AWARI
Vol. 5, 2024. 1-10, DOI: 10.47909/awari.644

Pro-Metrics

ORIGINAL ARTICLE

Web reference decay: Passing the HTTP errors instead of quality information

Akshaykumar Dundannanavar[I], **Gururaj S. Hadagali**[II]

[I] Department of Library and Information Science, Karnatak University, India.
Corresponding author.
Email: akshaykumardundannanavar@gmail.com
[II] Department of Library and Information Science, Karnatak University, India.

ABSTRACT

Web resource links routinely fail, and the content of links might change suddenly and without notice. When referrals to online sites offer proof or supporting data, these web dynamics are harmful. Any hyperlink's final content is susceptible to two phenomena: the link breaks (link rot) and the content changes from its original state (content drift). The combination of both effects is referred to as reference rot. This paper aims to determine the application of online citations, the number of lost web citations, and the process of recovering missing web citations through the application of the Internet Archive's Wayback Machine. According to the analysis, the *Webology* journal published 428 research articles between 2004 and 2020 that had 2,273 online citations. There are 1,260 (55.43%) accessible web citations out of the mentioned ones, whereas 1,013 (44.57%) are missing. There are 1,856 (81.65%) active web citations, up from 1,260 (55.43%) after an effort was made to retrieve 596 lost web citations using the Wayback Machine tool. That meant that the overall increase of the active URLs was 26.22%. The article presents a method for creating specific collections for online archiving in the future, even though web-archiving tools now take snapshots of websites in real time. The paper recommends that future curation standards for real-time web archiving shall include platform dynamics and cultural variations in link-sharing habits

Keywords: web decay; URLs rot; Wayback Machine; Internet Archives; HTTP errors.

1. INTRODUCTION

In an age where information is expanding exponentially in every facet of life, the academic sector is no exception. There is a boom of information in the shape of different kinds of documents available on the web in various formats. When many researchers prefer to read e-content, e-publishing of books, journals, and other documents is on the rise. Similarly, the number of citations from web resources has increased due to the influence of e-resources (Bansal & Parmar, 2020). Electronic resources have recently emerged as the preferred method for locating information on the Internet. The academic community's citation behavior has also changed as a result of this (Gul et al., 2014). They have begun to reference electronic resources in their scientific works, including their URL addresses or digital object identifiers (DOIs).

However, as online citations increased, the corrosion of URLs also rose. Link rot and content drift are two hallmarks of the web's dynamic nature. Link rot happens when linkages break over time, a problem that we have all

How to cite: Dundannanavar, A., & Hadagali, G. S. (2024). Web reference decay: Passing the HTTP errors instead of quality information. *AWARI; 5,* 1-10. DOI: 10.47909/awari.644.

Received: 28-07-2024 / **Accepted:** 18-09-2024 / **Published:** 30-09-2024
Copyright: © 2024 The author(s). This is an open access article distributed under the terms of the CC BY-NC 4.0 license which permits copying and redistributing the material in any medium or format, adapting, transforming, and building upon the material as long as the license terms are followed.

AWARI
Vol. 5, 2024. 1-10, DOI: 10.47909/awari.644

experienced on multiple occasions and that has been extensively investigated and quantified. Material drift or content drift occurs when the material at a certain web location changes over time, which we expect when acquiring information about the present state of affairs—for instance, weather, news, and stock prices (Jones et al., 2021). Because we frequently think that the most recent web material is the most desirable, the implications of content drift are not always obvious.

This paper focuses on reference rot in web-based references as a subset of the broader issue of web archives for knowledge storage/preservation. In such instances, a "404 not found" or "404 this page cannot be found" error message or, in more technical words, a HyperText Transfer Protocol (HTTP) error response may be displayed. Other notifications (with corresponding codes) may indicate restricted access or a problem in access: 400, bad request; 401, unauthorized; 403, disallowed; 404, not found; 405, method not permitted; 410, gone; 416, requested range not satisfied; 502, bad gateway; or 500, internal server error (Zittrain et al., 2014).

