Access and Citing Theses Literature: A Study of the Role of Repositories and Academic Search Engines

Vijayakumar P[1] and Vijayakumar M[2]

[1]Assistant Librarian, Ananda Rangapillai Library, Pondicherry University, Puducherry, India

[2]University Librarian, Ananda Rangapillai Library, Pondicherry University, Puducherry, India

Abstract

This work investigates the impact of citations to theses emanating through archives and search engines such as Shodhganga, NDLTD, and others and compares them with Google Scholar citation data. The study is quantitative and uses the Bibliometrics citation analysis technique to evaluate references in the Scopus Database. The citation pattern of theses and dissertations indexed in the database is compared with the citation frequency of the repositories.The results show little correlation between the domain of 'theses' output and its cited records. As expected, Scopus and Google Scholar have no correlation regarding citation frequencies. The findings help the authors understand the repositories' impact and provide attention to the high-impact repository sites.

Keywords: Bibliometrics, Citation Analysis, Google Scholar, Search Engines, Shodhganga

1. Introduction

Authors and researchers aim to increase the visibility of their publications and initiate efforts to make them available at many knowledge storehouses. A few mechanisms such as Shodhganga, NDLTD and search engines such as Google Scholar index the theses based on the deposit by the universities. Some other repositories, such as NDLTD, DART-Europe, EThOS, warrant the authors to submit copies of the theses. Knowing which kinds of storehouses facilitate easy access and consequently lead to citing 'theses' is essential. The role of the archives, repositories, and search engines is unequal, and only a few have a significant role in enabling access. In this work, we intend to find the impact of citations to theses emanating through archives and search engines such as Shodhganga, NDLTD and others and compare them with Google Scholar citations.

2. Earlier work

Citations from theses and dissertations are much more quickly and comprehensively collected than are citations voluntarily supplied by faculty, as indicated by (Zipp, 1996)

To identify effective citation strategies in theses, Bojana_Petriæ compared citation strategies in high- and low-rated master's theses by classifying them into two levels, where subject specialists award grades

Corresponding Author: Vijayakumar P, Email: vijay.lib@pondiuni.edu.in and Vijayakumar M, Email: librarian@pondiuni.ac.in

(Petriæ, B., 2007). (Kushkowski, J. D et al., 2003). in a longitudinal study of over 9100 citations from 629 master's and doctoral theses found that the length of theses increased over time and the number of citations in thesis references differs from one to another discipline. We want to address this kind of correlation in this work; the citation frequency and pattern differ from one domain to another.

If the theses are indexed in repositories, it leads to increased visibility and use. The theses can reach a wider audience when they have an access nature. Ferreras-Fernández et al. focussed on the benefits of accessibility, dissemination, visibility and impact of Ph.D. e-theses deposited in repositories in their seminal paper (Ferreras-Fernández, T et.al.,2016) Google Scholar citation counts with Mendeley reader counts and found that Mendeley reader counts were more useful for impact assessment purposes for dissertations that are less than two years old, while Google Scholar citations were more useful for older dissertations, especially in social sciences, arts and humanities (Kousha & Thelwall, 2019).

García-Pérez (2010) found that Google Scholar retrieved the most significant number of citations for all four Spanish psychologists, followed by PsycINFO and WoS. However, Google Scholar also had the utmost percentage of incorrect citations. WoS had the lowest percentage of incorrect citations but the fewest citations. PsycINFO had an intermediary number of citations and incorrect citations.

3. Objectives

With the reasons highlighted, we fix the purpose of this work in the below ways.

- To understand the impact of 'theses' repositories on citations
- To determine the citation pattern of theses and dissertations indexed in the Scopus and Google Scholar Database.
- Analyse both the cited and citing theses and the parameters that influence the citation process; and
- Study the role of different 'theses' repositories and their impact on users.

