International Journal of Information Science and Management
Vol. 22, No. 2, 2024, 193-210
DOI: https://doi.org/10.22034/ijism.2024.1977641.0 / DOR: https://dorl.net/dor/20.1001.1.20088302.2024.22.2.12.1

Original Research

Missing and Recovery of URLs Using Internet Archive: A Case Study on African Journal of Library, Archives and Information Science (AJLAIS)

Ahmed Olakunle Simisaye
Associate Prof., Department of Library & Information Science Tai Solarin University of Education, Ogun State. Nigeria.
Corresponding Author:
simisayeao@tasued.edu.ng
ORCID iD: https://orcid.org/0000-0002-8721-7100

Henry Chukwudi John
Lead Digital Library Services, African Leadership University, Kigali, Rwanda.
hjohn@alueducation.com
ORCID iD: https://orcid.org/0000-0003-0960-6861

Toluwani Joanna Iseyemi
Graduate Student, Department of Library & Information Science Tai Solarin University of Education, Ogun State. Nigeria.
iseyemitoluwani80@gmail.com
ORCID iD: https://orcid.org/0009-0002-5155-1411

Received: 21 December 2021
Accepted: 14 August 2023

Abstract

In recent times, authors of academic publications have depended on web resources. However, there have been some concerns raised on the permanency of e-resources on the web. In this context, this article investigated the availability, missing and recovery of Uniform Resource Locator (URL) citations cited in articles in the African Journal of Library, Archive and Information Science (AJLAIS) published between 2008 and 2017 using Internet Archive. A total of 986 URL citations in 129 research articles published in LIS journals were extracted. The finding showed that URL citations of the journal grew during the years 2008-2017 Finding further revealed that 986 represented 28.88% of the total 3414 references that appeared in the journal in the period under review. Of the total cited web citations 986 web citations 310(31.4%) were not accessible and 676 (68.6%) were accessible. This article adopted a W3C link checker to detect HTTP errors linked with missing URLs. HTTP 500 error messages (‘page not found’) were majorly the irresistible messages that denoted 20 per cent of all HTTP error messages. The highest percentage of missing URLs were associated with commercial domains. The attempt to recover 310 missing URL citations through Internet Archive increased the active web citations from 676 to 918 which accounted for 93.1% of active URL citations. This study further showed a negative correlation between path depth and missing URL citations. The statistical analysis of this research indicated that the number of citations and URL citations are positively correlated.

Keywords: Web Citations, Missing URL Citations, Recovery of URL citations, Internet Archive, African Journal of Library, Archives and Information Science, AJLAIS.

Introduction

Citation is an important aspect of research, without it, no single scholarly work could be accorded acceptance and prestige in the communities of scholars. Citations depict authors' acknowledgement of previously cited works done previously. Germain (2000) remarked that citations give credence to research and that is why scholars search for related literature to back up their studies. Initially, scholars were limited to searching for printed formats but nowadays, the emergence of electronic resources and the Internet have opened the horizon of unlimited materials that could be accessed by researchers from any part of the globe. This development makes consultation on the web more common among researchers, scholars and librarians (Casserly & Byrd, 2003). The convergence of e-content has made Internet resources the fastest and preferred channel of information (Riahinia, Zandian & Azimi, 2011).

The World Wide Web has liberalized scholarship by allowing everyone to publish all kinds of information using different media types such as News, blogs, wikis, encyclopaedias, photos, interviews, and public opinions just to mention a few examples. The web makes it possible for internet users, including scholars, to navigate the Net to search and retrieve numerous scholarly resources. This possibility increases the practice of using and citing Internet resources by scholars and researchers in various fields worldwide. This practice made Sellitto (2004) affirm that web resources are gradually replacing the traditional use of printed works among scholars.

A Uniform Resource Locator is a web address that specifies the location of a document and enables web browsers to reach it on any server. Generally, URLs consist of four parts-protocol, domain, directory, and file (Tajeddini, Azimi, Sadat-Moosavi & Sharif-Moghaddam, 2011). Over the years, academic researchers, including scholars in Library and Information Science are increasingly citing Universal Resource Locators (URLs) in their scholarly publications. The URL citations assist researchers in locating previous research works easily. However, there are challenges because of the nature of the Internet, Web resources are subject to modifications and changes (Casserly & Bird 2003), which make it a bit difficult to retrieve some online resources that have been used in previous research. This inability to locate online references used in a research article not only raises concerns about the necessity of citing URLs but also undermines the foundations of academic writing in which new research is built from published literature. This assertion was further corroborated by Costa, Gomes and Silva (2017) when they remarked that the Internet contains a large amount of information for research, however, many of them are lost as they could not be retrieved through their Uniform Resource Locator on the Internet.

Currently, one of the emergence areas of research receiving the attention of academic scholars and librarians is examining the decay and permanency of URLs of web resources cited in scholarly works (Dimitrova & Bugeja, 2007; Nagaraja, Joseph, Polen & Clauson, 2011; Gul, Mahajan & Ali, 2014) The increased use of web citations in scholarly literature posed a challenge to the researchers, academics, and librarians as they disappear often. It was discovered that 10 to 40 per cent of web-based citations and URLs often go missing (Markwell & Brooks, 2003; Sellitto, 2004; Tajeddini, Azimi, Sadat-Moosavi & Sharif-Moghaddam, 2011). When web citation links are missing, online citations of scientific research become less valid. Many reasons have been adduced for missing web resources. Tajedini, Sadatmoosavi, Ghazizade and Tajedini (2018) hinted that some websites get simply unreachable because the website domain names have expired. Other reasons are that the creators of web pages sometimes remove them from the internet, and sometimes they go missing because websites were redesigned with new structures. In addition, these scholars claimed that some web pages are moved to a new location, thus, displaying a redirect message or directly leading the browser to a new internet address which sometimes enjoys a different updated content. As a result of the disappearance of web resources, many scholars encountered many error messages when they searched the internet for web-based cited references in published research works. This assertion was confirmed by Dimitrova and Bugeja (2007) when they observed that irrespective of the reasons for such failures, most internet users have been presented with the famous error "404: Not Found". This type of error is prevalent in over a million entries in a search engine per day (ibid). This situation is quite worrisome to academic communities worldwide.

