Web Citations in Indian Theses: Exploring the Usage and Attributes

Rupesh Kumar A
Assistant Professor and Research Scholar
Department of Studies and Research in Library and Information Science
Tumkur University, Tumkur
email: a.rupeshkumar@gmail.com

ABSTRACT

The present study examined randomly selected electronic theses in Library and Information Science for prevalence of Web citations and explored various attributes such as type of content, accessibility, domain composition, path depth, file formats, and URL character length. 2125 Web citations were extracted from 285 ETDs. Majority of Web citations were for journal articles, and least for theses and dissertations. More than half (51%) of the Web citations had DOI. The rate of accessibility of Web citations was 85.14%.

Keywords: Library and Information Science, Electronic Theses,Web Citations.

INTRODUCTION

Citations in scholarly works act as a guide to explore the research already carried out on a particular area. They also reflect the sources from which the ideas originated. The primary purpose of citations is to enable the reader to access the original sources consulted by an author. A key element of citation is the Uniform Resource Locator (URL) or Internet address. With the increasing dependence on electronic resources for scholarly pursuits, web citations have assumed greater significance in research endeavours.

One of the problems of web citations is their dynamic nature. Online sources may change or disappear over time without notice. The availability, accessibility and permanence of links to online resources are a matter of great concern. "The accessibility of these references has become a great concern for researchers as the web is a constantly changing environment that provides no guarantee of permanence" (Carnevale and Aronsky, 2005). The process by which a website link or URL becomes inaccessible or broken over time, generally because the website which hosts the link disappears, changes content or moves to a new location is known as *web decay* or *link rot*. *Web decay* poses a serious threat to the continuous availability of online resources.

NEED FOR THE PRESENT STUDY

A systematic review of scholarly literature in various sources revealed that significant research has been carried out to measure the use of web citations (Vaughan and Shaw, 2002; Bhat and Sampath Kumar, 2008; Habibzadeh, 2013; Jalalifard, Norouzi, and Isfandyari-Moghaddam, 2013), web decay (Markwell and Brooks, 2003; Spinellis, 2003; Carnevale and Aronsky, 2005; Parker, 2007), recovery of inactive web citations and their half-life based on the research papers in journals (Sampath Kumar and Vinay Kumar, 2012; Sampath Kumar and Vinay Kumar, 2013; Sampath Kumar, Vinay Kumar and Prithviraj, 2015) . Some studies have also reflected the decay and recovery of web citations used in conference proceedings(Sampath Kumar and Prithviraj, 2012). However, there are only a few studieson measuring the decay of web citations used in electronic theses and

dissertations, and attempt their recovery, especially in the Indian context. The present study was designed to map the trends in the use of web citations cited in electronic theses and their various attributes. For the present study, randomly selected thesessubmitted during the year 2016 were downloaded from Shodhganga repository.

METHODOLOGY

Randomly selected theses in Library and Information Science were downloaded from Shodhganga repository (http://shodhganga.inflibnet.ac.in). Citations in these theses were extracted. Web citations were identified and segregated from print citations. W3C Link Checker (https://validator.w3.org/checklink) was used to check the accessibility of URLs. If the resource was found at the specific URL, it was marked as active citation. If the resource was not found, it was marked as inactive link. Other characteristics of URL citations such as domain, path depth, file formats and URL character length were also recorded. The study helped to measure the extent of accessibility of Web citations in Library and Information Science electronic theses.

RESULTS AND DISCUSSION

Table 1.Theses and Web Citations

Number of ETDs	Total Number of Citations	Number of Web Citations	Average Web Citations per thesis
285	18240	2125 (11.65%)	7.45

A total of 2125 Web citations were found in 285 theses. The average number of Web citations per thesis was 7.

Table 2. Distribution of Web Citations by Type of Content

Type of Source	**Number**	**Percentage**
Journals	1457	68.58
Books	39	1.85
Reports	307	14.43
Workshops/ seminars / Conference Proceedings	240	11.31
Theses / Dissertations	19	0.89
Newspaper articles	30	1.40
*Others	33	1.54
Total	**2125**	**100**

*Others: archives, working papers, white papers, brochures, pamphlets.

It may be noted from the above table that 68% of Web citations were related to journal articles. This indicates a high rate (68.58%) of prevalence of journal articles in electronic form on the Web as compared to other forms of resources. Theses and dissertations were the least (0.89%) found as Web resource among various forms of resources. Other forms of resources such as working papers, white papers, archives counted for 1.5% of the total number. The composition presented in the table indicates that there is diversity in the source of content of Web citations.

Table 3. Distribution of Web Citations by Top Level Domain (TLD)

Domain	Number	Percentage
DOI	1093	51.44
.com	473	22.24
.org	226	10.65
.ac	142	6.70
.edu	102	4.79
Other domains	29	1.35
.net	28	1.32
.info	28	1.32
.gov	4	0.19
Total	2125	100.00

In terms of top level domains, though the Digital Object Identifiers (DOI) URLs end in .org domain, they were classified separately as they as persistent identifiers. More than half of the total number of Web citations (51.44%) had DOIs. This indicates high level of use of DOIs by research scholars in Library and Information Science. This is an indirect indication of high level of awareness on DOIs. 22.44% of URLs had .com domain, followed by .org domain (10.65%). High prevalence of Web citations in .com domains may be due to the prevalence of majority of online scholarly content databases in .com domain.