4. Methods

Initially, we study how important 'theses' repositories, such as Shodhganga, ndtld and others, are organised and structured. We also measure the bibliographic data produced by them. Theses from Indian universities are indexed in Shodhganga, NDTLD and a few more. When they are indexed in different repositories, how do they have an impact on users who access and cite them? The access and download data are available only with the host, and the users may not be able to access the logs. The only open data available about use is the citations to the theses indexed. Using data sources such as Scopus, WoS or Dimensions is possible to use theses and citation impact.

The study is quantitative and uses the Bibliometrics citation analysis technique to evaluate references in the Scopus Database. The references cited were examined and organised using tabulation.

The citing data is analysed explicitly for the citations to the theses from different universities and subjects, and we measure the time of publication and the citing years.

5. Dataset

Shodhganga is the data source for this work, and the Scopus references contain 1610 records of Shodhganga. The thesis references are extracted and analysed in this work to track the use/citation pattern of the theses. If not indexed in Shodhganga, individual theses may not attract use and citations. The benefit of its indexing goes up as the indexing is carried out.

The data collection process for this study involved a two-step approach. In the initial step, we gathered bibliographic data from the theses. Initially, we determined the number of theses cited in Scopus. We conducted an advanced search using domain-specific criteria, resulting in approximately 1610 records. Subsequently, we refined these records to identify entries listed in Shodhganga ETD. Using the handle ID(URL) associated with the Scopus-listed records, we executed an additional search in Shodhganga, compiling bibliographic details such as URL, Title, Subject, University, Year of completion, and more for each thesis.

Moving on to the second step, we collected citation data for the selected records from Google Scholar. This dataset was then subjected to further analysis in alignment with our research objectives and subsequently presented in tabular formats.

6. Documents Indexed in Scopus

Various theses portals index and host the theses content and offer search facilities. Many users frequently access these portals to find the theses available on various domains. Not all 'theses' can cater for the requirements due to many issues. The predominant issues are quality and level besides relevance. A significant way to tap the quality content is to find how many are indexed in citation databases such as Scopus. By accessing Scopus. we have identified the number of titles available in it. Table 1 presents the number of source documents indexed in Scopus

Table1: Theses repositories and their records indexed

S.No	Thesis Portal	No. of source documents (Records) in Scopus
1	Networked Digital Library of Theses and Dissertations (Ndltd.org)	377
2	Shodhganga - A Reservoir of Indian Theses (Shodhganga.inflibnet.ac.in)	1610
3	DART-Europe E-theses Portal (dart-europe.eu)	77
4	EThOS - British Library Electronic Theses Online Services (https://ethos.bl.uk/Home.do)	1806
5	Open Access Theses and Dissertations(https://oatd.org/)	150
6	Thesis Commons (https://thesiscommons.org/)	18
7	National Academic Research and Collaborations Information System (NARCIS) of Netherlands (https://dans.knaw.nl/)	222
8	Worldwide ETD Index - Electronic Theses and Dissertations (https://etd.ohiolink.edu/)	3742

From Table 1, we can find the influence of ETD on Scopus than other theses portals. One reason could be its universal access and global content. The other two are the British Theses portal and Shodhganga, the Indian Theses portal.

Table 2: Top 10 Universities' Theses Cited in Scopus Database

S.No	Name of the University	No. of Theses Cited as from Scopus	Google Scholar Citations for (theses)
1.	Anna University	121	168
2.	Mahatma Gandhi University	53	77
3.	Cochin University of Science and Technology	41	229
4.	Pondicherry University	41	316
5.	University of Calcutta	39	55
6.	Aligarh Muslim University	38	58
7.	Gauhati University	38	69
8.	Punjabi University	37	177
9.	Savitribai Phule Pune University	35	77
10.	University of Mysore	33	59

6.1 Indian University theses which have an Impact on Citations

We tap the potential of specific university theses frequently cited in Scopus. This citation database has identified the theses most cited university are from Anna University. Other universities score less than half of Anna University. This data is reflected in Table 2. However, in Google Scholar, the theses from Pondicherry University are the most cited, followed by Cochin University of Science and Technology. The significance of Correlation Coefficient values between Scopus and Google Scholar stood as. 0.37with a p-value of 0.2883. The correlation is found to be less; however no negative correlation is observed.

