

# WEB CITATION AND DECAY OF URLS: A CASE STUDY OF LIBRARY AND INFORMATION SCIENCE THESES UPLOADED ON SHODHGANGA UNDER THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE

*Dr. Vijay Uttamrao Barphe*

**Dr. Vijay Uttamrao Barphe**  
(College Librarian  
Department of Library)  
Vasantrao Naik Arts, Science  
& Commerce College,  
Shahada, Nandurbar.

## ABSTRACT –

The present research study is aimed to identify the citation trends and decay range of URL citations used in Ph.D Theses. The data is collected from targeted theses uploaded on Shodhganga from 2012 to 2019 by Savitribai Phule Pune University Pune. Out of 17823 citations used in 79 Ph.D. theses 18% having web citations. All the web citations are checked by dead Link Checker ([www.deadlinkchecker.com](http://www.deadlinkchecker.com)) to test whether they were accessible or not. The study finds 64% of web citations were accessible and 36% of URLs are not accessible. It is found that the majority of HTTP error messages are due to the 404 "Page not found" which forms 46% of all HTTP error messages. Other "the server name or address could not be resolved" is 31.44% and "Time out" error messages are 5.39%. The '.org' domain is also found to be the 33.49%. The greatest number of web citations are of 'HTTP' having 98% and only 2% of web citations is 'https' URLs Scheme.

**Keywords:** Missing URLs; Savitribai PhulePune University; Shodhganga; URLs error messages; Web citation; Web decay.

## INTRODUCTION

Today's internet and World Wide Web has great impact on researcher and information retriever. It has marked a noteworthy change in the information-seeking approach as well as the mode of dissemination of information. The World Wide Web has now become one of the main sources of information processed, transmitted, stored, retrieved, and disseminated at speed and convenience. Henceforth innumerable users use the internet every time for research and learning. Development of web publications, e-journals, e-books, and e-databases; www is one of the main and best tools for research and collection of resources to researchers. Nowadays the numbers of researchers have been increased citing web resources in their scholarly publications.

This study analyzes Web Citation and Decay of URLs cited in Library and Information Science Ph.D. Theses uploaded at Shodhganga by The Savitribai Phule Pune University, Pune.

### **Shodhganga**

The Shodhganga is a repository of Indian Electronic Theses and Dissertations, developed by the INFLIBNET Centre. Shodhganga is using Open Source Software like DSpace Digital Library Software developed by the Massachusetts Institute of Technology. At Shodhganga Electronic Theses and Dissertations are open accessible to all Indian Universities and Institutions. Today it has 329922 full-text Electronic Theses and 43 MRPs reports by 500 plus Universities. Simultaneously 610 plus Universities signed MoU with INFLIBNET as on 1<sup>st</sup> December 2021. (<http://shodhganga.inflibnet.ac.in/newmoredetails/about.html>)

### **Savitribai Phule Pune University, Pune**

Savitribai Phule Pune University, Pune is one of the most popular educational centers in India. It was established on 10<sup>th</sup> February 1949 under the Poona University Act. The University has 46 academic departments, 307 recognized research Institutes, and 612 affiliated colleges running graduate and undergraduate Courses. ([http://www.unipune.ac.in/university\\_files/about](http://www.unipune.ac.in/university_files/about))

### **Web Decay**

- Decay is the state or process of rotting or decomposition.
- Web decay refers to a process whereby the web source disappears over a period of time.

### **OBJECTIVES OF THE STUDY**

- To identify the year wise distribution of web citations and print citations.
- To identify alive and dead web citation URLs.
- To know the error codes of missing URLs.
- To find out the domain-wise web citations.

### **METHODOLOGY**

The present study has been attempted to analyze the decay of web citations in theses uploaded on Shodhganga by Savitribai Phule Pune University, Pune from 2012 to 2019. For a collection of data, total no. of Ph.D. theses and total no. of citations in them are interpreted as a primary tool. All web citations are checked on the internet using dead Link Checker and analyzed data and then presented in a table and graph.

