Asian Journal of Information Science and Technology
ISSN: 2231-6108 Vol. 10 No.2, 2020, pp. 51-61

A Study of Availability and Recovery of URLs in Library and Information Science Scholarly journals

B. Niveditha[1] and Mallinath Kumbar[2]
[1]UGC-Junior Research Fellow, [2]Professor
[1&2]Department of Library and Information Science, University of Mysore, Manasagangotri, Mysuru, Karnataka, India
E-mail: niveditha.jb@gmail.com

Abstract **- The present study examines the availability and recovery of web references cited in scholarly journals selected based on their high impact factor published between 2008 and 2017. A PHP script was used to crawl the Uniform Resource Locators (URL) collected from the references. A total of 5720 articles were downloaded and 237418 references were extracted. A total of 33512 URLs were checked for their availability. Further the lexical features of URLs like file extension, path depth, character length and top-level domain was determined. The research findings indicated that out of 33512 web references, 20218 contained URLs, DOIs were found in 12799 references and 495 references contained arXiv or WOS identifier. It was found that 29760 URLs were accessible and the remaining 3752 URLs were missing. Most errors were due to HTTP 404 error code (Not found error). The study also tried to recover the inaccessible URLs through Time Travel. Almost 60.55% of inaccessible URLs were archived in various web archives. The findings of the study will be helpful to authors, publishers, and editorial staff to ensure that web references will be accessible in future.**
Keywords**: References, web references, URLs, DOIs, HTTP error, PHP, Time Travel.**

I. INTRODUCTION

The Internet has been portrayed as information superhighway as it contains vast repository of information. This has led to the use of electronic resources by the academic community and they have begun to reap the advantages of the Internet. It is in fact changing the ways in which the academic community seeks information and conduct research. Nevertheless, there is a gained momentum in authors of academic community citing URL links and DOIs in their research papers. The focus now is to determine the credibility of those cited web links or web references. The credibility of web references not only facilitate constant information transfer to other researchers, but also can result in enhanced academic productivity. In this context, the present study has tried to check the availability of URLs in five journals during the period 2008-2017 using a PHP script. It also aims at recovering inaccessible URLs cited in scholarly articles through Time Travel.

II. OBJECTIVES OF THE STUDY

1. To explore the proportion of URLs and DOIs used in scholarly LIS journals
2. To know the percentage of inaccessible URLs using PHP script.
3. To examine the lexical features of URLs like path depth, top-level domain, and character length.
4. To differentiate the lexical features of display URLs and destination URLs
5. To recover the inaccessible URLs through Time Travel.

III. HYPOTHESES

1. Web references are most cited in scholarly communications during 2008-2017.
2. URL permanence will increase as their age decrease.
3. The path depth and percentage of inaccessible URLs are positively correlated.

IV. METHODOLOGY

For the present study, data was drawn from five leading Library and Information Science scholarly journals. The journals were selected based on their high impact factor as per Clarivate Analytics' 2018 "Journal Citation Report." The journals selected for the current study are:

1. Journal of Informatics (JOI): This journal published by Elsevier focuses on quantitative aspects of Information Science. It has an impact factor of 3.484.
2. Journal of the Association for Information Science and Technology (JASIST): This journal published by Wiley Online Library has an impact factor of 2.835. The focus of this journal is to publish original research that covers storage, retrieval, dissemination and use of information.
3. Scientometrics: Scientometrics published by Springer Link mainly aims to publish works which endeavor quantitative features of scientific research. It has an impact factor of 2.173.
4. College and Research Libraries (CRL): This journal with impact factor 1.626 is the official research publication of Association of College and Research Libraries, which is a division of American Library Association.

5. Aslib Journal of Information Management (AJIM): This journal with impact factor of 1.461 published by Emerald covers broad range of topics including social media, information retrieval, digital libraries, etc.

