

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

2019

Accessibility and Decay of Web Citations in Computer Science journals

Mohsen Jalali

Institute of Higher Education of Urmia Science and Technology, Urmia, Iran, jalalimohsen750@gmail.com

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>



Part of the [Computer Engineering Commons](#), and the [Library and Information Science Commons](#)

Jalali, Mohsen, "Accessibility and Decay of Web Citations in Computer Science journals" (2019). *Library Philosophy and Practice (e-journal)*. 2968.

<https://digitalcommons.unl.edu/libphilprac/2968>

Accessibility and Decay of Web Citations in Computer Science journals

Mohsen Jalali

Institute of Higher Education of Urmia Science and Technology, Urmia, Iran

E-mail: jalalimohsen750@gmail.com

Abstract

The aim of this research is to scrutinize the accessibility and decay of web citations (URLs) used in refereed articles published by 27 Computer Science open access journals as indexed by Scopus. To do this, at first, we downloaded 1000 articles of Computer Science open access journals from 2009 to 2018. After acquiring articles, their web citations are extracted and analyzed from the accessibility and decay point of view. Moreover, for initially missed web citations complementary pathways such as using Google search engine are employed. Then, data collected are analyzed using descriptive statistical methods. Research findings indicated that 80.7% of articles have Web citations and rate of articles containing URL has increased from 72% in 2009 to 94% in 2018. The study revealed that at first check 78% of URLs are accessible, while 22% per cent have disappeared. It is notable that the rate of accessibility increased to 92% and the rate of decay decreased to 8% after using complementary pathways. The .edu/.ac domain, with an availability of 89% (a decay of 11%) has the greatest stability and persistence among all domains, while the most stable file format is PDF, with an availability of 85% (a decay of 15%).

Keywords: Web Citations, Accessibility, Decay, Computer Science journals

Introduction

According to Germain (2000) and Maharana, Nayak and Sahu (2006) Internet is one of the most important inventions in the history of mankind and Millions of users access the Internet every day for various beneficial purposes from hobby and communications to research and learning. Since the emergence of the Internet, there has been a continuous increase in the volume of scholarly resources in electronic form. These resources have provided a scope for researchers in various subject fields and stimulated their research productivity. Therefore, citation behavior of researchers has been influenced and they have been interested in Web citations (Isfandyari-Moghaddam, saberi and Mohammad Esmaeel, 2010). The main reason for the increase in the number of Web citations in papers is that the Internet has become the first choice for finding research information, scientific discoveries and keeping up with colleagues at other institutions (Zhao and Logan, 2002). Nevertheless, in this optimistic atmosphere the instability and shakiness of the Internet is a problem for researchers. They are worried about the decay and annihilation of the citations, which are gradually reducing accessibility to Internet citations (Germain, 2000; Dimitrova and Bugeja, 2007; Goh and Ng, 2007; Wagner et al., 2009). Therefore, this study attempts to delineate the accessibility and decay of web citations in the Computer Science journals.

Literature review

Harter and Kim (1996) in an article entitled "Electronic journals and scholarly communication: a citation and reference study" done one of the first studies on availability and permanency of URLs. The major purpose of their research was to study the effects of scholarly, peer-reviewed e-journals on formal scholarly and scientific communication, as measured by cited references. Accordingly, they extracted and examined 47 unique URLs of 39 scholarly, peer-reviewed e-journals published during 1993 to 1995 (roughly the end of 1995). Finally, they showed that one third of citations (31%) became inaccessible at the end of 1995 (Harter and Kim 1996).

Zhang (1998) investigated how much electronic sources have been used in formal scholarly communication, using a case study in the area of LIS during

1994–1996. As measured by e-resources cited, the study found that the impact of e-sources in scholarly communication in LIS is small as compared to that of print resources (Zhang, 1998).

Germain (2000) did a research to investigate the reliability of URLs in academic citation. To do her study, 31 randomly chosen academic journal articles, containing 64 citations with URLs, were reviewed. It is worth saying that the academic journals used were from a variety of disciplines. 13 citations were from information and library science, 10 from the hard sciences, 17 from computer science, 11 from the humanities, and 13 from the social sciences. The printed journals were published between 1995 and 1997. Results of this longitudinal study found an increasing decline in the availability of URL citations. Statistically, after a three-year period, almost 50 percent of the URL citations could not be accessed and two-thirds of the journal articles contained corroded citations (Germain, 2000).

