DLINE JOURNALS

Journal of Science and Technology Metrics

ISSN: 2582-6956

JSTM 2024; 5 (2)
https://doi.org/10.6025/jstm/2024/5/2/45-54

# Web Citation Analysis &Efficiency of Internet Archives in selected Civil Engineering journal: A case study

Shanthakumari K
Assistant Librarian
BMS School of Architecture
Research Scholar
Department of Studies & Research in Library & Information Science
Tumkur University, Tumkur
shanthakrs@gmail.com

Keshava
Professor
Department of Studies & Research in Library & Information Science
Tumkur University, Tumkur
keshtut@gmail.com

**Received: 18 May 2024,**
**Revised: 8 June 2024**
**Accepted: 14 June 2024**


**ABSTRACT**

*The study aimed to find the internet archive's efficiency and investigate the persistence and decay of URLs cited in articles from the "Journal of Composites for Construction" over a ten-year period (2013-2022). 23,725 citations from 867 articles were analysed, of which 7,843 (31.74%) were URLs. The main findings include persistence. Decay of URLs: 64.16% of the web citations were found to have vanished or become inaccessible. The main error was that the Error 404 Page was not found, accounting for 39.12% of the vanished URLs, followed by Error 403 at 35.30%. It states that the domain analysis of vanished URLs showed the .org (85.10%), and .com (14.71%). domains Accessibility of Citations extracted from scholarly articles excluding editorials and were verified. The study focused on one journal; hence, the generalizations of the findings are limited to "Journal of Composites for Construction". The timeframe was restricted to only ten years, which may not capture long-term URL persistence and decay trends.*

**Keywords:** Active Links, Vanished URL, Link Rot, Web References

## 1. Introduction

As a platform for accessing and sharing hyperlinked information, the World Wide Web has grown into a vital tool for research in recent years. The references appended to papers hold significant importance as cited works of publications and serve as essential resources for researchers to explore further within their respective fields. They provide a structured way to acknowledge e-sources and build upon existing knowledge, forming a crucial aspect of scholarly communication and advancing science & technology.(Ding et al., 2014).

The rapid expansion of the Internet and the Web in recent years has profoundly impacted how references are integrated into scholarly works. Authors now frequently include web citations, typically URLs, to provide direct access to online resources that support their arguments or findings (Germain, 2000). The practice enhances the accessibility of cited materials, allowing readers to easily locate and verify information. Web citations, often in the form of URLs of web pages, enhance accessibility to the resources mentioned in references. However, a significant challenge arises due to the transient nature of the Web.
However, a notable challenge stems from the transient nature of web content because. Web pages can be modified, moved, or deleted over time. This ephemeral quality risks the reliability and longevity of URL citations. Authors and researchers must consider strategies to mitigate this issue, such as using web archiving services or providing detailed descriptions of web content to ensure the integrity of their references.

Despite the convenience of web citations, the transient nature of online content poses a risk. Over time, the information on the Web becomes inaccessible or "rotten," rendering citations unavailable. As a result, researchers may encounter broken links or outdated content, diminishing the reliability and longevity of references. In summary, while web citations enhance accessibility to resources, the inherent volatility of online content underscores the need for sustainable archiving practices and alternative referencing methods to ensure the longevity and reliability of scholarly publications.

## 2. Literature Review

The study attempted by Loan et al. (2024) they are aimed to evaluate the condition of URLs cited in web references and to compare the performance of Chrome, Google, and the WayBack Machine in retrieving inactive URLs. The study used web references from the journal Library Hi Tech from 2004 to 2008 to assess the reliability of URL citations over time and the effectiveness of various tools in accessing archived or alternative versions of web content when original URLs are no longer functional. URLs were collected from articles and tested for persistence and decay using the three tools, and it was found that the Wayback Machine was the most effective, retrieving 72.6% of URLs, followed by Google (46.3%) and Chrome (37.1%). The study concludes that URL decay is a significant issue in web resource preservation and suggests using the WayBack Machine for effective retrieval. Niveditha et al., (2022), examined the use of URL references in Library and Information Science (LIS) and Communication and Media Studies (CMS) journals over ten years. A total of 20 journals, ten from each field, published between 2008 and 2017, were analysed. Using a PHP script, the study collected and assessed 12,251 articles and 555,428 references, focusing on 102,718 web citations. The findings reveal that many URLs were inaccessible, with 404 errors being the primary cause. The study also attempted to revive inaccessible URLs using "Time Travel." The results indicate differences in accessibility between the two disciplines, with CMS journals showing slightly higher accessibility rates. The study findings are useful to authors, publishers, and editorial staff, ensuring the accessibility of URL citations in future publications.