All the research articles published during the 10-year period, that is, from 2008 to 2017 were taken up for the study. Editorial notes, book reviews, short communication were excluded. The references that were adjoined at the end of each article were considered for the study. A total of 237418 references were selected from 5720 articles published in the five journals. The references that contained web links and DOIs were extracted as the study deals with their accessibility. The DOIs and arXiv identifier were first resolved to URLs using the syntax https://doi.org/. For example, a DOI name 10.1010.1234/567 would be resolved from the address https://doi.org/10.1010.1234/567. Similarly, arXiv identifier was resolved to URLs using the syntax https://arxiv.org/. A total of 33512 URLs were extracted for checking their availability.

A PHP script was developed to test bulk URLs. The script uses CURL library, a standard PHP extension to check for URL availability and documents the error code associated with inaccessible URLs. Apart from checking the URLs, the script obtains the destination URL. The web address which is displayed to the user regardless of the article's physical location is the Display URL. Destination URL is the URL that after multiple redirects goes to the landing page of the article or where the article resides, which is under the control of the publisher. The lexical features of URLs like their length, top-level domain, and path depth are also extracted.

The study used Time Travel (http://timetravel.mementoweb.org/) to find whether the URLs were archived or not. The Time travel recovers the inaccessible URLs that are archived in Internet Archive, Library of Congress Web Archive, Archive-it, Perma-cc, etc. The URLs that were not archived were considered as decayed URLs.

V. DATA ANALYSIS AND INTERPRETATION

A. Year-Wise Distribution of Articles, References and Web References

A total of 5720 articles published in five LIS scholarly journals during the period 2008-2017 were examined. The articles contained a total of 2, 37, 418 references with 14.12% (33,512) of references citing a web source. Table I summarizes the citation results for the 5720 research articles. The number of references and web references in journal articles are positively co-related and the relations is statistically significant ($r = .948$, $p = .000$). This was performed using Pearson's Correlation analysis. The average number of web reference per article has been increased substantially from 3.47 in the year 2008 to 9.40 in the year 2017. The percentage of web reference by year is varied from a low of 9.92 in the year 2010 to a high of 19.78 in the year 2017. This shows that the volume of web references in the research journals is not consistent during the 10-year period. The statistical relation shows that even though there is a negative correlation between the age and percentage of web references cited ($r = -0.884$) the relation is statistically significant ($p = 0.000$). This shows that the percentage of web references in articles has been continuously increased from 2008 to 2017.

TABLE I YEAR-WISE DISTRIBUTION OF ARTICLES, REFERENCES AND WEB REFERENCES

Year	Total number of articles	Total number of references	Average reference per article	Total number of web references	Average web reference per article	Percentage of web references
2008	384	13267	34.55	1332	3.47	10.04
2009	478	17155	35.89	1931	4.04	11.26
2010	517	18917	36.59	1878	3.63	9.93
2011	499	19019	38.11	2027	4.06	10.66
2012	541	20717	38.29	2664	4.92	12.86
2013	574	23574	41.07	2820	4.91	11.96
2014	657	27679	42.13	3540	5.39	12.79
2015	675	31162	46.17	4481	6.64	14.38
2016	697	32754	46.99	6277	9.01	19.16
2017	698	33174	47.53	6562	9.4	19.78
Total	5720	237418	41.51	33512	5.86	14.12

B. Journal-Wise Distribution of Articles, References and Web References

Table II reflects that a total of 5720 articles were published in the five journals during the years 2008-2017. More number of articles were published in Scientometrics (2575), followed by JASIST (1744) and JOI (647). The average reference per article was high in JASIST (50.15) and low in CRL (35.57). The percentage of web references also varied among journals. Highest percentage of web references were noticed in CRL (20.35%) and lowest percentage of web references were found in JASIST (11.78%).

