Jointly published by Akadémiai Kiadó, Budapest
and Kluwer Academic Publishers, Dordrecht

Scientometrics,
Vol. 59, No. 1 (2004) 29–41

Self-linking and self-linked rates of academic institutions on the Web

JUDIT BAR-ILAN

School of Library, Archive and Information Studies, The Hebrew University of Jerusalem (Israel)

In this paper we introduce two measures *self-linked* and *self-linking* that are the analogues of self-citing and self-cited rates for scientific journals. These rates are calculated for a sample of sites to assess their meaning and utility. Self-linked is the more meaningful measure for the sample sites. As a first step towards a better understanding of self-linking (linking within a site), a sample of pages from an academic site was characterized using the method of content analysis. Even though most of the links serve navigational or other technical purposes, the percentage of content-bearing links among the self-links is significant, and even the portion of research oriented links is non-negligible.

Introduction

One of the major features of the World Wide Web is the capability to create links between pages. These links are often viewed as analogues of citations in scientific publications (e.g. ROUSSEAU, 1997; BJORNEBORN & INGWERSEN, 2001; CRONIN, 2001). Links can lead to a different site or may lead to a page in the same site. Links for which the source and the target of the page are within the same site or domain are called *self-links*. Instead of counting self-links, we count *self-sitations* – the number of pages on the site having a link or links pointing to the same site. The difference between self-links and self-sitations is that for self-sitations a page with several self-links is counted with multiplicity one. The reason for counting self-sitations instead of self-links is technical: commercial search engines report the number of pages with self links and not the number of self-links. The term *sitation* has been coined by ROUSSEAU (1997). In this paper we develop the measures *self-linking* and *self-linked*, which correspond to the classical measures for journals: self-citing and self-cited. We calculate these rates for a number of academic institutions on the Web. Self-links seems to be more problematic than self-citations (of authors in scientific publications), because self-links are often inserted for navigational purposes. In order to learn more about self-links,

Received August 28, 2003

Address for correspondence:
JUDIT BAR-ILAN
School of Library, Archive and Information Studies, The Hebrew University of Jerusalem
P.O.Box 1255, Jerusalem, 91904, Israel
E-mail: judit@cc.huji.ac.il

0138–9130/2004/US $ 20.00

we downloaded all the self-sitations of a relatively small academic site and characterized a random sample of these pages. The results show that the percentage of content-bearing links is significant, and even the portion of research oriented links is non-negligible.

Background and literature review

One of the more widely studied aspects of citations in scientific publications is self-citations, either author or journal self-citations. Author self-citation means that the citing paper and the cited paper have one or more authors in common, while journal self-citations are references to articles published in the same journal in which the citing article appears (EGGHE & ROUSSEAU, 1990, p 220).

One of the earliest studies on author self-citation was carried out by TAGLIACOZZO (1977). Her aim was to determine the distinguishing features of self-citation based on a group of articles related to plant physiology and neurobiology. BONZI & SNYDER (1991) surveyed 51 authors in several natural science disciplines for reasons for citing and self-citing and found that the motivations for citing and self-citing were rather similar. In another study, SNYDER & BONZI (1998) found that there is no significant difference between the exposure in the text given to self-citations and citation to others. PERITZ & BAR-ILAN (2002) among other measures, calculated the author self-citation rates for papers published by the journal *Scientometrics* in the calendar years 1990 and 2000. They measured the synchronous author self-citation rates (the number of references in a citing paper with self-citations, (LAWANI, 1982)) and found that the average self-citation rate per paper (number of self-citation divided by the total number of citations of the given paper) was around 14%, with a few papers where the self-citation rate was above 50%. WHITE (2001) examined among other measures the self-citation rates of eight prominent information scientists, and reported that the self-citation rate (number of self-citations divided by the total number of references in the author's publications) ranged between 3% and 8%. CRONIN & SHAW (2001) studied the citation images of another three prominent information scientists, and found that the self-citation rates in this case were between 8% and 16%. WHITE (2000) reported that the self-citation rate of Eugene Garfield was around 23%, a finding supported by earlier reports of BAIRD & OPPENHEIM (1994).

