Recovery of vanished URLs: Comparing the efficiency of Internet Archive and Google

D. Vinay Kumar[1] and B. T. Sampath Kumar[2]
[1]Department of Library and Information Science, Kuvempu University,
JnanaSahyadri, Shivamogga, INDIA
[2]Department of Library and Information Science,
Tumkur University, Tumkur, INDIA
e-mail: vinay.86.kumar@gmail.com;
sampathkumar@tumkuruniversity.ac.in (corresponding author)

ABSTRACT

This article examines the vanishing nature of URLs and recovery of vanished URLs through Internet Archive and Google search engine. For that purpose study investigates the URLs cited in the articles of two LIS journals published during 2009-2013. A total of 226 articles published in two open access LIS journals were selected. Of 5197 citations cited in 226 articles, 21.05 percent were URLs (1094). Study found that 38.12 percent (417 out of 5197) URLs were found missing and remaining 61.88 percent of URLs were active at the time of URL check with W3C link checker. The HTTP 404 error message – "page not found" was the overwhelming message encountered and represented 54.2 percent of all HTTP error message. Internet Archive and Google search engine were used to recover vanished URLs. However, the Internet Archive recovered 66.19 percent of the total vanished URLs, whereas, Google manages to recover only 30.70 percent of the total vanished URLs. The recovery of vanishing URLs through Internet Archive and Google increased the active URL's rate from 61.88 per cent to 87.11 per cent and 73.58 per cent respectively. Study found that Internet Archive is a most efficient tool to recover vanished URLs compared to Google search engine.

Keywords: URL citation; Internet Archive; Google; Defunct URLs; Library & Information Science journals.

INTRODUCTION

Internet has been producing enormous quantity of electronic information and disseminates them to the needy at lightning speed via network. Perhaps, it has become a convenient tool for rapid information retrieval and sharing (Lawrence and Giles 1999). Information collections on the web are dynamic and beyond physical boundaries (Wang, Hawk and Tenopir 2000). The websites those are inherent in the web reinforces the scholarly communication. Many studies documented that use of Internet for scholarly information has become inevitable (Falagas, Karveli and Tritsaroli 2008; Goh and Ng 2007; Yang, Qiu and Xiong 2010).

Various studies such as those conducted by Sellitto (2004), Maharana, Nayak, and Sahu, (2006), Dimitrova and Bugeja (2007), Wu (2009), Moghaddam and Saberi (2011), Sampath Kumar and Prithviraj (2012) and Sampath Kumar and Vinay Kumar (2013) focused on the trends of using web resources in the scholarly communication and found that the scientific community is vigorously using web resources. The use of web resources has led scholars to

refer and cite more web resources as a part of their research activity (Jalalifard, Norouzi and Isfandyari-Moghaddam 2013).

Though the accessibility and citation trend to web resources is overwhelming, the cited URLs suffer from a problem of inconsistency. Spinellis (2003) opined the URLs become defunct due to changing web sites, relocating websites, restrictions to websites, incorrect citations, changes in organizational structure, hardware reconfigurations, file system reconfigurations. This problem is considered as the vanished URLs. Many studies documented the problems of vanishing nature of URLs as their age increases and tried to identify a solution to uphold the stability of URLs over time.

Internet based information is particularly useful for future researchers, since new conclusions and interpretations are based on previous research which also directs new research (Nagaraja et al. 2011). It is noteworthy for a researcher to confirm the permanency of web based information. However, web decay lacks the search and retrieval of scholarly information on the web. Despite, there is no universally accepted technical solution to control web decay one can use web archives to recoup missing web resources. Few studies have attempted to recover vanished URLs using web archives and search engines. This study focuses to examine whether any consistent tool that exist today to assist researchers in the recovery of missing URLs. The goal of this study is to compare the efficiency of Internet Archive and Google search engine and thereby, tries to identify the reliable tool to conquer the web decay phenomenon.