TABLE II JOURNAL-WISE DISTRIBUTION OF ARTICLES, REFERENCES AND WEB REFERENCES

Journal	Total number of articles	Total number of references	Average reference per article	Total number of web references	Average web reference per article	Percentage of web references
JOI	647	24901	38.49	3546	5.48	14.24
JASIST	1744	87468	50.15	10301	5.91	11.78
Scientometrics	2575	96137	37.33	14261	5.54	14.83
CRL	376	13376	35.57	2722	7.24	20.35
AJIM	378	15536	41.1	2682	7.1	17.26
Total	5720	237418	41.51	33512	5.86	14.12

C. Year-Wise Distribution of URLs and DOIs

The web reference permanence is of major concern to academicians and the DOIs are used to prevent their decay. The DOI is a character string that is used to identify a scholarly publication in the digital environment. Table III shows the distribution of URLs and DOIs in the five scholarly journals. It was found that out of the total 33,512 web references, 20,218 were URL links, 12,799 were DOIs and 495 were arXiv identifier and WOS identifier.

TABLE III YEAR-WISE DISTRIBUTION OF URLs AND DOIs

Year	URL		DOI		Others		Total web references
	Number	%	Number	%	Number	%	
2008	1303	97.82	12	0.9	17	1.28	1332
2009	1828	94.67	83	4.3	20	1.04	1931
2010	1710	91.05	122	6.5	46	2.45	1878
2011	1539	75.93	465	22.94	23	1.13	2027
2012	1928	72.37	709	26.61	27	1.01	2664
2013	1892	67.09	894	31.7	34	1.21	2820
2014	2141	60.48	1356	38.31	43	1.21	3540
2015	2292	51.15	2095	46.75	94	2.1	4481
2016	3099	49.37	3101	49.4	77	1.23	6277
2017	2486	37.88	3962	60.38	114	1.74	6562
Total	20218	60.33	12799	38.19	495	1.48	33512

D. Year-Wise Distribution of Accessible, Inaccessible and Recovered URLs

The DOIs and arXiv identifiers, which were resolved to URLs were tested for their availability and this is depicted in table IV. The result of the accessibility check by year indicated that of the 33512 URLs, 89.80% were accessible while the remaining 11.20% encountered accessibility error. The percentage of inaccessible URLs varied from a low of 5.15 in the year 2017 to a high of 27.03 in the year 2008. To know the correlation between the age and inaccessible URLs, Pearson's Correlation analysis was performed. It was found that there is positive correlation between the age and inaccessible URLs and the correlation was statistically significant ($r = 0.984$, $p = 0.000$). The table also depicts the percentage of recovered URLs by year through Time Travel. A total 60.55% of URLs were archived in various web archives. The percentage of recovered URLs varied from a low of 52.50 in the year 2008 to a high of 68.51 in the year 2013. The correlation analysis indicates that the percentage of recovered URLs and the age are negative correlated ($r = -0.418$, $p = 0.229$) and the co-relation is not statistically significant.

TABLE IV YEAR-WISE DISTRIBUTION OF ACCESSIBLE, INACCESSIBLE AND RECOVERED URLs

Year	Total URLs	Accessible URLs	%	Inaccessible URLs	%	Recovered URLs	%
2008	1332	972	72.97	360	27.03	189	52.5
2009	1931	1450	75.09	481	24.91	291	60.5
2010	1878	1478	78.7	400	21.3	235	58.75
2011	2027	1701	83.92	326	16.08	207	63.5
2012	2664	2222	83.41	442	16.59	241	54.52
2013	2820	2477	87.84	343	12.16	235	68.51
2014	3540	3217	90.88	323	9.12	217	67.18
2015	4481	4091	91.3	390	8.7	243	62.31
2016	6277	5928	94.44	349	5.56	203	58.17
2017	6562	6224	94.85	338	5.15	211	62.43
Total	33512	29760	88.8	3752	11.2	2272	60.55

E. Journal-Wise Distribution of Accessible, Inaccessible and Recovered URLs

The summary of journal wise accessible and inaccessible URL is presented in the Table V. 11.20% of URLs were inaccessible in the five journals. Further 18.98% of URLs were inaccessible in AJIM, followed by 13.32% in JASIST and 11.50% in CRL. Almost 70.26% of URLs from JASIST were archived from a total of 1372 inaccessible URLs. It was also noted that only 28.20% were archived in JOI.