The temporary or irreversible loss of access to a URL causes information discontinuity and, hence, compromises citation-related data as well as information integrity in all forms of reference rot. As a result, libraries, academic institutions, and periodicals must preserve information. Some anti-reference rot measures or initiatives allow URLs to be saved at the Internet Archives such as CLOCKSS (https://clockss.org/), Perma (https://perma.cc/), Webrecorder (https://webrecorder.net/; now Conifer: https://conifer.rhizome.org/), Portico (www.portico.org/), and Wayback Machine (https://archive.org/index.php). Nwala et al. (2021) emphasized that historians and researchers qualitatively examine major sources of information about recent events from social networks and keep them in the form of web-archived collections. Thus, link rot has academic and practical ramifications since it reduces the capacity to assess the reliability of papers with web-based information. The archive of links by authors is one solution. With this context in mind, the authors carried out this study to evaluate the availability and accuracy of references in articles published in *Webology* journal from 2004 to 2020, a Scopus-indexed journal.

2. REVIEW OF LITERATURE

Researchers have known that link rot is one of the most infamous issues on the web since its inception. Numerous investigations on diverse datasets have been carried out throughout history. There has been extensive research on the subject of reference rot in papers and online (Jacob, 1998). Tajeddini et al. (2011) discovered 4,562 online citations cited in 1,109 works published in six *Library and Information Science* scholarly publications. After using various strategies, the rate of inaccessible URLs was lowered to 5% from 34%. They extracted 11% of the unavailable URLs using the Internet Archive. Sife and Lwoga (2017) collected 574 web citations referenced in 822 papers published in four East African health science online journals between 2001 and 2015. The authors discovered that 253 (44.1%) of the site citations were unavailable. According to this investigation, just 36 (6.3%) web citations were recovered using the Wayback Machine.

Parmar and Pateria (2019) conducted a study to identify the current citation trends, the strength of web citations or URLs, and decay and half-life of URLs by analyzing the citations of articles published in the *Indian Journal of Agricultural Library and Information Services* between 2012 and 2016. A total of 980 references were reported in 94 papers, with 325 (33.16%) being web citations. At the time of testing, 202 (62.15%) of online citation URLs were accessible, whereas the remaining 123 (37.85%) of URLs were inaccessible. The irresistible error message 404 "page not found" arrived and comprised 51.22% of all HTTP error messages. URLs have an estimated half-life of 4.62 years. There was no correlation discovered between path depth and URL degradation. Articles that accessed using the domains .net, .org, .com, and .co were shown to be livelier than those accessed via other domains. Jalali (2019) investigated the accessibility and degradation of web citations in refereed publications from 27 computer science open access journals. The study extracted and analyzed 1,000 web citations from articles published between 2009 and 2018. The percentage of articles containing URLs increased from 72% in 2009 to 94% in 2018, with 80.7% of articles incorporating web citations. According to the study, 78% of URLs

were available during the initial check, while 22% disappeared. After employing complementary paths, the rate of accessibility climbed to 92%, and the rate of decay reduced to 8%. The .edu/.ac domain has the highest stability and perseverance, with 89% availability (11%) and an erosion rate of 11%. PDF was the most stable file format, with 85% availability.

Bansal and Parmar (2020) published a paper on the *Current Science Journal* during the period 2015 and 2016 about the accessibility, deterioration, and half-life of URLs. Out of 1,724 URLs, 977 (56.67%) were accessible, while 747 (43.33%) were decayed. The most common HTTP error message was 404, accounting for 59.03% of all HTTP error messages. The .org top-level domains have the most missing URLs. The average half-life of lost URLs was 1.76 years. This suggests that it will take around two years for 50% of the URL citations to vanish. This inconsistency of web references shows that the availability of online knowledge resources is not certain or assured. There are some ways that can slow the rate of URL decay, such as authors' obligation to avoid typing errors when inputting URLs, validating URLs before including them in the article, and keeping digital copies of mentioned web resources. DOIs can be mentioned in the reference list with URLs, databases, and e-journals of some specified domains (with a lower rate of URLs corrosion) for citations. Publications can also be cited via institutional repositories. According to the findings of O'Connor and Doherty (2020), link rot is a substantial problem for Irish academic library websites, with 1,954 of the 47,308 hyperlinks analyzed damaged. The most prevalent error was "404 not found," which accounted for 47.6% of all problems. "Timeout" (19.1%), "5XX Server Error" (9.1%), "403 Forbidden" (8.6%), and "Host Not Found" (6.5%) were other frequently encountered issues. "Connect Error" (2.3%), "SSL Error" (1.5%), "Invalid URL" (0.8%), and "Send/Receive Error" (0.1%) were among the less common issues. Other errors were used to make the remaining links. Link rot can be avoided and treated. Prevention entails linking to content with a persistent identifier, such as a DOI, rather than a URL. DOIs are typically used for scholarly journal publications. However, most content will never have a DOI due to the high cost of obtaining one. Web archives are another option.

