Investigation of the Currency, Disappearance and Half-Life of Urls of Web Resources Cited In Iranian Researchers: A Comparative Study

Oranus Tajedini
Assistant prof.,
Department of Knowledge and Information Science,
Shahid Bahonar University of Kerman, Iran.
Tajedini.o@gmail.com

Ali Sadatmoosavi
Assistant prof., Leishmaniosis Research Center,
Kerman University of Medical Sciences, Iran,
Corresponding Author, Moosavi56@gmail.com

Azita Ghazizade
PhD candidate in translation Studies,
Faculty of Literature and Foreign Languages, Allameh
Tabatabai University, Iran.
Azita.ghazizade@atu.ac.ir

Atefe Tajedini
Iran Public Libraries Foundation. Kerman Branch
at.tajedini@gmail.com

Abstract

This research was intended to comparatively investigate the currency, disappearance and half-life of URLs of web resources cited in Iranian researchers' articles indexed in ISI in information science, psychology and management from 2009 to 2011. The research method was citation analysis. The statistical population of this research was all articles by Iranian researchers in psychology, information science and management from 2009 to 2011 which were indexed in SSCI. In order to extract bibliographic information of articles, ISI database was searched and the titles of the articles were extracted. After investigating the currency and disappearance of cited URLs and calculating the half-life of web resources, collected data were analyzed in accordance with research questions by means of Excel Software.The results of this research revealed that in articles written by Iranian researchers indexed in ISI in information science, psychology and management there were 6152, 3639 and 8926 citations, respectively, of which 13.7, 44.8 and 14.23 percent were online citations, respectively. The most frequently used domain in all three fields was .org. The most stable and persistent domain in psychology was .com, in information science was .org and in management was for those domains other than the mentioned domains. The most frequent file format was pdf in all three fields. In information science, pdf. Files were the most stable while in management, rtf files and in psychology, ppt files were the most stable ones, respectively. In the initial search for online citations in psychology, information science and management, respectively, 58, 82 and 88 percent of citations were accessible which were even increased after second check with due measurements to 95, 98 and 97 percent, respectively. The research results also demonstrated that most accessible internet addresses in investigated articles of all three fields were found in the cited internet address. The status of inaccessible internet addresses in all investigated articles regarding error messages also indicated that in psychology and management 404 error message (Not found) was the most frequent error with 34 and 22 percent, respectively and in information science, 403 error message (forbidden) was the most frequent error message with 21 percent. The average half-life of online citations calculated in all investigated articles was 2.6 years which was calculated as 3 years and 4 months in information science, 2 years and 5 months in management and 1 year and 9 months in psychology. The results of this research showed that decay of internet addresses

should be regarded as a problem the most important reason of which is website reorganization and changes made to the names of internet domains. Some fields are more exposed to and affected by the consequences of decay of internet addresses. The influence of inactive links on the journals of a field is different based on the reliance of authors on internet based information. The absolute number of internet addresses also strengthens the problem of decay of internet addresses for the readers of the articles as compared with those journals whose authors have only cited a few online citations. The consequences of inactive links for those articles and resources which can be accessed through different ways or their print version is accessible are less serious. Tools like internet archives might make it possible to have a snapshot of the content of a site in a particular time. Google doesn't index dynamic pages or pages and sites which use robots.txt coding to prevent crawling. The best solution to improve the accessibility of internet resources is to request for all internet information be analyzed and recorded while examining the manuscripts. In so doing, the responsibility to archive information will be assigned to the publisher.

Keywords: Web Citation Availability, URL Persistence, Internet Archive, Half-life of Web References.

Introduction

It's been a long time that citations are considered as an important part of valuable scientific articles and are of prominent importance in scientific researches. Citations show authors' acknowledgment of previous works (Kochen, 1987), prove the validity and are witnesses confirming the research (Metcalfe, 2003); in fact, references determine the correctness of researcher's interpretations (Ticehurst & Veal, 2000) and connect them to a much wider range of other related works (Germain, 2000), they help document progression of knowledge (Grafton, 1999) and measure the influence of a scientific research (Garfield, 1979). Using internet and obtaining information from databases all over the world have made dissemination of information and knowledge faster and easier. Nowadays, researchers are able to search and retrieve their required information by means of computer technologies and information databases and can use such information in conducting studies in various scientific fields.

With the increased application and use of internet, researchers' citation behavior is also influenced and inclined towards citing web resources as the newest information sources, thus, the number of citations to URLs of web resources, is increased. The reason for such growth in the number of citations to URLs of web resources is that every day valuable information in digital format is added on the internet and the access to scientific databases, analytical-statistical data, new techniques and even projects under development and works and researches to which there is no other access and might not be available in print is possible via world wide web (Maharana, Nayak, & Sahu, 2006).

While internet has made data exploration easier and faster for researchers, it has also brought about problems like disappearance of information. Recent studies have revealed that the disappearance of URLs is not only a serious problem for web managers but also for academic researchers resorting to web citations in their researches. It was proved that usually between 10 to 40 percent of web citations disappear (Sadatmoosavi, Tajedini, & Moghadam, 2010). The results suggested that usually between 10 to 40 percent of citations to URLs disappear (Markwell & Brooks, 2003; Rumsey, 2002; Sellitto, 2004; Tyler & McNeil, 2003).

If citations and their links are not accessible, online citations of scientific productions become less valid. There are a number of reasons for disappearance of resources. Some websites get simply inaccessible since the website domain name has expired. Some other web pages are removed by their creators. Some web sites are redesigned with new structures and some others are moved to a new location, thus, displaying a redirect message or directly leading the browser to a new internet address which sometimes enjoys a different updated content. Therefore, server failure and connection problems will rise up. Irrespective of the reasons for such failures, most internet users have been presented with the famous error "404: Not Found", the phrase which is displayed in half a million or more entries in a search engine per day (Dimitrova & Bugeja, 2007a).

Today, since a great number of journals are indexed in ISI and have a high impact factor, authors are more interested in these journals having turned into valuable resources for scholars and researchers. The articles published in these journals usually contain the newest and the most reliable scholarly information. These articles which are mostly contributed by prominent researchers and scholars in different fields, are based on citations to previous scientific works. An investigation of journals indexed in ISI reveals that, due to the emergence of internet and web, resources published on internet are also resorted to for carrying out the research. This indicates the development of web resources as valid and documentary sources in academic and scientific fields (Tajeddini, Azimi, Sadatmoosavi, & Sharif-Moghaddam, 2011). Thus, according to the point that the reference list of an article is one of its most important components and citations to URLs (web resources) are in danger of decay and disappearance due to their changing and unstable nature, the researchers attempted to investigate the currency and disappearance of web resources cited in Iranian articles indexed in ISI in information science, psychology and management fields. Hence, in this research, the disappearance rate of cited URLs, the most cited domain, the most stable file formats, the average citations to URLs of web resources for each article, the process of citing web resources and half-life of URLs cited in Iranian researchers' articles indexed in ISI in information science, psychology and management fields were determined and some recommendations were provided to reduce the decay of online resources.