The reasons for journal self-citation have been less studied than author self-citation. GARFIELD (1974) discusses the difference between two measures related to journal self-citation: self-citing and self-cited rates. The self-citing rate is the percentage of journal self-citations divided by the total number of citations (references) that appeared in the

journal during a given period of time. The self-cited rate is the percentage of journal self-citations divided by the total number of citations the journal received during a given period of time. The JCR (ISI's *Journal Citation Reports*) online-help (2000) clearly explains the steps of these calculations. ROUSSEAU (1999) states that high self-cited rate is a sign of the journal's low visibility (it receives most of its citations from itself – authors in other journals are not aware of its existence); while high self-citing rates may be a sign isolation of the field covered by the journal (the articles appearing in the journal rarely refer to other journals – other journals are not highly relevant to the specific field). Low visibility and isolation are not the only reasons for journal self-citations; the journal itself may encourage this practice (see, for example, the instructions for authors of *The Information Society* journal: "It helps if you relate your article to recent relevant articles published in TIS" – retrieved February 3, 2003 from http://www.slis.indiana.edu/TIS/contributors/authors.html) or the prospective authors may feel that the chances of their paper being accepted to the specific journal may increase by adding some references to the journal the paper is being submitted to. NISONGER (2000) studied the effect of journal self-citations on the ranking of journals in the areas of information and library science, and genetics. The results showed that the removal of journal self-citations had little effect on journal rankings, thus he concluded that it was not necessary to exclude journal self-citations for serials management decisions based on the JCR.

On the Web one considers link counts instead of citation counts. INGWERSEN (1998) introduced the Web Impact Factor (WIF). The WIF is the number of pages with links pointing to a site divided by the number of pages in that site. Note that for the WIF the creation date of the pages is not limited to a certain time period (as in the case of the regular Impact Factor). An additional difference is that not the actual links are counted but the number of pages with links to the site. A page may contain a large number of links to the site, still it is counted with multiplicity one. For the regular Impact Factor, we count the number of citations and not the number of papers containing citations to the given paper. The reason for defining the WIF this way, is technical, this is the data provided by the general search engines (e.g. AltaVista and Fast) – they do not count the number of links to a site, but the number of pages with a link or links to the given site. For the search engines, the page is the unit of information. Ingwersen introduced several variants of the WIF including Self-link and External-link WIFs (counting only links from the site itself and counting only links where the source page does not reside on the site, respectively). SMITH (1999) and THELWALL (2000) pointed out technical

difficulties with the WIF calculations due to unequal coverage of the search engine AltaVista used for WIF calculations. Link counts produced by search engines are not always reliable as demonstrated by BAR-ILAN (2002).

Links are not exact analogues of citations in scientific publications. We have already pointed out the extensive use of hypertext links (especially self-links) for navigational purposes. EGGHE (2000) discusses some differences between hypertext links and citations. A citation always points from a later work to an earlier one, and this is not the case for links on the Web. He also raises the problem of counting the number of pages having links to a certain site, instead of the actual number of links. The parallel of this would be that if a paper refers to several different publications that appeared in a given journal, for example in this paper there are four references to different papers published in *Scientometrics*, for the calculation of the Impact Factor this would count as a single citation, exactly as for *College and Research Libraries* which is only referenced once in the current paper. Egghe, however, acknowledges that currently there are no better ways to count linkage. BJORNEBORN & INGWERSEN (2001) also point out there are often links in both directions between Web pages, "a situation not possible in traditional paper-based citation world". We cannot say that such situation is entirely impossible, rather it occurs very rarely for citations of scientific publications (only if and when authors of both papers are aware of ongoing work of the other party). An additional point raised by Bjorneborn and Ingwersen is that there is no convention for linking as there is for citing in the scientific world. According to PRIME et al. (2002) mirrors and duplicates are very problematic when applying traditional citation and co-citation analysis techniques to Web pages. The unit of analysis (site, directory or page) is not as clear-cut either as for classical citation studies.