The citation analysis study in the field of civil engineering conducted by Shanthakumari & Keshava (2021) analyzes the persistence of cited URLs in the International Journal of Pavement Engineering over ten years (2010-2019) and computes the half-life period of these URLs in journal articles. A total of 901 URLs cited in 299 research articles were examined, and their accessibility was verified using the W3C Link Checker. The study found that 30.63% of the articles contained URL citations, increasing steadily from 2.28% in 2010 to 5.07% in 2018 before decreasing to 2.54% in 2019. The half-life period of the cited URLs was computed using a specific formula. These findings provide insights into the prevalence and persistence of URL citations in the journal and trends in their usage over time. A similar kind of citation analysis study in the field of Library & Information science was conducted by (Loan & Shah, 2020). A study investigated the persistence and decay of Uniform Resource Locators (URLs) associated with web references, particularly analysing their age, domain, technical errors, and error codes. The research concentrates on web references from the *Journal of Informetrics* published between 2007 and 2011. 7,409 citations from 221 articles were examined, with 358 (4.8%) identified as web citations. Findings reveal that 32.12% of URL references had vanished, with Error 404 (Page not found) being the most common error. Domain analysis indicates that .gov domains accounted for the most missing URLs. The study suggests that

URL decay poses a significant challenge to preserving and citing web resources and advocates for collaborative efforts to address this issue. This research adds to the existing literature on URL persistence and decay and provides insights relevant to scholarly publishing and information preservation practices. Another study was conducted to know the efficiency of Internet Archive and Google by (Saberi & Abedi, 2012). This study focused on recovering disappeared URLs using two primary tools Internet Archive and Google. Previous research by various scholars has already employed these tools to retrieve vanished URLs. Similarly, the present study utilised the Internet Archive and Google for the same purpose. All 12,423 vanished URLs were entered simultaneously into Internet Archive and Google search boxes. This involved copying the exact URL of each URL citation into the respective search fields of these platforms.

## 3. Research design

### 3.1 Objectives

The research focuses on studying the status and characteristics of web citations in the "Journal of Composites for Construction", a publication by the American Society of Civil Engineers (ASCE). The specific objectives of this study are outlined as follows:

· To determine how the number of articles and the frequency of URL citations have evolved within the journal.

· Analyze the various types of top-level domains associated with URLs cited in civil engineering literature.

· To investigate how often URLs cited in the journal remain accessible over time and identify patterns of URL decay.

· To find the types of Errors and Error Codes in Vanished URL Citations

· To Study the distribution of domains and the depth of URLs cited in the journal.

· To find the availability of URLS &DOIs cited in selected journal

· To compare the efficiency of Internet Archive and Pandora of Internet archive to retrieve vanished URL citations in a civil engineering journal

## 4. Methodology

The study analyses citations from articles published in the "Journal of Composites for Construction" from 2013 to 2022. The main objective is to evaluate the accessibility and reliability of URL citations, specifically those containing URL links and Digital Object Identifiers (DOIs).

Using the W3C Link Checker, the study identifies vanished URL citations within these articles. It then employs Internet archive tools such as the Wayback Machine and Pandora of the Internet Archive to retrieve these vanished URLs. These tools are evaluated for their efficiency in recovering vanished URL citations, thereby assessing their effectiveness in preserving and maintaining the accessibility of scholarly references over time.