In his MA thesis entitled "the investigation of web citations of articles published in open access journals in social sciences", Saberi and Abedi (2012) investigated web citations. The research results revealed that from among all analyzed journals, 62 percent of articles had web citations and the average citation to web pages for each article was 5. The investigation of internet addresses regarding domain and file format also demonstrated that .uk, .net and .edu were more enduring and stable than .org, .gov and .com domains. Moreover, pdf files were the most stable ones. The analysis of accessibility of internet addresses also indicated that 67 percent of web citations were accessible from the cited internet addresses.

Dellavalle et al (2003) in a study entitled "Going, going, gone: Lost Internet references" investigated internet references and citations to URLs of web resources (accessible) in medical and scientific periodical from 2000 to 2003. To conduct this research, more than 1000 articles published in "New England Journal of Medical", "The Journal of the American Medical Association" and "Science" were selected and analyzed. The research results revealed that from among 25548 references, 672 of them were citations to URLs of web resources and 2.6 percent of all references were internet references. Investigating the stability and maintenance of internet addresses indicated that 87 percent of internet references were

stable and active and 13 percent of them yielded an error message and were inaccessible.

McCown, Chan, Nelson, and Bollen (2005b) investigated citations to URLs in a study titled "The availability and persistence of web references in D-Lib magazine" from July 1955 to August 2004. In this research, 4387 web citations were extracted from 453 articles and their accessibility was checked three times a week for 25 weeks from September 2004 to February 2005. The research results demonstrated that 28% of the URLs failed to resolve initially, and 30% failed to resolve at the last check. A majority of the unresolved URLs were due to 404 (page not found) and 500 (internal server error) errors. The half-life of a URL referenced in a D-Lib Magazine article was 10 years proving the web references in D-Lib Magazine as persistent. Moreover, 99 percent of URLs used the "http" scheme while 1 percent relied on schemes like "gopher" and "ftp". The results also showed that html and slash files (URLs that end with a slash as in http://foo.edu/) were the most persistent.

Dimitrova and Bugeja (2007a) conducted a research entitled "The half-life of internet references cited in communication journals" and examined the URL reference addresses in journalism and communication journals. 1126 online citations were selected from the articles of 5 leading journals between 2000 and 2003. The research results demonstrated that only 61 percent of URL citations remained accessible in 2004 and 39 percent were inaccessible. The results also showed that the average half-life for communication journals internet references was 3.17. In other words, it would take about three years for half of the internet citations to disappear. Moreover, .org and .gov were the most stable domains with 70 percent and 67 percent active links, respectively. Furthermore, the research results indicated that 86 percent of links didn't contain an error in the URL while 14 percent of citations were not hyperlinked correctly.

Wagner et al (2009) investigated the accessibility and decay of URL citations in healthcare management journals from 2002 to 2004. To accomplish this, they extracted 2011 unique internet addresses from 5 leading journals and examined them. The analysis of accessibility and decay of internet addresses revealed that 50.7 percent of internet addresses were accessible while 49.3 percent of internet addresses contained an error message and were inaccessible. The results also indicated that .edu and .net were the most stable domains with 68.4 and 61.5 percent accessibility, respectively.

Research questions

1. How were the articles and citations of ISI articles by Iranian researchers in information science, psychology and management distributed?

2. To what extent was the currency of URLs cited in ISI articles by Iranian researchers in information science, psychology and management?

3. What was the status of URLs cited in ISI articles by Iranian researchers in information science, psychology and management regarding the domain?

4. What was the status of URLs cited in ISI articles by Iranian researchers in information science, psychology and management regarding the file format?

5. To what extent was the disappearance rate of URLs cited in ISI articles by Iranian researchers in information science, psychology and management in the initial check?

6. To what extent was the disappearance rate of URLs cited in ISI articles by Iranian researchers in information science, psychology and management in the second check?

7. What was the status of accessible URLs cited in ISI articles by Iranian researchers in

information science, psychology and management?

8. What was the status of error messages in disappeared URLs in ISI articles by Iranian researchers in information science, psychology and management?

9. What was the status of half-life of URLs cited in ISI articles by Iranian researchers in information science, psychology and management?

Methods

This applied research was of descriptive and quantitative measurement type. The research method was citation analysis. Citation analysis has to do with mathematical formulas, calculations and statistical analyses and it's most significant advantage is that it can be practically and objectively carried out away from the influence of subjective ideas and biased directions on research results which is also the characteristic of some other research types like questionnaire method. Citations are interesting topics to undergo studies since they are easily accessible and don't undermine the credibility of data. Unlike data obtained from interview and questionnaire, citations are not intrusive criteria and don't require the cooperation of respondents. They don't mislead the response i.e. don't show any reactions (Asareh, 1999) The statistical population of this research was all articles by Iranian researchers in psychology (148 articles), information science (121 articles) and management (285) in ISI database from 2009 to 2011 which were indexed in SSCI (Social Science Citation Index). All these articles from the beginning of 2009 to the end of 2011 were investigated and their resources were recorded. The reason for selecting and investigating articles in 2009-2011 was the existence of a time gap to correctly calculate the currency, disappearance and half-life of the articles.

In this research, data were collected from the internet. First, to extract bibliographic information of articles, ISI database was searched on August the 22nd, 2016 as follows and the titles of the articles were extracted. In order to do the research, an advanced search was performed and the following formula was made use of:

(CU= Iran) *AND* **DOCUMENT TYPES:** (Article)
Refined by: RESEARCH AREAS: (PSYCHOLOGY OR OPERATIONS RESEARCH MANAGEMENT SCIENCE OR INFORMATION SCIENCE LIBRARY SCIENCE)
Timespan: 2009-2011. **Indexes:** SSCI.

The search was restricted to Social Sciences Citation Index (SSCI) -1983-present. Then, the Refine option was selected to restrict the results by topic into three fields of information science and library science, management and psychology. The results of each topic were separately saved as Tap-delimited and Save to Endnote.

Thereafter, all articles whose information was extracted from ISI database in the previous stage, were downloaded from the archive of the related journals and their bibliographic information, total number of citations and total number of online citations of each article were separately written down in notes. Therefore, editorial notes, reports, book introductions and criticisms were excluded and only research articles and review articles including a reference list –since they would be examined by the scholars and experts and were expected to entail a complete list of references – were investigated. Then, Firefox Browser 38.0 through which it was possible to have access to internet addresses cited in reference list of articles, was utilized to analyze the currency and disappearance of URLs cited in the articles. At first, it was

