

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

---

April 2021

## Pattern of use and characteristic features of web citations in Scholarly journals

Niveditha B.

*University of Mysore*, [niveditha.jb@gmail.com](mailto:niveditha.jb@gmail.com)

Mallinath Kumbar

*University of Mysore*, [mallinathkumbar@yahoo.com](mailto:mallinathkumbar@yahoo.com)

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



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

---

B., Niveditha and Kumbar, Mallinath, "Pattern of use and characteristic features of web citations in Scholarly journals" (2021). *Library Philosophy and Practice (e-journal)*. 5272.  
<https://digitalcommons.unl.edu/libphilprac/5272>

# **Pattern of use and characteristic features of web citations in Scholarly journals**

**Niveditha B.<sup>1</sup>, Mallinath Kumbar<sup>2</sup>**

<sup>1</sup>UGC-Senior Research Fellow, Department of Library and Information Science,  
University of Mysore, Manasagangotri, Mysuru - 570006  
Email: [niveditha.jb@gmail.com](mailto:niveditha.jb@gmail.com)

<sup>2</sup>Professor, Department of Library and Information Science, University of Mysore,  
Manasagangotri, Mysuru – 570006

## **Abstract**

The present study examines the reliability of web citations in scholarly journals of Library and Information Science and Communication and Media Studies. The journals were selected based on their high impact factor published between 2008 and 2017. A total 5,55,428 references were extracted, out of which 1,02,718 had web citations. The research findings indicated that there more number of URLs in CMS journal articles than in LIS journal articles. While examining the characteristic features of the URLs cited, it was found that .html files and organizational (.org) top-level domain were the most cited in both the disciplines. Moreover, URLs with path depth 2 and character length 41-50 were frequently cited in both the disciplines. The display and redirect URLs associated with the characteristic features are also determined in this study.

**Keywords:** Web citations, Library and Information Science, Communication and Media Studies, URLs, File Format, Top-Level Domain, Path Depth, Character Length

## **1. Introduction**

References identify and acknowledge previous research works that are used by the scholarly community while conducting or presenting their own research. The adoption of the Web and the Internet to disseminate information has resulted in the use of Web Citations in scholarly papers. The authors of all disciplines are using web citations as they are aware that the Web has tremendous amount of information and has become the primary choice of seeking information. The present study has tried to compare the use of web citation in two disciplines, Library and Information Science and Communication and Media Studies, during the period

2008-2017. The study also examines and differentiates the characteristic features of web citations like their file format, top-level domain, path depth and character length.

## 2. Objectives of the study

The study has been conducted with the following objectives:

1. To know the percentage of web references percentage of web citations used in the journal articles of Library and Information Science and Communication and Media Studies
2. To examine the characteristic features of URLs like file format, top-level domain, path depth and character length.
3. To differentiate the characteristic features of display URLs and redirect URLs.

## 3. Methodology

### 3.1 Selection of journals

For the present study, data was drawn from 20 journals, ten each from Library and Information Science and Communication and Media Studies. The journals were selected based on their high impact factor as per *Clarivate Analytics*' 2018 "Journal Citation Report." The journals selected for the present study are presented Table 1.

**Table 1: Journals selected for the study**

Library and Information Science		Communication and Media Studies	
Journal	Impact factor	Journal	Impact factor
Journal of Informetrics (JOI)	3.484	Journal of Computer-Mediated Communication (JCMC)	4.000
Information Processing and Management (IPM)	3.444	Journal of Communication (JOC)	3.729
Journal of the Association for Information Science and Technology (JASIST)	2.835	Communication Research (CR)	3.391
Scientometrics	2.173	New Media and Society (NMS)	3.121
College and Research Libraries (CRL)	1.626	Information, Communication and Society (ICS)	3.084
Journal of the Medical Library Association (JMLA)	1.541	Journal of Advertising (JOA)	2.880
Portal: Libraries and the Academy (Portal)	1.473	Political Communication (PC)	2.738

Aslib Journal of Information Management (AJIM)	1.461	Communication Theory (CT)	2.733
Journal of Academic Librarianship (JAL)	1.459	Media Psychology (MP)	2.570
Library and Information Science Research (LISR)	1.372	Public Understanding of Science (PUOS)	2.452

### 3.2 Selection of articles and references

All the research articles published during the 10-year period, that is, from 2008 to 2017 were taken up for the study. Editorial notes, book reviews, short communication were excluded. The references that were adjoined at the end of each article were considered for the study. A total of 12,251 articles that were published in 20 journals were downloaded, and 5,55,428 references were extracted to Microsoft Excel.

