in press). Compared with other main search engines, Google has good coverage of HTML and non-HTML documents (for instance, PDF, DOC, PPT, XLS, PS and RTF). The following method was applied, as shown below for an article from *Cybermetrics*, which matches (1) hyperlinks to the article if the URL appears in the link anchor, and (2) inclusions of the URL in the page, even if it is not hyperlinked. Thus, it retrieves precisely what we have defined to be URL citations.

Article title: LOTKA: A program to fit a power law distribution to observed frequency data
URL of the html file: http://www.cindoc.csic.es/cybermetrics/articles/v4i1p4.html

Google search example: www.cindoc.csic.es/cybermetrics/articles/v4i1p4.html/ **-site:**www.cindoc.csic.es

It was necessary to use *–site:* after the URL of an e-article, in order to exclude links from the same domain (www.cindoc.csic.es/ in the above example), many of which will be for navigational purposes. Note that our syntax does not retrieve URL citations from the articles in the same journal (journal self citation) because they are hosted by the same domain.

This method of data collection is intentionally different to using the *link:* command, as used in previous Webometrics research. However, compared with the title search method of other studies (Vaughan & Shaw, 2003; Vaughan & Shaw, in press), it also has both limitations and advantages in coverage. Since it does not return links unless the URL is also in the text of the links, it excludes links where the URL is not explicitly mentioned, for instance a link which is only embedded in title of an article in hypertext format. Nevertheless, it seems that in most formal citation styles the URL of the cited online article appears in the text of the links, hence the method has the potential to identify formal scholarly communication. Perhaps more importantly, there are Web pages with text URLs targeting online articles (without mentioning the title of articles), for instance from e-mail, discussion groups and e-archives. No previously used data collection method would have included these.

The names of each OA journal was searched for in the "Cited Reference Search" field in the ISI Web of Science to find the possible number of citations received in the year 2000. ISI searches were carried out in September 2004. Since few selected OA journals were indexed in the ISI (only *Information Research*); the aim was to find the number of citations to them in the reference sections of traditional journals indexed by the ISI.

A limitation in using this method related to different names (abbreviations) of cited sources in the ISI databases entered in the citation information. For example, four different abbreviations (J DIG INFO, J DIG INFORM, J DIGITAL INFORMATIO, and J DIGITIAL INFORMATI) for the "*Journal of Digital Information*" were used. Consequently, using truncation other possible abbreviations were searched and through a manual checking process unrelated names or abbreviations were excluded. In order to prevent possible similarity between abbreviations for different journals in different fields, the volume and issue of each retrieved article was checked against the original OA article.

Classification of URL Citations

All URLs were manually checked and based upon the initial classification scheme all creation motivations were classified into four broad categories and 15 sub-classes, as shown below:

URLs for formal scholarly reasons equivalent to citation

This type of URL creation motivation was attributed to formal citations in the reference sections or footnotes of the other documents on the Web, either in the text or hypertext format if the citing document was one of the following:

- Journals article/online paper;
- Conference/workshop paper;
- Research/project report;
- Thesis/dissertation;
- Book/book chapter
- Conference/workshop presentation slides (as a reference)

URLs for informal scholarly reasons

The following URL sources were characterized as relating to informal scholarly activities (For example, URLs in reading lists for an academic course).

- Class reading list/course syllabus
- Author's CV; or departmental Web site
- E-mail/discussion group e-archive/email alerts/current contents
- Records in online bibliographic database
- Annotated online bibliography

URLs for navigational/gratuitous reasons

- URLs in Web directories, subject indexes and "selected articles" pages (for example URLs from the "Metadata" sub-category in the Open Directory (www.dmoz.org) to e-articles in the same subject area. Although such URLs may be important and useful, they are not created by scholars, as far as we know, and are neutral with respect to scholarly communication in the way that a library is.
- URLs in mirror sites

Others

- Missing pages (not found or inaccessible at the time of this study)
- Not clear (the URL citation is missing from the page)

A one person inductive content analysis methodology was chosen for manual assignment of all URL creation motivations. But, two LIS PhD students were consulted for the initial classification of a sample of 100 URLs and for possible adding or modifying of predefined categories. The consistency between the two classifiers (based on the initial scheme) was 81%. The results confirm that classification of links/URLs motivations is a subjective issue, although improving on the agreement of Wilkinson et al. (2003). However, the major purpose of this study was to discover and identify types of "apparent" motivations for URL creation equivalent to formal citation to gain some evidence of formal scholarly communication on the Web. The initial results showed that there was no disagreement on identification and classification of motivations for formal URL citation (39 of 100 Web links) between the two classifiers. Most disagreement concerned motivations for creating URLs for navigational reasons and other categories. The use of one person's perception and interpretation of URL creation motivations (for all 1,313 URLs) is therefore the main limitation of the current study.