Niveditha and Kumbar's (2020) study investigated the availability of web citations in *Library and Information Science* scholarly journals. The journals were chosen for their high impact factor and publication between 2008 and 2017. A PHP script was used to check the availability of the URLs. A total of 7,986 articles were downloaded, with a total of 3,24,636 references extracted. A total of 51,839 URLs were examined for availability. The accessibility check revealed that over 39,866 (76.90%) of the URLs were accessible, while the remainder 11,973 (23.10%) had accessibility issues. The majority of failures were caused by the HTTP 404 error code (not found error) and the file extensions .doc and .pdf., and there is a higher chance of error in increasing the path depth of the URL.

Niveditha and Kumbar's (2022) study investigated the degeneration of web citations in scholarly journals of *Library and Information Science* and *Communication and Media Studies*. A PHP script was used to crawl URLs from references, resulting in 8,767 articles downloaded and 3,97,140 references extracted. A total of 71,289 URLs were checked for their obtainability, with 58,755 active and 12,534 inactive. The findings indicate that .html possesses the highest decay rate, with more URLs in *Library and Information Science* journal articles decaying than in *Communication and Media Studies*. The majority of errors in both disciplines were HTTP 404 error codes. Their findings suggest that publishers, authors, and editorial staff should ensure web resource availability before citing them.

There are many studies on HTTP errors, failure of top-level domains, file-type errors, and other reasons for the decay of URL links or web references. The majority of the research determines the irresistible error types, file types, domain names. This study focuses on the error codes associated with the web references and the reason behind the URL decay and recovery of the inactive URLs. It examines the web archives such as Wayback Machine and author and journal guidelines for the quality production of the scientific communications. These resources help researchers and journals for their effective performance.

AWARI
Vol. 5, 2024. 1-10, DOI: 10.47909/awari.644

3. NEED FOR THE STUDY

The authors carried out this study to uncover the following difficulties that have been reported in previous studies:

1. The relevance of presenting complete bibliographic information in the required format style rather than partial information.
2. The significance of adopting appropriate file extensions.
3. The active URLs associated with the top-level domain.
4. The significance of web archives in restoring missing citations.

4. OBJECTIVES

The study was carried out with the following goals in mind:

1. To identify the total production and citations used in *Webology* journal articles from 2004 to 2020 (i.e., from the year when the journal started its publication to the COVID period).
2. To determine the year-by-year distribution of web and print citations.
3. To sort out active/inactive web citations.
4. To identify error codes related to missing web citations.
5. To discover domains related to web citation persistence.
6. To discover file extensions related to web citation persistence.

5. RESEARCH METHODOLOGY

5.1. Selection of journal

The study was carried out to determine the accessibility and decay of online citations mentioned in *Webology* journal articles between 2004 and 2020. Every issue of the journal was carefully consulted for data collection. The researcher chose this journal to study since it covers a wide range of topics such as web and information science, and *Webology* is an international open access, peer-reviewed English journal devoted to the topic of the World Wide Web that serves as a place for discussion and experimentation. It is a platform for innovative research in information transmission and communication processes in general and, specifically, in the context of the World Wide Web.

5.2. Selection of articles and references

All research articles published in *Webology* journal between 2004 and 2020 were considered for the study, while *Webology* journal did not publish any articles or issues in 2014. Book reviews, editorial remarks, and short communications were excluded. The references at the conclusion of each article were considered. A total of 9,899 references were chosen from 428 papers published in the *Webology* journal.