### 3.3 Extraction of URLs and distinguishing their characteristic features

The references that contained web links and DOIs were extracted. To study the distribution of file format and other characteristic features of web citations, the DOIs and arXiv were first resolved to URLs using the syntax <https://doi.org/>. and <https://arxiv.org/>., respectively. For example, a DOI name 10.1010.1234/567 was resolved from the address <https://doi.org/10.1111.4321/789>. Similarly, arXiv identifier was resolved to URLs using the syntax <https://arxiv.org/>. There were a total of 1,02,718 URLs, and this data was pruned for accuracy, and year-wise duplicates were removed. The web address which is displayed regardless of the article's physical location is the Display URL. The URL after multiple redirects reaches where the article resides that is under the control of the publisher and is called the redirect URL. The display and redirect URLs associated with these characteristic features are distinguished in this study.

## 4. Results and Discussion

### 4.1 Year-Wise Distribution of Web Citations

Table 2 summarizes the year-wise distribution of web citations in Library and Information Science and Communication and Media Studies journal articles. The table shows that the number of web citations in Library and Information Science journal articles (51,839) is more than in Communication and Media Studies journal articles (50,879). The percentage of web citation by year in both disciplines has increased from a low of 11.21 and 7.22 in the year 2008 to a high of

22.80 and 37.22 in the year 2017. The percentage of web citations from the total number of references is high in Communication and Media Studies journal articles (22.05%) than in Library and Information Science journal articles (15.97%).

The data also indicates that the least percentage of web citations is noted during the year 2008 (11.21%) and 2010 (11.58%) in Library and Information Science journal articles. In Communication and Media Studies journal articles, the least percentage of web citations are noted in 2008 (7.22%) and 2011 (8.70%).

**Table 2: Year-Wise Distribution of Web Citations**

Year	Library and Information Science			Communication and Media Studies		
	Total number of references	Total web citations	%	Total number of references	Total web citations	%
2008	22064	2474	11.21	18226	1316	07.22
2009	23566	2965	12.58	20615	1933	09.38
2010	25556	2960	11.58	19707	1823	09.25
2011	26859	3452	12.85	20823	1811	08.70
2012	28495	3919	13.75	21399	3734	17.45
2013	32661	4676	14.32	22384	5673	25.34
2014	35440	5327	15.03	25622	7303	28.50
2015	41811	6942	16.60	26353	7815	29.66
2016	43662	8974	20.55	29089	9581	32.94
2017	44522	10150	22.80	26574	9890	37.22
<b>Total</b>	<b>324636</b>	<b>51839</b>	<b>15.97</b>	<b>230792</b>	<b>50879</b>	<b>22.05</b>

#### 4.2 Journal-Wise Distribution of Web Citations

Table 3 reflects the distribution of web citations in journals. The total web citations among all journals are found to be the highest in the journal Scientometrics (14261), followed by Information Communication and Society (13233) and Journal of the Association for Information Science and Technology (10301). A low number of web citations are noted in the Journal of

Advertising (940) and Journal of Medical Library Association (2255). It can also be seen from the table that three journals (two from LIS and one from CMS) had web citations of more than 10,000; three journals (one from LIS and two from CMS) had web citations between the range 5000 to 10,000, and the remaining 14 journals (seven each from LIS and CMS) had web citations below 5000.

The table also depicts that the highest percentage of web citations are cited in a Communication and Media Studies journal, Information Communication and Society with 38.05% references citing a web source, followed by the Journal of Medical Library Association (35.83%) and Journal of Communication (33.44%). A low percentage of web citations are noticed in the Journal of Advertising (4.84%), followed by Information Processing and Management (8.87%), and Public Understanding of Science (9.80%).