LITERATURE REVIEW

Vanished URLs

The studies on web citation analysis in the past have witnessed the most frequent use of URLs (Casserly and Bird 2003; Spinellis 2003; Sellitto 2005; Maharana, Nayak, and Sahu 2006; Falagas, Karveli and Tritsaroli 2008; Goh and Ng 2007; and Moghaddam and Saberi 2011; Sampath Kumar and Manoj Kuma, 2012; Gul, Mahaj and Ali 2014). Studies focused on web resources have witnessed the loss of URLs. URL decay is found to be significant phenomena in the past studies. In a study on URL decay, Casserly and Bird (2003) analyzed 500 URLs which are cited in library and information science journals between 1999 and 2000. They found that only 56.4 per cent of those URLs were permanent and the rest found to be missing. McCown et al. (2005) have analyzed About 4,387 unique URLs referenced in 453 articles published in D-Lib Magazine. They have checked the availability during September 2004 to February 2005. They reported a loss of 28 per cent URLs at initial check and 30 per cent failed to resolve at the last check. Sampath Kumar and Manoj Kumar (2012) in their article studied 2,890 URLs cited in two LIS journals viz., Information Research and DESIDOC Journal of Library and Information technology. Their study confined to a period of 14 years (1996-2009). They found that 26.08 per cent of all citations were not accessible during the time of testing and the majority of errors were due to HTTP 404 error code (not found). Nagaraja et al. (2011) observed that "URL decay" rate was on the increase over time. They compared the loss of URLs for research articles (14.5%) and non-research articles (18%). In a recent study by Gul, Mahaj and Ali (2014) found the loss of URLs cited in Ariadne web magazine. Their study confined to a span of three years (2010-2012). They reported that out of 1184 URLs a loss of 7.85% of URLs was identified.

Recovery of Vanished URLs

Many studies documented the use of Internet Archives and Google search engine to recover the vanishing URLs cited in journal article and conference proceedings. Wu (2009) studied web references cited in two key Chinese academic journals. Out of 1637 web references found 776 were inaccessible. Seventy eight inaccessible web references were randomly selected to know the rate of recovery. He used search engine and Internet Archive to recover the missing web references. Search engines found 62.8 percent, and Internet Archive (IA) found 24.4 percent of the missing web citations. The other strategies used to recover missing web citations includes searching keywords extracted from content descriptions (67.9%), searching URLs (27%), and 5.1 percent by navigating in the original website. Study revealed a fact that search engines in combination with Internet Archive retrieved 98% of the retrievable web references. In the same year, Wagner et al. (2009) examined a total of 2011 URLs were found cited in five medical science journals, of which 49.3 percent were found missing. They tried to find the missing online content and used Internet Archive and Google search engine. The most successful tool for finding the originally cited content at the 992 inactive URLs was using the Internet Archive's Internet Archive, which located 59.8 percent (593 items), followed by Google, which had links to 48.8 percent (484) of the missing material.

Saberi and Abdedi (2011) studied five open access journals indexed by the Institute for Scientific Information (ISI) from 2002-2007. This study witnessed 3734 web citations in which 1002 (27%) were inaccessible. Their study witnessed that the rate of accessibility increased to 89 per cent and rate of decay decreased to 11 per cent after searching gone URLs through the Internet, Internet archive and Google. Similarly, Tajeddini et al. (2011) explored the availability and decay of URLs cited in six Library and Information Science journals published during 2005-2008. Study identified 4562 (11%) of URLs cited in 1109 articles of which 1561 (34%) found inaccessible. The recovery strategies such as found at cited URL (64%), use of Internet archive (11%), use of Google search engine (12%), vanishing URLs found at another URL (7%), depth reduction method (3%) and editing URLs (1%) and searching vanishing URLs in the Internet (2%). His study identified the main strategies which revived more dead URLs were using the Google search (12%) and Internet Archive (11%).

In 2010, Moghaddam and Saberi's study on URLs cited articles published in Information research journal during 1995-2008 found that of 1761 cited URLs. Among these URLs, 1290 (73%) were accessible and 471 (27%) were found inaccessible. Internet Archive and Google search engine was used to retrieve inaccessible URLs. Internet Archive retrieved 28% per cent of inaccessible URLs whereas, Google search increased the rate of accessible URLs from 1290 (73%) to 1518 ones (86%). Conversely, inaccessible URLs decreased from 471 (27%) to 243 cases (14%). Similarly, Sadat-Moosavi, Isfandyari-Moghaddam, and Tajeddini (2012) studied four LIS journals published by Emerald to understand the web decay and recovery. Their study yielded a total of 2886 online citations of which 1028 were found missing. Original accessibility of studied online resources was 64 per cent. It was improved to 95 per cent by using recovery strategies through Internet Archive and Google, which revived online resources by 17 percent and 12 percent respectively. Table 1 summarizes studies that used Internet Archives and Google search engine to recover the vanishing URLs.