Calculation of Web Impact Factors was not the only case where bibliometric methods were applied to study Web linkage. LARSON (1996) applied techniques of classical cocitation analysis in order to examine the intellectual structure of the Web. ROUSSEAU (1997) showed that for the query "bibliometrics OR informetrics OR scientometrics" the distribution of links between Web sites followed a Lotka function. THELWALL studied the linkage between UK universities in a series of papers (2001, 2002, 2003). In the 2001 paper he shows that a variant of the original WIF: "the ratio of Web pages with links pointing at research-based pages to faculty members" was the measure that best correlated with research rankings. Thelwall considered external links only "because of the large impact that HTML design issues can have on internal links". In the 2002 paper Thelwall demonstrated the existence of geographic trends in university Web site interlinking, while in the 2003 paper he analyzed the linkage to the top 100 linked-to pages on UK university Web sites. THOMAS & WILLETT (2000)

carried out a webometric analysis of departments of librarianship and information science in the UK, while CHU (2001) studied the linkage to some ALA accredited LIS schools' websites. Linkage to journal sites was analyzed as well: BAR-ILAN (2002) carried out a content analysis of pages linking to the homepage of the e-journal Cybermetrics, and VAUGHAN & THELWALL (2003) studied the linkage to the homepages of a large number of journals in library and information science and in law.

The new measures

In this paper we introduce two new measures, self-linked and self-linking rates which are the Web analogues of the classical self-citing and self-cited rates.

Before we present the formal definitions, let us clarify how we count self-sitations. A self-sitation is a page in which there is a link or there are links that point to the same site on which the page resides. Thus it may be the case that from a single page there are several links to different pages or objects that reside on the same site, still we count this as a single self-sitation. The reason for this definition is technical; this is the data the general search engines can provide. Note that the linked-to object is not necessarily a textual page; it can be a binary resource (sound, image, a downloadable application, etc.) or an email address (using the "mailto" formalism). We have not defined what is meant by "site" on purpose, this has to be stated specifically each time a calculation is carried out: this could be a top-level domain, a country code, the domain for a specific institution or entity (e.g harvard.edu or cnn.com) or a sub-domain (e.g. slis.indiana.edu).

The self-linked and self-linking rates are based on the number of pages with links to/from the site and not on the actual link counts:

$$\text{self-linked} = \frac{\#\text{ self-sitation}}{\#\text{ pages with links pointing to site}},$$

$$\text{self-linking} = \frac{\#\text{ self-sitation}}{\#\text{ pages with links of any kind on site}}.$$

The interfaces of the general search engines do not allow us to specify a query that retrieves only pages on a specific site that have outgoing links on them (and not pages without any links on them). Therefore we have to estimate the denominator of

self-linking. We estimate the number of pages with links on the site by the number of pages residing on the site (this is only an estimate, since there are pages on the Web that do not have any outgoing links (e.g. text files)). Thus:

$$\text{estimated self-linking} = \frac{\text{\# self-sitation}}{\text{\# pages on site}} \quad .$$

In order to assess estimated self-linking versus self-linking, we downloaded the first 4000 pages retrieved by AlltheWeb from the site of the Ben-Gurion University in Israel (a sample of self-sitations from this sites was analyzed in a later section of this paper) out of the 22,578 pages reported by the search engine on August 10, 2003. Out of the 4000 URLs 495 were not existing or inaccessible. Out of the remaining 3505 pages, 744 pages had no outlinks on them, thus the percentage of pages in our sample without any links on them was around 20%. This single example seems to indicate that our estimate is reasonable. In order to increase our confidence in this statement, further tests should be carried out.

It has been pointed out before (e.g., INGWERSEN, 1998; SMITH, 1999 and THELWALL, 2001) that self-links should be handled with care, because a large number of these links may be navigational links (e.g. home, back).

Instead of counting the actual number of links, we count the pages having links (one or more) to the examined site. As we pointed out before, we do this for technical reasons. MURRAY & MOORE (2000) in an extensive study of Web pages found that the average number of internal links (these are what we call self-links) is 23, but the median is only 4 – thus there are a few pages with extremely large number of self-links. BRODER et al. (2003), based on crawl of 784 million pages in the second half of 2002, found that the average number of links was 62.35, while the median was 23 – links of all types internal an external, including relative links. Relative links indicate the position of the page to be linked relative to the current page (e.g. the html tag <a href="b.htm"> links to the file b.htm that resides in the same directory as the current file). Relative links are necessarily self-links. In their dataset 81.5% of the links were relative links. In this paper we characterized the self-sitations of a sample of pages from a university site (the Ben-Gurion University in Israel – see section on Analyzing the links of a specific academic sites). For the characterization we only considered absolute links. The median and average number of self-links for this sample were calculated, the median was 2, while the average was 20.4. Thus we see that especially for self-links there is considerable difference between counting the number of pages with self-links and the number of self-links.