TABLE V JOURNAL WISE DISTRIBUTION OF ACCESSIBLE, INACCESSIBLE AND RECOVERED URLs

Journal	Total URLs	Accessible URLs	%	Inaccessible URLs	%	Recovered URLs	%
JOI	3546	3163	89.2	383	10.8	108	28.2
JASIST	10301	8929	86.68	1372	13.32	964	70.26
Scientometrics	14261	13086	91.76	1175	8.24	713	60.68
CRL	2722	2409	88.5	313	11.5	206	65.81
AJIM	2682	2173	81.02	509	18.98	281	55.21
Total	33512	29760	88.8	3752	11.2	2272	60.55

F. Distribution of Http Error Codes Associated with Inaccessible and Recovered URLs

The various error codes that are encountered for inaccessible URLs are presented in the Table VI. The HTTP 404 error message "page not found" represented 74.55% of all HTTP error message and it is followed by HTTP 403 "forbidden error" (13.38%), HTTP 500 "internal server error" (5.22%) and HTTP 400 (3.38%). The 3572 inaccessible URLs showing various HTTP errors were entered in the search box of Time Travel. Nearly half of the inaccessible URLs (60.55%) were archived in various web archives and were recovered from Time Travel. A total of 2272 URLs could be retrieved successfully. The percentage of recovered URLs with respect to various HTTP errors is shown in Table 6. It was interesting to note that 58.60% URLs recovered were from HTTP 404 error message, 59.56% URLs recovered were from HTTP 403 error message, 74.49% URLs recovered were from HTTP 500 error message and 63.78% of them were recovered from HTTP 400 error message.

TABLE VI DISTRIBUTION OF HTTP ERROR CODES

Error Codes	Inaccessible URLs	%	Recovered URLs	%
400	127	3.38	81	63.78
401	2	0.05	1	50
403	502	13.38	299	59.56
404	2797	74.55	1639	58.6
406	5	0.13	3	60
408	3	0.08	1	33.33
409	2	0.05	1	50
410	30	0.8	26	86.67
412	2	0.05	1	50
416	5	0.13	5	100
418	3	0.08	3	100
429	5	0.13	5	100
463	2	0.05	0	0
479	1	0.03	1	100
500	196	5.22	146	74.49
502	11	0.29	9	81.82
503	54	1.44	46	85.19
504	2	0.05	2	100
521	1	0.03	1	100
530	2	0.05	2	100
Total	3752	100	2272	60.55

G. Distribution of Archived URLs in Time Travel

Table VII shows the distribution of archived URLs in Time Travel. The Internet Archive recovered the highest percentage of inaccessible URLs with almost 2148 URLs archived, followed by Arquivo.pt which has archived 533 URLs and Archive .is which has archived 503 URLs.

TABLE VII DISTRIBUTION OF ARCHIVED URLs IN TIME TRAVEL

Year	Total no. of recovered URLs	Internet Archive	LOC	Archive it	perma.cc	archive.is	arquivo.pt	Stanford web archive	Icelandic web archive	UK web archive	Web citation memento	Bibliotheca Alexandrina	Canadian Archive Memento	UK Government web archive
2008	189	167	25	18	4	29	28	6	7	9	32	17	4	6
2009	291	269	39	25	9	44	47	10	12	12	37	24	8	9
2010	235	220	35	18	4	35	47	6	9	6	35	22	4	11
2011	207	188	26	21	7	46	36	9	11	7	36	21	6	11
2012	241	231	49	41	32	55	53	31	33	29	59	38	29	31
2013	235	234	54	46	39	64	72	36	39	34	59	42	34	36
2014	217	204	32	28	16	38	43	17	19	16	46	19	16	15
2015	243	233	72	59	46	69	74	47	44	44	70	52	42	43
2016	203	205	73	70	62	77	85	60	60	57	71	61	57	58
2017	211	197	48	38	36	46	48	32	34	31	52	42	29	30
Total	2272	2148	453	364	255	503	533	254	268	245	497	338	229	250

