Library Progress International
Vol.42, No.2, July-December 2022: P.412-420
DOI: 10.5958/2320-317X.2022.00039.3

Print version ISSN 0970 1052
Online version ISSN 2320 317X

Original Article

Available online at www.bpasjournals.com

Web Citation Analysis on Journal of Travel Research: A Study

***D. Satyanarayana, **Dr. P. Damodar**

Author's Affiliation:
*Research Scholar, Osmania University, Hyderabad, Telangana 500007, India and Library Asst., Telangana University, Nizamabad, Telangana 503322, India
**Librarian, University College of Engineering, Osmania University, Hyderabad, Telangana 500007, India

Corresponding Author: D. Satyanarayana, Research Scholar, Osmania University, Hyderabad, Telangana 500007, India and Library Asst., Telangana University, Nizamabad, Telangana 503322, India
E-mail: satyamdantala@gmail.com

(Received on 18.08.2022, Revised on 29.09.2022, Accepted on 28.11.2022, Published on 15.12.2022)

How to cite this article: Satyanarayana D., Damodar P. (2022). Web Citation Analysis on Journal of Travel Research: A Study. *Library Progress International,* 42(2), 412-420.

ABSTRACT

A substantial amount of research has focused on the persistence or availability of Web citations. The present study analyzes Web citation distributions. Web citations are defined as the mentions of the URLs of Web pages (Web resources) as references in academic papers. The present paper primarily focuses on the analysis of the URLs of Web citations. Initially, all issues of the Journal of Travel Research published during 2012 to 2021 were downloaded directly from their publisher websites. Afterwards, all the journals' citations in either print or Web formats were calculated manually. Then, availability and/or decay of individual cited URLs were examined in the Web environments. Research findings indicated that, a total of 2939 URLs cited in the 740 articles were examined. The percentage of URLs increased from 58 percent in 2012 to 90 percent in 2021. The study found that 40 percent of URLs were not accessible at the time of testing and the remaining 60 per-cent of URLs were accessible. The HTTP 404 error message – ''file not found'' was the overwhelming message encountered and represented 53.29 percent of all HTTP error messages. The study also noticed that the average half-life of URLs of missing URLs was estimated to be 4.94 years.

KEYWORDS: *URLs, Web citations, half-life, Travel Research, HTTP error, Journal of Travel Research*

INTRODUCTION

The Internet is a great resource that is available to anyone who has access to it. The WWW contains information from all over the world and it is very useful for everyone. The information contained on the Web can also be useful for academic research. The use of Web sources in scholarly publishing has increased. Scholars make use of more and more Web resources in their research work.

In this study, we investigate how web citations are used in Journal of Travel Research articles published during 2012 to 2021 and analyze the results. This study assessed the overall rate of web citations and the frequency of in accessible URLs in Journal of Travel Research.

REVIEW OF LITERATURE

Manjunatha, G., Kumar, B. T. S., & Lakshmana, H. (2020). study investigated the use of URLs as citations and their longevity, based on 966 articles published in selected LIS journals published from 2011 to 2015. It is found that there are 36,968 references in 966 articles with an average of 38.26 per article. Of the 36,968 references, there are 5,867 URL references. The study also found that 46.53% URLs (2,730 out of 5,867) remained active while the remaining 3,137 (53.46%) were found to be missing. The largest number of missing URLs (83.22%) cited in LIS articles are published in the year 2012. The HTTP error 404 - 'file not found' error message was the common error 40.36% In this regard the study suggested that the author(s) should check the accessibility of URL citation before it is used in the reference list of the article.

Prithviraj, K. R., & Kumar, B. T. S. (2014). Conducted a study on to analyze the accessibility, corrosion and half-life of URLs cited in the articles of Indian LIS conference proceedings published during 2001 to 2010. A total of 5,698 URLs cited in the 1,700 articles were examined. The percentage of URLs increased from 39.10 percent in 2001 to 73.47 percent in 2009. The study found that 50.09 percent of URLs were not accessible at the time of testing and the remaining 49.91 percent of URLs were accessible. The HTTP 404 error message – "file not found" was the overwhelming message encountered and represented 53.29 percent of all HTTP messages. The study also noticed that the average half-life of URLs of missing URLs was estimated to be 4.94 years. Even though there are various retrieval tools being used to recover vanished URLs, still there is a need to improve such tools.