The non-accessibility of cited URLs is a serious issue bordering academic communities and web administrators. This concern has been shown by many scholars, for instance, Germain (2000) remarked that the constant decay of URLs is a real challenge affecting the usage and accessibility of URLs cited in research while Dimitrova and Bugeja (2007) lamented that the vanishing of URLs references affects the accessibility and usability of such missing web sources. Studies that have registered this concern of unreachable URLs cited in Library and Information Science (LIS) literature include (McCown, Chan, Nelson & Bollen, 2005, Sampath Kumar & Prithvi Raj, 2012, Gul et al., 2014). As a result of this challenge of the decay of cited URLs, several studies have been done to recover disappeared URLs referenced in LIS. Based on this background, this study investigated URLs cited in the African Journal of Library, Archive and Information Science (AJLAIS) published between 2008 and 2017 with the view of determining their accessibility, missing and retrieving URLs using Internet Archive.

The AJLAIS with an International Standard Serial Number (ISSN) of 07954778 came into existence in 1991. The journal was chosen for this study as it is one of the pioneer Journals in Library and Information Science (LIS) in Africa and is the influential LIS journal in terms of impact factor in Africa (Onyancha, 2009) Similarly, AJLAIS was rated the best with the highest impact factor among the journals of LIS in Africa (Aina, 2002). AJLAIS is indexed in highly rated indexing services such as Web of Science, Library Literature, Scopus, Social Science Citation Index Scopus, Academic Search Premier, Library, Information Science & Technology Abstracts (LISTA), Information Science and Technology Abstracts, Library Literature and Information Science. Presently, the journal is among the Q4 indexed in Scopus with H-Index=10. In the latest Scimago ranking, AJLAIS is in quartile 4 (www.scimagojr.com) AJLAIS was chosen for this study based on its popularity and reputation in Africa. In addition, several searches were done by researchers and the result shows that this type of study was not carried out in this journal, hence the need for this study.

The present study is significant as it may assist researchers in knowing the significance of providing a complete web link or address without any typing error while giving web citations in a document. In addition, this study will show the importance of providing complete bibliographical information rather than only URL or partial bibliographical information. This present study may assist any researcher in identifying i) the importance of providing complete web links or addresses without any typing error while giving web citations in the document published; ii) knowing the domains which have more durability; iii) the importance of providing complete bibliographical information rather than only URL or partial bibliographical information.

Literature Review

Various studies have been carried out on the missing and recovering of URLs in scholarly publications, Spinells (2003) revealed that 28% of the URLs sources referenced in Computer and Communication of the ACM articles during 1995- 1999 could not be retrieved, which implies that the rate of unavailable URLs was 34% and reduced to 5% after adopting various techniques. In an attempt to investigate the frequencies of the web- references cited in two key Chinese academic journals, Wu (2009) discovered that out of 1637 web references, 776 were unreachable. The author also adopted a search engine and Internet Archive to recover the missing web-references. Search engines recovered 62.8%, and Internet Archive found 24.4% of missing web citations.

Moghaddam, Saberi and Esmaeel (2010) also studied the URLs of cited articles published in the information research journal from 1995-2008 and found 1,761 cited URLs. Internet Archive and Google search engines were employed to retrieve unavailable URLs. Out of these 1761 cited URLs, 1,290 (73%) were accessible and 471 (27%) were inaccessible. They showed that the Internet Archive recovered 28% of unreachable URLs. Studies also have been carried out on the missing and recovering of URLs in scholarly publications, with most concentration evidence drawn from the study conducted by Tajeddini, Azimi, Sadat-Moosavi and Sharif-Moghaddam (2011) conducted a study on the availability and decay of URLs cited in most articles in six LIS scholarly journals published by Emerald, Science Direct and Sage. The study showed that the original accessibility of online sources was 66% which improved to 95% when the Wayback Machine and Google were employed to recover the missing web citations in the library and information science journals.

In exploring search engine access, Saberi and Abedi (2012) employed the use of the Google search engine to increase the accessibility rate of online citations from 75% to 94% and the rate of decay decreased from 25 to 6%. another study submitted accessibility of online sources referenced in library and information science academics revealed that 1028 (36%) out of 2886 URLs cited disappeared when they were searched (Sadat-Moosavi et al. 2012). Several studies reported decay rates, for instance, Dimitrova & Bugeja (2007) reported that the .edu domain ending had the most active links in contrast to Goh and Ng (2007) who submitted that, .org had the most active links. Concerning the half-lives of online resources, studies have established different values ranging from 1.6 years (Rumsey, 2002) to 11.5 years (Sampath Kumar & Manoj Kumar, 2012).

As reported by Saberi & Abedi (2012) disclosed that 73% (2732) of URLs were accessible and the remaining 27% (1002) of URLs were unavailable of the total 3734 online citations. Internet Explorer, Internet Archives, and Google were used to discover online citations that disappeared by using different keywords to search. After searching, accessibility of the disappeared URLS increased from 73 to 89%, while, the related decay of online citations reduced from 27 to 11%. In the same vein, Sampath Kumar and Vinay Kumar (2013) investigated 1290 URLs cited in 472 research articles published in Indian LIS journals for the period of 9 years (2002–2010) This study found that out of 6820 sources cited, only 1290(18.91%) URLs were cited these journal articles. The study further revealed that 39.84% of URL citations were not accessible while the remaining 60.15% of URL citations were still accessible. The usage of Wayback Machine assisted in recovering 229(44.55%) from the total of 514 missing online sources cited.