attempted to avoid typographical, copy and paste errors by directly clicking on the internet addresses cited in the online version of each article. The URLs were divided into two groups in this stage as accessible (without any error messages when accessed by an internet browser) and inaccessible (showing one error message when accessed by an internet browser). Those URLs which were redirected were classified as accessible while those websites which were inaccessible due to lack of password were classified as inaccessible. When errors were encountered in a direct access to online citations, attempts were made to determine if the intended content was still available on internet or not. First, the cited URLs were entered into the internet browser. Errors which occurred due to URL formatting were corrected using a set of heuristic rules as well as manual editing. Improper spaces in URLs were not considered as errors. Heuristic correction was required since a great number of errors were in one class of several available classes, for example, improper spaces in URLs, using backward slash (\\) instead of forward slash (//), non-alphanumeric characters (usually from non-English web sites) and incorrect characters and signs like http://(/), http://www+++, or http@. Although these errors were non-standard, they worked when typed in a browser, thus, were not considered as errors. If the URL didn't respond after a period of 60 seconds or an error message was displayed showing that it was not possible to restore the page (e.g. "404 not found", "page was unavailable", "file not found" and etc.) the URL was considered as inaccessible. If server error was faced – this error is displayed due to temporary failure of server – web pages were visited two times and if the cited content was not found, then the content was retrieved by means of an internet archive. To do this, the most famous online archive i.e. Wayback Machine was employed. In the homepage of this archive[1], the exact copies of missed URLs were inserted in search section and "Take me back" button was clicked to search for the online archive. If URL was found by Wayback Machine, "found in internet archive" section was marked in checklist. If it was not even found in the archive, it was attempted in a 15-minute period to find the missed information via search engine. The online search engine chosen for this research was Google. Its long-lasting popularity was the main reason for its preference over other more advanced search options. Another reason was the great number and rich variety of documents to which Google provides access. The other reason for selecting this search engine was that a lot of researchers believe that ranking based on connection to Google is more effective(Calishain & Dornfest, 2005). Therefore, up to 5 times, different phrases like titles, keywords, authors' names and the information relevant to cited resources were searched in Google. If the cited resource was not found in the first 20 retrieved results, that resource was regarded as inaccessible. Then, the currency or disappearance of URLs were written down in separate notes. In the second stage, URLs were recorded in separate note cards based on domain and file format. After taking notes and transferring information, collected data were analyzed in accordance with research questions by means of Excel Software and required tables and figures were created.

In order to determine the half-life of URLs of web resources, half-life formula applied in foreign researchers (Dimitrova & Bugeja, 2007a; Koehler, 1999; Tyler & McNeil, 2003)was utilized so as to be also used by Iranian researchers and librarians and provide ground for calculating the half-life of URLs used in online resources in different journals.

Half-life of URLs is defined as the time required for half of the URLs cited in a journal to be missed and decayed. This time, of course, may be different in different fields and different years. The half-life of URLs is different from the terms cited half-life (the half-life of citations

to a journal) and citing half-life (the half-life of citations in a journal) used in JCR. Cited half-life is a criterion to compare the life span of received articles and citing half-life is a criterion to compare life span of reference articles. Cited half-life determines the time when half of citations are calculated and half of citations are out of calculation (Tyler & McNeil, 2003). The following formula is used to calculate half-life of URLs cited in journal articles:

$$w(t) = w(0)e^{at}$$

In this formula, "ln" is logarithm to the base "e" (Neper is 2.7). "W (0)" is the number of online URLs at work during the distribution, "W (t)" is the number of online URLs at work after the time "t" and "a" is the rate of stability and persistence calculable from accessible URLs. Half-life of URLs (t_h) is calculated as follows:

$$t_h = [t \ln(0.5)]/[\ln w(t) - \ln w(0)]$$

Here, t_h is the number of years during which 50 percent of URLs stop performing. Based on the above formula, the half-life of URLs cited in psychology, information science and management fields were calculated.

Date acquired from this research were analyzed based on descriptive statistics including average and percent. The collected data – after investigating currency and disappearance rates of cited URLs and calculating half-life of online resources - were analyzed with accordance to research questions via Excel 2013 Software and the required tables and figures were also produced.

Research results

The first research question: How were the articles and citations of ISI articles by Iranian researchers in information science, psychology and management distributed?
Concerning psychology, as depicted in table 1, from among all journals indexed in ISI, 148 articles were written by Iranian researchers from 2009 to 2011 and these articles had a total number of 6152 citations.

Table 1
frequency distribution of articles and citations in psychology

Year	articles		citations	
	Frequency	Percent	Frequency	Percent
2009	38	26%	1464	24%
2010	56	38%	2198	36%
2011	54	36%	2490	40%
Total	148	100%	6152	100%

In information science, as shown in table 2, from among all journals indexed in ISI, 121 articles were written by Iranian researchers from 2009 to 2011 and these articles had a total number of 3639 citations.

Table 2
frequency distribution of articles and citations in information science

Year	Articles		Citations	
	Frequency	Percent	Frequency	Percent
2009	31	26%	1023	28%
2010	47	39%	1263	35%
2011	43	36%	1353	37%
Total	121	100%	3639	100%

Regarding management, as presented in table 3, from among all journals indexed in ISI, 285 articles were written by Iranian researchers from 2009 to 2011 and these articles had a total number of 8926 citations.

Table 3
frequency distribution of articles and citations in management

Year	Articles		Citations	
	Frequency	Percent	Frequency	Percent
2009	27	9%	804	9%
2010	53	19%	1467	16%
2011	205	72%	6655	75%
Total	285	100%	8926	100%

The second research question: To what extent was the currency of URLs cited in ISI articles by Iranian researchers in information science, psychology and management?

The research results suggested that in psychology, 742 of all studied articles had online citations and these citations were more in 2011 (40 percent) than in other years (table 4).

Table 4
frequency distribution of print citations and online citations in psychology

Year	Online citations		Print citations	
	Frequency	Percent	Frequency	Percent
2009	134	18%	1330	25%
2010	267	36%	1931	36%
2011	341	46%	2149	40%
Total	742	100%	5410	100%

According to the research results, in information science, 1127 of all studied articles had online citations. Data presented in table 5 also reveals that in information science the articles published in 2009 with 307 online citations (27 percent) and the articles published in 2011 with 452 online citations (40 percent) had the lowest and the highest number of online citations, respectively.

Table 5
frequency distribution of print citations and online citations in information science

Year	Online citations		Print citations	
	Frequency	Percen	Frequency	Percent
2009	307	27%	716	29%
2010	368	33%	895	35%
2011	452	40%	901	36%
Total	1127	100%	2512	100%

According to the research results, in management, 1112 of all studied articles had online citations. Data depicted in table 6 also shows that in management the articles published in 2010 with 94 online citations (8 percent) and the articles published in 2011 with 903 online citations (81 percent) had the lowest and the highest number of online citations, respectively.

Table 6
frequency distribution of print citations and online citations in management

Year	Online citations		Print citations	
	Frequency	Percent	Frequency	Percent
2009	115	10%	689	9%
2010	94	8%	1373	18%
2011	903	81%	5752	74%
Total	1112	100%	7814	100%

The third research question: What was the status of URLs cited in ISI articles by Iranian researchers in information science, psychology and management regarding the domain?