Sampath Kumar, B.T., & Vinay Kumar, D. (2013). investigates the availability, persistence and half-life of URL citations cited in two Indian LIS journals articles published during 2002 - 2010. Total 472 research articles published from 2002 to 2010 with 1290 URLs cited. Study found that only 18.91% (1290 out of 6820) of URLs cited in these journal articles. 39.84% of URL citations were not accessible and remaining 60.15% of URL citations were still accessible. Most of the error message from HTTP 404 "Page not found" with 54.86% of all HTTP error messages. However 51.06% URLs were recovered from HTTP 404 error message. Study also noticed that the half-life of URL citations was increased from 6.33 years to 13.85 years after recovering missing URLs from Wayback machine.

Many studies have estimated the "half-life period" (time required for half of all online citations in journal to disintegrate) of web citations in various disciplines and researchers found that this amount of time may differ. Goh and Ng (2007) investigated the link decay phenomenon in three leading Information Science journals. Articles spanning a period of seven years (1997-2003) were downloaded and their links were extracted. From these, a measure of link decay, the half life, was computed to be approximately 5 years. Kumar and Kumar (2012) found the half-life of URL citations cited in LIS scholarly journals was approximately 11.5 years. Moghaddam and Saberi (2011) observed that as the age of URL citations increases, the number of inaccessible URLs increases.

OBJECTIVES OF THE STUDY

1. To identify the web citations has been used by the authors in their articles.
2. To find out top-level domains constitute the source of the web-located citations.
3. To explore different file formats of the cited Web documents.
4. To rectify HTTP errors linked with missing URL citations.
5. To calculate half-time of Web resources referred in JTR.

METHODOLOGY

The data for the present study has been drawn from Journal of Travel Research published by Sage publications during the year 2012-2021. References that are appeared as a list at the end of the each articles under the bibliography or reference section are considered. The expanded bibliographies, endnotes and footnotes and annotations are counted in our dataset. After selecting all the references appended to the articles published

in Emerald LIS journals, URLs are extracted for further analysis

A total number of 58687 citations, as obtained from the references of 740 articles presented, have been analyzed and the necessary interpretations were made. After lengthy manual verification, total number of the research articles published during the 2012-2021, number of articles with web resources, number of total citations and number of citations to web resources were recorded into separate Excel files. The URL of each web resource was also noted for determining the nature of the web resource and its availability in 2022.

FINDINGS

Distribution of articles, citations and Web citations

The below Table shows that the web citation rates have increased over the past ten years. As it can be seen from the table, the web citations exhibit an exponential increase from 2012 to2021 in Journal of Travel Research journal.

Table 1 : Year Wise Distribution of Articles

Year	Issues	Total Articles	Total Ref.	Percentage	Total Web Citation	% of URL per articles
2012	6	61	3594	58.91803279	147	2.41
2013	6	59	3974	67.3559322	190	3.22
2014	6	57	3916	68.70175439	168	2.95
2015	6	53	4398	82.98113208	205	3.87
2016	8	77	7305	94.87012987	341	4.43
2017	8	74	4986	67.37837838	328	4.43
2018	8	76	5801	76.32894737	307	4.04
2019	8	87	7165	82.35632184	365	4.20
2020	8	91	8091	88.91208791	408	4.48
2021	8	105	9457	90.06666667	480	4.57
Total	**72**	**740**	**58687**	79.30675676	**2939**	3.97

The lowest number of articles recorded in the year 2015 (53 articles) and highest number was recorded in the year 2021(105 articles). In 2012, 147 web citation are recorded and it is constantly increasing in the following years 480 web citation are used by the authors. Based on the above table it can be said that authors have been increasingly making use of Web-based resources in their studies. There were total number of 2939 citations in 740 articles, and average 3.97 citations was calculated per article

.

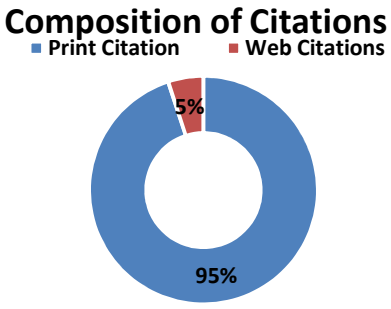

Figure 1: Composition of Citations

According to Table 1, total number of citations (58687), citing to printer and Web resources in Journal of Tourism Research articles is 55748(94.99%) and 2939 (5.01%), respectively. It demonstrates that paying less attention to Web citations in Journal of Tourism Research articles is considerable compared to Web citations used in Information Research journals, as reported by Isfandyari-Moghaddam, A., Esmaeel, M., & Saberi, M. K. (2010).

