Death of web citations: a serious alarm for authors

Oranus Tajeddini[1], Ali Azimi[2], Ali Sadatmoosavi[3] and **Hadi Sharif-Moghaddam[4]**
[1]Department of Library and Information Studies,
Islamic Azad University, Sciences and Research Branch, Tehran, IRAN
[2]Tehran University of Medical Sciences, Tehran, IRAN.
[3]Institute of Biochemistry and Biophysics, University of Tehran, Tehran, IRAN.
[4]Department of Library and Information Science, Payame Noor University, Tehran, IRAN.
e-mail: tajedini.o@gmail.com[1]

ABSTRACT

The study explores availability and/or decay of URLs cited in articles of six Library and Information Sciences (LIS) journals published by Emerald, Science Direct and Sage. The research was performed using a descriptive survey method. Initially, all issues of the six journals including Information Processing & Management, Library & Information Science Research, Journal of Librarianship and Information Science, Journal of Information Science, Online Information Review, and Journal of Documentation from 2005 to 2008 were downloaded directly from their publisher websites. Afterwards, all the journals' citations in either print or Web formats were calculated manually. Then, availability and/or decay of individual cited URLs were examined in the Web environments. Two groups of URLs were identified as accessible (without any accessibility error) or inaccessible (with accessibility errors). Two groups of accessible URLs were "accessible through first-check" and "accessible through second check". Research findings indicated that 66% of articles had web citations. Original accessibility of web citations was 66% which improved to 95% by second check availability using the Wayback Machine and the Google. Overall, from 4562 cited URLs 34% had error messages mostly related to "File error" type. The study recommends that the best solution to prevent decay or disappearance of Web citations and diminish URLs decay is to check availability of citations from while they are being published. The Wayback Archive and the Google can revive the decayed citations.

Keywords: Citation analysis; Web citations; URL accessibility; URL decay; Library and Information Science journals.

INTRODUCTION

Since the quasi-miraculous emergence of the Web in 1990s, there has been a continuous increase in the volume of scholarly resources in electronic forms, such as e-books, e-journals, e-databases, e-theses and dissertations, e-prints of research papers, and the like. (Maharana, Nayak and Sahu 2006). Consequently, citation behaviour of researchers has become influential in the continuous process of research. As Zhao and Logan (2002) have indicated, the Web has become the first choice for seeking information, breaking scientific discoveries and for keeping up with colleagues at other institutions. In spite of the fact that the internet has eased the accessibility of information resources and citations, we are witnessed that the online citations disappear at increasing rates over time (Dimitrova and Bugeja 2007). Missing online citations are become a controversial issue for researchers and web managers. According to Rumsey (2002) and Tyler and McNeil (2003) one-third of online citations vanish from original web locations for several reasons. Hence, the present study aims to perform a study on the use of web citations and citation behaviour of the

authors of articles published in Library and Information Science (LIS) journals. Maharana and his colleagues declared that citing web resources properly according to an established style is important in most of the subject fields and it is different from citing traditional resources.

Apart from the style of web citations, quality, authenticity and sustainability are the issues with documents on the Web, demanding the immediate concern of the information professionals (Maharana, Nayak and Sahu 2006). Several studies have dealt with this general problem of "URL decay". The idea that the internet-based information, unlike the printed media, can suddenly disappear because of various reasons is often called 'URL decay', or 'link rot' (Wren 2008). This study aims to explore availability and persistence of URLs cited in articles published in LIS journals of Emerald, Science Direct and Sage publications. In the literature we have a thorough review of studies regarding the issue URL decay as follows.

LITERATURE REVIEW

After the emergence of the Internet, Web citations have been frequently considered, used and studied (e.g. in Harter and Kim 1996; Zhang 1998; Koehler 1999, 2002, and 2004; Germain 2000; Davis and Cohen 2001; McCown et al. 2001; Markwell and Brooks 2002 and 2003; Casserly and Bird 2003; Dellavalle et al. 2003; Spinellis 2003; Sellitto 2004; Wren 2004, 2008; Maharana, Nayak and Sahu 2006; Wren et al. 2006; Zhao and Logan 2002; Dimitrova and Bugeja 2007a; Falagas, Karveli and Tritsaroli 2007; Goh and Ng 2007; Wagner et al. 2009; and Wu 2009).