Table 1: Tools Used for Recovery of Vanished URLs and Percentage of Recovered URLs in Previous Studies

No	Author	Year	Tools used for recovery of vanished URLs	Percentage of recovered URLs
1	Wu	2009	Google	62.8
			Internet Archive	24.4
2	Wagner et al.	2009	Internet Archive	59.8
			Google	48.8
3	Saberi and Abedi	2011	Internet Archive	62
			Google	
4	Tajeddini et al.	2011	Internet archive	11
			Google	12
5	Sadat-Moosavi, Isfandyari-Moghaddam, and Tajeddini	2012	Internet Archive	17
			Google	12
6	Moghaddam and Saberi	2010	Internet Archive	28
			Google	14

OBJECTIVE

The objective of this study is to compare the efficiency of Internet Archive and Google search engine and thereby, tries to identify the reliable tool to conquer the web decay phenomenon. The research questions posed are as follow:

a) What is the proportion of articles, citations and URL citations cited in LIS journals published during 2009-2013?
b) What is the trend in URL citations and rate of URLs loss encountered?
c) What are the top level domains and file types that are associated with missing and retrieved URLs?
d) Which is the efficient tool to recover missing URLs among Internet Archive and Google search engine?

METHOD

Study considers two open access journals viz., *Webology* and *Annals of Library and Information Studies* (ALIS). *Webology* is an Iran based LIS journal whereas, ALIS is Indian based LIS journal. Since, these journals are open access in nature and available to the LIS readers and research community at free of cost, study considered these two LIS journals for collecting the required data.

All citations were extracted from 226 articles published in *Webology* (49) and ALIS (177) during 2009-2013. The URLs (URLs) cited in the list of references appended to each article published in the said journals are extracted and then tested to know their accessibility using W3C link checker (http://validator.w3.org/checklink). W2C link checker tests a submitted web page for broken or non-valid hypertext links and reports the exact HTTP error messages encountered against to the missing URL.

Recovery of Vanished URLs
Keeping in view the objectives of the study, the vanished URLs that are recognized by W3C link checker were then put into the Internet Archive (https://archive.org/index.php) and Google search engine (http://www.google.com) for recovery. The URLs that are not archived

and found missing has been considered as decayed URLs. The present study aimed to investigate the availability, decay of URLs and the recovery of vanishing URLs using Internet Archive and Google search engine. Further, study aims to compare the efficiency of Internet Archive and Google search engine in recovering missing URLs.

Efficiency measurement

Study is confined to measure the efficiency of Internet Archive and Google search engine in recovering the missing URLs cited in the articles published in *Webology* and ALIS. The 'Efficiency' in recovering the missing URLs in the LIS articles published in *Webology* and ALIS is based on the number of missing URLs recovered and support to the increased accessibility of cited URLs.

RESULTS AND DISCUSSION

From a total of 5197 citations were extracted from 226 articles in which 1094 are URLs. Table 2 summarizes journal-wise distribution of articles, citations and URLs.

Table 2: Year-wise Distribution of Crticles and citations

Year	Webology			Annals of Library and Information Studies		
	Total articles	Total citations	URL citations	Total articles	Total citations	URL citations
2009	7	161	41	34	675	39
2010	9	217	65	43	1039	202
2011	11	312	123	36	819	134
2012	10	344	167	27	465	138
2013	12	328	75	37	837	110
Total	**49**	**1362**	**471**	**177**	**3835**	**623**
	Average Citations per article - **27.80**			Average Citations per article - **21.67**		
	Average URLs per article - **9.61**			Average URLs per article - **3.52**		

Table 2 depicts the distribution of total articles, total citations and URLs that are appended to *Webology* and ALIS during the year 2009-2013. URLs have seen a steady growth during 2009-2012 in *Webology* and had a slight shortfall in the year 2013.The URL citation trend has seen variations in ALIS. The volume of URLs in ALIS is neither consistently increased nor decreased, but, varied. The dependency on the information sources of preceding years is high among LIS scholars.

The average citations per article are highest in *Webology* (27.80%) than Annals of Library and Information studies (21.67%). In case of URLs *Webology* accounted an average of 9.61 URLs per article, whereas, articles appeared in ALIS obtained an average URLs of 3.52. It is evident that the reliance on web resources is highest among International authors.