Another study on vanished cited URLs in two Indian LIS Journals reported that 1,290 URLs were referenced in 472 articles published during 2002-2010. The study revealed that 39.84% of the cited URLs disappeared and the remaining 60.15% of Web citations were still accessible. Internet Achieve was used to recover 44.55% of missing URLs and this increased the percentage of accessible URLs from 39.84% to 77.90% (Sampath Kumar & Vinay Kumar, 2013). Prithviraj and Sampath Kumar (2014) investigated Indian LIS conference proceedings published from 2001 to 2010 and they disclosed that 50.09 per cent of URLs of web citations were not working and, the average half-life of missing URLs was estimated to be around five years. Similarly, Burnhill, Mewissen and Wincewicz (2015) studied 46,000 cited URIs appended to 6,400 e-theses downloaded from institutional repositories of five US institutions. They reported that 36.7% of URIs are not available on the live web, however, time travel was used to recover 18.3% of rotten links.

Vinay Kumar, and Sampath Kumar (2017), reported that 38.12 % (417 out of 5197) URLs vanished when the URLs were searched, but Internet Archive and Google were used to recover these vanished URLs from 61.88% to 87.11% and 73,58% respectively the study confirmed that Internet Archive is more effective than Google search engine in recovering of vanished URLs. In a study published in the web magazine ARIADE during 2010-2012, Gul et al. (2014) revealed that the ‘HTTP 404-file not found’ top error, representing 52.68 per cent of all HTTP error codes. The study further showed that many web citations were in “HTML” and “HTMLS” formats. The “ppt” files were most durable with 100 per cent accessibility while the domain “. com/.co” was the most stable and persistent domain with 95 per cent accessibility. In terms of missing web sources, a study of 822 articles published in four East African health science online journals, Sife and Lwoga (2017) reported that 574 were the web sources in the journals studied. These scholars submitted that 253 (44.1%) web citations were unreachable when searched. Out of the 1109 articles published in LIS journals, 4562 web citations were appended. The rate of unavailability of the URLs cited reduced from 34% to 5% after adopting various recovery techniques (Tajeddini, Azimi & Moghaddam, 2017).

In the same vein, Król (2019) studied the link rot in the websites of rural tourism facilities. In all, 919 websites were extracted and revealed that 464 sites have broken links. More recently, Manjunatha, Sampath Kumar and Lakshmana (2020) investigated the use of URLs as citations and their longevity in articles published in selected library journals between 2011 and 2015. Out of 36,968 references in 966 papers with an average of 38.26 per article, there were 5,867 URL references. W3C link checker was used to ascertain their accessibility. The study submitted that 46.53% URLs (2,730 out of 5,867) remained active in contrast, the remaining 3,137 (53.46%) disappeared.

It was reaffirmed in the literature reviewed that citation is an important aspect of research. The existence of the World Wide Web significantly contributes to research as scholars make use of web resources and such resources are cited in references. The reviewed literature also emphasized the challenges posed by the disappearances of URLs cited when researchers attempt to retrieve them for use. The missing URLs are attributed to the nature of the internet among scholars of the literature reviewed. The ever-increasing decay of URLs is worrisome to scholars and as a result of these concerns, the literature reviewed showed that many studies have been done by scholars worldwide especially by library and information scientists to recover some of these inaccessible URLS cited using different techniques and retrieve tools which include search engines, Internet Archive, Wayback Machine, Internet Explorer, Time Travel just to mention a few.

Based on the reviewed literature, it could be deduced that the Internet Archive and other tools are capable of recovering decayed URL citations. It was also revealed that the Internet Archive could be an indispensable asset for the authors of LIS research. It is therefore concluded that the existence of the recovery tools such as Internet Archive, and other web archives allow researchers to recover some of the vanished URLs on the web thereby making them available for use.

Although many studies have been carried out by scholars in library and information science, the majority of the investigations used library and information science journals out of Africa as case studies. This study, therefore, focused on the African Journal of Library, Archives and Information Science (AJLAIS), one of the library and information science journals in Africa that are ranked in Scimago. The objective of this study is to investigate the disappearance and retrieval of missing URLs using the Internet Archive in AJLAI (2008-2017).

Research Questions

The main objective of this study is to analyze and recover missing URLs of the African Journal of Library, Archives, and Information Science (AJLAIS) using the Internet Archive from 2008 to 2017. This research is guided by the following research questions:

1. What is the extent of URL Citations in articles published in AJLAIS?
2. How accessible and Missing URLs are in the AJLAIS in the period covered?
3. What are the types of HTTP error messages accounted for the miss3ng URL citations in AJLAIS?
4. What are the URL domains with active, missing, and recovered URL citations in AJLAIS in the period covered?
5. Is there a significant correlation between the path depth and URL citation decay of the AJLAIS?
6. What is the extent of recovered missing URLs through the Internet Archive by year of the AJLAIS?

Hypotheses

The hypotheses formulated for this study are:

$H0_1$: There is no significant relationship between the growth in number of citations and the growth of URL citations.

$H0_2$: There is no significant relationship between URL path depth and the missing URL citations.

Materials and Methods

This study employed a quantitative approach as data were collected to investigate the missing and recovery of URLs cited in AJLAIS between 2008- 2017. A descriptive research design was used to carry out the study.

Selection of Articles and References

The study was undertaken to analyze and recover missing URLs of articles published in the African Journal of Library, Archives and Information Science (AJLAIS) for 10 years (2008 to 2017) using the Internet Archive. The 2011 volumes of the Journal were not included as all efforts by the researchers to get the copies for the year did not yield results. All the references appended as a list at the end of each article under the heading references/bibliography sections were collected. The editorial articles, abstracts, expanded bibliographies, endnotes, footnotes, e-mail links, annotations and book reviews, short communications, and documentaries were not considered citations and thus not counted in the data collected for further study. In some articles, citations referring to print sources and web-located citations which are listed twice in the references/bibliography sections were counted as single citations. Researchers manually counted the references cited at the end of each article in this African Journal of Library, Archives and Information Science (AJLAIS).

Selection of Web Citations

Web citations [URLs)were extracted from the list of references/bibliography of the African Journal of Library, Archives and Information Science (AJLAIS). In contrast, listing/selecting/counting the URL citations, the duplicate or repeat URL in the same article twice or more than twice was considered/calculated as a single URL only, irrespective of the number of times it was cited in the reference list/bibliography of that article. However, if the same URL was cited in other articles, it was considered an independent URL for statistical analysis. The URLs so extracted/selected were listed separately yearly in the Journal (AJLAIS).