Figure 1 shows the distribution of cited internet addresses based on domains as well as the percent of inaccessible internet addresses in all investigated articles in psychology. The presented data indicate that internet addresses with .org domain had the highest number of citations i.e. from among 742 online citations, the domain of 156 citations was .org. However, the domain of 147 internet addresses was .com, the domain of 111 citations was .edu/.ac, the domain of 126 citations was .net and 47 citations had other domains. The percent of inaccessible internet addresses demonstrates that from among all investigated articles in psychology, .org and .com were more stable and persistent than .net, .edu and etc. (figure 1). As it is evident, .com/.co was the most frequent domain in accessible internet addresses. In other words, from 397 accessible internet addresses, the domain of 99 citations was .com/.co. Nevertheless, from 345 inaccessible internet addresses, citations with .edu/.ac domain were the most frequent (67 citations).
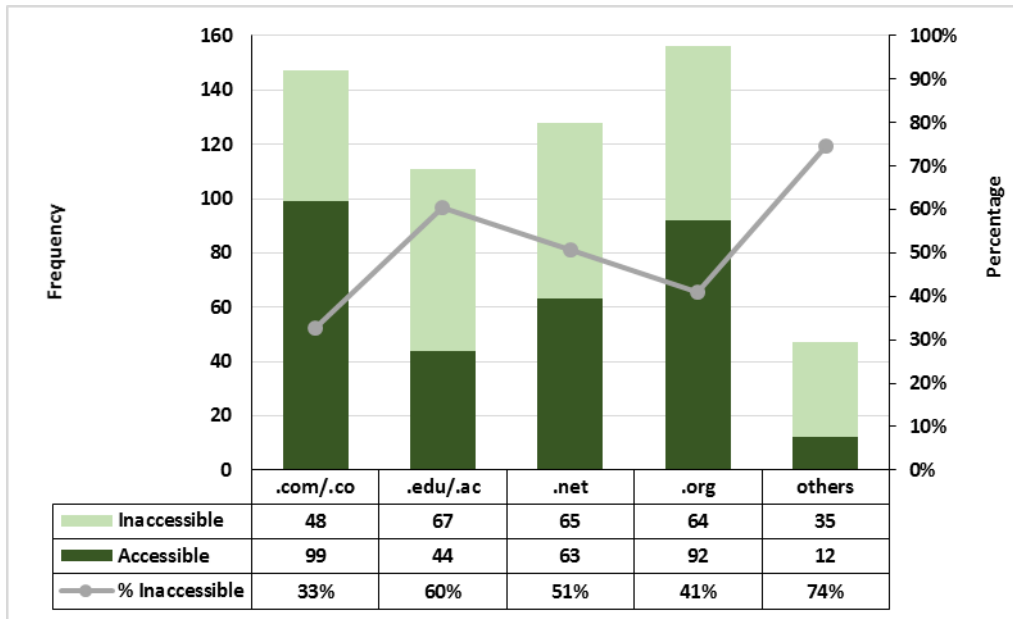
Figure 1: distribution of internet addresses based on domain in all investigated articles in psychology

Figure 2 shows the distribution of cited internet addresses based on domains as well as the percent of inaccessible internet addresses in all investigated articles in information science. The presented data indicate that internet addresses with .org domain had the highest number of citations i.e. from among 1127 online citations, the domain of 309 citations was .org. However, the domain of 279 internet addresses was .com, the domain of 135 citations was .edu/.ac, the domain of 193 citations was .net and 71 citations had other domains. The percent of inaccessible internet addresses demonstrates that from among all investigated articles in information science, .org and .com were more stable and persistent than .net, .edu and etc. In accessible internet addresses, .org was the most frequent domain. In other words, from 909 accessible internet addresses, the domain of 273 citations was .org Nevertheless, from 218 inaccessible internet addresses, citations with .net domain were the most frequent (67 citations).
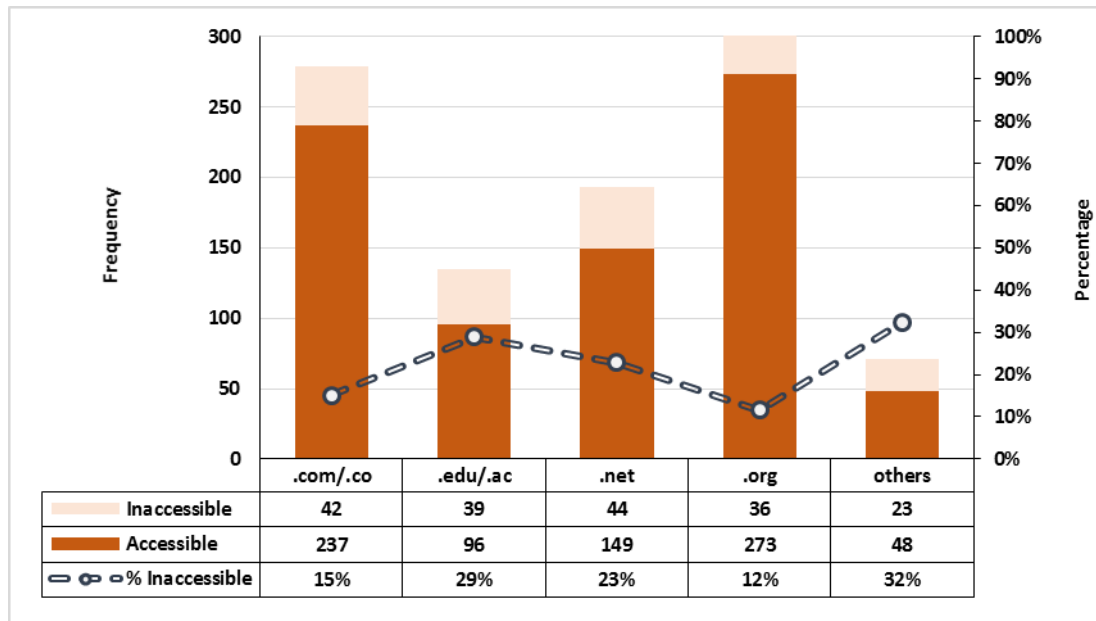
Figure 2: distribution of internet addresses based on domain in all investigated articles in information science