Exploring Source Characteristics for Formal URL Citations

One of the key questions of this study was related to the characteristics of the sources of formal URL citations, those equivalent to formal citation. Five characteristics for each source of citation were manually extracted and recorded, including:

1 *Language* (English or other languages). What is the predominant language of formal scholarly communication on the Web?
2 *Publication year* (2000-2004). How long did it take for an OA article to be formally cited on the Web?
3 *File format* (PDF, HTML, DOC, PostScript, and etc). What is the predominant file format of URL citation sources?
4 *Content level* (full text or bibliographic). What is the content level of the majority of URL citation sources?
5 *Type of URL citation* (text URL or hyperlink citation). How are URLs in the reference section or footnotes of citing sources typically displayed, in text or hypertext format?

Findings

URL Creation Motivations

The results of the URL citation motivation study are summarized in Table 1. It shows that 282 articles published in 2000 in 15 OA LIS journals have been targeted by 3045 URLs during the time of this study. As shown in Table 1, 43% of URLs (1313 URLs) were related to formal scholarly communication equivalent to citation, 18% of URLs (547 links) were created for informal scholarly reasons, 33% of URLs (995 URLs) were created for navigational purposes and 6% of the URLs (190 URLs) for other reasons. Table 1 also shows that the most formal and informal motivations for creating URL citations to OA journals respectively related to "journals and online papers" (20%) and "mailing lists/discussion groups" (5.6%).URL citations from Web directories, subject indexes and selected articles pages (32%) as well as those from journal mirror sites (0.7%) were categorised as navigational.

Table 1. Classification of URL citation creation motivations
(3045 OA LIS articles from the year 2000)

Broad creation reason	%	Sub class	Number of URLs	%
Formal Scholarly Communication (equivalent to formal citation)	**43% (1313 URLs)**	Journal/online paper	**620**	**20**
		Conference/workshop paper	302	9.9
		Project report	213	7
		Conference/ workshop presentation slides	74	2.4
		Book/chapter	58	1.9
		Thesis	46	1.5
Informal Scholarly Communication	**18% (547 URLs)**	Bibliography	54	1.8
		Database	99	3.3
		Author's CV	65	2.1
		Syllabus/readings	159	5.2
		Mailing list/discussion group	**170**	**5.6**
Navigational	**33% (995 URLs)**	Web Directory	973	32
		Mirror site	22	0.7
Others	**6% (190 URLs)**	Not clear	145	4.8
		Not found	45	1.5
Total	**100%**		**3045**	**100%**

Characteristics of URL Citation Sources

Five characteristics of sources of formal URL citations were manually examined, including, the language, publication year, content level, file format, and type. The results are summarized in Table 2. Of the 1313 URL citations equivalent to formal scholarly communication, 74 were from conference/workshop presentation slides in Power Point format to OA articles. Although this kind of link creation motivation can be considered as an implicit way to cite OA articles, for instance to present background information about the research, little is known about such citations, and their creation motivations may differ from those of journal articles. For the purpose of this study, URLs from conference/workshop presentations slides to OA articles were excluded to present a more explicit picture of characteristics of formal URL citations, (1313-74= 1239 URL citations).

Table 2. Characteristics of sources of URL citations to OA LIS articles (2000)

Characteristics of Sources of URL citations	Classification of Characteristics	Number of URL Citations	%
Language	**English**	1010	**82%**
	Other	229	18%
Content Level	**Full Text**	1096	**87.5%**
	Bibliographic	143	12.5%
Type of Web Citation	**Text**	746	**60%**
	Hypertext	493	40%
Publication Year	**2000**	226	18.2%
	2001	377	**30.4%**
	2002	317	25.6%
	2003	241	19.5%
	2004	78	6.35%
File Format	**PDF**	611	**49.31**
	HTML	514	41.49
	DOC	100	8.07
	RTF	9	0.73
	PS	5	0.40

About 82% of sources of URL citations were in English; 88% were from the full text documents and 12% from references of papers with bibliographic information. Manual checking of URLs in the reference sections/footnotes of citing sources showed that about 60% were in text format and 40% were hyperlinked. As shown in Table 2, about half of the sources of URL citations were published during 2000-2001. This shows the rapid impact of LIS OA journals in receiving the majority of

citations within about one year after their publication on the Web. The classification of file formats of URL citations indicated that about 59% of URL citations were non-HTML and 41% were HTML (Table 2).