5.3. Testing of URLs

The authors collected 2,273 web references (URLs) from a total of 9,899 references. Duplicate URLs were treated as a single citation. The bulk URLs were tested using the Web FX HTTP Status Tool (https://www.webfx.com/tools/http-status-tool/). The tool checks for the availability of URLs and records the error code associated with missing URLs. In addition to checking URLs, the script gathers lexical aspects of URLs such as file extensions, top-level domain, and half-life duration of web references. The remaining sorting out of the different top-level domains and the file extensions and their averages were done in MS Excel.

6. ANALYSIS AND RESULTS

The authors retrieved 428 articles from *Webology* journal between 2004 and 2020 and saved them in a folder. From the 428 articles, 9,899 references were collected, and 2,273 web references or URLS were sorted out using the URL extractor (https://www.convertcsv.com/url-extractor.htm). Only references were taken into account by the authors; footnotes, endnotes, and annotations were not considered in this study. The bulk URLs were tested with the Web FX HTTP Status Tool (https://www.webfx.com/tools/http-status-tool/).

Table 1 reveals that among the 428 papers published in the *Webology* journal between 2004 and 2020, a total of 9,899 citations were reported, with only 2,273 (22.96%) constituting web citations and 7,626 (77.04%) having

print citations. The year 2020 had the most articles published (198), followed by 2019 (43) and 2018 (35). The use of URLs in references is increasing on average. Because of the COVID pandemic, there has been an increase in the publication of papers since 2018, and their average online references have been used. In addition, no issues of journal *Webology* were released in 2014. Furthermore, the largest proportion of web referrals were reported in 2004 (50.39%), 2012 (47.59%), and 2011 (38.73%). Table 1 clearly indicates that the authors need to know the importance of using a greater number of web references.

Publication year of the journal	Number of articles	Total number of citations	Number of print citations	Percentage of print citations (%)	Number of URL citations	Percentage of URL citations (%)
2004	8	129	64	49.61	65	50.39
2005	13	235	145	61.70	90	38.30
2006	14	388	279	71.91	109	28.09
2007	14	337	250	74.18	87	25.82
2008	16	454	334	73.57	120	26.43
2009	7	158	117	74.05	41	25.95
2010	9	215	152	70.70	63	29.30
2011	12	315	193	61.27	122	38.73
2012	11	374	196	52.41	178	47.59
2013	12	292	213	72.95	79	27.05
2014	Not published	0	0	0.00	0	0.00
2015	11	390	338	86.67	52	13.33
2016	10	225	174	77.33	51	22.67
2017	15	539	469	87.01	70	12.99
2018	35	741	650	87.72	91	12.28
2019	43	1,118	1,007	90.07	111	9.93
2020	198	3,989	3,045	76.33	944	23.67
Total	428	9,899	7,626	77.04	2,273	22.96

Table 1. Year-wise distribution of articles and references.

Table 2 displays the stature of active and missing web citations as a percentage of total web citations indicated in the total articles for the period from 2004 to 2020. It is obvious that 1,260 (55.43%) of total online citations were active, while the remaining 1,013 (44.57%) were inactive or not functional. Articles published in 2005 had just 37.78% active web citations, followed by 2008 (39.17%) and 2012 (41.01%). However, publications produced in 2019 had higher active web citations (68.47%), trailing 2015 (67.31%). It is obvious that roughly half of all web references are likely to be dead or decayed. Authors and publishers shall use archives to store data and include the archived URLs in web references.

Table 3 lists the various error codes associated or related to missing URLs. Out of a total of 1,013 inactive URLs, error type 404 "file not found" accounted for more than half (52.81%), followed by other HTTP errors (33.96%), HTTP 303 "Redirect" (6.42%), and HTTP 403 "Forbidden" (3.45%). Rest HTTP failures accounted for the minimum share of missing URLs. This result is comparable to the findings of Parmar and Pateria (2019), who discovered HTTP 404 errors accounted for 51% of all URL failures in their investigation. There are several theories

Publication year of the journal	Number of URL citations	Number of active URL citations	Percentage of active URL citations (%)	Number of missing URL citations	Percentage of missing URL citations (%)
2004	65	31	47.69	34	52.31
2005	90	34	37.78	56	62.22
2006	109	54	49.54	55	50.46
2007	87	38	43.68	49	56.32
2008	120	47	39.17	73	60.83
2009	41	21	51.22	20	48.78
2010	63	30	47.62	33	52.38
2011	122	58	47.54	64	52.46
2012	178	73	41.01	105	58.99
2013	79	39	49.37	40	50.63
2014	0	0	0.00	0	0.00
2015	52	35	67.31	17	32.69
2016	51	28	54.90	23	45.10
2017	70	40	57.14	30	42.86
2018	91	60	65.93	31	34.07
2019	111	76	68.47	35	31.53
2020	944	596	63.14	348	36.86
Total	2,273	1,260	55.43	1,013	44.57