Figure 3 shows the distribution of cited internet addresses based on domains as well as the percent of inaccessible internet addresses in all investigated articles in management. The presented data indicate that internet addresses with .org domain had the highest number of citations i.e. from among 1112 online citations, the domain of 237 citations was .org. However, the domain of 227 internet addresses was .com, the domain of 161 citations was .edu/.ac, the domain of 160 citations was .net and 136 citations had other domains. The percent of inaccessible internet addresses demonstrates that from among all investigated articles in management, domains other than the ones mentioned were more stable and persistent than .net, .org, .com, .edu and etc. In accessible internet addresses, .org was the most frequent domain. In other words, from 934 accessible internet addresses, the domain of 199 citations was .org Nevertheless, from 178 inaccessible internet addresses, citations with .gov domain were the most frequent (45 citations).
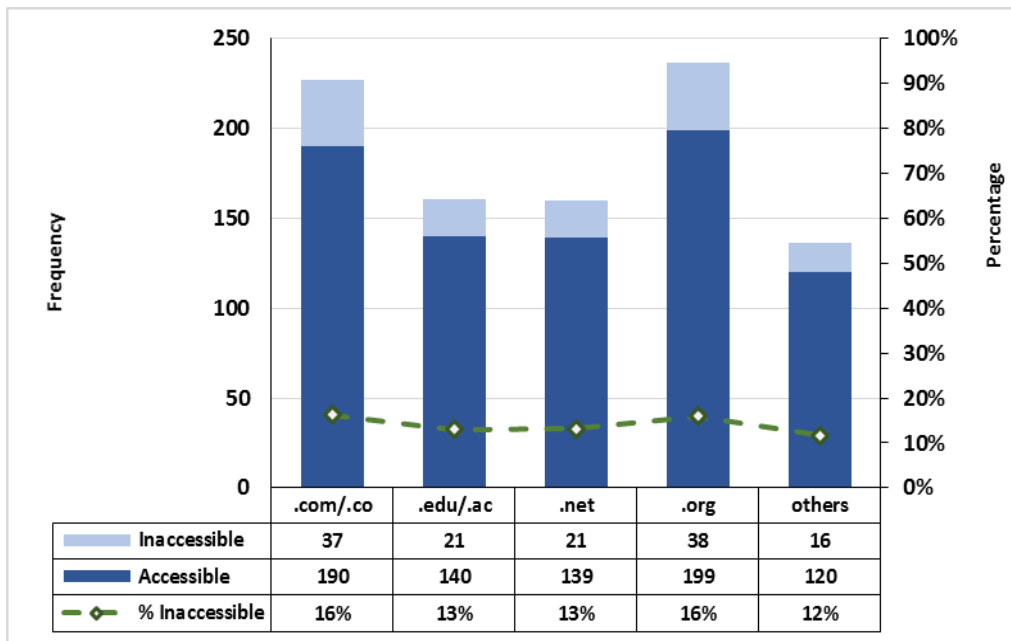
Figure 3: distribution of internet addresses based on domain in all investigated articles in management

The fourth research question: What was the status of URLs cited in ISI articles by Iranian researchers in information science, psychology and management regarding the file format?

Figure 4 indicates the distribution of cited internet addresses based on file format as well as the percent of inaccessible internet addresses in all investigated articles in psychology. Data provided in this figure reveal that internet addresses with pdf file format had the highest number of citations i.e. from among 742 online citations, 309 citations were pdf files and internet addresses cited as txt files had the lowest number of citations (4 online citations). The percent of inaccessible internet addresses in figure 4-3-4 shows that from among all investigated articles in psychology, pdf and html files were more stable and persistent than other files. In accessible internet addresses pdf files were the most frequent. In other words, from among 397 accessible online citations, 175 citations were pdf files. Furthermore, from among 345 inaccessible internet addresses, pdf files were the most frequent (134 citations).
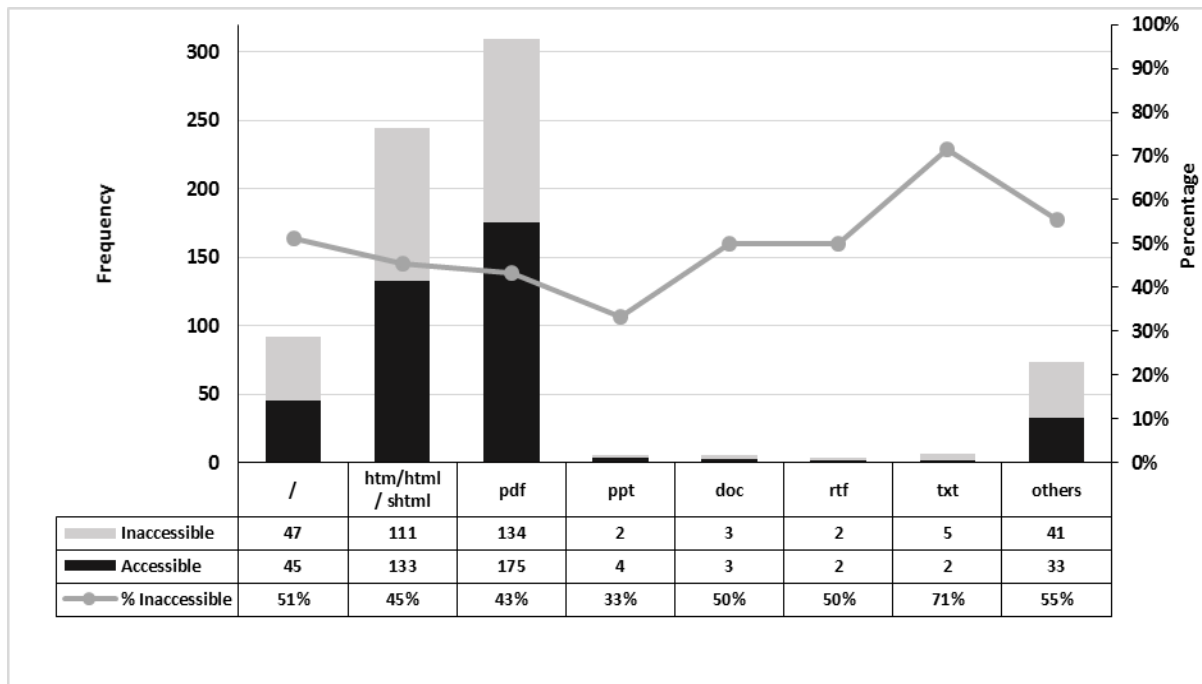
	/	htm/html / shtml	pdf	ppt	doc	rtf	txt	others
Inaccessible	47	111	134	2	3	2	5	41
Accessible	45	133	175	4	3	2	2	33
% Inaccessible	51%	45%	43%	33%	50%	50%	71%	55%

Figure 4: distribution of internet addresses based on file format in all investigated articles in psychology

Figure 5 indicates the distribution of cited internet addresses based on file format as well as the percent of inaccessible internet addresses in all investigated articles in information science. Data provided in this figure reveal that internet addresses with pdf file format had the highest number of citations i.e. from among 1127 online citations, 432 citations were pdf files and internet addresses cited as rtf files had the lowest number of citations (3 online citations). The percent of inaccessible internet addresses shows that from among all investigated articles in information science, pdf and html files were more stable and persistent than other files. In accessible internet addresses pdf files were the most frequent. In other words, from among 909 accessible online citations, 377 citations were pdf files. Furthermore, from among 218 inaccessible internet addresses, html files were the most frequent (64 online citations).
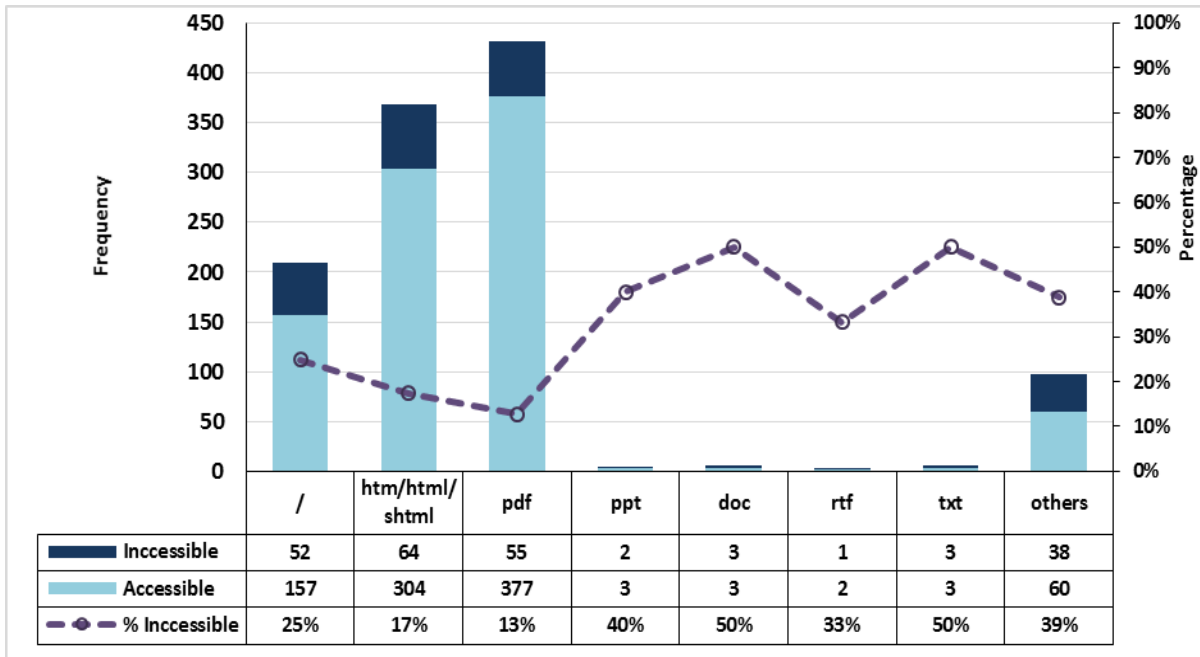
	/	htm/html/ shtml	pdf	ppt	doc	rtf	txt	others
Inccessible	52	64	55	2	3	1	3	38
Accessible	157	304	377	3	3	2	3	60
% Inccessible	25%	17%	13%	40%	50%	33%	50%	39%