### **REVIEW OF LITERATURE**

Moghaddam and Saberi (2011) examine the life and Death of web citation URLs of articles published in the Journal of the Medical Library Association from 2002-2008. A total number of 231 articles and 1049 citations are web citations, 69 percent of URLs are accessible and only 31 percent of URLs are not accessible at the time of testing.

Gull; Mahajan and Ali (2014) examine the growth and web decay rate of URL citations cited in one of the eminent information web magazines ARIADNE. This study covers 1392 citations identified from 73 articles out of 1184 web citations, and it found 9.73% of citations are not accessible.

Parmar and Pateria (2019) studied citations from the Indian Journal of Agricultural Library and

Information Services from 2012 to 2016, a total number of 980 citations were 94 articles and 325 web citations published in this journal. 62 % of web citations URL were accessible at the time and 38% of URLs were not accessible. Page not found- HTTP 404 Error was the majority of missing URLs is percent 51%.

Bansal and Parma (2020) examine the accessibility, deterioration, and half-life of URLs of cited documents in the Current Science Journal published from 2015 to 2016. A total number of 1564 articles and 1724 web citations. It was found 57% of web citations were accessible and

43% of URLs were not accessible at the time of testing. Page not found HTTP-404 ERROR Message is 59 percent.

## DATA ANALYSIS AND INTERPRETATION

For the present study, total number of citations collected from 79 Ph.D. theses of Library and Information Science submitted to the Savitribai Phule Pune University and uploaded by itself on Shodhganga are sources of data. The total number of citations are 17823 and web citations are 3120.

**Table 1 : Guide wise Theses uploaded on Shodhganga under The Savitribai Phule Pune University, Pune**

Sr. No	Guide name	No. of Theses Upload	Percentage
1	Dr. Patil S.K.	14	17.72
2	Dr. Deshpande Nilla J.	13	16.45
3	Dr. Singh Surya Nath	10	12.65
4	Dr. Prasad ARD	08	10.12
5	Dr. Mahajan S.G.	07	8.86
6	Dr. Konnur M.B.	07	8.86
7	Dr. Ganpule S.R.	05	6.32
8	Dr. Riswadkar M.R.	05	6.32
9	Dr. Pange B.M.	03	3.79
10	Dr. Kumbhar Rajendra	02	2.53
11	Dr. Sengupta Sivani	02	2.53
12	Dr. Ravichandra I.K.	02	2.53
13	Dr. Andersom B.	01	1.26
<b>Total</b>		<b>79</b>	<b>100</b>

Table 1 shows the Guide wise theses uploaded on Shodhganga by Savitribai Phule Pune University, Pune. It is found that the highest 17.72 percent theses are under the guidance of Dr. S.K.Patil, followed by 16.45% theses uploaded under the guidance of Dr. Deshpande Nilla J. 5 to 10 theses are uploaded under the guidance of Dr.

Riswadkar M.R., Dr. Ganpule S.R., Dr. Konnur M.B., Dr. Mahajan S.G., Dr. Prasad ARD, and Dr. Singh Surya Nath and lastly 1 to 4 theses uploaded under the guidance of Dr. Andersom B., Dr. Ravichandra I.K., Dr. Sengupta Sivani, Dr. Kumbhar Rajendra, and Dr. Pange B.M.

**Table 2 : Year-wise distribution of theses uploaded in Library and Information Science**

Sr. No.	Upload Year	No. of Theses Uploaded	No. of Citation	No. of web Citation	Average Web Citations per Article
1	2019	02	750	124	62
2	2018	03	790	198	66
3	2017	67	14659	2411	36
4	2016	01	390	11	11
5	2012	06	1234	376	63
<b>Total</b>		<b>79</b>	<b>17823</b>	<b>3120</b>	<b>39</b>

A total number of 17823 citations are identified from 79 Ph.D. theses; out of 3120 are web citations and 14703 are print citations. Table 2 shows the distribution of the total no. of citations and web citations by year. The table also shows

the average number of web citations per thesis, a maximum number of web citations is 66 in 2018 and a minimum of 11 in 2016. The average number of web citations are 39 across all uploaded theses in shodhganga.