Table 3 and Table 4 separate the HTML and non-HTML citing sources. Table 3 shows that of 746 (60.2%) of sources of citations with text URL citations to OA articles, 524 (42.3%) were in PDF format. It is interesting that of 493 (39.8%) of sources of URL citation with hyperlink citations, 344 (27.8%) were HTML documents and only 149 (12%) were non-HTML documents. It shows that the overall results were mainly influenced by text URL citations from non-HTML documents.

Table 3. Separating out the HTML and non-HTML URL citation sources

Characteristics of sources	PDF	HTML	DOC	Other	Total
Sources with text URL citation	**524 (42.3%)**	170 (13.7%)	39 (3.1%)	13 (1.0%)	746 (60.2%)
Sources with hyperlink URL citation	87 (7.0%)	**344 (27.8%)**	61 (4.9%)	1 (0.1%)	493 (39.8%)
Total	611 (49.3%)	514 (41.5%)	100 (8.1%)	14 (1.1%)	1239 (100%)

Table 4 shows that of 1084 (87.5%) full-text citing sources, 711 (57.4%) were from non-HTML sources and 373 (30.1%) from HTML documents; non-HTML documents are more important for creating citation networks among full-text documents and less significant for bibliographic documents.

Table 4. Separating out the HTML and non-HTML URL citation sources in terms of full-text / bibliographic documents

Characteristics of sources	PDF	HTML	DOC	Other	Total
Full text sources	**600 (48.4%)**	373 (30.1%)	97 (7.8%)	14 (1.1%)	1084 (87.5%)
Bibliographic sources	11 (0.9%)	**141 (11.4%)**	3 (0.2%)	0 (0%)	155 (12.5%)
Total	611 (49.3%)	514 (41.5%)	100 (8.1%)	14 (1.1%)	1239 (100%)

Correlation between ISI and URL Citations

Table 5 shows the number of OA research articles published by each OA journal in the year 2000, URL and ISI citation counts to them, and the average number of ISI and URL citations for journals in 2000 (the total number of ISI/URL citations an OA journal received for the year 2000, divided by the number of papers in that journal in the same year). The number of times an OA journal has been cited by the journals indexed by ISI between 1997 and September 2004 is shown in the last column of Table 5. It is clear that the URL citation counts (1,313 formal citations) are much higher than ISI citation counts (280 formal citations). Correlation tests between ISI citation counts and URL citation counts were preformed in order to study the relationship between the two variables. The Spearman correlation coefficient test was used because frequency distributions for data were found to be skewed.

Results showed that there was a slight statistically significant correlation between ISI and URL citation counts ($r=0.592$, significant at the 0.05 level). It is interesting that a higher correlation ($r=0.681$, significant at the 0.01 level) was found between URL citations counts to OA journals in 2000 and ISI citations to them during 1997 and September 2004. The results indicate that LIS OA journals receiving many URL citations also receive high numbers of ISI citations. The relationship between average numbers of URL citations (Web Impact Factors) and average numbers of ISI citations (ISI Impact Factors) showed a correlation (Spearman) between the two variables ($r=0.586$, significant at the 0.05 level).

Table 5. URL and ISI citation counts and Web and ISI Impact Factors

OA journal title	Number of OA research articles in 2000	URL citation to OA articles in 2000	ISI citation to journals in 2000	ISI citation to journals 1997- 2004	Average number of URL citations for journals in 2000	Average number of ISI citations for journals in 2000
D-Lib Magazine	47	695	132	1168	14.8	2.8
Ariadne	19	246	16	202	12.9	0.8
Cybermetrics	4	27	28	186	6.8	7.0
Journal of Electronic Publishing	26	128	27	196	4.9	1.0
Journal of Digital Information	10	28	2	112	2.8	0.2
Journal of Information, Law, and Technology	25	59	0	5	2.4	0.0
Information Research	16	37	40	197	2.3	2.5
First Monday	76	78	28	97	1.0	0.4
Information Technology and Disabilities	10	5	0	4	0.5	0.0
LIBRES	3	1	0	11	0.3	0.0
Journal of academic media librarianship	3	1	0	1	0.3	0.0
Issues in Science & Technology Librarianship	18	5	0	1	0.3	0.0
Jour. of Southern Academic & Special Librarianship	8	2	0	0	0.3	0.0
School Library Media Research	7	1	2	9	0.1	0.3
Library Philosophy and Practice	10	0	5	15	0.0	0.5
Total	**282**	**1313**	**280**	**2204**		