For the current research we used the Fast (http://www.alltheweb.com) search engine, because it is currently the second largest search engine with respect to coverage and it allows submitting queries of bibliometric nature (it is a little less flexible than AltaVista). This search engine was recommended by AGUILLO (2001) for bibliometric queries. The appropriate query for retrieving the self-sitations (pages containing self-links) of a site xxx.yyy.zzz using the advanced search page of Fast is:

Must include: xxx.yyy.zzz in the host name

Must include: xxx.yyy.zzz in the link to the URL

The inspection of the results reveals that the Fast search engine retrieves only pages with absolute links to the site, relative links (of the type aaa.html or ../bbb/ccc.htm) are ignored. Since most of the navigational links are usually relative ones, the self-linked pages retrieved by Fast ignore most pages on which all the links to the site are navigational ones. Therefore it is still meaningful to explore the self-linking and self-linked rates.

Data collection

As mentioned before, data for the current study was collected using the search engine Fast. We have already presented the query for self-sitations in the previous section. The query for retrieving the number of pages pointing to the site:

Must include: xxx.yyy.zzz in the link to the URL ,

and the query for retrieving the number of pages on site:

Must include: xxx.yyy.zzz in the host name .

We chose to study the self-linking and self-linked rates of academic institutions. Our sample consisted of fifteen top-level institutions in the US and the list of all universities in Israel. The US universities were picked based on the graduate school rankings of the *US News and World Report* (2003). We tabulated the list of the top five schools for every ranked discipline, and included all schools that appeared in more than one category. For institutions using several or alternative domain names we only considered the domain of the largest size. We had to exclude Johns Hopkins University, since we were unable to decide on a main domain name for it and its departments.

There are eight universities in Israel (see COUNCIL FOR HIGHER EDUCATION IN ISRAEL, 2002). The largest and best-know universities are the Tel-Aviv University and

the Hebrew University. These two and the fast-developing Ben-Gurion University cover most disciplines. The Bar-Ilan University's mission is to integrate Jewish tradition and science. It does not have a medical school and just recently opened a school of engineering. The University of Haifa puts an emphasis on the humanities and the social sciences and does not have a medical school or a faculty of science. The Technion is the Israel Institute of Technology, besides engineering, architecture and exact sciences it also has a medical school. The Weizmann Institute of Science only teaches graduate students and it is an outstanding research institution in the exact and life sciences. The Open University has a similar structure to other open universities in the world; it is accredited to award bachelor's degrees and masters's degrees only.

Results and discussion

The data was collected on February 4 and 6, 2003. The results are displayed in Tables 1 and 2. Relative rankings for self-linking (estimated) and self-linked are in ascending order, since for citations it is customary to assume that lower values are better. Low self-cited rate indicates high visibility and high self-citing rate may be a sign of isolation in the field. In each table, the institutions appear in alphabetic order.

Note that for the US data the estimated self-linking rate is between 61% and 77% for all the listed institutions. The average estimated self-linking rate is 69.55%, seems to be extremely high, especially when compared with the data for Israel, where the average estimated self-linking rate is only 48.40%, and even the highest value (57.57%) is lower than the lowest value for the US. At this point we can only speculate about reasons for these differences – these may be cultural or technological (applying different techniques for creating web-pages and sites, differences in the use of relative versus absolute links in websites). This point should definitely be further explored in the future.

The average self-linked rate for the US universities is 13.91% versus 17.40% for the Israeli universities. If, as for self-citing, we interpret the lower self-linked as more highly visible, than the most highly visible academic institution in the list is MIT. For Israel it is not easy to interpret the data, since as we said before the largest and best-known universities in the country are the Tel-Aviv University and the Hebrew University of Jerusalem, and the Hebrew University is ranked 7th out of 8. The high visibility of the Weizmann Institute is not surprising; it is an outstanding research institution.