Davis and Cohen (2001) made a citation analysis of undergraduate term papers in microeconomics and revealed a significant decrease in the frequency of scholarly resources cited between 1996 and 1999. Web citations checked in 2000 revealed that only 18% of URLs cited in 1996 led to the correct internet document (Davis and Cohen 2001). It should be reminded that Davis (2002) in an update to the 1996-1999 citation analysis mentioned above concluded that 65% of the citations pointed directly to the cited document, up from 55% in 1999 (Davis 2002).

Casserly and Bird (2003) examined 500 internet citations randomly chosen from scholarly articles published in library and information science journals. The average web citation for each paper was 2.5. They found that only 56.4% of those URLs were permanent, while the rest had disappeared from the original web address. Moreover, the error message "not available" was the most frequent message found by this study. They most frequent cited URLs belong to the domains ".org" and ".edu" (Casserly and Bird 2003).

Wren et al. (2006) explored URL decay in dermatology journals. To do this, they considered URLs in articles published between January 1, 1999, and September 30, 2004, in the 3 dermatology journals with the highest scientific impact. The percentage of articles containing at least 1 URL increased from 2.3% in 1999 to

13.5% in 2004. Of the 1113 URLs, 81.7% were available (decreasing with time since publication from 89.1% of 2004 URLs to 65.4% of 1999 URLs) ($P < .001$) (Wren et al. 2006).

Wagner et al (2009) studied the accessibility and decay of health care management journals during 2002-2004. They found that only 50.7% of those URLs were permanent, while the remaining 49.3% had disappeared from the original web address and results in error messages. The “.edu” domain with 68.4% accessibility, appear as the most stable domain (Wagner et al. 2009).

Saberi et al. (2011) explored the accessibility and decay of web citations URLs used in refereed articles published by Journal of Artificial Societies and Social Simulation (JASSS). For doing their research, they extracted 558 unique URLs referenced in 241 articles published from 1998 to August 2007. The study revealed that at first check 75% of web citations are accessible while 25% disappeared. It is notable that rate of accessibility increased to 94% and rate of decay decreased to 6% after using complementary pathways. The .edu/.ac domain with the accessibility of 98% (decay, 2%) has the most stability and persistency among all domains while the most stable file format is PDF with the accessibility of 99% (decay, 1%) (Saberi et al., 2011).

Objectives

The study has been conducted with the aim of the accessibility and decay of web citations (URLs) used in refereed articles published by Computer Science journals. The main purposes of the study include as follows:

- To determine distribution of articles, citations and Web citations;
- To determine distribution of articles which have Web citations;
- To study accessibility or availability status of URLs;
- To find out the types of domains from where the authors have cited (i.e. to find out the distribution of URLs by type of domain);
- To explore different file formats of the cited Web documents (i.e. to find out the distribution of URLs by type of file formats).

Methodology

This study is based on web citations of 27 Computer Science open access journals as indexed by Scopus. These journals are:

Rank	Title	Issn	SJR	SJR Quartile	H index	Country
1	Journal of Statistical Software	15487660	13,802	Q1	103	United States
2	SoftwareX	23527110	3,724	Q1	5	Netherlands
3	Journal of Machine Learning Research	15324435, 15337928	1,271	Q1	160	United States
4	Transactions on Data Privacy	20131631, 18885063	0,427	Q2	16	Spain
5	Journal of Cloud Computing	2192113X	0,412	Q2	15	United States
6	Diagnostyka	16416414	0,356	Q2	8	Poland
7	International Journal of Advanced Robotic Systems	17298814, 17298806	0,327	Q3	34	Austria
8	Kybernetika	00235954, 1805949X	0,321	Q3	31	Czech Republic
9	International Journal of Computer Games Technology	16877055, 16877047	0,287	Q3	16	Egypt
10	Journal of Computer Virology and Hacking Techniques	22742042, 22638733	0,271	Q3	8	United States
11	Modeling, Identification and Control	3327353	0,234	Q3	20	Norway
12	International Journal of Communication Networks and Information Security	20760930, 2073607X	0,227	Q3	15	Pakistan
13	Scientific Programming	10589244	0,215	Q3	28	Egypt
14	E-Informatica Software Engineering Journal	20844840, 18977979	0,213	Q3	4	Poland
15	Journal of Intelligent Systems	03341860, 2191026X	0,193	Q4	14	Germany
16	Journal of WSCG	12136972, 12136964	0,177	Q4	7	Czech Republic
17	Periodica polytechnica Electrical engineering and computer science	20645260, 20645279	0,170	Q4	9	Hungary
18	International Journal of Software Engineering and its Applications	17389984	0,162	Q4	14	South Korea
19	Journal of Object Technology	16601769	0,158	Q4	32	Switzerland
20	Journal of Computer Science	15493636, 15526607	0,147	Q4	21	United States
21	CrossTalk	21601577, 21601593	0,142	Q4	16	United States
22	Journal of Communications Software and Systems	18456421	0,134	Q4	9	Croatia
23	Webology	1735188X	0,126	Q4	12	Iran
24	Electronic Letters on Computer Vision and Image Analysis	15775097	0,117	Q4	6	Spain
25	Inteligencia Artificial	19883064, 11373601	0,117	Q4	10	Spain
26	International Journal of Soft Computing	18169503	0,109	Q4	12	Pakistan
27	Telfor Journal	23349905, 18213251	0,102	Q4	2	Serbia

The selection of these journals is based on their international validation through Scopus and Scimago Journal & Country Rank, the use of English language and the field of Computer Science, Software subject category.

Data collections were performed using the Internet. At first, all the published articles during 2009–2018 are downloaded from the journal archives. Accordingly, the total number of 4907 unique web citations is selected and then the accessibility of these citations is checked. When we could not directly access an e-address (link) inserted in any article by the author(s), we tried to view the referred web site. If this attempt appeared to be inadequate, the search engine “Google” was employed to access the missing reference(s).

Findings

In the present study, the Web resources referred to by authors in articles of computer science open access journals in the references section of their papers have been studied. A total number of 26112 citations, as obtained from the bibliographies of 1000 papers, have been analyzed and the necessary interpretations were made. Here, in line with research objectives mentioned earlier, research findings are provided as below:

Distribution of articles, citations and Web citations

As shown in Table 2, there were total number of 26112 citations in 1000 articles, and average "26.11 citations" was calculated per paper.

Additionally, among all citations (26112), there are 4907 Web citations with the average "4.90 Web citations" per paper. Comparing these results with findings of Casserly and Bird (2003) reveals that average of citations (26.11) and Web citations (4.90) per paper in computer science open access journals is more than average of citations (25) and Web citations (2.5) per paper in the journals studied by them.

Table2. Distribution of articles, citations, and Web citation

Year	Articles evaluated	Total citations	Average citations per paper	Total web citations	Average web citations per paper
2009	100	2465	24.65	351	3.51
2010	100	2624	26.24	380	3.8
2011	100	2987	29.87	412	4.12
2012	100	2543	25.43	478	4.78
2013	100	2822	28.22	498	4.98
2014	100	2756	27.56	460	4.6
2015	100	2398	23.98	571	5.71
2016	100	2169	21.69	511	5.11
2017	100	2497	24.97	594	5.94
2018	100	2851	28.51	652	6.52
All years	1000	26112	26.11	4907	4.90

Distribution of articles with Web citations

As can be seen in Table 3, of 1000 published articles 807 ones have Web citations. In a word, 80.7% of articles, at least, have one Web citation.

Table3. Distribution of articles with Web citation

Year	Articles evaluated	Articles with Web citation	Percentage of articles with Web citation
2009	100	72	72%
2010	100	69	69%
2011	100	78	78%
2012	100	71	71%
2013	100	84	84%
2014	100	82	82%
2015	100	83	83%
2016	100	85	85%
2017	100	89	89%
2018	100	94	94%
All years	1000	807	80.7%

Based on findings included in Table 3, it can be said that computer science authors have been increasingly making use of Web-based resources in their studies. Also, it is of importance that these results are considerably different from Wren and colleagues' (2006) because they reported that 13.5% of dermatology articles have Web citations.