Harter and Kim (1996) performed one of the oldest studies on accessibility and decay of URLs. From e-journals published during 1993-1995, they examined 47 unique URLs and reported 31% as unavailable URLs. Casserly and Bird (2003) examined 500 internet citations randomly chosen from scholarly articles published in LIS journals. They found that only 56.4% of those URLs were accessible, while the rest were disappeared from the original web addresses. Furthermore, the study showed that more than half of the online citations contained incomplete information and the majority did not include a retrieval date. The study showed that "file not found" was the most frequent error message and close to half of the online citations were initially unavailable.

Koehler (2004) studied Web page availibility and reported that a static collection of general Web pages tends to 'stabilize' somewhat after it has 'aged'. The study concluded that the Web documents are not particularly stable media for the publication of long-term information and the maintenance of individual objects or items.

Wren et al. (2006) carried out an investigation into URL decay in dermatology journals. They considered URLs in articles published between January 1, 1999, and September 30, 2004, in the 3 dermatology journals with high impact factor. 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). They concluded that URLs are increasingly used and lost in dermatology journals. Loss will continue until better preservation policies are adopted.

Dimitrova and Bugeja (2007a) studied cited URLs in journalism and communication field. They reported availible URLs as 61%. The .org domain with 70% active links was the most available domain.

Falagas, Karveli and Tritsaroli (2007) in a study explored accessibility of online resources of *Lancet* and *New England Journal of Medicine*. They found that 3.9% of the *Lancet* and 2.5% of the *New England Journal of Medicine* references were web resources. The two journals' inaccessible online resources was totally 62.2% which reduced to 35% after searching missed URLs into Google,.

Goh and Ng (2007) studied accessibility and decay of URLs of 3 LIS journals during 1997-2003. They reported decayed URLs as 31%. 56% of unavailable URLs brought 404 errors. The .edu with 36% active links was the most persistent domain. Accordingly, the half-life of online resources was 5 years.

In an article by Aronsky et al. (2007), web citations from a 20% random sample of all publications released in PubMed during a one-month observational study period (Feb 21 to Mar 21, 2006) were identified. The study included 4,699 publications from 844 different journals. Among the 141,845 references there were 840 (0.6%) web citations. From the 840 Internet references, 11.9% were already inaccessible within two days after an article's release to the public.

Wagner et al. (2009) studied accessibility of online resources of medical healthcare management journals from 2002 to 2004. They extracted 2011 unique URLs from 5 dominant journals in the field. The accessibility analysis of URLs showed that only 50.7% of URLs were accessible while the rest were unavailable from their original web addresses.

METHODOLOGY

The objectives of the present study are to:
   a) determine the ratio of print and web citations, total and per journal articles, and total and per journal web citations;
   b) specify decay or availability of URLs;
   c) determine the URL accessibility/decay of URLs per type of domains and file formats; and
   d) study the resulted error messages at inaccessible URLs.

The study was performed during a six-month period from September 2010 to February 2011. The articles of six LIS journals published by Emerald, Science Direct and Sage appeared from the beginning of 2005 to the end of 2008 were studied. Selected journals were as follows:
   a) *Information Processing & Management* (INFORM PROCESS MANAG) (IF= 1.783)
   b) *Library & Information Science Research* (LIBR INFORM SCI RES) (IF= 1.236)
   c) *Journal of Librarianship and Information Science* (J LIBR INF SCI) (IF= 0.581)
   d) *Journal of Information Science* (J INF SCI) (IF= 1.706)
   e) *Online Information Review* (ONLINE INFORM REV) (IF= 1.423)
   f) *Journal of Documentation* (J DOC) (IF= 1.405)

From each publisher two journals based on their JCR impact factor (IF) rankings, 2009) were selected. Therefore, all 2005-2008 issues of the selected journals were downloaded to a local disk. Only publications which had citation list were considered for analysis. Editorials, brief communications, special reports, book reviews, etc were excluded, if they had no citations. Eventually, a unique set of 40,133 citations were recorded in a spread sheet. Web or print citations were identified manually.

At the next stage, all web citations were extracted and their URLs hyperlinks were tested by examining their URLs' functionality. Initially, accessibility was tested by directly click on the URLs' hyperlinks. Afterwards, two groups of URLs were identified as accessible (without any accessibility error) and inaccessible (with accessibility errors). Two sets of accessible URLs were "accessible through first-check" and "accessible through second check". All URLs which were accessible through the first examination (without error) and all URLs that retrieved messages indicating redirection (e.g. 'you are being redirected', 'this page has been moved', etc.) and returned the right citation content were considered "accessible URLs through first-check". Other URLs which were available by adopting heuristic strategies were included in the "accessible through second check". Availability examinations were carried out all weekends and for unavailable URLs which returned 5** errors (server errors), the availability examination was repeated four times for four weekends. If then the URL was unavailable, it was recruited for heuristic URL refinement.