The research is significant due to the journal's high impact factor and its focus on civil engineering. It highlights the importance of maintaining accurate and accessible URL citations in scholarly publications. This study's findings aim to provide insights into improving the reliability and longevity of online references in academic research.

## 5. Data Analysis & Discussion

The data were analysed from 867 articles published in the "Journal of Composites for Construction," with 23,725 references scrutinized. This extensive dataset is the foundation for several key findings and insights from the analysis.

### 5.1. Distribution of URL and print citations in civil engineering journals

Over ten years, from 2013 to 2022, the Journal of Composites for Construction published 867 scholarly articles, with an average of 29 articles per year. The peak publication year was 2016, with 126 articles, closely followed by 2017, which shows 106 articles. These articles collectively referenced scholarly literature 23,725 times, with the highest number of citations occurring in 2020, a total of 4,085 citations. On average, each article contained approximately 28.91 citations. Of 23725 citations, 15882 (68.26%) are print citations, and 7843 (31.74%) are URL citations. The data shows that the majority of authors cited print references in their publications, compared to URL references in this journal article. Also, URL citations gradually increased from 2013 to 2022(0.80% to 67.51%).

This data states the journal's robust output and significant contribution to the field of composites for construction over the past decade. The citation patterns reflect active engagement with existing research, indicating a strong foundation of scholarly discourse and knowledge dissemination within the journal's domain.

| Year | Total number of Articles | Total number of citations | Average citations per article | Print citations | % | URL citations | % |
|---|---|---|---|---|---|---|---|
| **2013** | 84 | 2609 | 31.06 | 2588 | 99.20 | 21 | 0.80 |
| **2014** | 73 | 2060 | 28.22 | 2045 | 99.27 | 15 | 0.73 |
| **2015** | 80 | 2154 | 26.93 | 2145 | 99.58 | 9 | 0.42 |
| **2016** | 126 | 609 | 4.83 | 600 | 98.52 | 9 | 1.48 |
| **2017** | 106 | 3751 | 35.39 | 3710 | 98.91 | 41 | 1.09 |
| **2018** | 69 | 2549 | 36.94 | 1253 | 49.16 | 1296 | 50.84 |
| **2019** | 70 | 2929 | 41.84 | 1030 | 35.17 | 1899 | 64.83 |
| **2020** | 91 | 4085 | 44.89 | 1542 | 37.75 | 2543 | 62.25 |
| **2021** | 68 | 1951 | 28.69 | 635 | 32.55 | 1316 | 67.45 |
| **2022** | 100 | 1028 | 10.28 | 334 | 32.49 | 694 | 67.51 |
| **Total** | **867** | **23725** | **28.91** | **15882** | **68.26** | **7843** | **31.74** |

**Table 1. Distribution of URL and print citations in civil engineering journals**

### 5.2. Growth rate of URL citations in civil engineering journal articles

Table 5.2 reveals the growth rate of URL citations in research articles published in the Journal of Composites for Construction from 2013 to 2022. The trend shows fluctuations over the years, with the related growth rate (RGR) indicating a decline over this period. In 2014, the RGR was 0.79, suggesting a relatively high growth rate of URL citations during that year. Over the subsequent years, the RGR shows a decreasing trend, falling to 0.29 by 2022.

This decreased trend in RGR indicates that while URL citations in the journal's articles fluctuated in absolute numbers, their growth rate relative to the total number of citations decreased steadily over the years. This trend may reflect changes in citation practices, evolving preferences for different types of sources, such as DOIs over URLs, or challenges related to the persistence and reliability of online-based references over time.

