The International Information & Library Review (2012) 44, 202–211

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/iilr

# Decay and half-life period of online citations cited in open access journals

B.T. Sampath Kumar [a,*], K.S. Manoj Kumar [b]

[a] Dept. P.G. Studies and Research in Library and Information Science, Tumkur University, Tumkur 572103, Karnataka, India
[b] Government Home Science College, Woman's First Grade College Building, Hassan, Karnataka, India

**Abstract** This study investigates the decay and half-life of online citations cited in four open access journals published between 2000 and 2009. A total of 1158 online citations cited in 1086 research articles published in two science and social science journals spanning a period of 10 years (2000–2009) were extracted. Study found that 24.58% (267 out of 1086) of articles had online citations and these articles contained a substantially very less number of online citations (2.98%) compared to previous study results. 30.56% (26% in Science and 52.73% in Social Science) of online citations were not accessible and remaining 69.44% of online citations were still accessible. The 'HTTP 404 error message-page not found' was the overwhelming message encountered and represented 67.79% of all HTTP message. Domains associated with .ac and .net had higher successful access rates while .org and .com/.co had lowest successful access rates. The half-life of online citations was computed to be approximately 11.5 years and 9.07 years in Science and Social science journal articles respectively.


## Introduction

The World Wide Web is growing at an enormous speed, and has become an indispensable source for information and research (Bar-Ilan & Peritz, 2008). The Internet age has supplemented and will probably eventually supplant the printed media for disseminating scientific information. It is not only more versatile in the types of information that can be presented (e.g. multimedia), but is cheaper to create and quicker to disseminate (Wren, 2008). It also facilitates scientific communication in many ways. Formal references to information on the web are becoming increasingly common (Lawrence et al., 2001). Many early studies (Hester et al., 2004; Moghaddam et al., 2010; Olfson & Lawrence, 2005; Sellitto, 2004; Veena & Sampath Kumar, 2008; Wu, 2009; Yang et al., 2010) have proved that authors of scientific articles include references accessible only through the web.

Even though URLs are Internet addresses that provide the reader with a unique reference to online information, the transient nature of the Internet and in the absence of a permanent digital library (Dellavalle et al., 2003) citing URLs has the disadvantage that accessibility and content stability are not guaranteed (Koehler, 2004). Previous studies have also showed that many URLs cited in research articles have disappeared because of several reasons

* Corresponding author.
*E-mail address:* sampathbt_2001@rediffmail.com (B.T. Sampath Kumar).

1057-2317/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.iilr.2012.09.002

(Aronsky, Madani, Carnevale, Duda, & Feyder, 2007; Carnevale & Aronsky, 2005; Goh & Ng, 2007; Lopresti, 2010; McCown et al., 2005; Sellitto, 2004). Keeping in view the disappearance nature of URL citations, this study made an attempt to investigate the availability and decay online citations cited in Science and Social Science open access journals published during 2000—2009. It also aimed calculate the half-life period of online citations cited in these journals articles.

## Research hypotheses and questions

The web has significantly improved the access to scientific information and resources, providing a way of making data and applications accessible and sharable, with the added convenience of information retrieval and extraction (Lawrence & Giles, 1999). Prior studies have also documented the web popularity among the academics, researchers and students. Based on this we have formulated the following hypotheses and research questions for this study.

Hypotheses

- Number of references and online citations cited in the scholarly articles are positively correlated.
- The availability of online citations varies across the science and social science journal articles.
- Number of missing online citations varies across the science and social science journal articles.
- Path depth is positively correlated to decay of online citations (URLs).

Based on these assumptions the present study made an attempt to address the following questions:

- To what extent the online citations are used in science and social science journal articles?
- Is there any association between the availability of online citations and the journal's discipline?
- What percentage of online citations cited in science and social science journal articles are missing?
- Is there any correlation between the path depth and decay of online citations? And what is the half-life of online citations cited in science and social science journal articles?

## Review of literature

During the last decade, a considerable body of research on availability and persistence of web citations exits. During the last few years several studies have also carried out to estimate the half-life of web citation. This section highlights a comprehensive review of literature related to decay and half-life period of online citations.

### Decay of online citations

In recent years Web sources have gained momentum, which played an important role in the increase of citations to Web resources in scholarly publication in all discipline. The review of literature clearly showed that the web sources are most likely to be cited by research community in their scholarly work. Even though web citations have become more popular among the scholars, the impermanence nature of web citations questions their authenticity. In this context many studies have been conducted regarding the decay of online citations. This section highlights the results of earlier research works on availability and decay of online citations.

Dimitrova and Bugeja (2007), in an exploratory study, examined the use of online citations, focusing on five leading journals in Journalism and Communication. They analyzed 1126 URL reference addresses in citations of articles published between 2000 and 2003. The results showed that only 61% of the online citations remain accessible in 2004 and 39% do not. The content analysis also demonstrated that .org and .gov were the most stable domains. McCown et al. (2005) explored the availability and persistence of URLs cited in articles published in D-Lib Magazine. About 4387 unique URLs referenced in 453 articles published from July 1995 to August 2004. The availability was checked three times a week for 25 weeks from September 2004 to February 2005. Approximately 28% of those 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. Aronsky et al. (2007), study included 4699 publications from 844 different journals. Among the 141,845 references there were 840 (.6%) web citations. From the 840 web citations, 11.9% were already inaccessible within two days after an article's release to the public. In the same year Goh and Ng (2007) investigated the link decay phenomenon in three leading Information Science journals. It was found that approximately 31% of all citations were not accessible during the time of testing and the majority of errors were due to missing content. Citations from the .edu domain were found to have the highest failure rates of 36% when compared with other popular top-level domains.

Ducut et al. (2008) conducted a survey on MEDLINE URL's and found 10,208 URL addresses. All URL addresses were checked for errors during the initial run. A total of 2245 URL's were not accessible during the initial run of this group. A total of 163 URL's redirected to another page and the updated addresses were used in the study run. Wagner et al. (2009) found that the percentage of inactive URL's ranged from 39.2% for articles published in 2004 to 61.1% for articles published in 2002. The highest percentage of inactive URL's was found in the .com, top-level domain followed by the .gov and .org domains. Lopresti (2010) examined citations in five leading Environmental Science journals for accuracy. As many as 647 (24.41%) of the 2650 citations checked were found to contain errors. Of the five journals, Conservation Biology had the lowest percentage of citations with errors and Climatic Change had the highest.

A study by Bugeja and Dimitrova (2010) focused on nine leading journals in the area of Journalism and Communication. Using longitudinal data and a content analysis methodology, they analyzed the use of online footnotes in refereed journal articles over a four-year period (2000—2003). Authors warned that the decay of footnotes