We tried to modify unavailable URLs. Therefore, in case we faced with errors of URLs, we checked to find if the URL content is yet available through the web. Thus, as the first employed strategy, unavailable URL was entered into the Internet Explorer 7 (IE7) and if the URL worked, was considered accessible and saved into "accessible through second-check" records. Otherwise, if it did not respond within 60s or returned an error message it was considered as "missed URL". For avoiding unwanted errors, the URL was directly copied and pasted into the browser.

Missed URLs were rechecked for their likely errors in their strings. Therefore, as said by Wren (2004) non-standard signs, if any, because of space, %, \\ instead of //, http:/, ++, http@, non-alphanumeric characters (usually from non-English websites) or other rare misspelling in the URL were corrected manually, and then the corrected URLs were tested again for accessibility status. If the nonstandard URL worked into the IE7 browser, the URL was regarded as accessible and was saved in accessible through second-check records. Once more, if after a period of 60 seconds yet-inaccessible URL resulted no content or returned errors (e.g. "404 (not found)", "page was unavailable", "file not found" etc, errors), was regarded as "missed URL". Otherwise, was recorded in "accessible through second-check" list. String editing was not saved to the unavailable URLs.

At the next stage, path depth reduction strategy was used for unavailable or missed URLs. Based on the assumption that the lengthy URLs could be erroneous, a unit by unit depth reduction was performed. Unavailable URL strings sustained depth reduction in several steps. URL path depth was specified by a "/" after the top domain. Accordingly, an URL with just a top domain string (e.g. http://emeraldinsight.com) has a path depth of 0. Comparably, a string like http://emeraldinsight.com/journals/aslip.html has a path depth of 2.

Missed URLs were examined through a unique and a unit by unit path reduction operation. A unique operation was performed for every single of missed URLs. Therefore, after 1 unit path reduction, the URL was tested for availability. The reduction operation would be continued until either the path depth was=1 or the broken URL responded. If the URL worked in any depths ≥1, the operation would be finished and the URL was marked as available through the second-check. Otherwise (URL with depth=1 and yet unavailable), the URL was considered as unavailable. Path reduction was not saved to the missed URLs since they should be recruited for the next URL recovery strategy that was searching through an Internet Archive.

Thereafter, another availability check was established for the missed URLs using Wayback Machine[1] and then Google search. Wayback Machine is an old Internet Archive (IA) and likely is the most popular one (Klein 2008). The Google is also the most popular search engine. Therefore, the missed URLs were entered in the Wayback Machine by copying the exact URL given in the online citation. If the URL was found in Wayback, the URL was recorded in the "accessible through second-check records". If the URL content could not be found even via the Wayback Machine, it was recruited for Google search strategy stage. Up to 5 keywords extracted from the citation's author(s) name(s), title, and resource were entered to Google and the first 20-retrieved results were reviewed to find the extinct content.

Finally, if the adopted strategies yielded no results, the inaccessible URL was considered as "decayed" and the related errors were recorded on specific related notes. Then, the web citations of the studied journals' (either accessible or inaccessible) were classified based on their top domains and file formats. Using Microsoft Excel 2007 the collected data were analyzed and suitable related tables and figures were drawn.

RESULTS

The Ratio of Print and Web Citations, Total and per Journal Articles, and Total and per Journal Web Citations

According to Table 1, 1109 inspected articles had 40133 citations, in which 4562 (11%) were web citations. Average 30.21 citations were calculated for each paper. Moreover, among all citations (10242), there were 1761 Web citations with the average 5.19 Web citations" per paper.

Table 1: Distribution of Total and per Journal Articles, Total and per Journal Web Citations

Journal	Articles	Citations	Web Citations (%)		Articles Citing Web Citations(%)		Mean of Citations per Article	Mean of Web Citations per Article
Information Processing & Management	406	13407	754	6%	173	43%	33	4
Library & Information Science Research	109	5029	528	10%	104	95%	20	5
Journal of Librarianship and Information Science	72	2268	537	24%	62	86%	26	9
Journal of Information Science	185	7602	1076	14%	125	68%	41	9
Online Information Review	176	5362	974	18%	153	87%	30	6
Journal of Documentation	161	6465	693	11%	116	72%	40	6
Total	1109	40133	4562	11%	733	66%	36	6

Accessibility and Decay of Web Citations

Initially, from 4562 URLs, 66% (3001) were accessible and 34% (1561) were decayed. A total of 72% of *Journal of Information Science* URLs' were originally accessible (best

---

[1] . Available at: http://www.archive.org/web/web/php

performance), while 53% of *Library & Information Science Research* web citations were originally active.