Table 3: Top 10 Subjects theses most cited

S.No	Subject	No. of Theses Cited as from Scopus	Google Scholar Citations for (theses)
1.	Library and Information Science	98	136
2.	Economics	90	264
3.	Commerce	84	263
4.	Chemistry	81	98
5.	Management	81	400
6.	Computer Science	78	253
7.	Physics	71	111
8.	Electrical and Electronics Engineering	49	54
9.	Botany	41	180
10.	Electronics and Communication Engineering	36	139

This kind of analysis provides an understanding of domain-specific access and use. The use is measured in terms of citations. There is not much correlation between the domain of 'theses' output and its citations. A small discipline, Library and Information Science theses are most cited in Scopus, followed by Economics. Other subjects follow the suit, which are major domains.

Table 4. Citation frequency of Shodhganga Thesis (cited with more than four frequency)

S.No	Title of the Thesis	No. of Times Cited	Google Scholar Citations
1.	Unorganised manufacturing sector in India during postliberalisation period, Punjabi University	9	54
2.	Application of multi proxy tree ring parameters in the reconstruction of climate vis a vis glacial fluctuations from the eastern Himalaya, University of Lucknow	8	100
3.	Optimal, power dispatch and pricing for deregulated power industry, Pondicherry University	7	112
4.	Investigations on power system operation and management in restructured market, Pondicherry University	7	42
5.	Use of library classification schemes in the ICT environment in selected libraries in national capital region a study, Savitribai Phule Pune University	6	0
6.	Development of titanium nanotube Arrays for orthopaedic applications, Anna University	6	60
7.	Clinical librarianship in Kolkata with special reference to private hospitals, University of Calcutta	6	55
8.	An analysis of financial performance of state road transport corporation in Gujarat, Saurashtra University	6	295
9.	Computer studies of silicon carbide gallium nitride and indium phosphide based IMPATT devices operating in MM wave and terahertz region and corresponding studies on the photo sensitivity of the devices, University of Calcutta	5	15
10.	Study of institutions of social work education and role of social work educators in developing indigenous knowledge, Tata Institute of Social Sciences	5	25
11.	Efficient analysis of satellite image denoising and resolution enhancement for improving classification accuracy, Anna University	5	40

12.	Design and development of cluster algorithms for power system problems, Jawaharlal Nehru Technological University	5	5
13.	Studies on Litter Production Decomposition and Nutrient Recycling in Tea Agro Ecosystem of Cachar District Southern Assam, Assam University	5	25
14.	Scalable and enhanced remote services for power system reliability estimation, Anna University	4	0
15.	Wavelet Image Fusion Approach For Classification of Ultrasound Placenta Complicated By Gestational Diabetes Mellitus, Mother Teresa Womens University	4	0
16.	Feasibility studies on the removal of iron and fluoride from aqueous solution by adsorption using agro based waste materials, Anna University	4	16
17.	Influence of additives on the characteristics of Stone Matrix Asphalt, Cochin University of Science and Technology	4	0
18.	Synthesis, characterization and anticancer evaluation of some new Triazole derivatives, Manipal University	4	4

A few theses are more frequently cited from the Shodhganga repository and are named with their corresponding title in Table 4. Eighteen theses are cited three or more than four times in Scopus and are listed in the decreasing order of citation frequency. The corresponding Google Scholar citations are also presented in the third column in the table. There is no correlation, as expected, between Scopus and Google Scholar regarding citation frequency. The citation frequency in Google Scholar is uneven, and some theses are not indexed.