Journals ranked based on their average number of URL citations

Discussion and Conclusion

One interesting question is related to the difference between the proportions of URL creation motivations for LIS open access and LIS journal Web sites indexed by the ISI. Comparing the results of Vaughan & Shaw (2003) with current research showed that while about 30% of the Web citations to LIS ISI journals were created for formal citation reasons (online papers), 43% of URL citations to OA LIS journals were related to formal scholarly communications equivalent to citation. Although the two studies used different methodologies for data collection and sample years[*], both used Google searches and focused on intellectual impact of citations in the field of LIS.

The results show that OA LIS research papers have higher average citation impact than papers only available for payment in print or online. It seems that free online availability substantially increases a paper's impact as tested by Lawrence (2001) for computer science. Of sources of URL citations, 49% targeting OA articles in 2000 were published during 2000-2001 indicating that during 2000-2001, OA articles have received about half of the formal citations on the Web. Studying the distribution of URL citations during 2000-2004 (Table 2) showed that the majority of sources of citations were published in 2000 (30.4%) and number of citations decrease in the subsequent years.

Of sources of URL citations, 59% were in non-HTML and 41% in HTML indicating that non-HTML documents, especially in PDF format, are the predominant format for scholarly communication on the Web for this study. Thus, search engines that don't index non-HTML documents (especially PDF files) would likely be inappropriate for scientific data mining and comprehensive study of scholarly communication trends on the Web. More study of the file format and other characteristics of sources of Web citations could be useful for design and development of scholarly search tools for locating and ranking the OA documents on the Web. For this, Google Scholar (http://scholar.google.com) could help; it crawls many scholarly publishers' archives and preprint

[*] In this study URL of OA articles published in 2000 were searched in Google (described in the Methods section) and Vaughan & Shaw (2003) used title search strategy in Google for LIS ISI journals in 1997.

servers and uses networks of citing and cited references based upon the link structure among OA documents on the Web.

Of citing addresses to OA articles, 60% in the reference sections of Web documents were in text (text URL citation) and 40% in hypertext (hyperlinked URL citation) showing that using only link command search for locating the sources of Web citations is not a comprehensive method for studying trends of scholarly communications on the Web

Separating out the HTML and non-HTML in terms of content level and type of URL citation (text and hyperlinked URL) showed that overall results were influenced by non-HTML documents, especially PDF files. It seems that the methodology which used in this study to collect citations had a significant influence on the total results. Text URLs, for instance, were more commonly used in non-HTML documents for targeting OA LIS articles. In fact, text URLs, especially in non-HTML sources, are more influential than hyperlinks for creating citation network on the Web for this study.

Acknowledgments

Liwen Vaughan for her helpful comments on drafts of this paper.

References

Almind, T. C. & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: Methodological approaches to "Webometrics". Journal of Documentation, 53(4), 404-426.

Bar-Ilan, J. (2004). The use of Web search engines in information science research. Annual Review of Information Science and Technology, 38, 231-288.

Borgman, C. & Furner, J. (2002). Scholarly communication and bibliometrics. Annual Review of Information Science and Technology, 36, Medford, NJ: Information Today Inc., pp. 3-72.

Brody, T., Stamerjohanns, H., Vallières, F., Harnad, S., Gingras,Y. & Oppenheim, C. (2004). The effect of open access on citation impact. Retrieved November 13, 2001, from http://www.ecs.soton.ac.uk/~harnad/Temp/OA-TAadvantage.pdf/

DOAJ (Directory of Open Access Journals). Retrieved March 30, 2005, from http://www.doaj.org/

Goodrum, A.A., McCain, K.W., Lawrence, S. & Giles, C.L. (2001). Scholarly publishing in the Internet age: a citation analysis of computer science literature. Information Processing & Management, 37(5), 661-676.