Table 4. File Formats associated with Web Citations

Domain	Number	Percentage
jsp	400	18.81
html	369	17.35
pdf	347	16.35
asp	217	10.23
txt	206	9.70
ppt/pptx	175	8.26
xls/xlsx	170	7.98
cfm	89	4.18
php	76	3.59
cgi	44	2.08
doc/docx	31	1.48
Total	**2125**	**100**

Data on various file formats associated with Web citations is presented in Table 4. It is evident that jsp(18.81%) is the most widely found file format, closely followed by html (17.35%) and pdf (16.35%). Further, Powerpoint presentation files (8.26%), Excel Spreadsheet files (7.98%) and text files (9.70%) also have found considerable use in ETDs.However, doc/docx files were the least found among all file formats. High occurrence of scripting file formats such as asp and jsp may be due to proliferation of online databases programmed using ASP and JSP scripting languages.

Table 5. Path Depth of URLs in Web Citations

Path Depth	Number	Percentage
0	337	15.85
1	326	15.36
2	280	13.19
3	259	12.17
4	213	10.02
5	183	8.61
6	158	7.45
7	134	6.30
8	121	5.68
9	73	3.43
10 or more	41	1.94
Total	**2125**	**100**

Path depth of URLs in Web citations was also identified and recorded. Path depth refers to the levels in the URL structure. For instance, the url*www.tumkuruniversity.ac.in* has path depth 0, while *www.tumkuruniversity.ac.in/index.php* has path depth of 1. Path depth may have an impact on the accessibility of URLs. 15.85% of URLs had path depth of 0, while 1.94% URLs had path depth of 10 or more.

Table 6. Character Length of URLs in Web Citations

Character Length	Number	Percentage
<=20	373	17.55
21-30	349	16.43
31-40	268	12.62
41-50	231	10.87
51-60	209	9.82
61-70	197	9.29
71-80	192	9.05
81-90	179	8.45
91-100	88	4.12
>100	38	1.79
Total	2125	100

Number of characters in the URL is an important attribute considered for the study. URLs were classified on the basis of character length. 17.5% of URLs consisted of 20 or less characters, followed by 16.43% URLs with up to 30 characters. There were 38 (1.79%) URLs with 100 or more characters.

Table 7. Active and Inactive Web Citations

Total number of Web citations	Number of Active Web Citations	Percentage	Number of inactive Web Citations	Percentage
2125	1809	85.14	316	14.86

The accessibility of URLs in Web citations was checked using W3C Link Checker. 1809 out of 2125 (85.14%) Web citations were accessible, while 14.86% of Web citations were inaccessible.

CONCLUSION

In the present study, an attempt has been made to identify the prevalence of Web citations in ETDs in Library and Information Science. Further, an analysis of different characteristics of Web citations has also been presented. The study can be expanded further to include an in-depth analysis of inactive Web citations and their recovery.

REFERENCES

1. Bhat, V. R., & Kumar, B. T. S. (2008). Use of web based sources in scholarly electronic journals in the field of library and information science: A citation analysis. *ALIS Vol.55(2) [June 2008]*. Retrieved from http://nopr.niscair.res.in/handle/123456789/1776
2. Carnevale, R., &Aronsky, D. (2007).The life and death of URLs in five biomedical informatics journals.*International Journal of Medical Informatics*, *76*(4), 269–273. https://doi.org/10.1016/j.ijmedinf.2005.12.001
3. Habibzadeh, P. (2013). Decay of References to Web sites in Articles Published in General Medical Journals: Mainstream vs Small Journals. *Applied Clinical Informatics*, *04*(04), 455–464. https://doi.org/10.4338/ACI-2013-07-RA-0055
4. Jalalifard, M., Norouzi, Y., &Isfandyari-Moghaddam, A. (2013).Analyzing web citations availability and half-life in medical journals: A case study in an Iranian university.*Aslib Proceedings*, *65*(3), 242–261. https://doi.org/10.1108/00012531311330638
5. Markwell, J., & Brooks, D. W. (2003). "Link rot" limits the usefulness of web-based educational materials in biochemistry and molecular biology. *Biochemistry and Molecular Biology Education*, *31*(1), 69–72. https://doi.org/10.1002/bmb.2003.494031010165
6. Parker, A. (2007). Link rot: How the inaccessibility of electronic citations affects the quality of New Zealand scholarly literature. *Coda*, 16.
7. Sampath Kumar, B. T., &Prithvi Raj, K. R. (2012).Availability and persistence of web citations in Indian LIS literature.*The Electronic Library*, *30*(1), 19–32. https://doi.org/10.1108/02640471211204042
8. Sampath Kumar, B. T., & Vinay Kumar, D. (2013). HTTP 404-page (not) found: Recovery of decayed URL citations. *Journal of Informetrics*, *7*(1), 145–157. https://doi.org/10.1016/j.joi.2012.09.007
9. Sampath Kumar, B. T., Vinay Kumar, D., &Prithviraj, K. R. (2015).Wayback machine: Reincarnation to vanished online citations. *Program*, *49*(2), 205–223. https://doi.org/10.1108/PROG-07-2013-0039
10. Shaw, D., & Vaughan, L. (2003).*Bibliographic and Web citations: What is the difference?*https://doi.org/10.1002/asi.10338
11. Spinellis, D. (2003). The decay and failures of web references.*Communications of the ACM*, *46*(1), 71–77. https://doi.org/10.1145/602421.602422

Acknowledgement

The author gratefully acknowledges the guidance of his Research GuideDr. B T Sampath Kumar, Associate Professor, Department of Studies and Research in Library and Information Science, Tumkur University, Tumakuru in conducting the present study.