Table 2: Availability of URLs at the First Check

Journal	Total number of URL (%)		Total
	Accessible [a]	Inaccessible [b]	
Information Processing & Management	502 (67%)	252 (33%)	754(100%)
Library & Information Science Research	280 (53%)	248 (47%)	528(100%)
Journal of Librarianship and Information Science	310 (58%)	227 (42%)	537(100%)
Journal of Information Science	771 (72%)	305 (28%)	1076(100%)
Online Information Review	647 (66%)	327 (34%)	974 (100%)
Journal of Documentation	491 (71%)	202 (29%)	693 (100%)
Total	3001 (66%)	1561 (34%)	4562(100%)

[a] without any accessibility error
[b] with accessibility errors

After passing adopted refinement strategies including considering IE7 browse, manual editing, path depth reduction, searching into Wayback Machine and the Google, the URL accessibility rate increased from 66% (3001) to 94% (648) and inaccessibility decreased from 34% (1561) to 6% (45). This means 28% improvement in web citations accessibility. Table 3 illustrates the improvement results per adopted strategies.

Table 3: Final Accessibility and Decay of URLs

Journal	Total number of URL (%)		Total
	Accessible	Decayed	
Information Processing & Management	719 (95%)	35 (5%)	754(100%)
Library & Information Science Research	483 (91%)	45 (9%)	528(100%)
Journal of Librarianship and Information Science	513 (96%)	24 (4%)	537(100%)
Journal of Information Science	1027 (95%)	49 (5%)	1076(100%)
Online Information Review	940 (97%)	34 (3%)	974(100%)
Journal of Documentation	648 (94%)	45 (6%)	693(100%)
Total	4330 (95%)	232 (5%)	4562(100%)

Figure 1 demonstrates how accessible URLs were found. Accordingly, of the total number of accessible URLs, 2699 (64%) were found at cited URL, 495 (11%) were found at Internet archive, 517 (12%) were accessed using Google search engine, 302 (7%) were found at another URL other than cited URL, 109 (3%) were found through depth reduction, 42 (1%) were found by editing URLs and 66 (2%) were accessed through searching missing URLs in the Internet.
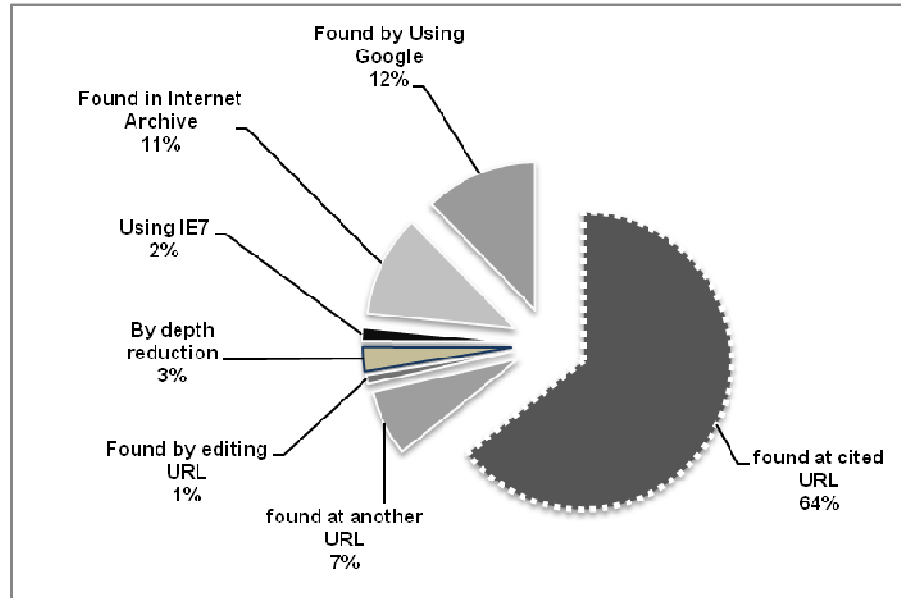

Figure 1: Percentages of Final Accessible URLs

The HTTP protocol defines 24 different errors that can occur within an HTTP exchange. (Spinellis 2003). In practice, whenever a URL is inaccessible an error message (HTTP code) appeares. In general, when URLs were checked we were faced with the following errors that were similar to previous studies (Wu 2009).The error massages have been categorized into three different types as classified in Table 4.