**Table 3 : Domain Types of web citations**

Sr. No.	Domain	No. of web citations in the year					Total	Percentage
		2019	2018	2017	2016	2012		
1	.org	32	74	750	03	186	1045	33.49
2	.com	29	71	404	02	65	571	18.30
3	.gov	13	5	250	02	9	279	8.94
4	.net	01	2	134	01	11	149	4.77
5	.edu	7	10	238	0	34	289	9.26
6	.ac.in	4	13	71	0	11	99	3.17
7	.ac	1	3	187	0	17	208	6.66
8	.in	2	0	33	0	3	38	1.21
9	.rec.in	17	0	20	01	2	40	1.28
10	.ernet	1	0	15	0	0	16	0.51
11	.nic.in	9	0	64	02	2	77	2.46
12	.co	2	0	10	0	1	13	0.41
13	other	6	20	235	0	35	296	9.48
<b>Total</b>		<b>124</b>	<b>198</b>	<b>2411</b>	<b>11</b>	<b>376</b>	<b>3120</b>	

Table 3 shows Domain Types of web citations. In the present study mainly 12 domain types such as .org, .com, .gov, .net, .edu, .ac.in, .ac, .in, .rec.in, .ernet, .nic.in, and .co domains are taken into

consideration. However, the domains which did not fall under these 12 categories are included under other categories. It is cleared from the table that the highest number of web citations is

domain type .org i.e. 33.49% followed by .com in the range of 5% to 10 %. Less than 5% of web domain type at 18.30%. Web citations domain citations domain are .net, .ac.in, .in, .rec.in, .edu, .gov, .ac, and other domain types are found .nic.in, .ernet and .co.

**Table 4 : Active and missing web citations**

Sr. No.	Uploaded year	Total No. Web citation	Active Web citation	%	Missing Web Citation	%
01	2019	124	75	60	49	40
02	2018	198	120	61	78	39
03	2017	2411	1633	68	778	32
04	2016	11	11	100	00	00
05	2012	376	168	45	208	55
<b>Total</b>		<b>3120</b>	<b>1872</b>	<b>60</b>	<b>1248</b>	<b>40</b>

Table 4 shows active and missing web citation URLs. Out of total web citations in Ph.D. theses, 40% of web citations are missing URLs, and 60 percent web citations are active. The highest number of active web citation URLs is 100 percent in the uploaded theses in 2016 and almost

every year 60 to 68 percent of URLs are active. The lowest number of active web citations found in 2012 is percent is 45%. The maximum number of missing URLs found during the year 2012 percent of which is 55%, followed by 2019 (40%), 2018 (39%), and 2017 (32%).

**Table 5 : URLs Scheme-wise yearly distribution of Web Citation**

Sr. No.	URL Scheme	Uploaded Year					Total	Percentage
		2019	2018	2017	2016	2012		
1	HTTP	111	176	2401	11	369	3068	98
2	HTTPS	13	22	10	00	07	52	02
<b>Total</b>		<b>124</b>	<b>198</b>	<b>2411</b>	<b>11</b>	<b>376</b>	<b>3120</b>	<b>100</b>

Table 5 shows the distribution of the URLs scheme in the Ph.D. Theses. The Majority of the URLs using the HTTP scheme percentage of

which is 98% and a small number 2 percent HTTPS URLs scheme are used.