Table 1. Self-linking and self-linked rates for 16 top US academic institutions

Name	Domain	Self-links	Links to site	Size of site	Self linked rate	Sited rank	Estim. self linking rate	Siting rank
CalTech	caltech.edu	152,516	922,532	219,616	16.53%	13	69.45%	9
Columbia	columbia.edu	175,819	1,448,269	281,547	12.14%	4	62.45%	2
Harvard	harvard.edu	480,489	2,523,499	622,731	19.04%	15	77.16%	15
Indiana-Bloomington	indiana.edu	241,831	1,769,623	366,963	13.67%	9	65.90%	4
MIT	mit.edu	302,040	3,652,318	479,157	8.27%	1	63.04%	3
Princeton	princeton.edu	118,489	893,187	165,009	13.27%	8	71.81%	12
Stanford	stanford.edu	421,874	2,525,771	610,417	16.70%	14	69.11%	8
Syracuse	syr.edu	91,944	717,605	138,008	12.81%	5	66.62%	6
U. Chicago	uchicago.edu	175,638	1,148,042	246,093	15.30%	12	71.37%	11
U. Illinois - Urbana-Champaign	uiuc.edu	329,968	2,554,782	430,345	12.92%	6	76.68%	14
U. Michigan-Ann Arbor	umich.edu	359,800	2,592,956	461,447	13.88%	10	77.97%	16
U. N. Carolina-Chapel Hill	unc.edu	181,078	1,830,194	244,279	9.89%	2	74.13%	13
U. Penn	upenn.edu	222,931	1,708,110	336,222	13.05%	7	66.30%	5
UC-SF	ucsf.edu	80,157	410,487	129,847	19.53%	16	61.73%	1
UC Berkeley	berkeley.edu	494,117	3,399,346	697,929	14.54%	11	70.80%	10
Yale	yale.edu	130,391	1,180,232	190,962	11.05%	3	68.28%	7

Table 2. Self-linking and self-linked rates for Israeli public universities

Name	Domain	Self-links	Links to site	Size of site	Self linked rate	Sited rank	Estim. self linking rate	Siting rank
Bar-Ilan U.	biu.ac.il	13,317	62,209	36,904	21.41%	6	36.09%	2
Ben Gurion U.	bgu.ac.il	6,804	75,260	21,308	9.04%	1	31.93%	1
Hebrew U.	huji.ac.il	72,268	313,555	128,888	23.05%	7	56.07%	7
Open U.	openu.ac.il	3,351	21,218	6,045	15.79%	3	55.43%	6
Technion	technion.ac.il	26,817	154,964	54,182	17.31%	5	49.49%	4
Tel-Aviv U.	tau.ac.il	29,918	175,225	51,957	17.07%	4	57.58%	8
U. Haifa	haifa.ac.il	24,901	99,790	51,901	24.95%	8	47.98%	3
Weizmann Inst	weizmann.ac.il	14,229	134,999	27,021	10.54%	2	52.66%	5

To summarize our findings: the self-linked rate seems to be the more meaningful measure, however further studies are needed in order to try to draw some correspondence between classical measures and the measures introduced in this paper.

Analyzing the links of a specific academic sites

As we saw in the previous section it is not easy to interpret the data regarding self-sitations. In order to enhance our understanding, we decided to characterize a sample of self-sitations from one of the academic sites for which we calculated the self-linked and the estimated self-linking rates. The search engine Fast retrieves at most 4000 results for a query, regardless of the number of results it reports to have found. Other search engines (e.g. Google and AltaVista) retrieve at most 1000 results. The results are ranked, therefore taking a sample of the retrieved results (in case it is not the whole set) will not constitute a random sample. An additional feature of Fast (limiting the results before or after a date set by the user) allows one to retrieve up to 8000 results (if the date is defined appropriately).

The two sites with the smallest number of self-sitations were the Open University of Israel and the Ben Gurion University (in Israel). We did not feel that the Open University is representative in its structure of the universities, thus we picked the self-sitations of the Ben Gurion University to be our test case. We were able to download all pages with self-sitations. The actual number of pages retrieved on January 22, 2003 was 6423 (slightly less than the number initially reported by the search engine). All pages were downloaded on the same day. Out of the URLs retrieved by the search engine Fast, 567 pages were either non-existent of unreachable. Out of the remaining 5856 pages we picked a random 10% sample. The self-links on these 581 pages were characterized using the method of content-analysis (see KRIPPENDORFF, 1980). The analysis was a multi-faceted content analysis, characterizing the source page, the link context and the target page. The content categories were developed inductively based on the analyzed pages. Reliability of the process was checked by a second coder, who coded 10% of the pages in the sample. Inter-coder agreement was satisfactory; it was between 82% and 96% depending on the facet. Out of the 581 pages in the sample, 550 actually contained self-links, on more than 70% of the examined pages we located more than one self-link.