Availability of cited URLs at first check

Figure 1 demonstrates that of 4907 cited URLs, 3827 (78%) were accessible, while only 1080 (22%) were inaccessible. In other words, 3827 URLs directly led to referred resources. Yet, in order to access 1080 URLs we encountered errors. Consequently, as mentioned earlier, when we failed to directly access the reference in the electronic address provided by the authors, the referred website was visited. If this attempt was also insufficient, search engine "Google" was employed to access the missing reference(s).

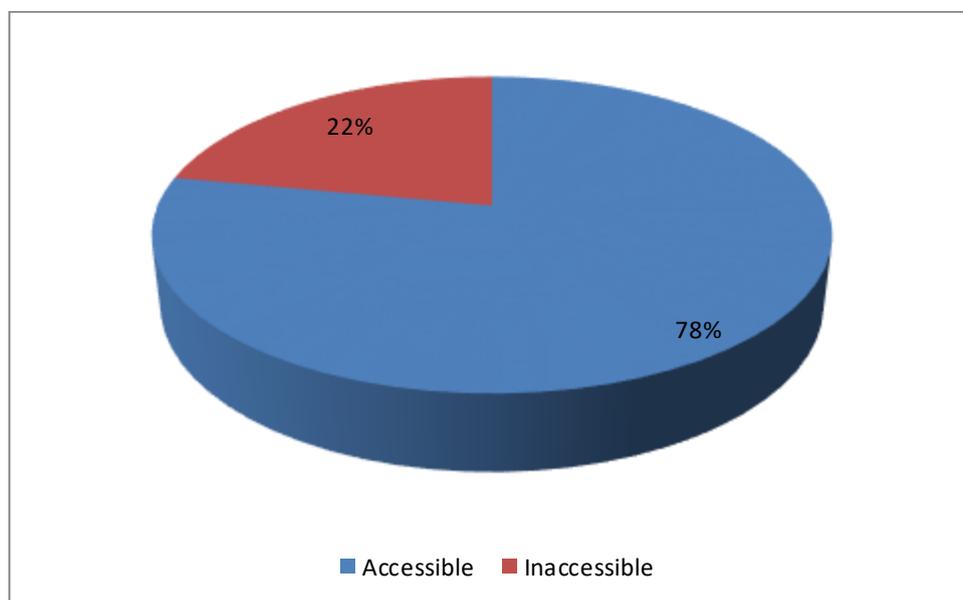


Figure1. Availability of URLs at first check

The status of the availability of cited URLs after searching through Google has been illustrated in Figure 2.

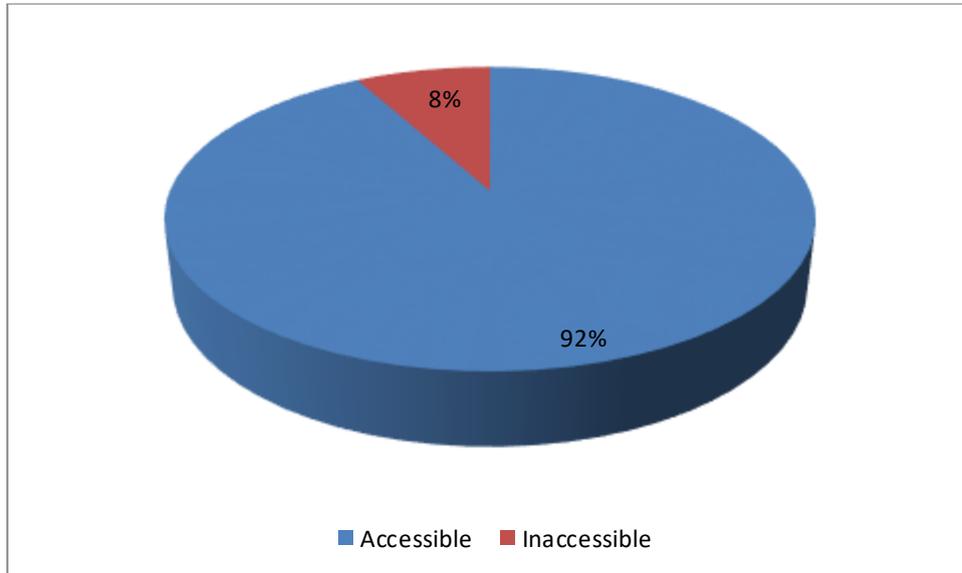


Figure2. Availability of URLs after searching at Google

As reflected in Figure 2, after searching missing URLs via Google, accessible URLs increased from 3827 (78%) to 4514 ones (92%). Conversely, inaccessible URLs decreased from 1080 (22%) to 393 cases (8%).