Testing of URLs

The web citations (URLs) extracted were tested to determine whether they were active or missing on the Web. Researchers checked all of them one by one in the World Wide Web Consortium's (W3C) Link Checker, at first instance to test the links (URLs) associated with a cited web resource. The Link Checker is a freely available online service (http://validator.w3.org/checklink) that tests a submitted URL for broken or non-valid hypertext links and reports the type of HTTP error messages encountered for unavailable URLs. The Link Checker tool was selected for its unique features to test the persistence and accessibility of URLs. Upon completion of the checking of the URLs using W3C Link Checker, those URLs that lead directly to the web source were categorised as active web citations, and those unavailable URLs with HTTP error messages were classified as inactive/missing web citations. Multiple reasons for citation failure exist. The exact error message was recorded and then classified according to the types of errors (ex: HTTP 403, HTTP 404, HTTP 500, HTTP 502, HTTP 503, etc.) as used by Kumar and Sushmitha (2019). Furthermore, the URL for each online citation was coded for a top-level domain (ex: .com, .org, .gov, .edu, etc.,), and URL path depth (ex: 0, 1, 2, 3, etc.,). Two round follow-up checks of Journal (AJLAIS) URLs were done. The results of such tests were noted and compiled to form statistical tables and analyzed to examine the reasons for the non-accessibility and to calculate the half-life value of web citations in this journal. Statistical Package for Social Sciences (SPSS) software, version 20, was used for the analysis, while Product Moment Correlation was used for the testing of the hypotheses at a 0.05 level of significance.

Recovery of missing URLs

The missing URL citations which gave HTTP error messages were entered in the search box of Internet Archives (http://www.archive.org/web/web.php) by copying the exact URL as they appeared in the references. The "Take Me Back" button was clicked to search for the accessibility of the submitted missing web citation. The missing URL submitted was retrieved/found, then it was considered as an active web citation: otherwise, it was considered a missing web citation. The error messages shown by the Wayback Machine for the inactive/missing URLs were noted and categorized.

Results

Analysis of the data collected from the AJLAIS is presented by using tables and diagrams to visualize the data at a glance. The results were drawn based on analysis and interpretation and explanations were also given at the very end of every table.

Research question 1: What is the extent of use of URL Citations in articles published in AJLAIS (2008-2017)?

Table 1 indicates that 129 articles were published in AJLAIS from (2008-2017) and contained 3414 citations, out of which 986 were URLs (28.88%). The percentage of the URL citations varied. The highest URL citations occurred in 2016 with 297 URLs (51.03%) of the total citations for that year, on the other hand, the most negligible percentage of URLs was in 2009 with 71 citations (16.13%).

Table 1
Distribution of Articles, Citations and URLs Citation by year in AJLAIS

Year	Total articles	Total Citations	Percentage	Total URLs	Percentage	Average URL citation per article
2008	16	354	10.37	76	21.46	4.75
2009	15	440	12.88	71	16.13	4.73
2010	14	318	9.31	77	24.21	5.5
2012	13	272	7.96	90	33.08	6.92
2013	15	370	10.83	85	22.97	5.67
2014	17	399	11.68	102	25.56	6.0
2015	15	453	13.26	99	21.85	6.6
2016	17	582	17.04	297	51.03	17.5
2017	7	226	6.61	89	39.38	12.7
Total	129	3414	100	986	28.88	7.6

Research Question 2: What is the URL accessibility and Missing URL in AJLAIS?

The URLs were checked using the W3C link checker to determine the URLs accessible and missing at the given URLs used in the references. The result is presented in Table 2. Table 2 gives the distribution of active and missing URLs by year. The result of the URL accessibility test in W3C link checker indicated that of the 986 URL citations, 47(61.8%) of URL citations were still accessible in the year 2008, 28(39.4%) 2009, 52(67.5%) 2010, 90(71.1%), 49(57.6%) 2013, 83(81.4%) 2014, 78(78.8%) 2015, 204(68.7%) 2016 and 71(68.6%) 2017. The total of active URL citations was 676(68.6%).

The URLs were also checked using the W3C link checker to determine the URLs missing at the given URLs used in the references. Table 2 gives the distribution of missing URLs by year. The result of the URL accessibility test in the W3C link checker indicated that of the 986 URL citations, 310(31.4%) of URL citations encountered were missing. The range of missing URLs also varied. The percentage of missing URL citations decreased significantly from 29(38.2%) in the year 2008 to 18(20.2%) in the year 2017. The table revealed that 29(38.2%) of missing URL citations was 2008, 43(60.6%) 2009, 25(32.5%) 2010, 26(28.9%) 2012, 36(42.4%) 2013, 19(18.6) 2014, 21(21.2%) 2015, 93(31.3%) 2016 and 18(20.2%) 2017

Table 2
Distribution of Active and Missing URL in AJLAIS (2008-2017)

Year	URL Citations	Active URL citations	% of Active URL citations	Missing URL citations	% of Missing URL citations
2008	76	47	61.8	29	38.2
2009	71	28	39.4	43	60.6
2010	77	52	67.5	25	32.5
2012	90	64	71.1	26	28.9
2013	85	49	57.6	36	42.4
2014	102	83	81.4	19	18.6
2015	99	78	78.8	21	21.2
2016	297	204	68.7	93	31.3
2017	89	71	76.8	18	20.2
Total	**986**	**676**	**68.6**	**310**	**31.4**

Research Question 3: What are the types of HTTP error messages encountered for the missing URL citations?

The summary of various Hypertext Transfer Protocol (*HTTP*) error codes is presented in Table 3. Of the 310 missing URL citations, 62(20.0%) HTTP 500 Internal Server Error, 59(19.0%) HTTP 404 Not Found, 46(14.8%) HTTP 400 Bad Request response, 43(13.9%) HTTP 410 Gone, 32(10.3%) HTTP 415 Unsupported Media Type, 23(7.4%) HTTP 502 Bad Gateway server error, 15(4.8%) HTTP 403 Forbidden, 13(.42%) HTTP 302 Found redirect status, 9(2.9%) HTTP 503 Service Unavailable and 8(2.6%) HTTP 300 error message- “page not found”.