H. File Extension Associated with Inaccessible and Recovered URLs

The data as illustrated in Table VIII indicates that the greatest numbers of cited URLs citations are .html files. Out of 33512 URLs, 26740 are .html files, followed by 5026 .php files, and 425 are .asp files. File format having the highest percent of inaccessible URLs was the .pdf (35.00%), followed by .cgi (28.24%). Low level of loss was associated with the .html (7.92%) and .cfm (12.92%). Table 8 also indicates the percentage of recovered URLs with respect to their file format. 64.62% of .html files, 61.29% of .cfm files, 58.23% of .asp files and 55.56 percent of .jsp files were recovered from Time Travel.

.

TABLE VIII FILE EXTENSION ASSOCIATED WITH INACCESSIBLE AND RECOVERED URLs

File Extension	Total URLs	Accessible URLs	%	Inaccessible URLs	%	Recovered URLs	%
.asp	425	346	81.41	79	18.59	46	58.23
.cfm	240	209	87.08	31	12.92	19	61.29
.cgi	85	61	71.76	24	28.24	11	45.83
.html	26740	24623	92.08	2117	7.92	1368	64.62
.jsp	85	67	78.82	18	21.18	10	55.56
.pdf	140	91	65	49	35	24	48.98
.php	5026	3716	73.94	1310	26.06	725	55.34
Others	771	647	83.92	124	16.08	69	55.65
Total	33512	29760	88.8	3752	11.2	2272	60.55

I. Path Depth of Display and Destination URLs

Table IX summarizes the path depth of display and destination URL. Out of 33512 URLs, display URLs with path depth 2 (54.56%) were frequently cited, followed by URLs with path depth of 3 (15.33%) and path depth 4 (9.88%). Unlike the display URL, only 27.26% destination URLs had a path depth of 2. There was an increase in destination URLs having path depth of 3 (21.37%) and path depth 4 (20.04%).

TABLE IX PATH DEPTH OF DISPLAY AND DESTINATION URLs

Path Depth	Display URL	%	Destination URL	%
PD = 0	899	2.68	396	1.18
PD = 1	2457	7.33	3468	10.35
PD = 2	18284	54.56	9135	27.26
PD = 3	5138	15.33	7163	21.37
PD = 4	3310	9.88	6717	20.04
PD = 5	1736	5.18	4040	12.06
PD = 6	909	2.71	1436	4.29
PD = 7	445	1.33	797	2.38
PD>7	334	1	360	1.07
Total	33512	100	33512	100

J. Path Depth Associated with Inaccessible and Recovered URLs

Table X shows that out of 33512, URLs with path depth 2 (18284) were most frequently cited, followed by URLs with path depth 3 (5138) and path depth 4 (3310). Studies in the past[15,28] have indicated that increased URL depth was associated with accessibility problems. But this study showed that highest percentage of inaccessible URLs (19.45%) were found in the URLs associated with path depth 1, followed by URLs with path depth 6 (18.59%) and path depth 3 (18.88%). The Table also indicates the percentage of recovered URLs from the Time Travel. It indicates that URLs (74.60%) with path depth 0 were recovered the most, followed by URLs with path depth 1 (67.15%) and path depth 2 (63.92%). To know the relationship between the path depth of the URLs and the percentage of inaccessible URLs, Pearson's Co-relation analysis was performed. It is found that the path depth and the percentage of inaccessible URLs are positively correlated ($r = 0.049$, $p = 0.900$), but the relation is not statistically significant. The percentage of recovered URLs and the path depth are negatively correlated ($r = -0.784$, $p = 0.012$), and the relation is statistically significant.