### 5.3. Accessibility of Active and Vanished URL Citations

Table 5.3 summarises the accessibility status of URL citations referenced in articles from the Journal of Composites for Construction. The findings reveal that a total of 7,843 URL citations were analysed. 35.84% are active citations, and 2,784 were found to be accessible or active at the time of analysis. The majority, 64.16% (5,059) of the URL citations, had decayed or become inaccessible over time.

| Year | Total citations | web citations | Cumulative Web citations | RGR | Doubling Time (Dt) |
|---|---|---|---|---|---|
| 2013 | 2609 | 21 | 21 | 0.00 | 0 |
| 2014 | 2060 | 15 | 36 | 0.88 | 0.79 |
| 2015 | 2154 | 9 | 45 | 1.61 | 0.43 |
| 2016 | 609 | 9 | 54 | 1.79 | 0.39 |
| 2017 | 3751 | 41 | 95 | 0.84 | 0.82 |
| 2018 | 2549 | 1296 | 1391 | 0.07 | 9.80 |
| 2019 | 2929 | 1899 | 3290 | 0.55 | 1.26 |
| 2020 | 4085 | 2543 | 5833 | 0.83 | 0.83 |
| 2021 | 1951 | 1316 | 7149 | 1.69 | 0.41 |
| 2022 | 1028 | 694 | 7843 | 2.42 | 0.29 |
| **Total** | **23725** | **7843** | **25757** | 1.07 | 0.65 |

**Table 2. Growth rate of URL citations in civil engineering journal articles**

| Year | Total URL citations | Active URL citations | % | Vanished URL citations | % |
|---|---|---|---|---|---|
| 2013 | 21 | 8 | 38.10 | 13 | 61.90 |
| 2014 | 15 | 6 | 40.00 | 9 | 60.00 |
| 2015 | 9 | 2 | 22.22 | 7 | 77.78 |
| 2016 | 9 | 3 | 33.33 | 6 | 66.67 |
| 2017 | 41 | 15 | 36.59 | 26 | 63.41 |
| 2018 | 1296 | 310 | 23.92 | 986 | 76.08 |
| 2019 | 1899 | 484 | 25.49 | 1415 | 74.51 |
| 2020 | 2543 | 967 | 38.03 | 1576 | 61.97 |
| 2021 | 1316 | 613 | 46.58 | 703 | 53.42 |
| 2022 | 694 | 376 | 54.18 | 318 | 45.82 |
| **Total** | **7843** | **2784** | 35.84 | **5059** | 64.16 |

**Table 3. Accessibility of Active and Vanished URL Citations**

### 5.4. Availability of Top-level domains

Table 5.4 provides a summary of the top-level domains) associated with URL citations analysed in the Journal of Composites for Construction. The analysis covered a total of 7,843 URL citations. Most URL citations comprise domain .org 5,550 citations, accounting for 70.76% of the total. Commercial domain (.com/.co) 2,204 citations, representing 28.10% of the total. These educational (.edu/.ac) contributed a minimal percentage, with .edu at 0.10% (8 citations) and .ac at 0.04% (3 citations) and other domains such as .gov, .net, .int, .info: Each of these domains contributed to a very small fraction of the citations, ranging from 0.01% to 0.13%.

| Domain type | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | Total | % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| .COM | 14 | 8 | 6 | 2 | 28 | 307 | 528 | 704 | 343 | 264 | 2204 | 28.10 |
| .EDU | 1 | 1 | | | 1 | | 1 | 3 | 1 | | 8 | 0.10 |
| .ORG | 1 | | 2 | 3 | 4 | 969 | 1359 | 1822 | 961 | 429 | 5550 | 70.76 |
| .AC | 1 | 2 | | 1 | 2 | | | 4 | | | 10 | 0.13 |
| .GOV | 2 | | 1 | 1 | | | | 1 | | | 5 | 0.06 |
| .NET | | | | | 1 | | | 1 | 1 | | 3 | 0.04 |
| .INT | | | | | | | | 1 | | | 1 | 0.01 |
| .INFO | | | | | | | | 2 | 2 | | 4 | 0.05 |
| OTHERS | 2 | 4 | | 2 | 5 | 19 | 11 | 6 | 8 | 1 | 58 | 0.74 |
| **Total** | 21 | 15 | 9 | 9 | 41 | 1295 | 1899 | 2544 | 1316 | 694 | 7843 | 100.00 |