Table 3
HTTP Error-wise distribution of missing URL citations in AJLAIS

HTTP Error	Count	Percentage
HTTP 500	62	20.0
HTTP 404	59	19.0
HTTP 400	46	14.8
HTTP 410	43	13.9
HTTP 415	32	10.3
HTTP 502	23	7.4
HTTP 403	15	4.8
HTTP 302	13	4.2
HTTP 503	9	2.9
HTTP 300	8	2.6
Total	**310**	**100**

Research Question 4: What are the URL domains associated with active, missing and recovered missing URL citations in AJLAIS between 2008 and 2017?

The domains associated with cited URLs, missing URL citations, recovery of missing URLs, and total URL active citations are summarized in Table 4. Of the 986 URL citations, the organizational domain is the top domain accounted for (21.5%) followed by the commercial domain (19.4%). The study also examined the extent of missing URLs by domain, with a total number of 310 missing URL citations, the highest percentage of missing URLs were associated with the commercial domain (17.1%) followed by the education domain (14.1%) and geographical domain (13.5%). The available data also indicated that very few numbers of missing URLs were associated with domains such as .mil (2.9%), .net (2.9%), and .res (3.5%). Table 4 also indicates the recovery of missing URL citations concerning their domains. The URLs associated with .com (19.8%) were the highest, followed by .edu (16.9%), geo (12.8%) and .gov (12.4%) while the least recovery was made for ERNET (2.1%).

Table 4

Distribution of URL domains associated with active, missing and recovered missing URL citations in AJLAIS between 2008-2017

URL Domains	Total URL active citation		Missing URL citations		Recovered URL citation	
	Number	%	Number	%	Number	%
CO/COM	191	19.4	53	17.1	48	19.8
EDU	153	15.5	45	14.1	41	16.9
ORG	212	21.5	31	10.0	25	10.3
AC	87	8.8	23	7.4	17	7.02
GOV	128	13.0	34	11.0	30	12.4
NIC	19	1.9	13	4.2	7	2.9
ERNET	6	0.6	9	2.9	5	2.1
NET	21	2.1	15	4.8	13	5.3
RES	10	1.01	11	3.5	8	3.3
INT	9	0.9	12	3.9	10	4.1
GEO DOMAINS	122	12.3	42	13.5	31	12.8
MIL	5	0.5	9	2.9	6	2.5
INFO	23	2.3	13	4.2	1	2.5
Total	986	100	310	100	242	100

Hypotheses

H0$_1$: There is no significant relationship between the growth in number of citations and the growth of URL citations.

Table 5 shows that 129 articles were published from 2008-2017. The table revealed that 354(22.1%) numbers of citations was 2008, 440(29.3%) 2009, 318(22.7%) 2010, 272(20.9%) 2013, 370(24.7%) 2014, 399(23.5%) 2015, 582(34.2%) 2016, 226(32.9%) 2017. The highest average citations per article were in the year 2016 (34.2 citations per article), followed by 2006 (17.06 citations per article) and 2017 (32.9 citations per article). The URL citations grew during the years 2008-2017.

Table 5

Growth in the number of citations and the growth of URL citations

Year	Total articles	Total citation	Average citation per article	URL citations	Average citation per article
2008	16	354	22.1	76	4.75
2009	15	440	29.3	71	4.73

2010	14	318	22.7	77	5.5
2012	13	272	20.9	90	6.92
2013	15	370	24.7	85	5.67
2014	17	399	23.5	102	6
2015	15	453	30.2	99	6.6
2016	17	582	34.2	297	17.5
2017	7	226	32.9	89	12.7
Total	129	3414	26.5	986	7.6

The assumption of Figure 1 was that the growth in the number of citations and the growth of URL citations do not correlate with the growth of URL citations.

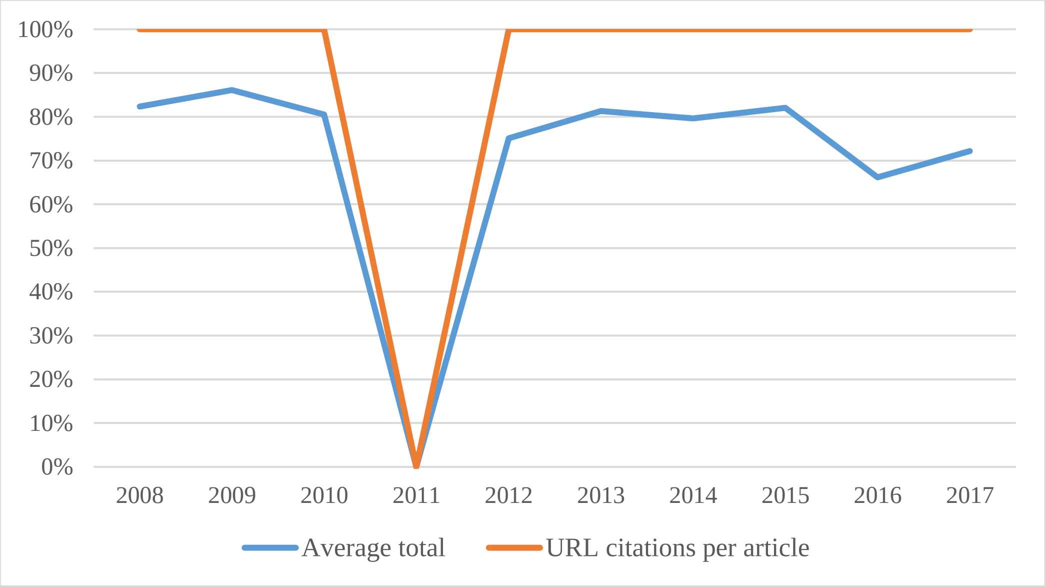

Figure 1: Active URL citations per article and Missing URL citations per article

The statistical analysis shows that the number of citations and URL citations do not correlate (r = .543; p = .105), hence the hypothesis that states that there is no significant relationship between the growth in number of citations and the growth of URL citations is accepted (Table 6).