TABLE X PATH DEPTH ASSOCIATED WITH INACCESSIBLE AND RECOVERED URLs

Path Depth	Total URLs	Accessible URLs	%	Inaccessible URLs	%	Recovered URLs	%
PD = 0	899	773	85.98	126	14.02	94	74.6
PD = 1	2457	1979	80.55	478	19.45	321	67.15
PD = 2	18284	17253	94.36	1031	5.64	659	63.92
PD = 3	5138	4168	81.12	970	18.88	561	57.84
PD = 4	3310	2728	82.42	582	17.58	326	56.01
PD = 5	1736	1450	83.53	286	16.47	156	54.55
PD = 6	909	740	81.41	169	18.59	98	57.99
PD = 7	445	376	84.49	69	15.51	43	62.32
PD>7	334	293	87.72	41	12.28	14	34.15
Total	33512	29760	88.8	3752	11.2	2272	60.55

K. Character Length of Display and Destination URLs

Table XI shows the URL length and it can be found that a total of 12408 display URLs had length 41-50, 7311 URLs had length of 31-40, and 3894 URLs had a length of 51-60. When the landing page of URL is reached, a total of 7111 URLs had character length of 61-70, followed by 6954 URLs with length 51-60 and 4994 URLs with character length 41-50.

TABLE XI CHARACTER LENGTH OF DISPLAY AND DESTINATION URLs

Character length	Display URL	%	Destination URL	%
<20	300	0.9	360	1.07
21-30	1822	5.44	1664	4.97
31-40	7311	21.82	3504	10.46
41-50	12408	37.03	4994	14.9
51-60	3894	11.62	6954	20.75
61-70	2719	8.11	7111	21.22
71-80	1840	5.49	2920	8.71
81-90	1251	3.73	3489	10.41
91-100	721	2.15	918	2.74
>100	1246	3.72	1598	4.77
Total	33512	100	33512	100

L. Character Length Associated with Inaccessible and Recovered URLs

Table XII shows the percentage of accessible, inaccessible, and recovered URLs. URLs with URL length 61-70 were found to be inaccessible more (23.43%), followed by URLs with character length 91-100 (20.53%) and 51-60-character length (20.49%). To know the relation between percentage of vanished URLs and the character length, Pearson's correlation analysis was performed.

It was found that there is positive correlation between percentage of inaccessible URLs and the character length, and this relation is statistically significant ($r = 0.704$, $p = 0.023$). This clearly indicates that a greater number of characters in an URL leads to its decay. The table also illustrates the percentage of recovered URLs. Majority of missing URLs were recovered from those URLs having 21-30-character length (77.29%), followed by 31–40-character length (73.29%) and URLs having less than 20 characters (66.67%). The statistical relation shows that even though there is a negative correlation between the percentage of recovered URLs and the character length ($r = -0.922$), the relation is statistically significant ($p = .000$).

M. Top-Level Domain of Display and Destination URLs

The top-level domain associated with the display and destination URL is summarized in table XIII.
A total of 20083 display URLs had the organizational top-level domain, followed by 4510 having the commercial top-level domain. On the other hand, a total of 14957 destination URLs have commercial top-level domain followed by 9392 organizational top-level domains.

N. Top-Level Domain Associated with Inaccessible and Recovered URLs

The analysis of total and inaccessible URLs by type of top-level domain is shown in table XIV. Six main types of top-level domain have been considered in this study. They are .com, .edu, .gov, .info, .net, and.org. The top-level domain like .int, .mil and all the country top-level domains were considered in the "Others" category. The top-level domain having the greatest number of inaccessible URLs was the information top-level domain (.info) (46.09%) followed by educational (.edu) top-level domain (20.39%).