**Table 4. Availability of Top-level domains**

### 5.5. Association of top-level domains with the Vanished URL citations

Table 5.5 indicates that among the 5059 vanished URL citations analysed, domains ending in ".org" had the highest percentage of vanished URLs at 85.10%. This was followed by ".com" domains, which accounted for 14.71% of vanished URLs. The domains such as ".edu", ".net", and ".ac" had a significantly lower percentage of vanished URLs.

| Domain type | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | Total | % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| .COM | 9 | 6 | 5 | 1 | 9 | 159 | 267 | 185 | 70 | 33 | 744 | 14.71 |
| .EDU | 1 | 1 | 0 | 0 | 0 | | 0 | 1 | 0 | | 3 | 0.06 |
| .ORG | 1 | | 2 | 2 | 3 | 809 | 1142 | 1383 | 679 | 284 | 4305 | 85.10 |
| .AC | 0 | 0 | | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0.04 |
| .GOV | 0 | | 0 | 1 | 1 | | | 0 | 0 | 0 | 2 | 0.04 |
| .NET | 0 | | | 1 | 1 | 0 | | | 0 | 0 | 2 | 0.04 |
| .INT | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | | 0 | 1 | 0.02 |
| .INFO | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00 |
| Total | 11 | 7 | 7 | 6 | 14 | 968 | 1409 | 1571 | 749 | 317 | 5059 | 12.50 |

**Table 5. Association of top-level domains with the Vanished URL citations**

This finding is significant because it suggests a discrepancy in the persistence and reliability of URL content across different organisational domains. Traditionally, domains like ".org" are associated with organisations that publish accurate and authoritative documents. However, the high percentage of vanished URLs in ".org" domains raises queries about the long-term accessibility and reliability of URL links.

### 5.6. Association of error codes with the Vanished URL citations in the civil engineering journal

In analysing the error messages associated with vanished URL citations in the Journal of Composites for Construction, the most common errors, such as 404 Page Not Found this error was the most prevalent, accounting for 39.12% of the total vanished URL citations, equivalent to 1,979 citations. It indicates that the referenced web page could not be located, and 403 Error was the second most encountered error, representing 35.30% of the vanished URLs with 1,786 citations, which signifies access denied or forbidden.

| Error codes | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | Total | % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HTTP 301 | | | | | | | | 1 | | | 1 | 0.02 |
| HTTP 302 | | | | | | 96 | 165 | 352 | 186 | 79 | 878 | 17.36 |
| HTTP 400 | | | | | | 75 | 120 | 205 | | | 400 | 7.91 |
| HTTP 403 | 1 | 1 | 1 | 1 | 1 | 252 | 334 | 587 | 430 | 178 | 1786 | 35.30 |
| HTTP 404 | 5 | 5 | 4 | 60 | 15 | 500 | 718 | 577 | 58 | 37 | 1979 | 39.12 |
| HTTP405 | | | | | | 2 | 1 | | | | 3 | 0.06 |
| HTTP 418 | | | | | | | | | 3 | | 3 | 0.06 |
| HTTP 500 | | | | | | | | | 1 | | 1 | 0.02 |
| HTTP 503 | | | | | 1 | | | | | | 1 | 0.02 |
| HTTP 520 | 1 | | | | 2 | | 2 | 2 | | | 7 | 0.14 |
| **Total** | **7** | **6** | **5** | **61** | **19** | **925** | **1340** | **1724** | **678** | **294** | **5059** | **100.00** |

**Table 6. Association of error codes with the Vanished URL citations in civil engineering journal**

### 5.7. Path Depth associated with URL web citations

The study shows that the highest number of vanished URLs, 4855 (95.97%), occurred at Path Depth Level-2 and Path Depth Level-3, followed by a much lower percentage of 2.10% of vanished URLs.

Goh and Ng (2007) state that longer URL path depths can contribute to link failure. The study observed this phenomenon, with URLs with deeper path levels exhibiting higher rates of disappearance.