Table 6
Relationship between the growth in the number of citations and the growth of URL citations

		$\bar{x}$	SD	Number of citations	Growth of URL citations
Number of citations	Pearson Correlation			1	.543
	Sig. (2-tailed)	12.90	5.36		.105
	N			10	10
Growth of URL citations	Pearson Correlation			.543	1
	Sig. (2-tailed)	98.60	75.41	.105	
	N			10	10

H02: There is no significant relationship between URL path depth and the missing URL citations.

Path depth

Goh and Ng (2007) submitted that the URLs path depth could be associated with link failure due to increasing complexity as the length of a URL increases. In the present study, each of the URLs path depth was categorized into different levels 0, 1, 2, 3...6.7,8, 9, 10. The purpose of this method is to find out the association between web-link path depth and missing URLs. The data presented in Table 7 reveals that of the 986 URL citations, the highest number of cited URLs belonged to the path depth level '9' (14.0%) followed by path depth level '7' (13.7%). Very few numbers of URL sources with a path depth level of 8 and above have been cited. One of the objectives of this study was to understand the relationship between path depth and missing URL citations. The highest percentage of missing URLs (15.5%) were found in the missing URL citations associated with Path depth level 2 (15.5%) followed by Path depth level 5 (13.5%) and Path depth level 10 (11.0%). The correlation between the path depth and missing URLs was measured and found a negative correlation between these two. However, the data presented in Table **7** indicates that the highest percentage of recovery was achieved for the missing URLs with path depth level 6 (95.2%) followed by path depth level 5 (92.7).

Table 7
URL path depth and the missing URL citations

Domains	Total URL citation		Missing URL citations		Recovered URL citation	
	Number	%	Number	%	Number	%
PD=0	111	11.3	33	10.6	21	63.6
PD=1	127	12.9	22	7.1	18	81.8
PD=2	43	4.4	48	15.5	35	72.9
PD=3	89	9.0	25	8.1	18	72.0
PD=4	83	8.4	16	5.2	14	87.5
PD=5	94	9.5	41	13.2	38	92.7
PD=6	67	6.8	21	6.8	20	95.2
PD=7	135	13.7	26	8.4	23	88.5
PD=8	26	2.6	25	8.1	22	88.0
PD=9	138	14.0	19	6.1	13	68.4
PD=10	73	7.4	34	11.0	20	58.8
Total	986	100	310	100	242	78.1

The highest percentage of missing URLs (15.5%) were found in the disappeared URL sources associated with Path depth level 2 (15.5%) followed by Path depth level 5 (13.5%) and Path depth level 10 (11.0%). The correlation between the path depth and missing URLs was measured and found a negative correlation between these two. (r= -.332; p = .319), this relationship is substantiated by Table 8, hence the hypothesis that there is no significant relationship between URL path depth and the missing URL citations is accepted.

Table 8
Relationship between URL path depth and the missing URL citations

		$\bar{x}$	SD	URL path depth	Missing URL citations
URL path depth	Pearson Correlation			1	-.332
	Sig. (2-tailed)	89.64	36.56		.319
	N			11	11
Missing URL citations	Pearson Correlation			-.332	1
	Sig. (2-tailed)	28.18	9.81	.319	
	N			11	11

Discussion

Although citations in printed materials formed the central part of the sources in the journal with 71.1%, the finding however shows that URL citations rose from 2008-2017. The percentage of the URLs ranged from a low of 21.46 % of the total sources and it grew higher to 39.38% of the citations in 2007. This aligns with the findings of Kumar and Sushmitha (2019). This could be attributed to the extensive use of the web as an essential source of information. Similarly, this discovery is in line with the finding of Manjunatha, Sampath Kumar and Lakshmana (2020) which reported an increasing trend in the URLs in LIS journal articles studied. A similar result earlier presented by Yang, Qiu and Xiong (2010) revealed a noticeable impact of web-based sources on scholarly journals in humanities and social sciences in China and web citations have grown significantly in the social sciences from 2006 to 2007.

The finding also indicates that the URL accessibility test in the W3C link checker showed that of the 986 URL sources, (68.6%) of URL citations were still accessible. This discovery is in line with the finding of Vinay Kumar and Sampath Kumar (2017) reported that the result of the URL accessibility test in the W3C link checker indicated that of the 2133 URL citations, 61.42% of URL sources were still accessible while the remaining 38.58% of URL citations encountered access errors. The percentage of missing URL citations decreased significantly from 78% in the year 2006 to 12.90% in the year 2015.

The finding indicates that URL citations in the articles in AJLAIS were 986 in the period covered. The average URL source per article in the journal was 7.6. There were inconsistencies in the development of the URL citations per article in the journal. This is in agreement with Kamar and Sushmitha (2019). Although previous studies such as Dimitroval and Bugeja (2007), and Sife and Bernard (2013) have established that web citations are more popular among scholars in research, this present study submitted some variance because of the inconsistencies in the growth of URL citations. This may be attributed to challenges facing the use of web resources in less developed countries in the African continent.

The finding also indicates that HTTP error messages encountered for the disappeared URL citations are HTTP 500, HTTP 404 and HTTP 400. This discovery is contrary to the finding of Vinay, Divyashree & Sampath Kumar (2019) who reported that HTTP error 404- Forbidden accounted for a more significant part of the errors. This finding also disagrees with Parmar and Pateria (2019) who reported that HTTP Error 404 (Page not found) was the most missing URLs. This finding is however in agreement with Kumar, Vinay Kumar, and Prithviraj (2015) and Vinay Kumar, Sampath Kumar and Parameshwarappa (2015) who submitted that the highest

HTTP error in the study of two Library and Information Science (LIS) journal articles published by Emerald Publishers during 2008 and 2012. The finding also indicates that the domains associated with cited URLs, disappeared URL citations and recovery of missing URLs are organizational domain, commercial domain, education domain and geographical domain. This result is consistent with the results found in the previous studies conducted by Wren (2008); Janakiramaiah and Doraswamy (2011); Vinay et al. (2019). This study found that the Internet Archive could retrieve disappeared web citations which resulted in an overall recovery rate of 78.19%. This finding tallies with the discovery of Prithviraj and Sampath Kumar (2014) which reported that the Internet Archive resulted in an increase in the percentage of active URLs from 49.91% to 79.08%.