Figure 5: distribution of internet addresses based on file format in all investigated articles in information science

Figure 6 indicates the distribution of cited internet addresses based on file format as well as the percent of inaccessible internet addresses in all investigated articles in management. Data provided in this figure reveal that internet addresses with pdf file format had the highest number of citations i.e. from among 1112 online citations, 402 citations were pdf files and internet addresses cited as rtf and .ppt files had the lowest number of citations (2 online citations). In accessible internet addresses pdf files were the most frequent. In other words, from among 934 accessible online citations, 332 citations were pdf files. Furthermore, from among 178 inaccessible internet addresses, pdf files were the most frequent (70 online citations).
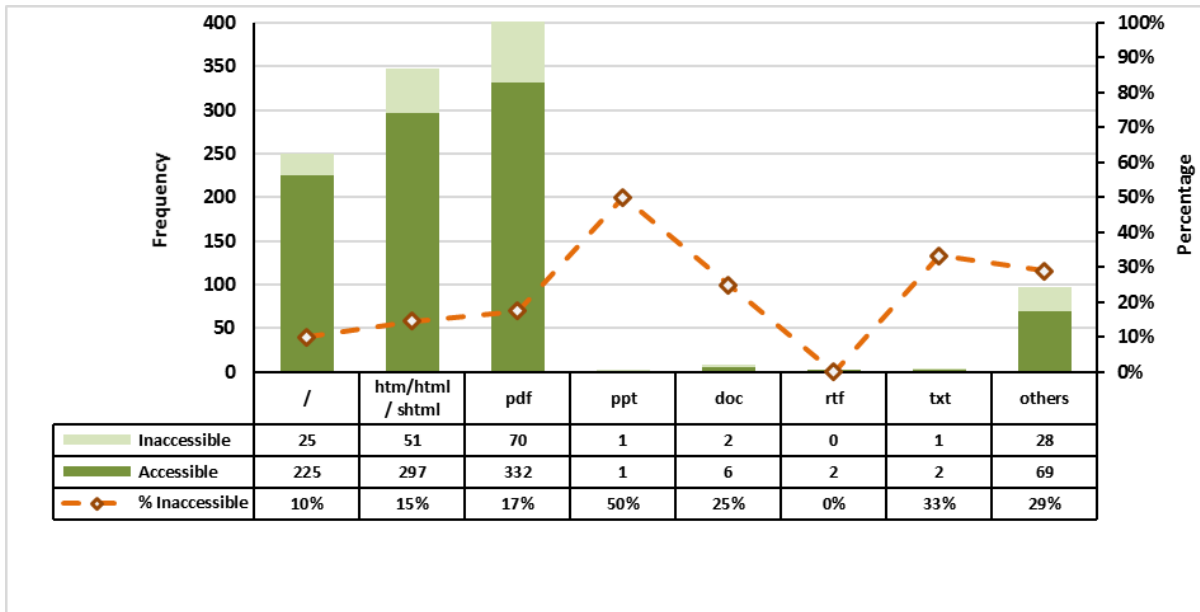

	/	htm/html / shtml	pdf	ppt	doc	rtf	txt	others
Inaccessible	25	51	70	1	2	0	1	28
Accessible	225	297	332	1	6	2	2	69
% Inaccessible	10%	15%	17%	50%	25%	0%	33%	29%

Figure 6: distribution of internet addresses based on file format in all investigated articles in management

The fifth research question: To what extent was the disappearance rate of URLs cited in ISI articles by Iranian researchers in information science, psychology and management in the initial check?

Table 7 shows the status of accessibility of internet addresses cited in all investigated articles in information science, psychology and management in the initial check.

Table 7

Accessibility of URLs cited in all investigated articles in initial check

The status of URLs	Accessible		Inaccessible		overall	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Psychology	397	58%	345	42%	742	100%
Information science	909	82%	218	18%	1127	100%
Management	934	88%	178	12%	1112	100%
Total	2240	75%	741	25%	2981	100%

The data presented in this table demonstrate that in all investigated articles, from among 2981 online citations, 75 percent of citations were accessible and 25 percent of them were inaccessible. In other words, from among 2981 internet addresses, 2240 of addresses directly led to cited information, nevertheless, in direct access to 741 internet addresses, errors were encountered.

The sixth research question: To what extent was the disappearance rate of URLs cited in ISI articles by Iranian researchers in information science, psychology and management in the second check?

In order to identify whether the intended contents and URLs cited in articles by Iranian researches in psychology, information science and management were still accessible or not, the following measures were taken. First, the internet address was entered into the internet browser and the cited internet address was searched for. When the cited content was not found, the displayed error was corrected if possible. Otherwise, the internet address was searched in the internet archive. If it was not even found in the archive, the lost information was sought in Google. Hence, up to 5 times, different phrases like titles, keywords, authors' names and the information relevant to cited resources were searched in Google. If the cited resource was not found in the first 20 retrieved results, that resource was regarded as inaccessible. The status of accessibility of cited internet addresses in the second check is presented in table 8.

Table 8

Accessibility of URLs cited in all journals in the second check

The status of URLs	Accessible		Inaccessible		Overall	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Psychology	704	95%	38	5%	742	100%
Information science	1104	98%	29	2%	1127	100%
Management	1089	97%	23	3%	1112	100%
Total	2897	97%	90	3%	100%	100%

As it is shown in this table, after entering the lost internet address into the internet browser, searching it in internet archive and via Google search engine, the accessible internet

addresses were increased from 75 percent to 97 percent. The inaccessible internet addresses were also decreased from 25 percent to 3 percent.