Table 3 reveals that out of 5197 citations, 1094 (21.05%) are URLs. The citations towards web resources has remarkably grown during the year 2009 (9.57%) to 2012 (37.70%). In the year 2013 the use of URLs in citations has a short fall and only 15.88% of URLs have been cited. Spinellis (2003) attributed that the shortfall of URLs is due to constraints on the URLs imposed by the journals editors.

The average citation per article has showed an increasing trend. An average of 4.84 URLs per article has been cited during 2009-2013. The year 2009 witnessed an average of 1.95 URLs

per article followed by 5.13 in 2010, 5.47 in 2011 and 8.24 percent in the year 2012. In 2013 the average URL citation per article has scored 3.78 per cent.

Table 3: Distribution of Articles, Ctations and URLs

Year	Total articles	Total citations	URL citations	Percentage of URLs	Average URL citations per article
2009	41	836	80	9.57	1.95
2010	52	1256	267	21.26	5.13
2011	47	1131	257	22.72	5.47
2012	37	809	305	37.70	8.24
2013	49	1165	185	15.88	3.78
Total	**226**	**5197**	**1094**	**21.05**	**4.84**

Missing and Active URLs

Table 4: Year-wise Distribution of Active and Missing URLs

Year	Total URLs	Active URLs	% of Active URLs	Missing URLs	% of missing URLs
2009	80	48	60.00	32	40.00
2010	267	145	54.31	122	45.69
2011	257	163	63.42	94	36.58
2012	305	191	62.62	114	37.38
2013	185	130	70.27	55	29.73
Total	**1094**	**677**	**61.88**	**417**	**38.12**

Table 4 shows the summary of active and missing URLs. The percentage of vanishing URLs varied from a low of 29.73 in the year 2013 to a high of 45.69 in the year 2010. A total of 38.12 percent of URLs were not accessible at the time of testing with W3C link checker.

HTTP Errors associated with Missing and Recovered URLs

URLs failure is connected with different HTTP errors. The type of errors may be HTTP 403, HTTP 404, HTTP 500, HTTP 502 and HTTP 503. The summary of various error codes associated with vanishing URLs are projected in Figure 1. The initial check with W3C link checker brings out the exact HTTP error message associated with missing URLs. Results has shown that the 'HTTP 404 -page not found' was the overwhelming message encountered and represented 54.20 per cent of all HTTP message and it is followed by HTTP 500 (33.57 per cent) and HTTP 403 (9.35%).

Recovery of Missing URLs

The loss of URLs cited in LIS journals has an effect on the imminent scholarly work. We tried to recover 417 vanishing URLs using Internet Archive and Google search engine. Of 417 missing URLs, 66.19 percent of URLs (276 out of 417) have been recovered by Internet Archive and the rate of recovery through Google is unsatisfactory which recorded only 30.70 per cent of recovery of vanishing URLs (128 out of 417).

Table 5: HTTP Errors associated with Missing and Recovered URLs

HTTP Errors	Missing URLs	*Percentage*	Recovery of vanishing URLs by Internet Archive	% of recovery to the total vanishing URLs by IA	Recovery of vanishing URLs by Google	% of recovery to the total vanishing URLs by Google
301	1	0.24	1	100.00	1	100.00
400	5	1.20	4	80.00	5	100.00
401	1	0.24	0	0.00	0	0.00
403	39	9.35	25	64.10	23	58.97
404	226	54.20	154	68.14	41	18.14
406	1	0.24	1	100.00	0	0.00
500	140	33.57	87	62.14	58	41.43
503	2	0.48	2	100.00	0	0.00
504	2	0.48	2	100.00	0	0.00
Total	417	100.00	276	66.19	128	30.70

Table 5 indicates the distribution of vanishing URLs based on the HTTP errors. HTTP 404 – "Page not found" is the utmost (54.20 per cent) HTTP error associated with missing URLs. The results of this study can be comparable with the studies by Spinellis (2003), Goh and Ng (2007), Sadat-Moosavi, Isfandyari-Moghaddam, and Tajeddini (2012) and Sampath Kumar and Manoj Kumar (2012), Saberi and Abedi (2012) and Sampath Kumar and Vinay Kumar (2013). Sadat-Moosavi, Isfandyari-Moghaddam, and Tajeddini (2012) stated that HTTP 404 occurs due to several reasons such as filtering, connection failure status, explorer malfunctioning due to the used proxies, and intrusion of alphanumeric characters. HTTP errors 500 – "Internal Server error" (33.57 per cent), 403 – "Forbidden" (9.37 per cent), and 400 – "Bad request" (1.20 per cent) are the most occurred HTTP errors next to HTTP 404.