The finding further shows that the most active domain is .org, which tallies with Kumar and Sushmitha (2019) and negates Vinay et al. (2019) which revealed that .org is the most missing domain. The study showed a negative correlation between between path depth and disappeared URL citations. This discovery is contrary to the submission of Goh and Ng (2007), and Prithviraj and Sampath Kumar (2014) but the finding is supported by a recent study of Vinay et al. (2019).

Limitation

This study did not study file formats associated with missing and recovered URLs published articles in this journal during the years covered. These could be investigated by other studies in future while similar studies could be carried out on this journal or other library and information science journals in Africa using different web Archives such as Bibliotheca Alexandria, DBpedia, NARA Archive, Website along Internet Archive.

Conclusion

The study has shown the analysis of accessibility, missing and recovery of missing URLs of articles published in the African Journal of Library, Archives and Information Science (AJLAIS) published during 2008-2017 using the Internet Archive The study was guided by seven research questions and two hypotheses. The finding indicates that 129 articles were published from 2008-2017, this excludes 2011 and volume II of 2017 as they were not published by AJLAIS. Print citations are still dominating web citations in the journal as only 28.88 percent of sources are reported as web citations in the journal. The result indicates that the URL accessibility test in the W3C link checker indicated that of the 986 URL citations, (68.6%) of URL sources were still accessible. The discovery shows that the percentage of missing URL citations decreased significantly from 29(38.2%) in the year 2008 to 18(20.2%) in the year 2017.

The result shows that HTTP error messages encountered for the missing URL citations are HTTP 500, HTTP 404 and HTTP 400. The conclusion indicates that the domains associated with cited URLs, missing URL citations and recovery of missing URLs are organizational domain, commercial domain, education domain and geographical domain. The finding shows that the URL citations grew from 2008-2017. The discovery shows a significant relationship between the percentage of missing URLs and the age of URLs. The study has not found any association between the path depth and the missing URLs.

Recommendations

Based on the findings of the study, the following recommendations were made;

1. The citations to web content should be complete and include full bibliographic information plus the date on which the sites were accessed by the authors.

2. The Editorial Board of the journal could institute a policy that will encourage authors to cite web citations in the subsequent articles that will be published as the percentage of URL citations is still low for sources in printed documents.

3. The Editor of the journal should check URLs referenced in manuscripts submitted by authors for peer review to minimize the number of missing URLs.

4. The editorial staff of journals should develop guidelines, for authors and references about the type of URL sources permissible, based on consideration permanency of the cited content and its scholarly importance.

5. Author(s) should confirm the accessibility of URL citation before it is used in the reference list of the manuscripts submitted for publication in the Journal.

6. Aside from the above, AJLAIS could be using the Digital Object Identifier (DOI) system.

7. There is also a need for authors to retain digital backup or printed copies of cited internet-only information to facilitate content recovery should a URL become unavailable from the Internet.

8. Editorial staff should work with authors to preserve and make available cited URL sources. One possible strategy would be to support the development and maintenance of the Internet Archive.

References

Aina, L.O. (2002). African Journal of Library, Archives and Information Science as a resource base for library and information science research in Africa, *Africa Journal of Library, Archives and Information Scienc*e 12(2), 167-176.

Burnhill, P., Mewissen, M. & Wincewicz, R. (2015). Reference rot in scholarly statement: Threat and remedy. *Insights,* 28(2), 55-61. http://dx.doi.org/10.1629/uksg.237

Casserly, M.F. & Bird, J.E. (2003). Web citation availability: analysis and implications for scholarship. *College* and *Research Libraries.* 64(7), 300-317.

Costa, M., Gomes, D. & Silva, M. J. (2017). The evolution of web archiving. *International Journal on Digital Libraries*, 18(3), 191-205. https://doi.org/10.1007/s00799-016-0171-9

Dimitrova, D.V. & Bugeja, M. (2007). The half-life of internet references cited in communication journals. *New Media & Society,* 9(5), 811-826. https://doi.org/10.1177/1461444807081226

Germain, C. A. (2000). URLs: Uniform resource locators or unreliable reliable resource locators? *College and Research Libraries,* 6(4), 359-365. https://doi.org/10.5860/crl.61.4.359

Goh, D. H. L. & 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

Gul, S., Mahajan, I. & Ali, A. (2014). The growth and decay of URLs citation: A case of an online library & information science journal. *Malaysian Journal of Library and Information Science*, 19(3), 27–39.

Janakiramaiah, M. & Doraswamy, M. (2011). *Measuring impact of web resources in Conference Proceedings: A citation analysis in CALIBER*, 541-549.

Król, K. (2019). The link rot phenomenon and its influence on the quality of the websites of rural tourism facilities in Poland. *Economic and Regional Studies/ Studia Ekonomiczne i Regionalne*, 12(1), 68-79. https://doi.org/10.2478/ers-2019-0007

Kumar, B. T. S., Vinay Kumar, D. & Prithviraj, K.R. (2015). Wayback machine: Reincarnation to vanished online citations, Program: electronic library and information systems, 49(2), 205-223. https://doi.org/10.1108/PROG-07-2013-0039

Kumar, D.V. & Sushmitha, M. (2019). Recovery of missing URLs cited in annals of library and information studies: A study of time travel, *Annals of Library and Information Studies (ALIS),* 66(1), 22-32. Retrieved from http://op.niscair.res.in/index.php/ALIS/article/view/22420

Manjunatha, G., Sampath Kumar, B. T. & Lakshmana, H. (2020). Longevity of URL citations Cited in LIS journal articles: A Webometric Study. *Library Philosophy and Practice (e-Journal)*. 3965. Retrieved from https://digitalcommons.unl.edu/libphilprac/3965