Harnad, S. (1990). Scholarly Skywriting and the Prepublication Continuum of Scientific Inquiry. Psychological Science 1: 342 – 343, Retrieved November, 12, 2004, from http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad90.skywriting.html

Harnad, S. (1991). Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge. Public-Access Computer Systems Review, 2 (1): 39 - 53. Retrieved November 12, 2004 from http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad91.postgutenberg.html/

Harnad, S. (1999). The Future of Scholarly Skywriting, in i in the Sky: Visions of the information future. Retrieved November, 12, 2004, from http://cogprints.org/1698/00/harnad99.aslib.html/

Harnad, S., Brody, T., Vallieres, F., Carr, L., Hitchcock, S., Gingras, Y, Oppenheim, C., Stamerjohanns, H., & Hilf, E. (2004). The access/impact problem and the green and gold roads to open access. Serials Review 30. Retrieved November, 12, 2004, from http://www.nature.com/nature/focus/accessdebate/21.html/

Harter, S. P. (1996). The impact of electronic journals on scholarly communication: A citation analysis. The Public-Access Computer Systems Review, 7. Retrieved November 13, 2001, from http://info.lib.uh.edu/pr/v7/n5/hart7n5.html/

Harter, S. & Ford, C. (2000). Web-based analysis of E-journal impact: Approaches, problems, and issues, Journal of the American Society for Information Science, 51(13), 1159-76.

Hawkins, D. T. (2001). Bibliometrics of electronic journals in information science. Information Research, 7(1), Retrieved November, 12, 2004, from http://InformationR.net/ir/7-1/paper120.html

Hitchcock, S., Carr, L., & Hall, W. (1996). A survey of STM online journals 1990–95: The calm before the storm. Retrieved November 20, 2004, from http://journals.ecs.soton.ac.uk/survey/survey.html/

Ingwersen, P. (1998). The calculation of Web Impact Factors. Journal of Documentation, 54(2), 236-243.

ISI press release essay on the impact of open access journals: A citation study from Thomson ISI. Retrieved November 13, 2004, from http://www.isinet.com/oaj/

Kim, H.J. (2000). Motivations for hyperlinking in scholarly electronic articles: A qualitative study. Journal of the American Society for Information Science, 51(10), 887-899.

Kousha, K. (2004). Motivations for Linking to Open Access Scholarly E-Journals: A Comparison of the Characteristics of Web Citations in Four Science and Social Science Disciplines. Approved PhD proposal, University of Tehran, LIS Department.

Kurtz, M.J. (2004). Restrictive access policies cut readership of electronic research journal articles by a factor of two, Harvard-Smithsonian Centre for Astrophysics, Cambridge, MA, Retrieved November 13, 2001, from http://opcit.eprints.org/feb19oa/kurtz.pdf/

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. Nature, 411, 521. Retrieved November 13, 2001, from http://www.nature.com/nature/debates/e-access/Articles/lawrence.html/

MacRoberts, M.H., & MacRoberts, B.R. (1989). Problems of citation analysis: A critical review. Journal of the American Society for Information Science, 40(5), 342–349.

Maguire, C. F. (2003). Library and information science (LIS) professionals' receptivity to an on-line, peer-reviewed, open-access LIS journal. Retrieved August, 20, 2004, from http://www.dankstreet.com/Downloads+index-req-getit-lid-34.html/

Rousseau, R. (1997). Sitations: An exploratory study. Cybermetrics, 1(1), Retrieved November 14, 2001, from http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html/

Shin, E.-J. (2003). Do Impact Factors change with a change of medium? A comparison of Impact Factors when publication is by paper and through parallel publishing. Journal of Information Science, 29(6), 527 - 533.

Smith, A.G. (1999). A tale of two Web spaces: Comparing sites using Web impact factors. Journal of Documentation, 55(5), 577-592.

Thelwall, M., Vaughan, L., & Björneborn, L. (2005). Webometrics. Annual Review of Information Science and Technology, 39, Medford, NJ: Information Today Inc. 81-135.

Vaughan, L. (2004). New measurements for search engine evaluation proposed and tested. Information Processing & Management, 40(4), 677-691.

Vaughan, L. & Hysen, K. (2002). Relationship between links to journal Web sites and Impact Factors. Aslib Proceedings: New Information Perspectives, 54(6), 356-361.

Vaughan, L. & Shaw, D. (2003). Bibliographic and Web citations: What is the difference? Journal of the American Society for Information Science and Technology, 54(4), 1313-1324.

Vaughan, L. & 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, 54(1), 29-38.

Vaughan, L. & Shaw, D. (in press). Measuring journal impact with Web citations: A comparison of four science disciplines. Journal of the American Society for Information Science and Technology.

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