Distribution of URLs Accessibility by Type of Domain

URL is an acronym for Uniform Resource Locator and is a reference (an address) to a resource on the Internet. The first part of a URL, just behind the two bars, contains the type of protocol used in the exchange of information, in this case http:. The second part of the URL, corresponds to the name of the domain where the data or the service is located and optionally its connection port.

The key portion of a URL address is the domain name. All those commercials that show how easy it is to create a website are web hosting companies. If a person uses their hosting service, that is where the website will live. It's easy to identify the domain name because it will often begin with a www, which stands for worldwide web, and end with a period and a collection of letters. In this study, only six domains viz. org, .net, .gov, .edu, .com, and .ac were taken into consideration; however, those domains not belonging to any of these categories were included in "others".

- .ac - academic.
- .com - a commercial organization
- .edu - an educational organization
- .gov – an official government site
- .net - it was for network organizations
- .org - mostly non-profit organizations

Table 2 : Distribution of URL Domain Type

Active URLs			**Inactive URLs**			**Total URLs**		
Domain	**Number**	%	**Domain**	**Number**	%	**Domain**	**Number**	%
.co/.com	501	29%	.co/.com	483	41%	.co/.com	984	33%
.org	482	27%	.org	268	23%	.org	750	26%
.gov	143	8%	.gov	179	15%	.gov	322	11%
.edu	76	4%	.edu	36	3%	.edu	112	4%
.net	33	2%	.net	37	3%	.net	70	2%
.ac	12	1%	.ac	11	1%	.ac	23	1%
Others	507	29%	Others	171	14%	Others	678	23%
Total	1754	100%	**Total**	1185	100%	**Total**	2939	100%

The findings of the above table, indicate that of the 1185 decayed web citations, the top-level domain .com had the greatest number (41%) of missing URLs followed by.org (23%). government sites (15%) and other top level domains (14%). Nevertheless, the findings indicate that there is little loss associated with .edu (3%), .net (3%) and .ac (1%). These findings support earlier studies (Dimitrova and Bugeja, 2007; Goh and Ng, 2007; Isfandiari and Saberi, 2010) which reported that .org, .edu and .net were the most persistent domains. The findings are also in line with those of Moghaddam et al. (2010) who reported that the .com domain was among those with poorer stability and persistence.

Distribution of URLs by type of file formats

Table 3: Distribution of URLs by file formats

File Format	Number	Percentage	Active Number	Percentage	Inactive Number	Percentage
HTML	1139	38.75%	821	46.81%	318	26.84%
PDF	1119	38.07%	746	42.53%	373	31.48%
PHP	48	1.63%	21	1.20%	27	2.28%
ASP	127	4.32%	63	3.59%	64	5.40%
JSP	22	0.75%	8	0.46%	14	1.18%
xml	3	0.10%	2	0.11%	1	0.08%
doc	6	0.20%	1	0.06%	5	0.42%
txt	2	0.07%	1	0.06%	1	0.08%
CGI	6	0.20%	5	0.29%	1	0.08%
Others	467	15.89%	66	3.76%	381	32.15%
TOTAL	2939	100.00%	1754	100.00%	1185	100.00%

The data as illustrated in above table indicate that the greatest numbers of cited web resources are HTML/HTM files. Out of 2939 Web citations, 1139 cases are HTML files, followed by 1119 PDF files, 127 ASP files, 22 JSP files, DOC and CGI 6 files and 2 PPT txt files. Some other file formats, 467 ones, which did not match these six categories, were included in the "other" category.

These findings are in agreement with findings of McCown et al (2001) and Maharana, Nayak and Sahu (2006) which reported that most of cited Web resources contain HTML/HTM files.

Error codes associated with the missing URLs

Error messages at inaccessible URLs. The HTTP protocol defines 24 different errors that can occur within an HTTP exchange. In addition, some errors can occur before the client and server get a chance to communicate (Spinellis, 2003). In practice, whenever a URL was inaccessible an error message (HTTP code) appeared. In general, when URLs were checked we were faced by the following errors:

Error 302 - Found (Previously "Moved Temporarily")
Error 308 - Permanent Redirect
Error 400 - Bad Request
Error 401 - Unauthorized
Error 403 – Forbidden
Error 404 - Not Found
Error 406 - Not Acceptable
Error410 - Gone
Error 500 - Internal Server Error
Error 502 - Bad Gateway
Error 503 - Service Unavailable
Error 504 - Gateway Timeout