Markwell, J. & Brooks, D. W. (2003). "Link rot" limits the usefulness of web-based educational materials in biochemistry and molecular biology. *Biochemistry and Molecular Biology Education*, 31(1), 69-72. https://doi.org/10.1002/bmb.2003.494031010165

McCown, F., Chan, S., Nelson, M. L. & Bollen, J. (2005). The availability and persistence of web references in D-Lib Magazine. Retrieved from https://arxiv.org/ftp/cs/papers/0511/0511077.pdf

Moghaddam, A. S., Saberi, M. K. & Esmaeel, S.M. (2010). Availability and half-life of web references cited in information research journals: A citation study. *International Journal of Information Science and Management*, 8(2), 57-75. Retrieved from https://ijism.isc.ac/article_698149_2f3b10898eabe799e626be2f8eeae18b.pdf

Nagaraja, A., Joseph, S. A, Polen, H. H. & Clauson, K. A. (2011). Disappearing act: Persistence and attrition of uniform resource Locators (URLs) in an open access medical journal, *Program: Electronic Library and Information Systems*, 45(1) 98–106. https://doi.org/10.1108/00330331111107420

Onyancha, O. B. (2009). A citation analysis of Sub-Sahara African library and information science journals using Google Scholar, *African Journal of Library, Archives Information Science*, 19(2), 101-116. Retrieved from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=63c089319a6474b8a918c31306fcffa64d9d0bfd

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)*. 3595. Retrieved from https://digitalcommons.unl.edu/libphilprac/3595

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

Rumsey, M. (2002). Runaway train: Problems of permanence, accessibility, and stability in the use of Web sources in law review citations. *Law Library Journal*, 94, 27–39.

Riahinia, N., Zandian, F. & Azimi, A. (2011). Web citation persistence over time: A retrospective study, *The Electronic Library*, 29(5) 609-620. https://doi.org/10.1108/02640471111177053

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

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). https://doi.org/10.1108/00012531211215196

Sampath Kumar, B. T. & Kumar, K. S. Manoj (2012). Persistence and half-life of URL citations cited in LIS open access journals. *Aslib Proceedings*, 64(4), 405-422. https://doi.org/10.1108/00012531211244752

Sampath Kumar, B.T. & Prithvi Raj, K.R. (2012). Availability and persistence of web citations in Indian LIS literature, *The Electronic Library*, 30(1)19-32. https://doi.org/10.1108/02640471211204042

Sampath Kumar, B.T. & Vinay Kumar, D. (2013). HTTP 404-page (not) found: Recovery of decayed URL citations. *Journal of Informetrics*; 7(1), 145–157. https://doi.org/10.1016/j.joi.2012.09.007

Sellitto, C. A (2004). Study of missing web-cites in scholarly articles: Towards an evaluation framework. *Journal of Information Science*, 30(6) 484-495. https://doi.org/10.1177/0165551504047822

Sife, A. S. & Bernard, R. (2013). Persistence and decay of web citations used in theses and dissertations available at the Sokoine National Agricultural Library, Tanzania. *International Journal of Education & Development Using Information & Communication Technology*, 9(2), 85–94. Retrieved from https://files.eric.ed.gov/fulltext/EJ1071354.pdf

Sife, A.S. & Lwoga, E.T. (2017). Retrieving vanished Web references in health science journals in East Africa. *Information and Learning Sciences*, 118 (7/8), 385-392. https://doi.org/10.1108/ILS-04-2017-0030

Spinellis, D. (2003). The decay and failures of web references. *Communications of the ACM*, 46(1), 71-77. https://dl.acm.org/doi/fullHtml/10.1145/602421.602422

Tajeddini, O., Azimi, A. & Moghaddam, H. S. (2017). Death of web citations: a serious alarm for authors. *Malaysian Journal of Library & Information Science*, 16(3), 17- 29.

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

Tajedini, O., Sadatmoosavi, A., Ghazizade, A. & Tajedini, A. (2018). Investigation of the currency, disappearance and half-life of urls of web resources cited in iranian researchers: A comparative study, *International Journal of Information Science and Management*, 16(1), 2018, 27-47. Retrieved from https://ijism.isc.ac/article_698265_ab1eb66d82d85b389042dea84ad15880.pdf

Vinay Kumar, D. & Sampath Kumar, B.T. (2017). Prevalence of URLs in Library and Information Science (LIS) Literature: A citation analysis. *COLLNET Journal of Scientometrics and Information Management.* 11(2), 287-297. https://doi.org/10.1080/09737766.2017.1299390

Vinay Kumar, D. & Sampath Kumar, B.T. (2017). Recovery of vanished URLs: Comparing the efficiency of Internet Archive and Google. *Malaysian Journal of Library & Information Science*, 22(2), 31-43. https://doi.org/10.22452/mjlis.vol22no2.3

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. https://doi.org/10.18329/09757597/2015/8105

Vinay, R. S., Divyashree G. N. & Sampath Kumar, B. T. (2019). Decayed URLs in LIS journal articles: An exploration. In *9th KSCLA National Conference on Library in the Life of the User* (pp. 540–544). Department of Studies and Research in Library and Information Science, *Tumkur University*. Retrieved from https://www.researchgate.net/profile/Vinay-R-S-2/publication/331554316_Decayed_URLs_in_LIS_Journal_articles_An_Exploration/links/5c7fd0d492851c69505a8385/Decayed-URLs-in-LIS-Journal-articles-An-Exploration.pdf

Wren, J. D. (2008). URL decay in MEDLINE-a 4-year follow-up study. *Bioinformatics*, 24(11), 1381-1385. https://doi.org/10.1093/bioinformatics/btn127

Wu, Z. (2009). An empirical study of the accessibility of web references in two Chinese academic journals. *Scientometrics*, 78(3), 481–503. https://doi.org/10.1007/s11192-007-1951-1

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 citations. *Scientometrics*, 84(1), 1–19. https://doi.org/10.1007/s11192-009-0142-7