The seventh research question: What was the status of accessible cited URLs?

Data provided in table 9 indicate that in all investigated articles, from among 2673 accessible internet addresses, 66 percent of addresses were accessible from the cited internet address, 7 percent from Google search engine, 8 percent from internet archive, 18 percent from an internet address other than the one cited, 2 percent from searching the lost internet address in internet browser, 3 percent from shortening the internet address and 2 percent were accessible by editing the errors of internet address.

Table 9
The status of accessible internet addresses in all investigated articles

Accessibility status	Frequency	Percent
Found in the cited internet address	1771	66%
Found by entering the internet address into internet browser	53	2%
Found in internet archive	221	8%
Found by error correction	66	2%
Found via Google search engine	190	7%
Found in another internet address	491	18%
Found by shortening internet address	76	3%
Total	6325	100%

The eighth research question: What was the status of error messages in disappeared URLs?

Table 10 shows the status of inaccessible internet addresses in all investigated articles regarding the error messages. Data of this table reveal that in all inaccessible internet addresses, 404 error message (Not found) was the most frequent error message (24 percent). Then, 400 error message (bad request) with 17 percent frequency, 500 error message (server error) with 16 percent, 410 error message (gone) with 12 percent, 401 error message (unauthorized) with 11 percent and 403 error message (forbidden) and 406 error message (unacceptable) with 10 percent frequency were encountered in inaccessible internet addresses.

Table 10
The status of error messages in all investigated articles

Error message	Frequency	Percent
400 (bad request)	15	17%
404 (not found)	22	24%
500 (server error)	14	16%
403 (forbidden)	9	10%
401 (unauthorized)	10	11%
410 (gone)	11	12%
406 (not acceptable)	9	10%
Total	90	100%

The ninth research question: What was the status of half-life of URLs cited in ISI articles by Iranian researchers in information science, psychology and management?

In order to calculate the half-life of online citations, half-life formula applied in foreign researches (Koehler, 1999; Tyler & McNeil, 2003; Dimitrova & Bugeja, 2007) was utilized so as to be also used by Iranian researchers and librarians and provide ground for calculating the half-life of online citations of different journals in future studies[2].

Data collected from the investigation of ISI articles carried out by Iranian researchers in psychology, information science and management in a one-year period are presented in table 11. These data show that the average half-life of online citations was 2.6 years. In other worlds, it lasts for about two years and a half for half of online citations in these articles to disappear. Moreover, based on this table, the highest value of average half-life of online citations was for the articles in information science with 3 years and the lowest was for the articles in psychology with 5 months.

Table 11
half-life of online citations in all investigated articles

Title of journal/year	2009	2010	2011	Average
Psychology	3.2	2.6	2.9	1.9
Information science	5.1	4.8	4.7	3.4
Management	4.6	4.4	3.7	2.5
				2.6

Discussion and Conclusion

Investigating the frequency distribution of articles, citations and online citations revealed that from among all ISI journals in three investigated fields from 2009 to 2011, the highest number of articles were in management (285 articles) and these articles had the highest number of citations with 8926 articles. However, the highest number of online citations belonged to information science (1127 online citations) which is due to the fact that the researchers in information science (despite fewer articles) are more familiar with online citations and their advantages.

The bibliographic citations in all investigated articles from 2009 to 2011 in psychology, information science and management suggested that online citations in all three fields were the most frequent in 2011 (46, 40 and 81 percent, respectively) than in other years. This might be due to the developments in modern technology and its consequent facilities as well as use of web resources in conducting researches by researchers of different fields.

The findings of this section of the research were compatible with those of the research by Saberi and Abedi (2012) and Maharana, Nayak & Sahu (2006). The results of the research by Saberi and Abedi (2012) demonstrated that 62 percent of articles in information science used internet resources. The results of the research by Sadatmoosavi et al (2012) also estimated online citations as forming 65.12 percent of all citations.

The status of cited internet addresses based on domain in all investigated articles from 2009 to 2011 in psychology, information science and management suggested that the cited internet addresses with .org domain had the highest number of citations in all three investigated fields. Furthermore, the stability of cited internet addresses in psychology and information science with .org and .com was more than that of other domains. Nonetheless, in

management, domains other than the ones mentioned were the most stable and persistent. The results related to psychology and information science in this section of the research were compatible with the results of the study by Casserly & Bird (2003). These researchers reported the stability and persistence of .org and .com domains more than other domains. However, the results of the research by Saberi and Abedi (2012) were not compatible with those of this research since the findings of this research showed that internet addresses with .edu/.ac domain had the highest number of citations and the .net and .gov were more stable than other domains.

The status of cited internet addresses based on file format in all investigated articles from 2009 to 2011 in psychology, information science and management suggested that pdf. Internet addresses had the highest number of citations in all three investigated fields. Moreover pdf and html files were more stable than other file formats in internet addresses in psychology and information science but in management, rtf and html files were more stable. The results of this section were compatible with those of the research by Wagner et al (2009), Sadatmoosavi et al (2012) and Germain (2000) but Dellavalle et al (2003) evaluated the stability of .rtf files so low.

In general, in all investigated articles of three fields, from 2981 cited internet addresses 2240 addresses directly led to cited information (75 percent) while in direct access to 741 internet addresses errors were encountered (25 percent). It is worth mentioning that articles in management were the most accessible ones in initial search. The findings of this section were compatible with those of previous studies (Casserly & Bird, 2003; Dimitrova & Bugeja, 2007b; McCown, Chan, Nelson, & Bollen, 2005a; McCown et al., 2005b; Saberi & Abedi, 2012; Wagner et al., 2009; Wren, 2004). In all investigated articles of all three fields, accessible internet addresses were increased from 75 percent to 97 percent and inaccessible internet addresses were decreased from 25 percent to 3 percent. It is worthy of consideration that articles in information science were the most accessible ones in second search. The results of this section of the research were compatible with those of the researches by Falagas, Karveli & Tritsaroli (2007) and Dellavalle et al (2003) in increasing the accessibility of online citations.

In all investigated articles of all three fields, from 2673 accessible internet addresses, 66 percent of addresses were accessible from the cited internet address, 7 percent from Google search engine, 8 percent from internet archive, 18 percent from an internet address other than the one cited, 2 percent from searching the lost internet address in internet browser, 3 percent from shortening the internet address and 2 percent were accessible by editing the errors of internet address. It is worth mentioning that the highest number of accessible articles (1104) were in information science and in all three fields, most accessible articles were available in the cited internet addresses. The results of this section were compatible with those of the research by Wagner et al (2009).

In general, in all investigated articles of all three fields, 404 error message (Not found) was the most frequent error message (24 percent) in all inaccessible internet addresses. Of course, most of the errors in information science and psychology articles were 404 (Not found) error messages and they were similar and 403 (forbidden) error message was the most frequent error message in all three fields. The results of this section were compatible with the results reported by McCown et al. (2005a) and Wren (2004).