Distribution of URLs by type of domain

As one can see from Figure 3, the percentage of accessibility and decay of domains which indicated in Figure 4, proved that .edu/.ac with 89% accessibility (11% decay) is the most stable and persistent domain while .net domain with merely 59% accessibility (41%) is the least stable.

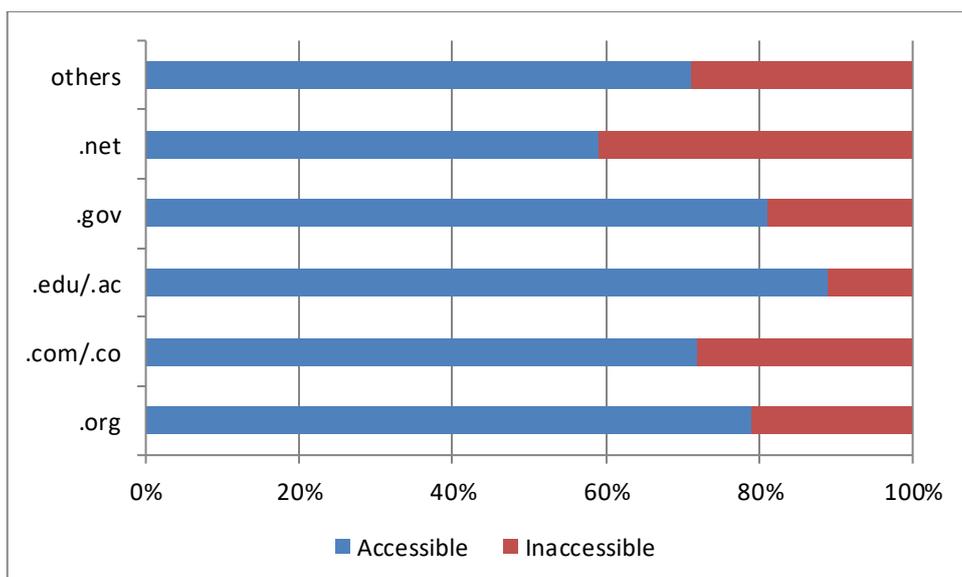


Figure3. Availability and decay of URLs by type of domain

Distribution of URLs by type of file formats

The percentage of accessibility and decay of file formats are indicated in Figure 4. Hence, PDF files appear as the most stable files (85% accessibility and 15% decay) and PPT files appear as the most instable files and susceptible to decay in the web citations (2% accessibility and 98% decay).