Table 5. Citation Frequency in a long period window

Year	No. of Theses	Google Scholar Citations for theses
1963	1	2
1965	1	3
1967	2	0
1969	1	2
1971	1	3
1972	1	0
1975	1	0
1977	1	1
1979	1	0
1982	1	0
1984	1	0
1985	4	5
1988	2	2
1989	2	24
1990	2	5
1991	5	5
1992	3	7
1993	5	10
1994	8	55
1995	8	5
1996	13	18

1997	8	10
1998	11	15
1999	7	11
2000	13	46
2001	12	10
2002	16	56
2003	17	23
2004	23	53
2005	24	142
2006	27	47
2007	41	50
2008	39	52
2009	52	130

2010	124	763
2011	176	630
2012	179	289
2013	198	402
2014	144	367
2015	107	99
2016	78	81
2017	52	13
2018	28	42
2019	18	41
2020	6	3
2021	1	1
2022	1	0

The number of citations recorded in different years for the theses is presented chronologically in Table 5 and Figure 1. The period between 2010-2015 has recorded more citations than the rest. The skewness in citation frequency may be due to the uneven coverage of the 'theses'.

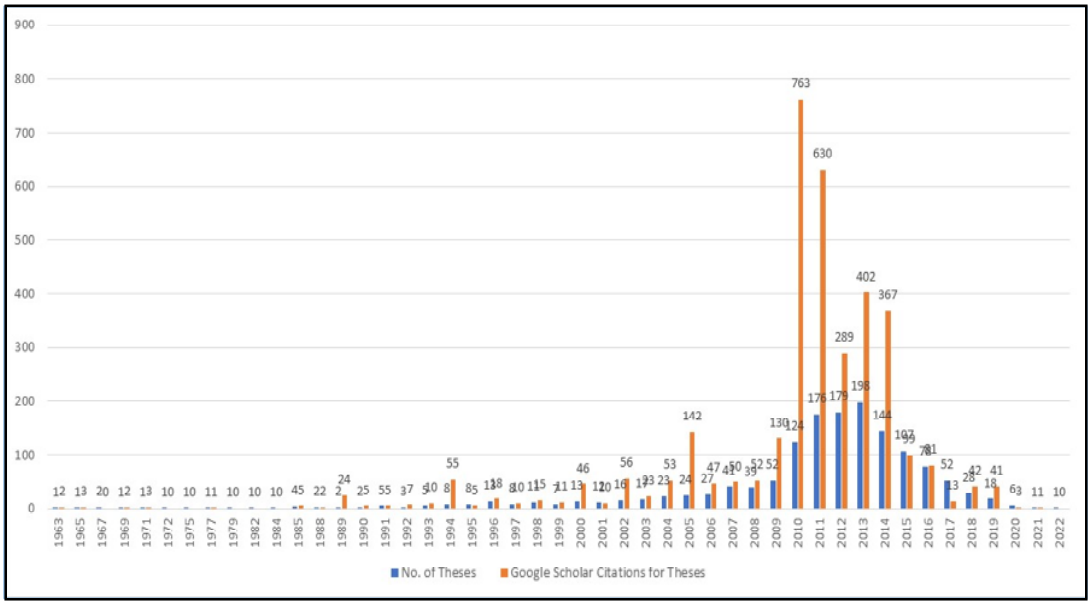

Figure 1: Citation distribution in periods

Table 6: Top 10 highly cited Theses

S.No	Theses	Google Scholar Citations for theses
1.	Impact of dividend policy on shareholders' value: a study of Indian firms, Kapoor, Sujata ,Jaypee Institute of Information Technology, 2010.	100
2.	An analysis of financial performance of state road transport corporation in Gujarat, Trivedi, Shilpa M ,Saurashtra University, 2010.	59
3.	Effective learning and classification using random forest algorithm, Kulkarni, Vrushali Yadunath, Savitribai Phule Pune University, 2014.	55
4.	Coal mining and its impact on environment of Nokrek biosphere reserve, Meghalaya, Sarma, Kiranmay , North-Eastern Hill University, 2002.	33
5.	Successful delivery of public- private partnerships for Infrastructure development, Agrawal, Ranjan, Jaypee Institute of Information Technology, 2011.	33
6.	Entrepreneurial Competency: a study with reference to socially and economically backward communities in Chennai City, Kochadai, M, Pondicherry University, 2011.	31
7.	Impacts of climate variability on agriculture in Kerala, Gopakumar, C S, Cochin University of Science and Technology, 2013.	30
8.	Impact of microfinance on poverty, employment and women empowerment in rural Punjab, Bansal, Deepty, Punjabi University, 2010.	30
9.	Impact of Information Communication Technology (ICT) on professional development and educational needs of library professionals in the universities of Kerala, Mathew, K S, Cochin University of Science and Technology, 2011.	29
10.	A study of productivity and financial efficiency of textile industry of India, Zala, Virambhai S , Saurashtra University, 2010.	26