Table 4: Error codes associated with missing URLs

Year	Error 302	Error 308	Error 400	Error 401	Error 403	Error 404	Error 406	Error 410	Error 500	Error 502	Error 503	Error 504	others	Total
2012				1	22	35	1	1			2		19	81
2013	1				25	50	1	1	1		1		25	105
2014					11	29		2			1		27	70
2015					21	27		2					39	89
2016		2	1		35	58		1			7	1	31	136
2017		1			41	49	1	1	1		1	1	20	116
2018	1		3		58	52		2	1		3		19	139
2019	1		2	2	65	60	1	1	1	2	2	1	24	162
2020	2	1	2	1	51	64			2				30	153
2021		3			62	55	1		1		1		11	134
TOTAL	5	7	8	4	391	479	5	11	7	2	18	3	245	1185

The above table represents the distribution of HTTP error codes associated with missing Web URLs. Out of total 1185 inactive URLs, error type 404 (Page not found) reported for half of missing citations 479, followed by other HTTP errors, HTTP 403 (Forbidden) 391 and rest of HTTP errors collectively accounted for 315 of missing URLs.

Distribution of Path depth in missing URLs

To determine how a URL's path length influences decay rates, we calculated the path depth for each accessible and inaccessible URL. To calculate path depth, we added 1 to the depth for every directory or file after the domain name. For example, http://foo.com/ has a path depth of 0, http://foo.com/bar.html has a depth of 1, http://foo.com/dir/bar.html has a depth of 2, etc. We also added 1 to the path depth for any existing query string in a URL (e.g., http://foo.org/cgi?bar=2 has a path depth of 2).

Table 5: Distribution of Path depth in missing URLs

			Active URLs			**Missing URLs**		
Path depth	**Number**	**%**	**Path depth**	**Number**	**%**	**Path depth**	**Number**	**%**
0	208	7%	0	179	15%	0	29	2%
1	723	25%	1	639	54%	1	84	7%
2	575	20%	2	314	26%	2	261	22%
3	516	18%	3	234	20%	3	282	24%
4	394	13%	4	177	15%	4	217	18%
5	293	10%	5	136	11%	5	157	13%
6	129	4%	6	47	4%	6	82	7%
7	54	2%	7	16	1%	7	38	3%
8	28	1%	8	9	1%	8	19	2%
9	12	0%	9	3	0%	9	9	1%
10	5	0%	10	0	0%	10	5	0%
above 10	2	0%	above 10	0	0%	11	2	0%
TOTAL	**2939**	**100%**	**TOTAL**	**1754**	**148%**	**TOTAL**	**1185**	**100%**

The table 5 indicates towards the association of path depth of web citations and their active and missing status. According to Goh and Ng (2007) "The URLs path depth could be associated with link failure due to increasing complexity as the length of a URL increases". In the present study each of 2939 URLs path depth was categorised into different levels-0, 1, 2, 3...10.

The purpose of this method is to find out the association between weblink path depth and decay of URLs. It is apparent from the table that more missing URLs (2%) were having 0 level path depths than those having 10 level path depth. Similarly, missing URLs having1 level path depth were more (7%) than those having 9 level path depth. The significant finding drawn from this table is contrary to the statement of Goh and Ng (2007);Prithviraj and Sampath Kumar (2014) and other many authors conducted same studies for other journals, as the present table shows no association between the path depth and decay of URLs but the finding is supported by a recent study of Vinay, Divyashree and Sampath Kumar (2019).

Half-life of Web citations in Information Research

Half-life is the time required for half of all online Web citations in a journal to disintegrate. In order to estimate the half-life of the URL citations examined in this study, the researcher adopted the procedure used by Koehler (1999); The following formula was used to calculate the half-life of online citations for each journal year. The following formula was used to calculate the half-life of Journals of Travel Research articles

$$\text{Half} - \text{Life}[t_{(h)}] = \frac{\text{t Ln}(0.5)}{\text{Ln } W_{(t)} - Ln\, W_{(0)}}$$

Where $W_{(0)}$is the number of working URLs at the time of publication, $W_{(t)}$is the number of working URL citations at some later time t. $t_{(h)}$is the estimated number of years it takes for 50% of the published Internet citations to stop working.