Table 2. Year-wise distribution of active and missing URLs.

for why this response was obtained in past studies. The issue may stem from an unresolved host name, or the target web server was not addressed after the DNS name resolution was successful. This issue occurs when the URL changes as a result of changing the directory or file name, relocating or deleting files, or not updating the site. Regretfully, this HTTP 404 error does not specify the exact nature of the issue. To overcome issues like these, authors can make use of some anti-reference rot measures or initiatives to allow URLs to be saved at the Internet Archives such as CLOCKSS (https://clockss.org/), Perma (https://perma.cc/), Webrecorder (https://webrecorder.net/; now Conifer: https://conifer.rhizome.org/), Portico (www.portico.org/), and Wayback Machine (https://archive.org/index.php).

HTTP error	Number of missing URL citations	Percentage of missing URL citations (%)
HTTP 303	65	6.42
HTTP 307	19	1.88
HTTP 401	7	0.69
HTTP 403	35	3.45
HTTP 404	535	52.81
HTTP 503	8	0.79
Others or N/A	344	33.96
Total	1,013	100

Table 3. Distribution of HTTP error codes.

Table 4 shows the evidence that .html files account for the majority of online citations. Out of 2,273 URLs, 488 are .html files, 269 are .pdf files, and 86 are .php files. The file type with the largest percentage of disappearing URLs was .asp (63.89%), which was followed by .php (58.14%). A minimal degree of loss was linked to the following URLs: .cgi—1 (14.29%) out of 7, .cfm—3 (20.00%) out of 15, and .doc—5 (33.33%) out of 15 URLs.

File extension	Number of URL citations	Percentage of URL citations (%)	Number of active URL citations	Percentage of active URL citations (%)	Number of missing URL citations	Percentageof missing URL citations (%)
.doc	15	0.66	10	66.67	5	33.33
.jsp	7	0.31	3	42.86	4	57.14
.html	488	21.47	255	52.25	233	47.75
.pdf	269	11.83	134	49.81	135	50.19
.txt	2	0.09	1	50.00	1	50.00
.asp	36	1.58	13	36.11	23	63.89
.cfm	15	0.66	12	80.00	3	20.00
.cgi	7	0.31	6	85.71	1	14.29
.php	86	3.78	36	41.86	50	58.14
Others	1,348	59.31	790	58.61	558	41.39
Total	2,273	100	1,260	55.43	1,013	44.57

Table 4. File extension associated with active and missing URLs.

Only seven domains were examined in this study: .org, .gov, .com, .edu, .info, .net, and .int. Domains that did not fit into any of these groups or categories were included in the "others" category. Table 5 shows that 588 (25.87%) of the 2,273 URLs had a connection to the .com domain, while 1,036 (45.58%) were related to the .org domain. Notably, online citations related to the domain .org had more active links (639 (61.68%) out of 1,036), followed by the domain .info (11 (57.89%) out of 19) and .com (318 (54.08%) out of 588) URLs. In addition, it demonstrates that 8 (61.54%) out of 13 URLs with the greatest missing percentage are related to the .int domain.

Top-level domains	Number of URL citations	Percentage of URL citations (%)	Number of active URL citations	Percentage of active URL citations (%)	Number of missing URL citations	Percentageof missing URL citations (%)
.com	588	25.87	318	54.08	270	45.92
.edu	201	8.84	93	46.27	108	53.73
.gov	87	3.83	45	51.72	42	48.28
.info	19	0.84	11	57.89	8	42.11
.org	1,036	45.58	639	61.68	397	38.32
.net	79	3.47	40	50.63	39	49.37
.int	13	0.57	5	38.46	8	61.54
Others	250	11	109	43.60	141	56.40
Total	2,273	100	1,260	55.43	1,013	44.57

Table 5. Top-level domains associated with active and missing URLs.