We only examined absolute links (links of the type http://...), since the retrieval of Fast is based only on these links. We categorized the intention as decided by us based on the information in the surroundings of the link. We looked at each self-link separately, sometimes there were several links on the page with the same intended

function, in which case all were assigned the same value. At other times, different self-links on the page were intended to have different functions, in which case several values were assigned for the page. Altogether 669 values were assigned in the category *intention of link*. The breakdown of these values appears in Table 3.

We see that nearly 60% of the pages had self-links which were categorized as either technical or superficial. What was included under technical? Most of the time when the intention of the link was defined as technical (210 times or 58% of the total for this value) it clearly served navigational purposes. In another 34% (121 entries) of the cases, technical meant that the email of the person creating/publishing the Web page appeared at the bottom of the page. We found navigational links on 210 out of the 550 pages: thus only about 38% of the self-sitations contained navigational self-links.

On the other hand, in more than 40% of the cases, the self-links were not inserted for technical purposes, but were content-bearing links. These type of self-links and self-sitations should not be ignored, and should be further explored. More than 10% of the content-bearing categorizations were defined as research oriented (linking to a research project the person/unit participates in, linking to a presentation/abstract/full text article, organization/active participation in scientific conferences, etc.).

In general, nearly 50% of the self-links were targeted toward e-mail addresses (creator of page, teacher of course, conference participant, list of members of academic units, as general information on personal homepages, CVs, etc.).

Table 3. Values assigned to the category intention of link for the sample of self-sitations in absolute numbers and in percentages

Value	Number of occurrences	Number of total (out of 669)
Technical	360	53.8%
General/informative	87	13.0%
Professional (work related)	68	10.2%
Educational (refers to courses, teaching material)	62	9.3%
Research oriented	30	4.5%
Superficial (appears in a logo)	29	4.3%
Administrative	26	3.9%
Social	4	0.6%
Other	3	0.4%

Concluding remarks

In this paper we introduced two measures self-linked and self-linking. We calculated the measures for a sample of academic institutions. The self-linking rate is extremely high, which can be partially explained by the fact that self-links are often used for navigational purposes.

On the other hand, a closer examination of a sample of self-sitations revealed that only 38% of the self-sitations in our sample contained navigational self-links. This finding warrants further exploration. Until now it was customary to assume that self-links are not of much value, because they exist mainly for organizational purposes. These assumptions should be reassessed in view of the initial findings in this paper.

In this paper we considered a small sample only from a single university, in order to strengthen the hypothesis that a considerable percentage of self-sitations on academic sites are content-bearing, the self-links of additional universities must be examined.

The measure self-linked seems to be the more promising metric. This measure, like the Web Impact Factor (WIF), assesses the visibility of a site and can be used as an additional measure for the evaluation of the Web presence of domains or institutions.

References

AGUILLO, I. (2001). More on search engines with scientometric purposes. *SIGMETRIC* discussion list, January 10, 2001. Retrieved February 2, 2003 from http://listserv.utk.edu/cgi-bin/wa?A2=ind0101&L=sigmetrics&P=R1366

BAIRD, L. M., OPPENHEIM, C. (1994). Do citations matter? *Journal of Information Science*, 20 (1) : 2–15.

BAR-ILAN, J. (2002). How much information search engines disclose on the links to a Web page? – A longitudinal case study of the 'Cybermetrics' home page. *Journal of Information Science*, 28 (6) : 455-466.

BJORNEBORN, L., INGWERSEN, P. (2001). Perspectives of Webometrics. *Scientometrics,* 50 (1) : 65–82.

BONZI, S., SNYDER, H. W. (1991). Motivations for citation: A comparison of self citation and citation to others. *Scientometrics*, 21 (2) : 245–254.

BRODER, A., NAJORK, M., WIENER, J. (2003). Efficient URL Caching for World Wide Web Crawling. In *Proceedings of the Twelfth International World Wide Web Conference*. Retrieved August 8, 2003, from http://www2003.org/cdrom/papers/refereed/p096/p96-broder.html

CRONIN, B. (2001). Bibliometrics and beyond: some thoughts on Web-based citation analysis. *Journal of Information Science*, 27 (1) : 1–7.