Table 4: Classification of error massages in case of URLs inaccessibility

Host name/sever Errors	File errors	Access restriction
1. The requested URL could not be retrieved. 2. Unable to determine IP address from host name. 3. Name Error: The domain name does not exist. 4.Here are some related websites for: qosforum.com 5. HTTP 400 – False request. 6. HTTP 500 – Internal server error. 7. Cannot find the server or DNS error. 8. 503 Service Unavailable	1. HTTP 404 – File not found. 2. Object not found! Error 404. 3. The document you requested does not exist on this server. The document may have existed previously, but was removed because it was out of date. 4. The website has been restructured. We have rearranged our website. 5. Not a very informative URL. 6. Redirect to a new URL and new web page without consistent content. 7. 410 Gone	1. 403 Forbidden. 2. Windows cannot visit the folder. 3. Your IP address is invalid for this session. 4. Your client does not have permission to access. 5. You do not have permission to access on this server. 6. Password needed for entry. 401 Unauthorized

Figure 2 summarizes three types of inaccessible web citations. "File error" was the biggest problem, in about 56% of cases, "Host name or Sever error" in about 17% of cases and those needing permission or password (access restriction) in 27%.
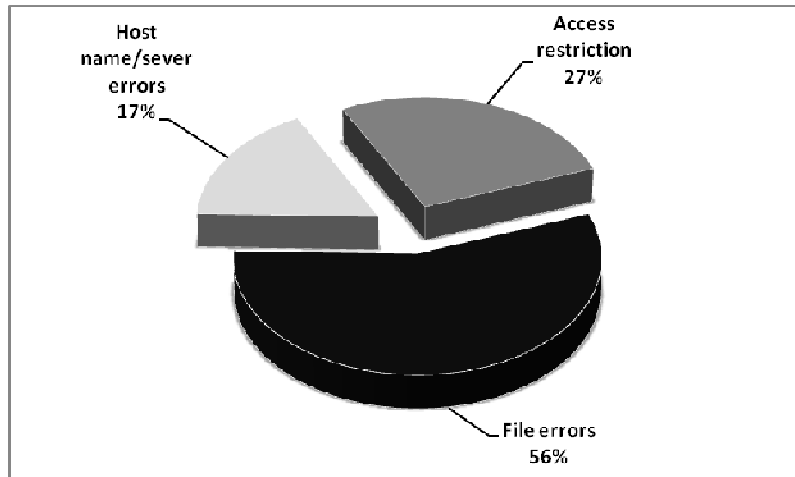
Figure 2: Percentages of HTTP Codes for the URLs' Accessibility Errors

Distribution of URLs Accessibility by Type of Domain and File Format

An URL is the address of the location of a digital document on the Web. A URL essentially has four parts: protocol, domain, directory and file. A domain name is the way to identify and locate computers connected to the Internet. No two organizations can have the same domain name. A domain name always contains two or more components separated by periods, which are called "dots". Some examples of domain names are: ibm.com, nasa.gov, utexas.edu and tcs.co.in. A domain name can often tell the user if it is a government site, an academic site or a commercial site. Some common top-level domain name endings are:

- .com or .co: a commercial organization;
- .edu or .ac: an educational organization;
- .gov: an official government site;
- .org: mostly non-profit organizations; and
- .net: traditionally it was for network organizations, but now can be used by anyone.

In this research, five different types of domain have been taken into consideration. They were .org, .edu/.ac, co/.com, .gov and .net, while those domains not falling into any of these categories fall into the "other" category.