A noteworthy finding is that proportionally low level of loss was associated with organizational (.org) top-level domain (4.58%). The Table also shows the top-level domains associated with the percentage of recovered URLs. The top-level domain having the greatest number of

recovered URLs was the governmental top-level domain (.gov) (67.42%), followed by organizational top-level domain (67.25%). The low level of recovered URLs is associated with network (.net) top-level domain (44.78%) and educational top-level domain (58.09%).

TABLE XII CHARACTER LENGTH OF INACCESSIBLE AND RECOVERED URLs

Character length	Total URLs	Accessible URLs	%	Inaccessible URLs	%	Recovered URLs	%
<20	300	267	89	33	11	22	66.67
21-30	1822	1593	87.43	229	12.57	177	77.29
31-40	7311	6873	94.01	438	5.99	321	73.29
41-50	12408	11758	94.76	650	5.24	417	64.15
51-60	3894	3096	79.51	798	20.49	463	58.02
61-70	2719	2082	76.57	637	23.43	387	60.75
71-80	1840	1511	82.12	329	17.88	184	55.93
81-90	1251	1040	83.13	211	16.87	103	48.82
91-100	721	573	79.47	148	20.53	72	48.65
>100	1246	967	77.61	279	22.39	126	45.16
Total	33512	29760	88.8	3752	11.2	2272	60.55

TABLE XIII TOP-LEVEL DOMAIN OF DISPLAY AND DESTINATION URLs

Top-level domain	Display URL	%	Destination URL	%
.com	4510	13.46	14957	44.63
.edu	2212	6.6	2377	7.09
.gov	972	2.9	991	2.96
.info	115	0.34	120	0.36
.net	906	2.7	847	2.53
.org	20083	59.93	9392	28.03
Others	4714	14.07	4828	14.41
Total	33512	100	33512	100

O. Testing of Hypotheses

Table XV illustrates the formulated hypotheses, statistical test applied to verify the hypotheses and the results. It can be seen from the table that only one hypothesis was not supported by the study results.

TABLE XIV TOP-LEVEL DOMAIN OF INACCESSIBLE AND RECOVERED URLs

Top-level domain	Total URLs	Accessible URLs	%	Inaccessible URLs	%	Recovered URLs	%
.com	4510	3777	83.75	733	16.25	430	58.66
.edu	2212	1761	79.61	451	20.39	262	58.09
.gov	972	840	86.42	132	13.58	89	67.42
.info	115	62	53.91	53	46.09	34	64.15
.net	906	772	85.21	134	14.79	60	44.78
.org	20083	19164	95.42	919	4.58	618	67.25
Others	4714	3384	71.79	1330	28.21	779	58.57
Total	33512	29760	88.8	3752	11.2	2272	60.55

TABLE XV TESTING OF HYPOTHESES

SL.No	Hypotheses	Statistical test	p value	Result
H1	Web references are most cited in scholarly communications during 2008-2017	Co-relation	0	Supported
H2	URL permanence will increase as their age decrease.	Co-relation	0	Supported
H3	The path depth and percentage of inaccessible URLs are positively correlated.	Co-relation	0.9	Not supported

VI. CONCLUSION

The internet has facilitated exchange of scientific information through the World Wide Web. URLs may contribute directly to research through rapid information transfer. The present study confirms the use of URLs in the references cited in five journals during the year 2008-2017. The stability of URLs has hampered information access in the Web. The URLs become worthless if they wane when they move to a new location or change their content. It is obvious from the present study that URL decay can be reduced to some extent using Digital Object Identifier (DOI).