The findings suggest that URLs with shorter path depths have better long-term accessibility than URLs with deeper path structures. Understanding the relationship between path depth and vanished URLs is crucial for researchers and publishers in civil engineering and other fields. It highlights the importance of the contents of URLs in a way that enhances their longevity and accessibility.

Publishers and authors may benefit from adopting practices that minimise path depth in URLs to ensure the reliability of web citations in scholarly publications. The study provides insights into how path depth influences the persistence of URL citations, emphasising the need for effective web archiving practices to maintain the integrity of scholarly citations over time.

| Path Depth | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | Total | % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 2 | 1 | 1 | 1 | 1 | 8 | 1 | 1 | 1 | 17 | 0.34 |
| 1 | 4 | 0 | 0 | 1 | 3 | 4 | 6 | 6 | 2 | 0 | 26 | 0.51 |
| 2 | 4 | 0 | 4 | 2 | 7 | 945 | 1384 | 1545 | 673 | 291 | 4855 | 95.97 |
| 3 | 2 | 4 | 1 | 2 | 6 | 33 | 13 | 11 | 14 | 20 | 106 | 2.10 |
| 4 | 1 | 1 | | | 6 | 0 | 1 | 4 | 6 | 0 | 19 | 0.38 |
| 5 | 1 | 1 | | | 1 | 1 | 1 | 7 | 6 | 0 | 18 | 0.36 |
| 6 | | | | 0 | | 1 | | 2 | 1 | 6 | 10 | 0.20 |
| 7 | 0 | 0 | 1 | | 2 | 1 | 2 | 0 | | | 6 | 0.12 |
| 8 | 1 | | | | | | | 1 | | | 2 | 0.04 |
| **Total** | **13** | **8** | **7** | **6** | **26** | **986** | **1415** | **1577** | **703** | **318** | **5059** | **11.11** |

**Table 7. Path Depth associated with vanished URL web citations**

### 5.8. Year wise distribution of DOIs

Another study conducted to know the availability of digital object identifier available in the citations. The data presented in the table 5.8 and found that only few (0.19%) DOIs present in the citations. It states from the study states that research scholars are not much aware about to include DOIs in their citations.

| Year | Total no of Citations | Print citation | URL citation | No. of DOI | % |
|---|---|---|---|---|---|
| **2013** | 2609 | 2588 | 21 | 0 | 0 |
| **2014** | 2060 | 2045 | 15 | 4 | 0.19 |
| **2015** | 2154 | 2145 | 9 | 0 | 0 |
| **2016** | 609 | 600 | 9 | 0 | 0 |
| **2017** | 3751 | 3710 | 41 | 0 | 0 |
| **2018** | 2549 | 1253 | 1296 | 0 | 0 |
| **2019** | 2929 | 1030 | 1899 | 0 | 0 |
| **2020** | 4085 | 1542 | 2543 | 0 | 0 |
| **2021** | 1951 | 635 | 1316 | 0 | 0 |
| **2022** | 1028 | 334 | 694 | 0 | 0 |
| **Total** | **23725** | 15882 | 7843 | 4 | 0.19 |

**Table 8. Year-wise distribution of DOIs**

### 5.9. URLs retrieved through Wayback machine and Internet archive

The study examined the effectiveness of two Internet Archive tools such as the Wayback Machine and Pandora, in recovering vanished URL citations from a selected civil engineering journal over ten years. This study presents a few key findings from the data presented in Table 5.9. Also, data represents the recovery rates of Internet archive Wayback Machine. Recovery rates varied significantly, from a low of 22.22% (2014) to a high of 53.45% (2018).

This indicates fluctuating effectiveness over the years. Pandora of Internet Archive recovery rates also varied, with a low of 7.69% (2017) and a high of 42.86% (2015). The tool showed variability in retrieving vanished URLs across different years.