**Table 6 : URLs Scheme-wise yearly distribution of Active and Missing Web Citation**

Uploaded Year	HTTP					HTTPS				
	Total No. Web citation	Active Web citation	%	Missing Web Citation	%	Total No. Web citation	Active Web citation	%	Missing Web Citation	%
2019	111	63	57	48	43	13	12	92	01	08
2018	176	102	58	74	42	22	18	82	04	18

2017	2401	1626	68	775	32	10	07	70	03	30
2016	11	11	100	00	00	00	00	00	00	00
2012	369	168	45.52	201	54	07	00	00	07	100
<b>Total</b>	<b>3068</b>	<b>1970</b>	<b>64</b>	<b>1098</b>	<b>36</b>	<b>52</b>	<b>37</b>	<b>71</b>	<b>15</b>	<b>29</b>

Table 6 shows the URLs scheme year-wise distribution of active and missing web citations. A total number of 3068 URLs on HTTP schemes, out of which 1970 (64%) were active and 1098

(36%) URLs are missing. HTTPS URLs scheme found in a very small in number i.e. 52 out of which 71 percent of URLs are active and 29 percent of URLs were missing.

**Table 7 : URLs Scheme distribution of Web Citation**

Sr. No.	URLs Status	HTTP	%	HTTPS	%	Total Web citation	%
1	Accessible	1970	60	37	71	2007	64
2	Inaccessible	1098	40	15	29	1113	36
<b>Total</b>		<b>3068</b>	<b>100</b>	<b>52</b>	<b>100</b>	<b>3120</b>	<b>100</b>

Table 7 shows the distribution of the URLs scheme in the set of 3120 web citations. A total number of 3068 URLs are using the HTTP

scheme of total 40% of URLs are inaccessible. And in case of HTTPS 29% of URLs are inaccessible.

**Table 8 : Description of web citations in Library and Information Science Theses**

Sr. No.	Domain	Total No. Web citation	Web citation With only URL	%	Web citation With URL and Partial Bibliographical Information	%	Web citation With URL and Complete Bibliographical Information	%
1	2019	124	31	25	76	61	17	14
2	2018	198	70	35	21	11	107	54
3	2017	2411	305	13	573	24	1533	63
4	2016	11	08	73	3	27	0	0
5	2012	376	7	2	56	15	313	83
<b>Total</b>		<b>3120</b>	<b>421</b>	<b>13.49</b>	<b>729</b>	<b>23.36</b>	<b>1970</b>	<b>63.14</b>

Table 8 shows the types of web citations in the Ph.D. Theses such as web citations with complete Bibliographical Information, partial Bibliographical Information and lastly only URL. Overall maximum of 63.14 percent of web

citations are URL and complete Bibliographical Information, 23.36% of web citations are URL and partial Bibliographical Information and lastly, only 13.49 percent of web citations are only URL Information available in Ph.D. Theses.

**Table 9 :HTTP Error Codes Associated with Missing URLs**

Sr. No.	Error Codes	Year					Total	%
		2012	2016	2017	2018	2019		
1	HTTP 400	00	00	02	00	00	02	0.17
2	HTTP 404	111	00	348	26	25	510	46
3	HTTP 403	4	00	08	05	01	18	1.61
4	HTTP 405	0	00	01	00	01	02	0.17
5	HTTP 410	00	00	01	00	00	01	0.08
6	HTTP 500	0	00	02	00	01	03	0.26
7	HTTP 502	00	00	01	00	00	01	0.08
8	HTTP 503	0	00	02	02	00	04	0.35
9	HTTP 522	00	00	01	00	00	01	0.08
10	Denied access	26	00	55	15	09	105	9.43
11	Too many redirections	1	00	10	10	00	21	1.88
12	The server name or address could not be resolved	49	00	272	19	10	350	31.44
13	The hostname in the certificate is invalid or does not match	02	00	05	00	00	07	0.62
14	The Certificate Authority is invalid or incorrect	00	00	01	00	00	01	0.08
15	Invalid URL	1	00	18	1	00	20	1.79
16	Timeout	13	00	45	00	02	60	5.39
17	The abnormal	1	00	01	00	00	02	0.17
18	No Mapping	00	00	02	00	00	02	0.17
19	Compressed grizp	00	00	03	00	00	03	0.26
<b>Total</b>		<b>208</b>	<b>00</b>	<b>778</b>	<b>78</b>	<b>49</b>	<b>1113</b>	<b>100</b>