Table 6: Half Life of Web Citation

Year	t	w(0)	Ac URL W(t)	log(0.5)	t ln(0.5)	ln W(t)-ln W(0)	t(h)
2012	10	147	66	-0.30103	-3.0103	-0.347773	8.6559236
2013	9	190	85	-0.30103	-2.70927	-0.349335	7.755514
2014	8	168	98	-0.30103	-2.40824	-0.234083	10.287966
2015	7	205	116	-0.30103	-2.10721	-0.247296	8.5210075
2016	6	341	205	-0.30103	-1.80618	-0.221001	8.1727409
2017	5	328	212	-0.30103	-1.50515	-0.189538	7.9411522
2018	4	307	168	-0.30103	-1.20412	-0.261829	4.5988777
2019	3	365	203	-0.30103	-0.90309	-0.254797	3.5443534
2020	2	408	254	-0.30103	-0.60206	-0.205826	2.9250857
2021	1	480	347	-0.30103	-0.30103	-0.140912	2.1363014
		2939	1754	-0.30103	0		64.538922
The half-life of URL is							**6.4538922**

The half-life period of URLs cited during 2012-2021 in Journal of Travel Research is presented in the following tables. the average half-life for cited Internet resources is estimated to be 6.45 years. This means that it will take about six and half years for half of the Internet citations in JTR to vanish.

SUGGESTION AND CONCLUSION

The researcher conducted a study on the citation of URLs used as citations in the field of LIS journal articles shows that there is a constant growth in the use of URLs as citations in Journal of Travel Research. Though the overall percentage of the web references in the Journal of Travel Research articles slowly and steadily increased, yet they are not a match to print references. The most common cause of failure was "404 not found", followed by the "403 forbidden error". The highest number of cited web resources belonged to HTML files. The URLs of the cited web resources in Journal of Travel Research were found more concentrated at the URL path depth of 1, 2 and 3.

Increasing use of web citations, the information accessibility problems have increased too. Some fields of study may be more prone to the effects of URL decay than others particularly if many of the scholarly materials utilized are available on the internet and rules permit the use of internet documents in scholarly materials. Web citation decay is a serious problem in scholarly communication (Isfandyari Moghaddam & Saberi, 2010; Sadat mosavi, 2010). As a result, some researchers have suggested that web-based resources cannot be considered as a suitable research resource (Sellitto, 2005; Falagas, Karveli & Tritsaroli, 2007). The citations to web content should be complete and include full bibliographic information plus the date on which the site was accessed by the author, the date on which cited web page was created and last revised.

In order to decrease the inactive URLs, it has been suggested that authors, editors, and publishers should work together through:

1. Checking URLs systematically before publication to minimize unavailability due to spelling errors or misprints.
2. Publisher and editorial staff should work with authors to preserve and make available cited URLs sources. One possible strategy would be to support the development and maintenance of the Internet archive by encouraging the authors to upload the cited URLs sources.
3. The editorial staff develops guidelines, for authors and references about the type of URLs sources permissible, based on consideration permanency / stability of the cited content and its scholarly importance.
4. In the long term, all research organizations must come to the realization that, due to the dynamic nature of the Internet, certain minimum standards must be maintained in posting scholarly information there (Altman and King 2007).
5. Research community should give a preference to the use of DOIs instead of URLs.
6. Use of citation and referencing software tools would avoid citation errors resulting from typing.
7. To keep the online information active is to archive the Web sources at more than one place such as in institutional repositories (IR), subject repositories, national repositories and academic networks of authors. However, authors need to take the copyright policies in view while archiving the information.

REFERENCES

1. Goh, D.H.L., and Ng, P.K. (2007). Link decay in leading information science journals. Journal of the American Society for Information Science and Technology, 58(1), 17. DOI:10.1002/asi.20513

2. Isfandyari-Moghaddam, A., Esmaeel, M., &Saberi, M. K. (2010). Availability and Half-life of Web References Cited in Information Research Journal: A Citation Study. International Journal of Information Management, 8(2), 58–75.

3. Kumar, B.T.S. and Kumar, K.S.M. (2012). Decay and half life period of online: Citations cited in open access journals. The International Information and Library Review, 44(4), 202-211.

4. Mardani, Amir. (2012). An investigation of the web citations in Iran's chemistry articles in SCI. Library Review. 61, 18-29. 10.1108/00242531211207398.

5. McCown, F., Chan, S., Nelson, L.M., and Bollen, J. (2005). The availability and persistence of web citations in D-Lib Magazine. Available at: http://www.iwaw.net/05/papers/iwaw05-mccown1.pdf

6. Prithviraj, K. R., & Kumar, B. T. S. (2014). Corrosion of URLs: Implications for electronic publishing. IFLA Journal, 40(1), 35–47. https://doi.org/10.1177/0340035214526529

7. 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

****