**Table 3: Journal-Wise Distribution of Web Citations**

Library and Information Science				Communication and Media Studies			
Journal	Total number of references	Total web citations	%	Journal	Total number of references	Total web citations	%
JOI	24901	3546	14.24	JCMC	17946	4548	25.34
IPM	31553	2798	08.87	JOC	25811	8631	33.44
JASIST	87468	10301	11.78	CR	24503	4781	19.51
Sciento	96137	14261	14.83	NMS	40663	6468	15.91
CRL	13376	2722	20.35	ICS	34782	13233	38.05
JMLA	6293	2255	35.83	JOA	19427	940	04.84
Portal	10412	2906	27.91	PC	14715	2396	16.28
AJIM	15536	2682	17.26	CT	15348	3274	21.33
JAL	22658	6999	30.89	MP	12495	4147	33.19
LISR	16302	3369	20.67	PUOS	25102	2461	09.80
<b>Total</b>	<b>324636</b>	<b>51839</b>	<b>15.97</b>	<b>Total</b>	<b>230792</b>	<b>50879</b>	<b>22.05</b>

### 4.3 Distribution of file format

File extensions indicate the native file format, which is used to store data files created with the software program. A total of 17 file formats were identified. The file formats .cgi, .doc, .jsp, .txt, xls, .ppt, .rtf, .exe, .nsf, .xml, .shtml, .xhtml were categorized in “Others” category.

**Table 4:** File Format of URLs in LIS and CMS Journal Articles

File format	LIS articles		CMS articles	
	Total URLs	Percentage	Total URLs	Percentage
.html	40401	77.94	44388	87.24
.pdf	7736	14.92	3624	7.12
.asp	1056	2.04	1022	2.01
.php	1213	2.34	1205	2.37
.cfm	626	1.21	314	0.62
Others	807	1.56	326	0.64
<b>Total</b>	<b>51839</b>	<b>100</b>	<b>50879</b>	<b>100</b>

The data illustrated in Table 4 indicates that the greatest numbers of cited URLs in both disciplines are .html files followed by .pdf files and the least cited are .cfm files. The results from the study of Saberi and Abedi (2012) and Jalalifard et al., (2013) also indicated that greatest number of cited URLs is .html and .pdf files. The file formats in the “Others” category just constituted 2% in LIS journal articles and 1% in CMS journal articles.

### 4.4 Distribution of display and redirect URLs by top-level domain

The distribution of display and redirect URLs by top-level domain type is shown in the following tables. Twelve main types of top-level domain have been identified in this study. They are .com, country top-level domains, .edu, .gov, .info, .int, .mil, .net, .org, .name, .design and .ngo. The top-level domains like .info, .int, .mil, .name, .design, and .ngo were considered in the “Others” category in the following tables.

It can be seen from Table 5 and Table 6 that in both the disciplines .org (organizational) is the leading top-level domain used followed by .com (commercial). It can also be observed from previous studies (Ducut et al., 2008; Sampath Kumar & Prithviraj, 2012) that .org is the most

common top-level domain cited. Some URLs tend to redirect, and a web page with different URL is opened. When the redirect URL is reached, there is a drastic change in the top-level domain when compared to the display URL. The Commercial top-level domain is the leading top-level domain when the redirect URL is reached, and it is followed by the organizational top-level domain.

Table 5 and Table 6 also show that the URLs having the organizational top-level domain have shown substantial reduction after multiple redirects. In Library and Information Science journal articles, the percentage of URLs with top-level domain .org has reduced from 59.14% to 29.82% with a difference of 29.32%. Similarly, in Communication and Media Studies journal articles, URLs with .org as their top-level domain has reduced from 71.56% to 22% with a difference of 49.56%.