In this investigation, an attempt was made to look through and retrieve the missing URLs via the Internet Archive's Wayback Machine. Table 6 reveals that pleasantly, 596 (58.84%) of the 1,013 missing URLs were located. Consequently, the sheer number of active URLs increased, and the number of lost URLs decreased. Stated alternatively, the percentage of active URLs increased from 1,260 (55.43%) to 1,856 (81.65%) following the recovery of lost URLs. As a result, the total increase in active URLs was 26.22%.

HTTP error	Number of missing URL citations	Number of recovered missing URLs	Percentage of recovered missing URLs (%)	Number of not recovered missing URLs	Percentage of not recovered missing URLs (%)
HTTP 303	65	41	63.08	24	36.92
HTTP 307	19	11	57.89	8	42.11
HTTP 401	7	5	71.43	2	28.57
HTTP 403	35	17	48.57	18	51.43
HTTP 404	535	317	59.25	218	40.75
HTTP 503	8	7	87.50	1	12.50
Others	344	198	57.56	146	42.44
Total	1,013	596	58.84	417	41.16

Table 6. Recovery of missing URLs.

As can be seen from the data, 417 (41.16%) URLs were still not recovered, while approximately 87.50% of HTTP 503 errors were recovered. Following that, HTTP 401 errors were recovered at the 71.43% level, and HTTP 303 errors were recovered at the 63.08% level. The recovery rates for HTTP 403 "Forbidden" (48.57%), HTTP 307 (57.89%), and HTTP 404 "File not found" (59.25%) were poor. Thus, it can be concluded that the Wayback Machine Tool is the most helpful in recovering the missing URLs, and the Internet Archive enables data archiving that offers permanent URLs.

7. OBSERVATIONS AND SUGGESTIONS

1. The authors may use Internet Archives to store data and include that archived link in the references.
2. The authors must adhere to the guidelines outlined in the reference style manual.
3. Researchers may comprehend the value of web references and URLs that lead to get the original source of information and facilitate the study of related material.
4. Many Internet Archives offer free services for storing webpages and snapshots. Every author should follow the author guidelines of the journals mentioned on their home pages so that every article and its reference formats can have association or equality between them.
5. When including URLs or online references, authors need to exercise extreme care because even a single additional space between characters might cause links to deteriorate.
6. Rather than using links to downloaded or duplicated folders, the author should provide unique, permanent links.
7. When citing sources, researchers should be mindful of whether they are referring to the original document or one that was downloaded from a database. This is because the primary cause of URL decay is permanent storage of the file; decay can also result from files being misplaced or changed.
8. The journal may also keep up an archive server to save data and ensure that it is always available without any problems or mistakes.
9. In order to enable authors to adhere to the format by citing the journal guidelines, the journal will include the required format for research papers as well as the reference format and specific style to follow, along with some samples in the author guidelines.
10. Providing permanent URLs or web references contributes to academic growth and enhances research metrics for both researchers and journals. In addition, web citations facilitate data verification, reduce instances of plagiarism, and support quality communication in academic writing.
11. After retrieving missing URLs from the Wayback Machine, the authors discovered that the majority of the authors or writers still need to know how to archive a webpage or file since, according to current archives, there is a bogus file archiving error in the Internet Archives.

8. DISCUSSION AND CONCLUSION

This study emphasizes how online citations were used between 2004 and 2020, as referenced in the *Webology* journal. It has been observed that the percentages of online citations in publications have been steadily rising. Likewise, link deterioration also increases in direct proportion to age. When referencing an online site, the stability of the URL is a crucial factor to take into account. Researchers consider inaccessible URLs useless because they are unable to obtain worldwide information on the Internet. The location and substance of the URLs determine their stability. Using DOIs can help reduce URL volatility. The authors of academic works should carefully verify the URLs before citing them in their publications to prevent the issue of degeneration. Additionally, the degraded URLs must be updated or removed. It is also essential to save a digital copy of the referenced URLs. It is the responsible duty of the editors and publishers to verify that the URLs are accessible before publishing. Thus, the publishers, authors, and editing staff should ensure that the academic work's referenced materials are easily available to upcoming scholars.