Table 9 show that 19 types of error codes associated with missing web citations were encountered during the study period. 'HTTP 404' is encountered in the majority of missing URLs presenting 46% of all error codes. A total number of 31.44 % missing URL citations shows 'The server name or address could not be resolved', 9.43% 'Denied access' error codes followed by 5.39% 'Timeout' error codes, 'Too many redirections' 1.88%, 'invalid' 1.79% and 'HTTP 403' is 1.61%, 'HTTP 503' is 0.35%, HTTP 500 and Compressed grizp error codes is 0.26%,

HTTP 400, HTTP 405, The abnormal and No Mapping error codes is 0.17%, HTTP 410, HTTP 502, HTTP 522 and The Certificate error codes is 0.08%.

### FINDINGS AND CONCLUSION

A total no. of 17823 citations of 79 theses extracted from Library and Information Science on Web Citation and Decay of URLs were analyzed and the following conclusions are drawn.

1. A total no of 17823 citations was identified from the theses, out of 82% are print citations, and the rest 18% are web citations.
2. The highest no. of theses (85%) was uploaded in 2017 and only 15% of theses uploaded during 2012, 2016, 2018, and 2019.
3. The highest number of web citations domain type .org is 33.49% followed by .com domain type at 18.30%.
4. The active and missing web citations URLs : A total number of 3120 URLs testing were found 60 percent of web citations are active and 40% of web citation URLs are missing.
5. The distributions of URL schemes like http and https : The majority of URLs schemes used the web citations are http scheme 98 percent and only 2 percent of web citations is https URLs scheme.
6. They also examine the active and missing URLs errors encounters, so it was found the majority of errors were due to missing content 46 percent of URLs are HTTP 404 (File not found) and followed by 'The server name or address could not be resolved' 31% errors code.

### SUGGESTIONS

- Based on the findings of the study, the following suggestions are put forward to ensure the research output.
- The presented research shows that 36 percent of the web citations are inaccessible, to further reduce this rate researchers check URLs before submitting research report and check spelling mistakes.
- 40 percent of the web citations in the presented research report are web citations with only the URL and web citations with the

URL and partial bibliographical information, therefore, the researcher mentions that the entire web content should be complete and include full bibliographic information as well as the date on which the website was accessed by researchers.

### REFERENCES

1. Bansal, S. and Parmar, S. (2020). Decay of URLs Citation: A Case study of Current Science. *Library Philosophy and Practice*, 3582. <https://digitalcommons.unl.edu/libphilprac/3582>
2. Gul, S. Iram M. and Asifa A. (2019). The Growth and decay of URLs citations: a case of an online Library and Information Science Journal. *Malaysian Journal of Library and Information Science*, 19(3), 27-39.
3. Moghaddm Alireza Isfandyari and Mohammad Karim Saberi (2011). The Life and Death of URLs: The Case of Journal of the Medical Library Association. *Library Philosophy and Practice*, <https://digitalcommons.unl.edu/libphilprac/592>
4. Parmar, S. and Pateria, R. K. (2019). Web Citations and Decay of URLs: A Case study of Indian Journal of Agricultural Library and Information Services. *Library Philosophy and Practice*, 3595. <https://digitalcommons.unl.edu/libphilprac/3595>
5. <http://shodhganga.inflibnet.ac.in/newmoredetails/about.html> Retrieved December 01, 2021.
6. [http://www.unipune.ac.in/university\\_files/about\\_university.htm](http://www.unipune.ac.in/university_files/about_university.htm) Retrieved December 01, 2021