Internet Archive achieved 100 per cent recovery of vanishing URLs associated with HTTP 301, HTTP 406, HTTP 503 and HTTP 504, followed by HTTP 400 (80 per cent), HTTP 404 (68.14 per cent) and HTTP 403 (64.10 per cent). On the other hand, Google has recovered 100 per cent of vanishing URLs with HTTP 301 and HTTP 400 followed by HTTP 403 (58.97 per cent), HTTP 500 (41.43 per cent), and HTTP 404 (18.14 per cent).

Among the 1094 URLs identified, .co/.com domain consisted of the largest percentage (27.24 per cent), followed by .edu (23.31 per cent) and .org (20.20 per cent). It is a notable finding that the Geo domains have scored 8.50 per cent of the total cited URLs (Table 6).

We also examined the rate of vanishing URLs having top level domain (Table 6). A high-level proportion of vanishing URLs have been pertaining to URLs with number (100 per cent) followed by international domains (70 per cent), academic domains (59.42 per cent) and organization domain (52.04 per cent). URLs with educational domains are found to be most stable. The URLs associated with academic domains and Government domains (38.98 per cent loss or URLs) tend to publish reliable and valuable information. But their credibility and consistency over time can be questioned due to their loss has scored 59.42 percent and 38.98 percent respectively.

Table 6: Top Level Domains associated with Missing and Recovered URLs

Domains	Total URL citations	% of total URLs	Missing URLs	Percent	Retrieved by Internet Archive	Percent	Retrieved by Google	Percent
.co/.com	298	27.24	87	29.19	62	71.26	32	36.78
.org	221	20.20	115	52.04	82	71.30	24	20.87
.edu	255	23.31	57	22.35	35	61.40	18	31.58
.ac	69	6.31	41	59.42	24	58.54	12	29.27
.gov	59	5.39	23	38.98	17	73.91	10	43.48
.nic	17	1.55	7	41.18	5	71.43	2	28.57
.ernet	8	0.73	4	50.00	3	75.00	1	25.00
.net	34	3.11	13	38.24	9	69.23	5	38.46
.res	11	1.01	3	27.27	2	66.67	2	66.67
.int	10	0.91	7	70.00	4	57.14	3	42.86
Numbers	5	0.46	5	100.00	0	0.00	2	40.00
.info	14	1.28	7	50.00	5	71.43	1	14.29
Geo domains	93	8.50	48	51.61	28	58.33	16	33.33
Total	**1094**	**100.00**	**417**	**38.12**	**276**	**66.19**	**128**	**30.70**

The top-level domain having the greatest number of recovered vanishing URLs through Internet Archive was the .ernet domain (75 per cent) and .gov (73.91 per cent) followed by .org (71.30 per cent) and .com (71.26 per cent). The recovery has achieved by IA towards URLs with academic domains is 58.54 per cent, whereas, it is only 29.27 per cent by Google. Google has not achieved more than 50 per cent of recovery for URLs any domains except .res (66.67 per cent). URLs with domains such as .edu, .ac, .org and .com those contain qualitative and reliable information seem to be less retrieved by Google. This may elevates the question of Google's efficiency in comparison with Internet Archive in retrieving missing URLs.

Table 7: File Types associated with Missing and Recovered URLs

File types	Total URLs	% to total URLs	Missing URLs	% of missing URLs	Retrieved by Internet Archive	% of recovery by IA	Retrieved by Google	% of recovery by Google
Htm/ Html/ Shtml	741	21.94	240	32.39	169	70.42	75	31.25
PDF	241	11.97	131	54.36	74	56.49	38	29.01
ASP	38	1.65	18	47.37	16	88.89	5	27.78
DOC	4	0.37	4	100.00	2	50.00	2	50.00
PPT	4	0.18	2	50.00	2	100.00	1	50.00
JSP	4	0.09	1	25.00	0	0.00	0	0.00
CFM	14	0.55	6	42.86	6	100.00	3	50.00
PHP	48	1.37	15	31.25	7	46.67	4	26.67
Total	1094	38.12	417	38.12	276	66.19	128	30.70

Total, missing and retrieval URLs by their file formats are projected in Table 7. Out of 1094 URLs, eight file formats were recognized. The URLs which end with only the slash (/) and

none of the file format was mentioned has been considered as HTML file. Some 741 out of 1094 URLs were HTM/HTML/SHTML files followed by PDF (241) and PHP (48).This finding matches with the previous findings of McCown et al. (2005), Maharana, Nayak, and Sahu (2006), Moghaddam and Saberi (2010), Saberi and Abedi (2012), Sampath Kumar and Vinay Kumar (2013).