**Table 5:** Top-level Domain of Display and Redirect URLs in LIS Journal Articles

Top-level domain	Display URL		Redirect URL	
	Total URLs	Percentage	Total URLs	Percentage
.org	30656	59.14	15456	29.82
.com	6728	12.98	21262	41.02
country code	6563	12.66	6690	12.91
.edu	4200	8.10	4542	8.76
.gov	1849	3.57	2111	4.07
.net	1458	2.81	1379	2.66
Others	385	0.74	399	0.77
<b>Total</b>	<b>51839</b>	<b>100</b>	<b>51839</b>	<b>100</b>

**Table 6:** Top-level Domain of Display and Redirect URLs in CMS Journal Articles

Top-level domain	Display URL		Redirect URL	
	Total URLs	Percentage	Total URLs	Percentage
.org	36351	71.45	32322	63.58
.com	6960	13.68	11187	22.00
country code	4112	8.08	3949	7.77
.edu	1871	3.68	1889	3.72
.gov	695	1.37	702	1.38
.net	687	1.35	619	1.22
others	203	0.40	211	0.42



<b>Total</b>	<b>50879</b>	<b>100</b>	<b>50879</b>	<b>100</b>
--------------	--------------	------------	--------------	------------

#### 4.5 Path depth of the display and redirect URLs

The number of directory levels in the URL constitutes the URL depth. The path depth of URLs was grouped into 10 path depths ranging from 0 to 8 and the remaining URLs with path depth more than 8 were categorized under “>8”.

It can be observed from Table 7 and Table 8 that URLs with path depth 2 are most frequently cited, followed by URLs with a depth of 3 and 4. This is consistent in both disciplines. A study done by Sampath Kumar et al. (2015) reported that online citations with path depth 2 were most frequently cited followed by path depth 3 and 4.

In Library and Information Science journal articles, the URLs with path depth 2 continued to be the highest after multiple redirects followed by URLs with path depth 3 and 4. In Communication and Media Studies journal articles, there is a slight difference as the URLs with path depth 3 are found to be the highest after URL redirection, and this is followed by URLs with path depth 2 and 4.

URLs with path depth 2 have shown a significant reduction after multiple redirects. In LIS discipline, the percentage of URLs with path depth 2 after multiple redirects has reduced from 52.78% to 30.28% with a difference of 22.5%. In Communication and Media Studies journal articles, URLs with path depth 2 has reduced drastically from 66.78% to 21.87% with a major difference of 44.91%. The URLs with path depth 0 have also shown a slight decrease in both the disciplines after the redirects. On the contrary, the remaining path depths have seen a surge after multiple redirects.

**Table 7:** Path Depth of Display and Redirect URLs in LIS Journal Articles

<b>Path Depth</b>	<b>Display URL</b>		<b>Redirect URL</b>	
	<b>Total URLs</b>	<b>Percentage</b>	<b>Total URLs</b>	<b>Percentage</b>
0	1403	2.71	589	1.13621
1	3910	7.54	6177	11.9157
2	27360	52.78	15698	30.2822

3	8021	15.47	12700	24.4989
4	5513	10.63	8544	16.4818
5	2767	5.34	3583	6.91178
6	1494	2.88	2175	4.19568
7	761	1.47	1691	3.26202
8	252	0.49	275	0.53049
>8	358	0.69	407	0.78512
<b>Total</b>	<b>51839</b>	<b>100</b>	<b>51839</b>	<b>100</b>

**Table 8:** Path Depth of Display and Redirect URLs in CMS Journal Articles

Path Depth	Display URL		Redirect URL	
	Total URLs	Percentage	Total URLs	Percentage
0	764	1.50	355	0.70
1	2456	4.83	3740	7.35
2	33975	66.78	11129	21.87
3	5610	11.03	15030	29.54
4	3134	6.16	10326	20.30
5	2635	5.18	3315	6.52
6	1279	2.51	5685	11.17
7	577	1.13	826	1.62
8	178	0.35	229	0.45
>8	271	0.53	244	0.48
<b>Total</b>	<b>50879</b>	<b>100</b>	<b>50879</b>	<b>100</b>

#### 4.6 Character Length of Display and Redirect URLs

Uniform Resource Locators are a combination of characters that consist of alphabets, numbers, and special characters. The character length of each URL is calculated, and the data is presented in the following tables. Table 9 and Table 10 shows that the URLs with character length 41-50 characters are cited highly, followed by character length of 31-40 and 51-60. This is found to be same in both the disciplines. The findings are similar to the results from previous studies of Sampath Kumar and Vinay Kumar (2013) and Sampath Kumar et al. (2015) where the highest number of URLs had character length of 41-50.

After multiple redirects, there is a change in the character length of the URLs. It is noted that after the URL redirection, there is a high number of URLs with character length 51-60 in Library and Information Science journal articles and in Communication and Media Studies journal articles character length of 61-70 is found to be the highest.