The calculation of half-life of internet resources indicated that the average half-life of

internet resources cited in all investigated journals was 3.9 years. Moreover, the highest value of average half-life of online citations was for the articles in information science (3 years and 4 months). It seems as if the reason for low half-life of online citations in articles investigated in this research was overuse of less stable internet citations (.org, .edu, .com) as well as less stable files with HTM/HTML/SHTML and (/).

Finally, if the observations and results of this study and other similar studies be acknowledged, the decay of internet addresses should be regarded as a problem the most important reason of which is website reorganization (Aronsky, Madani, Carnevale, Duda, & Feyder, 2007). The other reason can be the changes to the name of internet domains (Hu, 2004).

Some fields are more exposed to and affected by the consequences of decay of internet addresses (Davis & Cohen, 2001; Falagas et al., 2008; McCown et al., 2005b). The influence of inactive links on the journals of a field is different based on the reliance of authors on internet based information. The absolute number of internet addresses also strengthens the problem of decay of internet addresses for the readers of the articles as compared with those journals whose authors have only cited a few online citations. The consequences of inactive links for those articles and resources which can be accessed through different ways or their print version is accessible are less serious.

Although the decay of online citations is a serious problem, researchers and publishers might downplay this since they believe search engines like Google are able to locate resources in their new internet addresses. They must remember that these tools don't index all internet documents and can't locate the items deleted from the web. Tools like internet archives including Wayback Machine and website might make it possible to have a snapshot of the content of a site in a particular time. They don't, however, include all documents that are on the internet or are accessible through internet. Google doesn't index dynamic pages or pages and sites which use robots.txt coding to prevent crawling. In addition, ranking of a site or page in Google search results depends on the number of other pages connected to this page or site (Klein & Nelson, 2008). It is believed that the responsibility to archive the internet content used in an article is upon the author or the publisher of the article who has used that content. Dellavalle et al (2003) believed that the best solution to improve the accessibility of internet resources is to request for all internet information be analyzed and recorded while examining the manuscripts. In so doing, the responsibility to archive information will be assigned to the publisher.

Endnotes

1. http://archive.org/web
2. The method of calculating half-life of online citations was explained in methods section in details.

References

Aronsky, D., Madani, S., Carnevale, R. J., Duda, S., & Feyder, M. T. (2007). The Prevalence and Inaccessibility of Internet References in the Biomedical Literature at the Time of Publication. *Journal of the American Medical Informatics Association, 14*(2), 232-234. doi: 10.1197/jamia.M2243.

Asareh, F. (1999). Citation analysis. *National Studies on Librarianship and Information Organization, 9*(3:4), 34-48.

Calishain, T., & Dornfest, R. (2005). *Google hacks*: O'Reilly Media, Inc.

Casserly, M. F., & Bird, J. E. (2003). Web citation availability: Analysis and implications for scholarship. *College and Research Libraries, 64*(4), 300-317.

Davis, P. M., & 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, 52*(4), 309-314. doi: 10.1002/1532-2890(2000)9999:9999<::AID-ASI1069> 3.0.CO;2-P

Dellavalle, R. P., Hester, E. J., Heilig, L. F., Drake, A. L., Kuntzman, J. W., Graber, M., & Schilling, L. M. (2003). Going, Going, Gone: Lost Internet References. *Science, 302*(5646), 787-788. doi: 10.1126/science.1088234

Dimitrova, D. V., & Bugeja, M. (2007b). Raising the dead: Recovery of decayed online citations. *American Communication Journal, 9*(2), 371-382.

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

Falagas, M. E., Karveli, E. A., & Tritsaroli, V. I. (2008). The risk of using the Internet as reference resource: A comparative study. *International Journal of Medical Informatics, 77*(4), 280-286. doi: 10.1016/j.ijmedinf.2007.07.001

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics, 1*(4), 359-375.

Germain, C. A. (2000). URLs: Uniform resource locators or unreliable resource locators. *College and Research Libraries, 61*(4), 359-365.

Grafton, A. (1999). *The footnote: A curious history*: Harvard University Press.

Hu, J. (2004). Co-founders release Google owner's manual. Retrieved from http://news.com.com/2100-1038-5202090.html.

Klein, M., & Nelson, M. L. (2008). *Revisiting lexical signatures to (re-) discover web pages.* Paper presented at the International Conference on Theory and Practice of Digital Libraries.

Kochen, M. (1987). How well do we acknowledge intellectual debts? *Journal of Documentation, 43*(1), 54-64. doi: 10.1108/eb026801

Koehler, W. (1999). An analysis of web page and web site constancy and permanence. *Journal of the American Society for Information Science, 50*(2), 162-180.

Maharana, B., Nayak, K., & Sahu, N. K. (2006). Scholarly use of web resources in LIS research: A citation analysis. *Library Review, 55*(9), 598-607. doi: 10.1108/00242530610706789.

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

McCown, F., Chan, S., Nelson, M. L., & Bollen, J. (2005a). The availability and persistence of web references. *D-Lib Magazine*. arXiv preprint cs/0511077.

McCown, F., Chan, S., Nelson, M. L., & Bollen, J. (2005b). The availability and persistence of web references *D-lib magazine*. *5th International Web Archiving Workshop (IWAW05), Vienna, Austria*.

Metcalfe, M. (2003). Author(ity): The literature review as expert witnesses. *Forum Qualitative Sozialforschung, 4*(1). doi: http://dx.doi.org/10.17169/fqs-4.1.761.

Rumsey, M. (2002). Runaway train: Problems of permanence, accessibility, and stability in the use of Web sources in law review citations. *Law Library Journal, 94*(1), 27-39.

Saberi, M. K., & Abedi, H. (2012). Accessibility of web citations in five open access ISI journals. *Internet Research, 22*(2), 234-247.

Sadat-Moosavi, A., Isfandyari-Moghaddam, A., & Tajeddini, O. (2012). Accessibility of online resources cited in scholarly LIS journals: A study of Emerald ISI-ranked journals. *Aslib Proceedings: New Information Perspectives, 64*(2), 178-192. doi: 10.1108/00012531211215196

Sadatmoosavi, A., Tajedini, O., & Moghadam, H. S. (2010). Assessing accessibility in four library and information sciences journals: A comparative study. *Information Sciences and Technology, 25*(4), 617-633.

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

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

Veal, A. J. (2005). *Business research methods: A managerial approach*. Pearson Education Australia/Addison Wesley.

Tyler, D. C., & McNeil, B. (2003). Librarians and link rot: A comparative analysis with some methodological considerations. *Portal, 3*(2), 615-632.

Ticehurst, G. W., & Veal, A. J. (2000). *Business research methods: A managerial approach,* Addison Wesley Longman.

Wagner, C., Gebremichael, M. D., Taylor, M. K., & Soltys, M. J. (2009). Disappearing act: Decay of uniform resource locators in health care management journals. *Journal of the Medical Library Association, 97*(2), 122-130. doi: 10.3163/1536-5050.97.2.009.

Wren, J. D. (2004). 404 not found: The stability and persistence of URLs published in MEDLINE. *Bioinformatics, 20*(5), 668-672. doi: 10.1093/bioinformatics/btg465.