URLs with JSP file type (25%), PHP (31.25%) HTML/HTM/SHTML (32.39%) are the least missing compared to other file formats. DOC (100%), PDF (54.36%) and ASP (47.37%) are found to be the most unstable file types.

The greatest numbers of recovery of vanishing URLs was achieved by Wayback for URLs with different file types. The recovery for PPT and CFM file formats is 100 percent in Internet Archive which is only 50 percent in Google. Both Internet Archive and Google are not able to recover URLs with JSP file format. Google has shown a very low recovery rate for PHP (26.67 per cent), ASP (27.78 per cent), and PDF file formats (29.01 per cent) compared to Internet Archive. Internet Archive could able to recover 88.89 percent of ASP files, 56.49 per cent of PDF files and 46.67 percent of PHP files which is higher than Google.

Table 8 depicts the year-wise distribution of recovery of vanished URLs through Internet Archive and Google. The recovery of vanishing URLs has been achieved by Google which is counted 30.70 per cent which is very lower than the recovery achieved by Internet Archive (66.19 per cent). The recovery through IA to vanishing URLs cited in different years is varied from a low of 52.73 per cent (2013) and a highest of 80.70 per cent (2012). Google search engine fails to recover more than 40 percent of vanishing URLs for any year. Figure 2 indicates that the recovery of vanishing URLs via Internet Archive (IA) is highest for all years than Google search engine.

Table 8: Year-wise Recovery of Vanished URLs through Internet Archive and Google

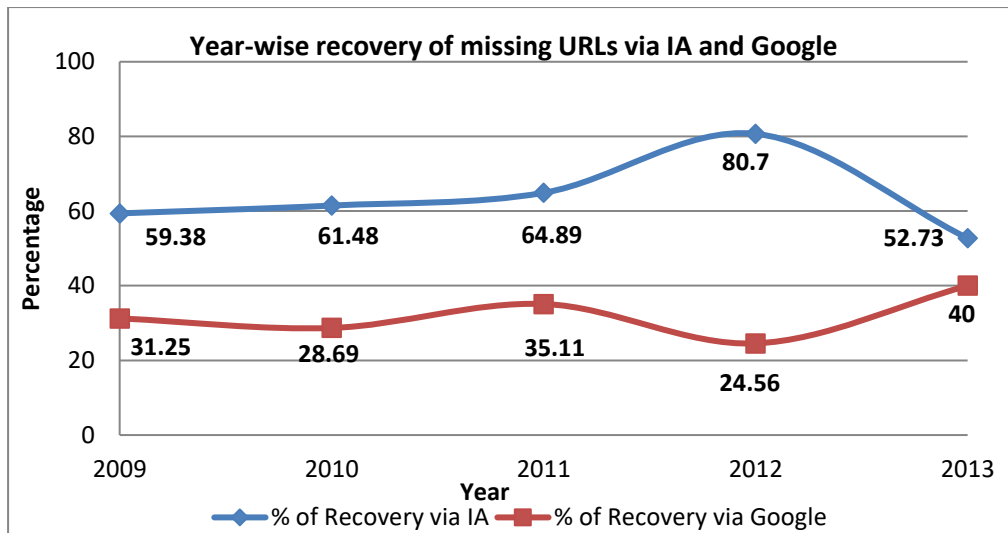| Year	Total URLs	Missing URLs	Retrieved through Internet Archive	% of Recovery through Internet Archive	Retrieved through Google	% of Recovery through Google
2009	80	32	19	59.38	10	31.25
2010	267	122	75	61.48	35	28.69
2011	257	94	61	64.89	33	35.11
2012	305	114	92	80.70	28	24.56
2013	185	55	29	52.73	22	40.00
Total	1094	417	276	66.19	128	30.70

Figure 2: Year-wise Recovery of Vanished URLs via Internet Archive and Google

An attempt has been made to recover vanishing URLs through Internet Archive and the Google which increased the rate of active URLs. The summary of active and vanishing URLs after the recovery process is presented in Table 9. The recovery through Internet Archive has decreased the decay of URLs from 38.12 percent to 12.89 percent and increased the active URLs rate from 61.88 percent to 87.11. On the other hand, Google search engine decreased the missing URL's rate to 26.42 percent and increased the active URLs rate to 73.58 percent (805) from 61.88 percent (677). Both Internet Archive and Google left 12.89 percent and 26.42 percent of URLs respectively to permanent inaccessibility.