As shown in Figure 3, the domains of the cited URLs mostly include the .com/.co and .org types. Accordingly, of 4562 Web citations, the highest number of domains, i.e. 1359 cases were of .org types and 999 ones were of .com/.co types. This revealed that the data sources of most of the Web citations in the present study were websites of various professional institutions or societies and commercial organizations, and the like. Also, 915 domains were of .edu/.ac, 273 domains (.net), 393 citations (.gov) and 623 cases belonged to domains other than the above types, categorized in the form of "others".
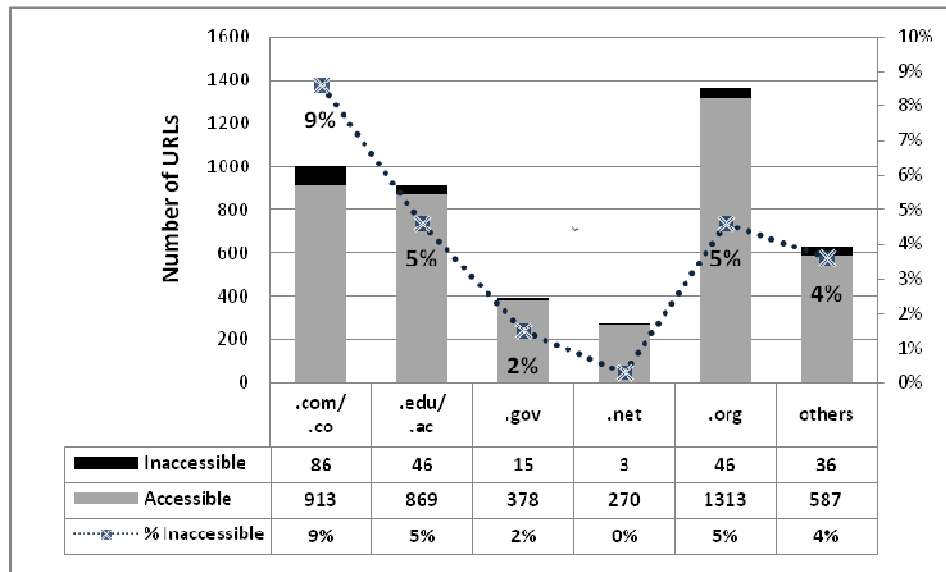
Figure 3: Accessibility and Decay of URLs Based on Their Top Domains

Similar to previous studies (McCown et al. 2001; Maharana, Nayak and Sahu 2006), the URLs were categorized into seven different file formats as follows:

- Slash files (/): URLs which end by / sign, for example, http://foo.edu/;
- HTM/HTML/SHTML (hyper text markup language): Web documents created in HTML scripting language;
- PDF (portable document format): the file format for documents created using Adobe Acrobat;
- PPT: PowerPoint presentations;
- DOC: documents created using MS-Word;
- RTF (rich text formats): a text file format that includes formatting features, such as bold, italic, and underlined texts; and
- Others.

The data as illustrated in Figure 4 indicate that the largest number of cited web resources were HTML/HTM/SHTML files. Of 4562 Web references, 1885 were HTML files, followed by1126 Slash files, 945 PDF files, 61 DOC files, 11 PPT files and 9 RTF files. 525 files, which did not match these six categories, were classified as the "other" category.

These findings are in agreement with that of McCown et al (2001) and Maharana, Nayak and Sahu (2006) which reported that most of cited Web resources contain HTML/HTM files. Additionally, as shown in Figure 4, PDF files are the most stable files. In McCown et al (2001), the most stable files were also PDF files.
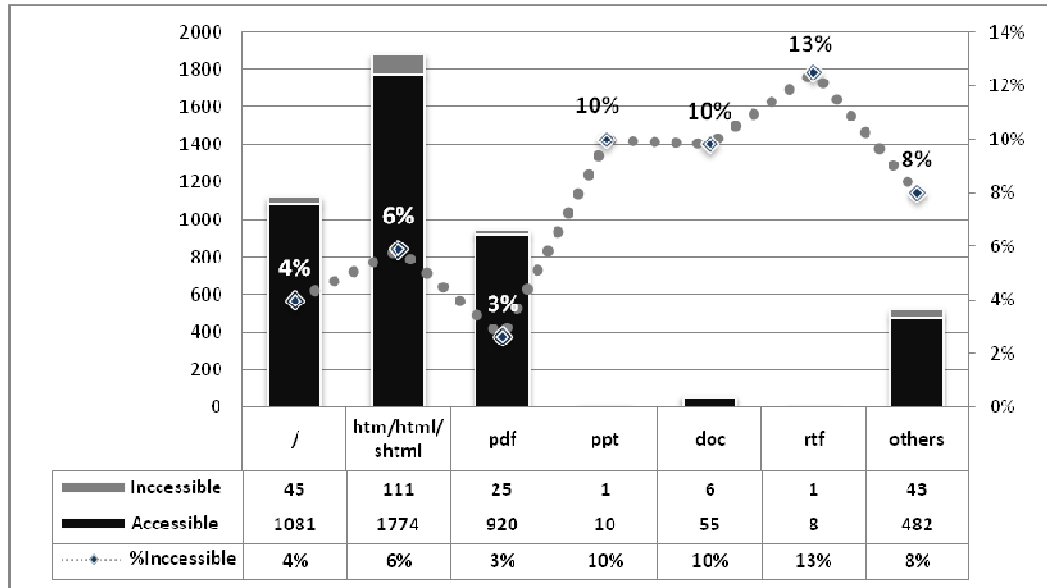
Figure 4: Distribution of URLs by Type of File Formats

DISCUSSION

"URLs decay" phenomenon is a relatively new topic studied highly in recent years basically because growing use of Web citations in scholarly papers (Zhao and Logan 2002; Maharan et al. 2006). Considering the Internet as the first choice of researchers is 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 (Wren 2004). Therefore, even though the authors may appreciate the risk of future inaccessibility of Web citations, they cannot easily avoid its use in their publications (Falagas, Karveli and Tritsaroli 2008). In spite of web advantages, Web resources have led us to a threatening challenge; citations are become decayed and disappeared.