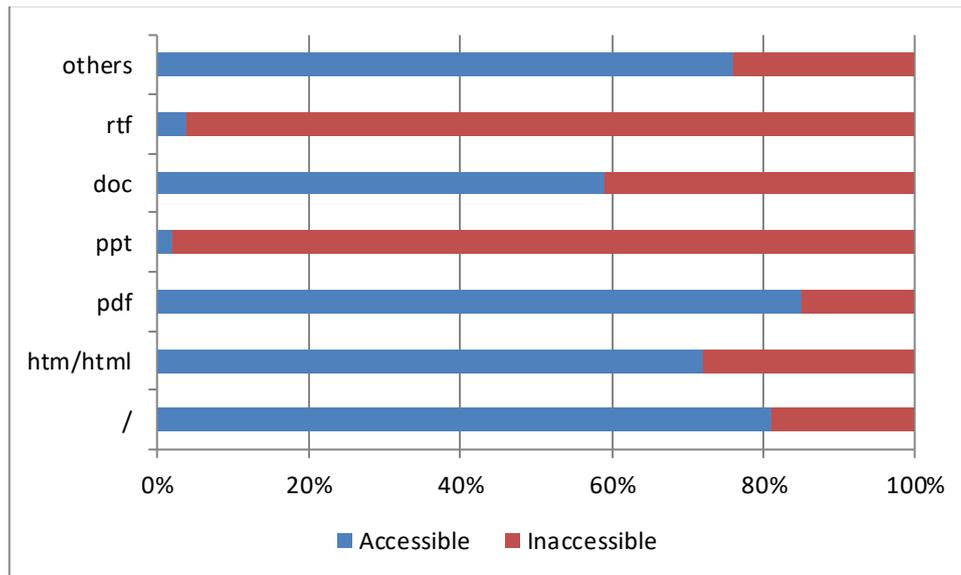


Figure4. Availability and decay of URLs by file format types

Discussion and conclusion

This research analyzed the web citations used in refereed articles published in Computer Science journals. Research findings indicated that among all citations (26112), there are 4907 Web citations with the average "4.90 web citations" per paper. Based the data collected, 80.7% of articles have Web citations and rate of articles containing web citations has increased from 72% in 2009 to 94% in 2018. In the study of accessibility and decay of web citations research findings showed that 78% (3827 URLs) are accessible and the remaining 22% (1080 URLs) are inaccessible at the first survey. It is notable that rate of accessibility increased to 92% (4514 URLs) and rate of decay decreased to 8% (393 URLs) after searching gone web citations through the internet and Google search engine.

Additionally, the .edu/.ac domain with the accessibility of 89% (decay, 11%) has the most stability and persistency among all domains while the most stable file format is PDF with the accessibility of 85% (decay, 15%).

At the end of this section it is valuable to investigate our results in a comparison context. Our results show that the accessibility of web citations in the Computer Science journals is 78% (22% decay) while the reported decay by other researches appear to be higher (Casserly & Bird (2003) 45.4%, Davis and Cohen (2001) 45%, Dimitrova and Bugeja (2007) 50%, Germain (2000) 50%, Wagner et al.(2009) (49.3%), Isfandyari-Moghaddam & Saberi (2011) 31%, Harter and Kim (1996) 31%, Goh and Ng(2007) (31%), Dellavalle et al.(2003) 13%, Saberi & Abedi (2012) 27%, Isfandyari-Moghaddam, et al.(2010), 27% , Saberi et al.(2011), 25%). Comparison findings indicated that, our obtained decay is lower maybe due to the effect of open access journal compared to non-open access ones which be under previous studies.

Today, the Internet has become a valuable, perhaps indispensable resource in conducting scientific research (Lawrence & Giles, 1999; Saberi & Ekhtiyari, 2019.), not just because of the added convenience of rapid information retrieval and sharing, but because it also provides a means of making resources available that the printed media simply cannot. Therefore, even though the authors may appreciate the risk of future inaccessibility of Internet references, they cannot easily avoid their use in their publications (Falagas, Karveli & Tritsaroli, 2007).

From what the present research showed us, it can be said that, as declared by Dimitrova and Bugeja (2007), the Internet may prove to be an inhospitable medium, especially for webbased research, because Web citations are speedily as well as constantly fading away. In spite of this, it should be accepted that the use of Internet for identifying valuable and timely information has become inevitable for most scientists as well as the public with access to the World Wide Web, since scientific and other work is created and added in digital format on the Internet every day. Such idea is reinforced by Dimitrova and Bugeja (2007) who declared as following:

"Internet research is vital to scholarship because the medium serves as a convenient electronic warehouse of data accessible at all hours and in great quantities, thereby increasing the scope and breadth of scholarship".

Due to the variety and rapid dissemination of information, scientific journals have more audiences compared to other information interfaces (Saber, 2009; Saber, M. K. and Barkhan & Hamzehei, 2019). In order to increase the rate of availability of URLs in scientific journals, Wren, Johnson, Crockett, Heilig, Schilling & Dellavalle (2006) suggested that publishers, editors, and authors should work together through:

1. Requiring authors to retain digital backup or printed copies of cited Internet-only information to facilitate content recovery should a URL become unavailable.
2. Advocating the inclusion of referenced Internet content in an online archive.
3. Checking URLs systematically before publication to minimize unavailability due to spelling errors or misprints.