Table 9: Year-wise Summary of Active and Vanished URLs after the Recovery from Internet Archive and Google

Year	Total URLs	Active URLs	Missing URLs	Active URLs after recovery from		Decay of URLs after recovery from	
				IA	Google	IA	Google
2009	80	48	32	67	58	13	22
2010	267	145	122	220	180	47	87
2011	257	163	94	224	196	33	61
2012	305	191	114	283	219	22	86
2013	185	130	55	159	152	26	33
Total	**1094**	**677**	**417**	**953**	**805**	**141**	**289**
Percentage (N=1094)		**61.88**	**38.12**	**87.11**	**73.58**	**12.89**	**26.42**

DISCUSSION AND CONCLUSION

In the present study, we found that URL references are found to have an increasing trend among LIS scholars. The *percentage* of URLs increased by approximately four fold during 2009 (9.57%) to 2013 (37.70%). The increasing trend of URLs has been documented by

Rumsey (2002), Sellitto (2004), Bar-Ilan and Peritz (2004), Wu (2009), Sampath Kumar and Prithviraj (2012), Jalalifard, Norouzi and Isfandyari-Moghaddam (2013). In the present study also it was found that the average URLs per article are 4.84, which indicate that each paper in *Webology* and ALIS has referred at least four web resources. This trend clearly indicates that the LIS researchers are highly depending on web resources for scholarly information.

Other significant findings of this study are that about 27.24 percent of URLs cited are associated with commercial domains. The widespread availability of web resources of commercial domains is a reason for their highest citation rate in LIS journals. The trend to cite web resources belonging to organizational (20.20%), academic (6.31%), Governmental (5.39%) domains is perhaps, found to be very less. It is also an interesting finding is that the rate of URL loss for educational domains is found to be less compared to .ac, .org and .com domains. Based on the findings of the study it is recommended that the LIS authors should prefer to cite web resources from educational domains.

The study has been confined to compare the efficiency of Internet Archive and Google in retrieving the missing URLs. We found that Internet Archive recovered 66.19 percent of missing URLs, whereas, Google could able to recover only 30.70 per cent. Efficiency in relation to the recovery of missing URLs through Internet Archive and Google was tested. Study found that Internet Archive is the efficient tool which had increased the accessibility rate of cited URLs from 61.88 percent to 87.11 percent. Google managed to increase the percentage of active URLs to 73.58 percent which is shorter by 13.53 percent to the rate achieved by Internet Archive. We found that Internet Archive recovered more vanished URLs than Google search engine. The Internet Archive is found to be the efficient tool compared to Google in recovering vanished URLs.

The vanished URLs could be dodging at the time of publication of e-content. The role of publishers is imperative and they can take responsibility to give careful consideration towards self-archive mechanism of the digital content which they publish (Anderson, Tarczy-Hornoch and Bumgarner 2006). The publishers, authors, bloggers, web masters may submit a copy of e-resource, which they possess, to Internet Archive after removing the hurdles of copyright issues. Further, it would be noteworthy step for all countries to have an international policy for Internet resource preservation in the light of copyright of scholarly work. Further, the institutional repositories of educational institutions, research institutions, government departments and commercial organizations which are connected to the Web can provide permanent access to digital scholarly work. However, the use of Internet Archive in conjunction with these institutional repositories along with national-based web archiving initiatives, and search engines could be radical measure to combat of web decay.

ACKNOWLEGEMENT

REFERENCES

Anderson, N.R., Tarczy-Hornoch, P. and Bumgarner, R.E. 2006. On the persistence of supplementary resources in biomedical publications. *BMC Bioinformatics*, Vol.7, no.1: 260-267, Avai;able at: www.biomedcentral.com/content/pdf/1471-2105-7-260.pdf.