The study results suggest that the Wayback Machine consistently outperformed Pandora of the Internet archive in recovering vanished URL citations throughout the study period 2013-2022. Wayback Machine's higher overall recovery rate (40.08%) compared to Pandora's (15.43%) indicates it is the more reliable tool for preserving and retrieving online citations. Researchers and publishers can choose the Wayback Machine to mitigate the impact of link rot and ensure the longevity of web citations in scholarly publications. Understanding the variability in recovery rates by year can inform strategies for improving web archiving practices and enhancing the reliability of web citations.

| | | Wayback Machine | | Pandora of IA | |
|---|---|---|---|---|---|
| Year | Vanished URLs | Recovered URLs | % | Recovered URLs | % |
| 2013 | 8 | 4 | 50 | 1 | 12.50 |
| 2014 | 9 | 2 | 22.22 | 1 | 11.11 |
| 2015 | 7 | 3 | 42.86 | 3 | 42.86 |
| 2016 | 6 | 3 | 50.00 | 1 | 16.67 |
| 2017 | 26 | 8 | 30.77 | 2 | 7.69 |
| 2018 | 986 | 527 | 53.45 | 195 | 19.78 |
| 2019 | 1420 | 707 | 49.79 | 216 | 15.21 |
| 2020 | 1576 | 567 | 35.98 | 183 | 11.61 |
| 2021 | 703 | 259 | 36.84 | 61 | 8.68 |
| 2022 | 318 | 92 | 28.93 | 26 | 8.18 |
| **Total** | **5059** | **2172** | **40.08** | **689** | **15.43** |

**Table 9. URLs retrieved through Wayback machine and Internet archive**

## 6. Conclusion

The citation analysis study discusses the issue of URL decay and proposes several strategies to mitigate this problem and ensure the long-term persistence of URL citations. It emphasises the importance of research scholars, authors, and subject experts' careful handling of URLs and suggests that they should be diligent while typing and verifying URLs before publication. Editors are encouraged to check URLs to thoroughly minimise link rot risk. A key recommendation is prioritising using Digital Object Identifiers (DOIs) over URLs whenever possible. DOIs provide a more stable and persistent way to locate digital content than URLs, which are prone to change or become obsolete over time.

It also suggests establishing multiple archives for URL citations to enhance their permanence. It proposes that national digital libraries take a proactive role in archiving open-access and copyright-free web content. By doing so, these libraries can contribute to preserving online information for future generations and prevent the decay of valuable web resources.

## References


[1] Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X., Zhai, C. (2014). Content based citation analysis: The next generation of citation analysis. *Journal of the Association for Information Science and Technology*, *65*(9), 1820–1833. https://doi.org/10.1002/asi.23256

[2] Germain, C. A. (2000). URLs: Uniform Resource Locators or Unreliable Resource Locators. *College & Research Libraries*, *61*(4), 359–365. https://doi.org/10.5860/crl.61.4.359

[3] Loan, F. A., Khan, A. M., Andrabi, S. A. A., Sozia, S. R., Parray, U. Y. (2024). Giving life to dead: Role of WayBack Machine in recovery of dead URLs. *Data Technologies and Applications*, *58*(2), 201–213. https://doi.org/10.1108/DTA-06-2022-0242

[4] Loan, F. A., & Shah, U. Y. (2020). The decay and persistence of web references. *Digital Library Perspectives*, *36*(2), 157–166. https://doi.org/10.1108/DLP-02-2020-0013

[5] Niveditha, B., Kumbar, M., Sampath Kumar, B. T. (2022). Rotten web citations cited in scholarly journals: Use of time travel for retrieval. *Aslib Journal of Information Management*, *74*(2), 225–243. https://doi.org/10.1108/AJIM-05-2021-0139

[6] Saberi, M. K., Abedi, H. (2012). Accessibility and decay of web citations in five open access ISI journals. *Internet Research*, *22*(2), 234–247. https://doi.org/10.1108/10662241211214584

[7] Shanthakumari, K., Keshava (2021). An analysis of persistence and obsolescence of web citations of pavement engineering literature citations of pavement engineering literature. *Library Philosophy and Practice*. https://digitalcommons.unl.edu/libphilprac/5923