The web citation decay rate was 34%. Previous studies reported the web citations decay as 50% (Germain 2000), 13% (Dellavalle et al. 2003), 45.4% (Casserly and Bird 2003), 39% (Dimitrova and Bugeja 2007a), and 49.3% (Wagner et al. 2009). The decay was decreased to 5% by adopting various strategies such as Google and Wayback Machine search, URL path reduction or truncation, and manual editing. The main strategies which revived more dead URLs were using the Google search (12%) and Wayback Machine (11%). Dimitrova and Bugeja (2007b) showed that the Wayback Machine performance in reviving unavailable URLs was largely better than the Google. They also showed that 64% of citations retrieved through the Google were also found in the Wayback Machine and only 36% of citations were uniquely available through the Google. In contrast, 67% of the citations found in the Wayback Machine did not overlap and 33% overlapped with the Google.

Investigation through the URL top domains showed that the (org) has received more citations than other domains. This finding was in agreement with Dimitrova and Bugeja (2007) and McCown et al. (2005). PDF with 97% accessibility were the most stable file format. This was not in agreement with McCown et al.'s (2005) study who reported HTML

files as the most stable format. There were three types of inaccessible web citations and the "file error" was the most prevalent error for about 56% of inaccessible citations.

Ultimately, we faced with 5% absolutely unavailable web citations even with employing useful services such as the Google, the Wayback Machine or the heuristic strategies. This should be considered as an issue of using web citations. Checking the availability of web citations prior to publishing submitted manuscripts could improve the availability status, since the authors are likely more informed with their used web citations and simply can modify the URL strings' errors or replace decayed URLs with live alternatives.

CONCLUSION

Internet may prove to be an inhospitable medium (Dimitrova and Bugeja 2007), especially for web-based research, because Web citations are speedily being used as well as constantly fading away. Nevertheless, it should be accepted that 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 (Dimitrova and Bugeja 2007).

In current study we examined some strategies for recovering dead citations in hope for patching the emerging hole in web citation area. However, in order to increase the rate of availability of URLs, it has been suggested that publishers, editors and authors should work together through:
a) Requiring authors to retain digital backup or printed copies of cited Internet-only information to facilitate content recovery should a URL become unavailable;
b) Advocating the inclusion of web citations in an online archive;
c) Checking URLs systematically before publication to minimize unavailability due to spelling errors or misprints (Wren et al. 2006; Dimitrova and Bugeja 2007).

In addition to the above recommendations using domains and files which are more stable is recommended. As a solution to prevent decay or disappearance of Web citations and diminish URLs decay use of *Wayback Machine* is recommended. According to Dimitrova et al (2007), it is "a part of the Internet Archive, a nonprofit organization devoted to preserving data, texts, audio, Web sites, and other digital materials since the early days of the online revolution. Nonetheless it had some limitations. For example, the Wayback Machine worked just for HTML based URLs. Therefore, we could not search the Wayback Machine for the FTP (file transport protocol) based URLs. Also, this reputable internet archive had some limitations in archiving dynamic pages and pages containing Java Scripts. The Wayback Machine did not archive the pages which had not external links to other websites. The Google was also unable to crawl the pages including active contents and pages using robots.txt codes. Using heuristic rules were rather tedious. The user is not always patient enough for manual editing of URL strings.

Citation decay means loss of data and this fact should be considered seriously for those who especially work with online resources such as open access journals or other free accessed materials on the web.

ACKNOWLEDGEMENT

REFERENCES

Casserly, M. and Bird, J.E. 2003. Web citation availability: analysis and implications for scholarship, *College & Research Libraries*, Vol. 64, no. 4: 300-317. Available at: http://crl.acrl.org/content/64/4/300.full.pdf+html

Davis, P.M and Cohen, S.A. 2001. The effect of the web on undergraduate citation behaviour 1996-1999. *Journal of the American Society for Information Science and Technology*, Vol. 52, no. 4: 309-314.