Even though Wayback Machine is an impressive tool used for recovering lost URLs, the issue of missing web citations still exists. Therefore, authors should take care to avoid typing errors when entering URLs, verify URLs before including them in the paper, and save digital copies of mentioned online resources as some options that might slow down the rate at which URLs deteriorate. In addition to URLs, authors may provide DOIs in their reference lists. Databases and e-journals belonging to certain domains (whose URL corrosion rates are often lower) may also be taken into consideration for citations. Citations for publications can also be made via institutional repositories.

Conflict of interests

The authors declare that there are no conflicts of interest.

Contribution statement

Mr. Akshaykumar Dundannanavar: Research conceptualization, literature review, data collection, analysis, and interpretation.

Dr. Gururaj S. Hadagali: Framing the study concept, need and methodology, discussion part, and overall supervising the manuscript.

Statement of data consent

The data generated during the study have been included in the article. ●

REFERENCES

Bansal, S., & Parmar, S. (2020). Decay of URLs citation: A case study of current science. *Library Philosophy and Practice (e-journal),* Article 3582. https://digitalcommons.unl.edu/libphilprac/3582

Gul, S., Mahajan, I., & Asifa, A. (2014). The growth and decay of URLs citations: A case of an online Library & Information Science journal. *Malaysian Journal of Library & Information Science, 19*(3), 27-39. https://mjs.um.edu.my/index.php/MJLIS/article/view/1781

Jacob, N. (1998). Fighting Linkrot. *Newsletter.* Neilsen Norman Group. https://www.nngroup.com/articles/fighting-linkrot/

Jalali, M. (2019). Accessibility and decay of web citations in Computer Science journals. *Library Philosophy and Practice (e-journal),* Article 2968. https://digitalcommons.unl.edu/libphilprac/2968

Jones, S. M., Klein, M., & Van de Sompel, H. (2021). Robustifying links to combat reference rot. *Code4Lib Journal, 50,* Article 15509. https://journal.code4lib.org/articles/15509

Niveditha, B., & Kumbar, M. (2020). Availability of web citations in Scholarly Library and Information Science Journals: A study. *PEARL-A Journal of Library and Information Science, 14*(2), 202-208. https://www.indianjournals.com/ijor.aspx?target=ijor:pjolis&volume=14&issue=2&article=011

Niveditha, B., & Kumbar, M. (2022). URL degeneration: A warning to scholarly community. *SRELS Journal of Information Management, 59*(1), 53-63. https://doi.org/10.17821/srels/2022/v59i1/163835

Nwala, A. C., Weigle, M. C., & Nelson, M. L. (2021, September). Garbage, glitter, or gold: Assigning multi-dimensional quality scores to social media seeds for web archive collections. In *2021 ACM/IEEE Joint Conference on*

Digital Libraries (JCDL) (pp. 80-89). IEEE. https://ieeexplore.ieee.org/abstract/document/9651761

O'Connor, C., & Doherty, A. (2020). 404 Not found: A study of hyperlinks on Irish academic library websites. *An Leabharlann: The Irish Library, 29*(1), 7-11. https://repository.rcsi.com/articles/journal_contribution/404_not_found_A_study_of_hyperlinks_on_Irish_academic_library_websites/12229044?file=22488372

Parmar, S., & 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 (e-journal),* Article 3595. https://digitalcommons.unl.edu/libphilprac/3595/

Sife, A. S., & Lwoga, E. T. (2017). Retrieving vanished web references in health science journals in East Africa. *Information and Learning Science, 118*(7/8), 385-392. https://www.emerald.com/insight/content/doi/10.1108/ILS-04-2017-0030/full/html

Tajeddini, O., Azimi, A., & Moghaddam, H. S. (2011). Death of web citations: A serious alarm for authors. *Malaysian Journal of Library & Information Science, 16*(3), 17-29. https://www.researchgate.net/publication/227344592_Death_of_web_citations_a_serious_alarm_for_authors

Zittrain, J., Albert, K., & Lessig, L. (2014). Perma: Scoping and addressing the problem of link and reference rot in legal citations. *Legal Information Management, 14*(2), 88-99. https://www.cambridge.org/core/journals/legal-information-management/article/perma-scoping-and-addressing-the-problem-of-link-and-reference-rot-in-legal-citations/15A59548BF9882B06D3064DA7E290859