To overcome the problem of inaccessibility of URLs some suggestions are needed to be implemented. The authors should check the URL links that they use in their references. The inaccessible URLs should be requisitely removed or updated by the authors. Apart from the authors, it is also the responsibility of the editors and publishers to check the availability of the URLs before their publication. The authors as well as publishers should use web archives to recover the inaccessible URLs. Hence, it is the responsibility of the authors, publishers, and editorial team to make sure that the cited resources in the scholarly work can be available to the future researchers without any impediment.

REFERENCES

[1] Dimitrova, D. V., & Bugeja, M. (2007). The half-life of internet references cited in communication journals. *New Media & Society*, 9(5), 811–826.

[2] Goh, D.H., & Ng, P.K. (2007). Link decay in leading information science journals. *Journal of the American Society for Information Science and Technology*, 58(1), 15–24. https://doi.org/10.1002/asi.20513

[3] Jalalifard, M., Norouzi, Y., & Isfandyari-Moghaddam, A. (2013). Analyzing web references availability and half-life in medical journals: A case study in an Iranian university. *Aslib Proceedings*, 65(3), 242–261.

[4] Prithvi Raj, K. R., & Sampath Kumar, B. T. Web Citation Trends in Indian LIS Journals: A Citation Analysis. (2015). *COLLNET Journal of Scientometrics and Information Management*, 9(2), 295–310.

[5] Saberi, M. K., & Abedi, H. (2012). Accessibility and decay of web references in five open access ISI journals. *Internet Research*, 22(2), 234–247.

[6] Sadat-Moosavi, A., Isfandyari-Moghaddam, A., & Tajeddini, O. (2012). Accessibility of online resources cited in scholarly LIS journals: A study of Emerald ISI-ranked journals. *Aslib Proceedings*, 64(2), 178–192.

[7] Sampath Kumar, B. T., & Manoj Kumar, K. S. (2012). Persistence and half-life of URL citations cited in LIS open access journals. *Aslib Proceedings*, 64(4), 405–422.

[8] Sampath Kumar, B. T., & Prithvi Raj, K. R. (2012). Availability and persistence of web references in Indian LIS literature. *The Electronic Library*, 30(1), 19–32.

[9] Sampath Kumar, B. T., Vinay Kumar, D., & Prithvi Raj, K. R. (2015). Wayback machine: reincarnation to vanished online citations. *Program*, 49(2), 205–223.

[10] Sampath Kumar, B.T. and Vinay Kumar, D. (2013). "HTTP 404-page (not) found: recovery of decayed URL citations". *Journal of Informetrics*, 7(1), 145-157.

[11] Spinellis, D. (2003). The Decay and Failures of Web References. *Communications of the ACM*, 46(1), 71–77.

[12] Tajeddini, O., Azimi, A., Sadat-Moosavi, A., & Sharif-Moghaddam, H. (2011). Death of web references: a serious alarm for authors. *Malaysian Journal of Library and Information Science*, 16(3), 17-29

[13] Vinay Kumar, D. & Sampath Kumar, B. T. (2017). Finding the unfound: Recovery of missing URLs through Internet Archive. *Annals of Library and Information Studies*, 64(3), 165-171.

[14] Vinay Kumar, D., Sampath Kumar, B. T., & Parameshwarappa, D. R. (2015). URLs Link Rot: Implications for Electronic Publishing. *World Digital Libraries - An International Journal*, 8(1), 59–66.

[15] Wu, Z. (2008). An empirical study of the accessibility of web references in two Chinese academic journals. *Scientometrics*, 78(3), 481–503.

[16] Yang, S., Qiu, J., & Xiong, Z. (2010). An empirical study on the utilization of web academic resources in humanities and social sciences based on web references. *Scientometrics*, 84(1), 1–19.

[17] Zhang, Y. (2007). The Effect of Open Access on Citation Impact: A Comparison Study Based on Web Citation Analysis. *Libri*, 56(3), 145–156.

[18] Zhao, D. & Logan, E. (2002). Citation analysis using scientific publications on the Web as data source: A case study in the XML research area. *Scientometrics*, 54(3), 449-472.