Bar-Ilan, J., and Peritz, B. C. 2004. Evaluation, continuity and disappearance of documents on a specific topic on the web: A longitudinal study of informetrics. *Journal of the American society for Information Science and Technology*, Vol.55, no.11: 980–990.

Casserly, M. and Bird, J. 2003. Web citation availability: analysis and implications for scholarship. *College and Research Libraries*, Vol.64, no.4: 300-17.

Dimitrova, D.V. and Bugeja, M. 2007. The half-life of internet references cited in communication journals. *New Media and Society*, Vol.9, no.59: 811-826.

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.

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.

Gul, S., Mahaja, I. and Ali, A. 2014. The Growth and decay of URLs citation: A case of an online Library and Information Science Journal. *Malaysian Journal of Library and Information Science*, Vol.19, no.3: 27-39.

Jalalifard, M., Norouzi, Y. and Isfandyari-Moghaddam, A. 2013. Analyzing web citations availability and half-life in medical journals: A case study in an Iranian university. *Aslib Proceedings: New Information Perspectives*, Vol.65, no.3: 242-261. DOI. 10.1108/00012531311330638

Lawrence, S., and Giles, C.L. 1999. Accessibility of information on the web. *Nature*, Vol.400, no.6740: 107–109.

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.

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

Moghaddam, A.I. and Saberi, M.K. 2010. Availability and half-life of web references cited in information research journal: A citation study. *International Journals of Information Science Management*, Vol.8, no.2: 57–75.

Nagaraja, A., Joseph, S.A., Polen, H.H. and Clauson, K.A. 2011. Disappearing act. *Program*, Vol.45, no.1: 98-106. Available at: http://dx.doi.org/10.1108/00330331111107420.

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: 27–39.

Saberi, M.K. and Abedi, H. 2012. Accessibility and decay of URLs in five open access ISI journals. Internet Research, Vol.22, no.2: 234-247. Available at: http://dx.doi.org/10.1108/10662241211214584.

Sadat-Moosavi, A., Isfandyari-Moghaddam, A. and Tajeddini, O. 2012. Accessibility of online resources cited in scholarly LIS journals: A study of Emerald ISI-ranked journals. *Aslib Proceedings*, Vol.64, no.2: 178-192.

Sampath Kumar, B.T. and Manoj Kumar, K.S. 2012. Persistence and half-life of URLs cited in LIS open access journals. *Aslib Proceedin*gs, Vol.64, no.4: 405 – 422.

Sampath Kumar, B.T. and Prithvi Raj, K.R. 2012. Availability and persistence of URLs in Indian LIS literature. *The Electronic Library*, Vol.30, no.1: 19-32.

Sampath Kumar, B.T. and Vinay Kumar, D. 2013. HTTP 404-page not found: Recovery of decayed URLs. *Journal of Informetrics*, Vol.7, no.1: 145-157. Retrieved from, http://www.sciencedirect.com/science/article/pii/S175115771200082X

Sellitto, C. 2004. A study of missing web-cites in scholarly articles: towards an evaluation framework. *Journal of Information Science*, Vol.30, no.6: 484-95.

Sellitto, C. 2005. The impact of impermanent web-located citations: a study of 123 scholarly conference publications. *Journal of the American Society for Information Science and Technology*, Vol. 56, no.7: 695-703.

Spinellis, D. 2003. The decay and failures of web references. *Communications of the ACM,* Vol.46, no.1: 71-78. Available at: http://onlinelibrary.wiley.com/doi/10.1002/asi.20159/full

Tajeddini, O., Azimi, A., Sadatmoosavi, A., and Moghaddam, H. S. 2011. Death of web citations: A serious alarm for authors. *Malaysian Journal of Library and Information Science*, Vol.16, no.3: 17–29.

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 Medical Library Association*, Vol.97, no.2: 122–130. Available at. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2670212/pdf/mlab-97-02-122.pdf

Wang, P., Hawk, W.B., and Tenopir, C. 2000. Users' interaction with World Wide Web resources: An exploratory study using a holistic approach. *Information processing and management*, Vol.36, no.2: 229-251.

Wu, Z. 2009. An empirical study of the accessibility of web references in two Chinese academic journals. *Scientometrics*, Vol.78, no.3: 481-503.

Yang, S., Qiu, J. and Xiong, Z. 2010. An empirical study on the utilization of web academic resources in humanities and social sciences based on URLs. *Scientometrics,* Vol.84, No.1: 1-19.