Dellavalle, R.P., Hester, L.F., Heiling, A., Drake, J.W., Kuntzman, M.G. and Schiling L.M. 2003. Going, going, gone: Lost Internet references. *Science,* Vol. 302, no. 5646: 787–88. Available at: http://www.sciencemag .org/content/302/5646/787.full.pdf

Dimitrova, D.V. and Bugeja M. 2007. Raising the dead: recovery of decayed online citations. *American Communication Journal,* Vol. 9, no. 2. Available at: http://www.quosafulltext.com/sc_ddm/sc_ddm.jsp

Falagas, M.E., Karveli, E.A., and Tritsaroli, V.I. 2008. The risk of using the Internet as reference resource: A comparative study. *International Journal of Medical Informatics*, Vol. 77, no. 4: 280-286. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1386-5056(07)00124-4

Germain, C.A. 2000. URLs: Uniform Resource Locators or Unreliable Resource Locators. *College and Research Libraries,* Vol. 61, no. 4: 359–65. Available at: http://crl.acrl. org/content/61/4/359.full.pdf.

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. Available at: http://onlinelibrary.wiley.com/doi/10.1002/asi.20513/pdf.

Harter, S.P. and Kim, H.J. 1996. Electronic journals and scholarly communication: a citation and citation study. *Information Research*, Vol. 2, no. 1: 9. Available at: http://InformationR.net/ir/2-1/paper9a.html

Klein, B. 2008. Google and the search for federal government information. *Against the Grain.* Vol. 20, no. 2: 30-34. Available at: http://www.against-the-grain.com/TOCFiles /20-2_Klein.pdf.

Koehler, W. 1999. An analysis of Web page and Web site constancy and permanence. *Journal of the American Society of information Science and Technology*, Vol. 50, no. 2: 162-180.

Koehler, W. 2002. Web page change and persistence-a four-year longitudinal study. *Journal of the American Society of information Science and Technology,* Vol. 53:162–171. Available at: http://www.onlinelibrary.wiley.com/doi/10.1002/asi.10018/pdf .

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. Available at: http://www.emeraldinsight.com/journals.htm? articleid= 1576422.

McCown, F., Chan, S., Nelson, L.M., and Bollen, J. 2005. The availability and persistence of web citations in D-Lib Magazine. Available at: http://www.iwaw.net/05/ papers/iwaw05-mccown1.pdf

Rumsey, M. 2002. Runaway train: Problems of permanence, accessibility, and stability in the use of web sources in law review citations. *Law Library Journal,* Vol. 94, no. 1: 27-39. Available at: www.aall.org/products/pub_llj_v94n01/2002-02.pdf.

Spinellis, D. 2003. The decay and failures of web citations. CACM, Vol. 46, no. 1: 71–77. Available at: http://www.dmst.aueb.gr/dds/pubs/jrnl/2003-CACM-URLcite/html/urlcite.html

Tyler, D.C. and McNeil, B. 2003. Librarians and link rot: a comparative analysis with some methodological considerations. *Portal: Libraries and the Academy,* Vol. 3, no. 4: 615–32.

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: 122-130. Available at: http://www.ncbi.nlm. nih.gov/pmc/articles/PMC2670212/?tool =pubmed.

Wren, J.D. 2004. 404 not found: the stability and persistence of URLS published in MEDLINE. *Bioinformatics.* Vol. 20, no. 5: 668-672. Available at: http://bioinformatics .oxfordjournals.org/ content/20 /5/668.full.pdf+html.

Wren, J.D. 2008. URL decay in MEDLINE—a 4-year follow-up study. *Bioinformatics*, Vol. 24:1381-1385.

Wren, J.D., Johnson, K.R., Crockett, D.M, Heilig, L.F., Schilling, L.M. and Dellavalle, R.P. 2006. Uniform resource locator decay in dermatology journals. *Archives of Dermatology,* Vol. 142: 1147-1152. Available at: http://archderm.ama-assn.org/cgi/reprint/142/9 /1147.pdf.

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,* Vol. 24, no. 4: 241–254. Available at: http://jis.sagepub.com /content/24/4/24.1

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. Available at: http://www.springerlink.com/content/P212416833U25L83/fulltext.pdf.

Zhiqiang, Wu. 2009. An empirical study of the accessibility of web citations in two Chinese academic journals. *Scientometric,* Vol. 78, no. 3: 481-503. Available at: http://www.springerlink.com/content/ dx77052378x8x203/fulltext.pdf