References

Casserly, M. and Bird, J.E. (2003). Web citation availability: analysis and implication for scholarship. *College & Research Libraries*, Vol. 64, No. 4: 300-317.

Davis, P.M. and Cohen, S.A. (2001), "The effect of the Web on undergraduate citation behavior 1996-1999", *Journal of the American Society for Information Science and Technology*, Vol. 52 No. 4, pp. 309-14.

Dellavalle, R.P. Hester, E.J. Heilig, L.F. Drake, A.L. Kuntzman, J.W. Graber, M. and Schilling, L.M. (2003). Going, going, gone: lost Internet references. *Science*, Vol. 302, No. 5646:787-788.

Dimitrova, D. V. and M. Bugeja. (2007). The half-life of internet references cited in communication journals. *New Media & Society*, Vol. 9, No. 9: 811–826.

Falagas, M.E., Karveli, E.A and Tritsaroli, V.I. (2007). The risk of using the internet as reference resource: A comparative study. *International journal of medical informatics*, Available at:
<http://dx.doi.org/10.1016/j.ijmedinf.2007.07.001>.

Germain, C.A. (2000). URLs: uniform resource locators or unreliable resource locators. *College and Research Libraries*, Vol. 61, No. 4: 359-65.

Goh, D.H. and Ng, P.K. (2007). Link decay in leading information science journals. *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 1: 15-24.

Harter, S.P. and Kim, H.J. (1996). Electronic journals and scholarly communication: a citation and reference study. *Information Research*, Vol. 2, No. 1.

Isfandyari Moghaddam, A. and Saberi, M. K. (2010). Availability and half-life of Web references cited in *Information Research Journal: A citation study*. *International Journal of Information Science and Management*, 8 (2), 57-75.

Isfandyari-Moghaddam A. and Saberi M. K. (2011). The life and death of URLs: the case of *Journal of the Medical Library Association*. *Library Philosophy and Practice*. 592, Available at: <https://digitalcommons.unl.edu/libphilprac/592/>

Lawrence, S. and Giles, C.L. (1999), "Accessibility of information on the Web", *Nature*, Vol. 400, No. 2, pp. 107-9.

Maharana, B. Nayak, k. and Sahu, N. K. 2006. Scholarly use of Web resources in LIS research: a citation analysis. *Library Review*, Vol. 55, No. 9: 598-607.

Saberi M. K., (2009). Open Access Journals with a view of journals covered in ISI. *Information Sciences and Technology*, 24 (2), 105-122.

Saberi, M. K., Isfandyari-Moghaddam, A. & Mohamadesmaeil, S. (2011). Web Citations Analysis of the JASSS: the First Ten Years. *Journal of Artificial Societies and Social Simulation*, Vol. 14, No. 4: 22.

Saberi, M. K. and Abedi, H. (2012). Accessibility and decay of web citations in five open access ISI journals. *Internet Research*, Vol. 22 No. 2: 234-247.

Saberi M. K., Barkhan, S and Hamzehei, R (2019) "A Bibliometric Study and Visualization of *Library Philosophy and Practice* during 1998-2018". *Library Philosophy and Practice (e-journal)*. 2565 (2019), pp. 1-18.
<https://digitalcommons.unl.edu/libphilprac/2565>

Saberi, M.K., and Ekhtiyari, F., (2019). Usage, captures, mentions, social media and citations of LIS highly cited papers: an altmetrics study. *Performance Measurement and Metrics*, 20 (1), pp.37-47.

Wagner, C. Gebremichael, M.D. Taylor, M.K. and Soltys, M.J. 2009. Disappearing act: decay of uniform resource locators in health care management journals. *Journal of the Medical Library Association*, Vol. 97 No. 2: 122-30.

Wren, J.D. 2004. 404 not found: the stability and persistence of URLs published in MEDLINE. *Bioinformatics*, Vol. 20 No. 5: 668-672.

Zhao, D. Z and Logan, E. 2002. Citation analysis using scientific publication on the Web as data source: a case study in the XML research area. *Scientometrics*, Vol. 5, No. 3: 449-472.

Zhang, Y. (1998). The impact of Internet based electronic resources on formal scholarly communication in the area of library and information science: A citation analysis. *Journal of Information Science*, 24 (4), 241–254.