CRONIN, B., SHAW, D. (2001). Identity creators and image-makers: Using citation analysis and thick description to put authors in their place. In: *Proceedings of the 8th International Conference on Scientometrics and Bibliometrics,* Sydney, Australia, July 16-20, 2001, 127–138.

COUNCIL FOR HIGHER EDUCATION IN ISRAEL (2002). Retrieved February 6, 2003 from http://www.che.org.il/eng.htm

EGGHE, L. (2000). New informetric aspects of the Internet: Some reflections, many problems. *Journal of Information Science*, 26 (5) : 329–335.

EGGHE, L., ROUSSEAU, R. (1990). *Introduction to Informetrics.* Amsterdam: Elsevier.

GARFIELD, E. (1974). Journal citation studies, XVII. Journal self-citation rated – there's a difference. *Current Contents*, #52, 5-7.

INGWERSEN, P. (1998). The calculation of Web Impact Factors. *Journal of Documentation*, 54 (2) : 236–243.

JCR (2000). *Analyze Self-Citations*. Retrieved February 2, 2003 from http://isi1.jcrweb.com/help/jcr_q6.htm

KRIPPENDORFF, K. (1980). *Content Analysis – An Introduction to its Methodology*. Beverley Hills: Sage Publications.

LARSON, R. R. (1996). Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of Cyberspace. *ASIS'96*. Retrieved February 2, 2003 from http://sherlock.berkely.edu/asis96/asis96.html

LAWANI, S. M. (1982). On the heterogeneity and classification of author self-citations. *Journal of the American Society for Information Science*, 33 : 281–284.

MURRAY, B. H., MOORE, A. (2000). Sizing the Web – A White Paper. *Cyvellaince.* Retrieved August 8, 2003 from http://www.cyveillance.com/web/forms/request.asp?form_type=wp&wp_name=sizing_internet

NISONGER, T. E. (2000). Use of Journal Citation Reports for serials management in research libraries: An investigation of the effect of self-citation on journal rankings in library and information science and genetics. *College and Research Libraries* 2000 : 263–275.

PERITZ, B. C., BAR-ILAN, J. (2002). The sources used by bibliometrics-scientometrics as reflected in references. *Scientometrics*, 54 (2) : 269–284.

PRIME, C., BASSECOULARD, E., ZITT, M. (2002). Co-citations and co-sitations: A cautionary view on an analogy. *Scientometrics*, 54 (2) : 291–308.

ROUSSEAU, R. (1997). Sitations: an exploratory study. *Cybermetrics*, 1(1). Retrieved February 2, 2003 from http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html

ROUSSEAU, R. (1999). Temporal differences in self-citation rates of scientific journals. *Scientometrics*, 44 (3) : 521–531.

SMITH, A. C. (1999). A tale of two Web spaces: comparing sites using Web impact factors. *Journal of Documentation*, 55 (5) : 577–592.

SNYDER, H., BONZI, S. (1998). Patterns of self-citation across disciplines (1980-1989). *Journal of Information Science*, 24 (6) : 431–435.

TAGLIACOZZO, R. (1977). Self-citation in scientific literature. *Journal of Documentation*, 33 (4) : 251–265.

THELWALL, M. (2000). Web impact factors and search engine coverage. *Journal of Documentation*, 56 (2) : 185–189.

THELWALL, M. (2001). Extracting macroscopic information from Web links. *Journal of the American Society for Information Science and Technology*, 52 (13) : 1157–1168.

THELWALL, M. (2002). Evidence for the existence of geographic trends in university Web site interlinking. *Journal of Documentation*, 58 (5) : 563–574.

THELWALL, M. (2003). The top 100 linked-to pages on UK university Web sites: High inlink counts are not usually associated with quality scholarly content. *Journal of Information Science*, 28 (6) : 483–491.

US News and World Report (2003). *Education: Best Graduate Schools 2003.* Retrieved February 2, 2003 from http://www.usnews.com/usnews/edu/grad/rankings/rankindex.htm

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.

WHITE, H. D. (2000). Toward ego-centered citation analysis. In: B. CRONIN, H. BARSKY ATKINS (Eds), *The Web of Knowledge* Medford, NJ: Information Today, pp. 475–496.

WHITE, H. D. (2001). Authors as citers over time. *Journal of the American Society for Information Science and Technology*, 52 (2) : 87–108.