In LIS journal articles, the percentage of URLs with character length 41-50 after multiple redirects has reduced from 36.69% to 12.59%. In Communication and Media Studies journal articles, URLs with character length 41-50 has reduced drastically from 43.19% to 5.72% with a difference of 37.47%. This similar trend is seen in URLs with a character length of 31-40 and URLs with a character length of more than 100. The URLs with the remaining character length have seen an increase after redirection. This is similar in both LIS and Communication and Media Studies journals.

**Table 9:** Character length of Display and Redirect URLs in LIS Journal Articles

Character length	Display URL		Redirect URL	
	Total URLs	Percentage	Total URLs	Percentage
<=20	454	0.88	534	1.03
21-30	2759	5.32	2870	5.54
31-40	10597	20.44	2624	5.06
41-50	19022	36.69	6524	12.59
51-60	6235	12.03	10980	21.18
61-70	4390	8.47	10196	19.67
71-80	2972	5.73	9771	18.85
81-90	2035	3.93	4379	8.45
91-100	1236	2.38	2539	4.90
>100	2139	4.13	1422	2.74
<b>Total</b>	<b>51839</b>	<b>100</b>	<b>51839</b>	<b>100</b>

**Table 10:** Character length of Display and Redirect URLs in CMS Journal Articles

Character length	Display URL		Redirect URL	
	Total URLs	Percentage	Total URLs	Percentage
<=20	202	0.40	309	0.61

21-30	1826	3.59	2351	4.62
31-40	13007	25.56	1303	2.56
41-50	21974	43.19	2909	5.72
51-60	3811	7.49	6453	12.68
61-70	2759	5.42	19204	37.74
71-80	2426	4.77	11344	22.30
81-90	1730	3.40	3096	6.09
91-100	1107	2.18	2620	5.15
>100	2037	4.00	1290	2.54
<b>Total</b>	<b>50879</b>	<b>100</b>	<b>50879</b>	<b>100</b>

## 5. Conclusion

The present study investigated the use of web citations cited in Library and Information Science and Communication and Media Studies scholarly journals during the year 2008-2017. It can be noted from the study that the percentages of web citations in articles have been continuously increasing in both disciplines. However, it can be seen that Library and Information Science journals have slightly more web citations than the Communication and Media Studies journals. The present study documents the characteristic features of the URLs cited like the file format, top-level domain, path depth and character length. It is found that the researchers prominently use the .html file format and .org top-level domain the most. The study also differentiates the characteristic features of display and redirect URLs and it was found that there were changes in the top-level domain, path depth and character length as soon as the landing page of the article is reached.

## References

1. Ducut, E., Liu, F., & Fontelo, P. (2008). An update on Uniform Resource Locator (URL) decay in MEDLINE abstracts and measures for its mitigation. *BMC Medical Informatics and Decision Making*, 8(1). <https://doi.org/10.1186/1472-6947-8-23>
2. Jalalifard, M., Norouzi, Y., & Isfandyari-Moghaddam, A. (2013). Analyzing web citations availability and half-life in medical journals: A case study in an Iranian university. *Aslib Proceedings*, 65(3), 242–261. <https://doi.org/10.1108/00012531311330638>

3. Saberi, M. K., & Abedi, H. (2012). Accessibility and decay of web citations in five open access ISI journals. *Internet Research*, 22(2), 234–247. <https://doi.org/10.1108/10662241211214584>
4. Sampath Kumar, B. T., & Prithviraj, K. R. (2012). Availability and persistence of web citations in Indian LIS literature. *The Electronic Library*, 30(1), 19–32. <https://doi.org/10.1108/02640471211204042>
5. Sampath Kumar, B. T., & Vinay Kumar, D. (2013). HTTP 404-page (not) found: Recovery of decayed URL citations. *Journal of Informetrics*, 7(1), 145-157.
6. Sampath Kumar, B. T., Vinay Kumar, D., & Prithviraj, K. R. (2015). Wayback machine: reincarnation to vanished online citations. *Program*, 49(2), 205–223. <https://doi.org/10.1108/PROG-07-2013-0039>