The most highly cited theses in Google Scholar are listed in Table 6. The thesis with the most citations, "Impact of dividend policy on shareholders' value: a study of Indian firms" by Kapoor, Sujata, Jaypee Institute of Information Technology, has 100 citations. The second most cited thesis is "An analysis of financial performance of state road transport corporation in Gujarat" by Trivedi, Shilpa M, Saurashtra University, with 59 citations.

The theses contain valuable content for academic and research growth. The exploitation of 'theses' is lesser than other forms of communication such as journal and conference papers. Indexing of 'theses' is restricted

to particular types of indexes. Not all are evenly indexed, and all are indexed. At the same time, 'theses' have uneven quality, leading to the skewed nature of its exploitation.

7. Conclusion

The findings help the authors understand the repositories' impact and provide attention to the high-impact repositories. The citation to the theses hosted in the repositories has an impact based on their availability and users' awareness of them. All repositories are not equal in visibility and impact. At the same time, the impact of the theses repositories is less when we compare with the other archives such as cornhill, eprints and others which as popular among the end-users

References

Ferreras-Fernández, T., García-Peñalvo, F., Merlo-Vega, J. A., & Martín-Rodero, H. (2016). Providing open access to PhD theses: visibility and citation benefits. Program, 50(4), 399–416. https://doi.org/10.1108/prog-04-2016-0039

García-Pérez, M. A. (2010). Accuracy and completeness of publication and citation records in the Web of Science, PsycINFO, and Google Scholar: A case study for the computation of h indices in Psychology. Journal of the American Society for Information Science and Technology, 61(10), 2070–2085. https://doi.org/10.1002/asi.21372

Kousha, K., & Thelwall, M. (2019). Can Google Scholar and Mendeley help to assess the scholarly impacts of dissertations? Journal of Informetrics, 13(2), 467–484. https://doi.org/10.1016/j.joi.2019.02.009

Kushkowski, J. D., Parsons, K. A., & Wiese, W. H. (2003). Master's and doctoral thesis citations: Analysis and trends of a longitudinal study. portal: Libraries and the Academy, 3(3), 459-479.3

Petriæ, B. (2007). Rhetorical functions of citations in high- and low-rated master's theses. Journal of English for Academic Purposes, 6(3), 238–253. https://doi.org/10.1016/j.jeap.2007.09.002

Scopus. (2023). Scopus. Retrieved September 1, 2023, from https://www.scopus.com

Shodhganga : a reservoir of Indian theses @ INFLIBNET. (2023). Inflibnet.ac.in. https://shodhganga.inflibnet.ac.in

Zipp, L. S. (1996). Thesis and Dissertation Citations as Indicators of Faculty Research Use of University Library Journal Collections. Library Resources & Technical Services, 40(4), 335–342. https://doi.org/10.5860/lrts.40n4.335