

Growth and Utilisation of Shodhganga

Ellora Barman¹, S. Ravikumar² and Bidyut Bikash Borauh³

¹Phd Research Scholar, North Eastern Hill University (NEHU),
Shillong, Meghalaya, India

²Associate professor, North eastern Hill University (NEHU), Shillong,
Meghalaya, India

³Phd Research Scholar, North Eastern Hill University (NEHU), Shillong,
Meghalaya, India

Abstract

The paradigm shift from print to digital has revolutionised the academic sphere, with electronic theses and dissertations (ETDs) gaining prominence. Shodhganga, a digital repository launched in 2009 by the University Grants Commission (UGC) of India, has emerged as a pivotal platform for archiving and accessing Indian theses and dissertations.

The study's objectives encompass comprehending Shodhganga's usage patterns, global and regional engagement, yearly update trends, and text analysis of titles using Latent Dirichlet Allocation (LDA). The Google Trends analysis revealed a rising interest in Shodhganga, attributed to growing online research, Indian student enrolment, and enhanced internet accessibility. The platform's popularity was underscored by global interest, with India, Nepal, and Pakistan leading and scholars from diverse countries seeking its resources.

Annual updating trends disclosed an escalating volume of contributions since 2010, peaking in 2019, indicating the repository's evolving prominence. Employing LDA, the study unveiled salient topics like women's empowerment, rural development, education, inorganic chemistry, and technology. This technique illuminated thematic growth and facilitated efficient information retrieval within Shodhganga's extensive content.

In an era of information deluge, the study highlights the prowess of text mining techniques, especially LDA, in extracting meaningful insights from voluminous text data. Shodhganga's role in fostering research collaboration, knowledge dissemination, and plagiarism detection was underscored, establishing its pivotal position in academia. As the repository continues to expand, it stands to significantly impact diverse research domains, serving as a comprehensive repository of Indian scholarly output. The article underscores the transformative role of digital repositories like Shodhganga in shaping the trajectory of research and education in the digital age.

Keywords: Electronic Theses and Dissertations (ETD), Latent Dirichlet Allocation (LDA), Shodhganga

Corresponding Author: Ellora Barman, E-mail: missellora@gmail.com, Dr S. Ravikumar, E-mail: ravikumarnehu@gmail.com and Bidyut Bikash Borauh, E-mail: boruahofficial2017@gmail.com

1. Introduction

An electronic thesis and dissertation (ETD) is a digital version of a thesis or dissertation submitted and stored electronically rather than in print. ETDs are becoming increasingly popular in universities and colleges in the United States and abroad. They are openly accessible to the public through online repositories managed by the respective institutions' libraries (Electronic et al. | The Graduate School, 2023). Dissertations (ETDs) - Sheridan Libraries, 2023) The Graduate School and the library typically work together to facilitate the transition to ETDs. The ETD website provides instructions and guidelines for formatting and submitting the final documents electronically.

Shodhganga is a digital repository of theses and dissertations submitted to universities in India (Shodhganga: A Reservoir of Indian Theses @ INFLIBNET, 2023), (Shodhganga - Wikiwand, 2023). It was launched by the University Grants Commission (UGC) in 2009 and is managed by the Information and Library Network (INFLIBNET) Centre, an autonomous inter-university centre of the UGC of India (What Is Shodhganga? A Comprehensive Guide (2023) - Worthleaf.Com, 2023). Shodhganga means "research and discovery" in Sanskrit. The repository has a collection of over 30,000 theses and 8,000 synopses (Shodhganga: A Reservoir of Indian Theses @ INFLIBNET, 2023). It provides a platform for research students to deposit their PhD theses and make them available to the entire scholarly community in open access (What Is Shodhganga? A Comprehensive Guide (2023) - Worthleaf.Com, 2023). Shodhganga allows researchers, academicians, and students across India to search, view, and download research papers, ideas, and dissertations from different subject areas (Electronic Theses & Dissertations, 2023). It also provides a platform for universities to deposit their ideas, creating a centralized and comprehensive database of Indian research output. The full text of all the documents submitted to Shodhganga is available to read and download in open access to the academic community worldwide (Kaur & Verma, 2022). The repository can capture, index, store, disseminate, and preserve ETDs submitted by the researchers (Electronic Theses & Dissertations, 2023). Shodhganga is a valuable resource for researchers, students, and academicians to access the latest research and findings in their respective fields. It is also a valuable tool for plagiarism detection, as the Ministry of Education, Government of India has initiated a program called "ShodhShuddhi" which provides access to Plagiarism Detection Software (PDS) to all universities/institutions in India since September 1, 2019 (What Is Shodhganga? A Comprehensive Guide (2023) - Worthleaf.Com, 2023), (Sinha & Purkayastha, 2018). In this age of information explosion, lots of documents in every branch of the universe of knowledge are available for information seekers. With increased text data, manual literature review and traditional information retrieval techniques are becoming challenging and time-consuming (Suresh et al., 2014; Asmussen & Møller, 2019). Text mining is a technique to identify valuable patterns from the unstructured text of available literature through automatic analysis (Gohil, 2015; Ravikumar et al., 2023). Latent Dirichlet Allocation (LDA) is the widespread text mining method that assumes each document is a mixture of topics and each topic is a mixture of words. It is a probabilistic model that represents topics by word probabilities. The words with the highest probabilities in each topic usually give a good idea about the topic (Jelodar et al., 2019).

2. Objective

The main objectives of the paper is as follows –

- ❖ To know the shodhganga uses pattern since its inception.
- ❖ To know what the regions are involved worldwide and within India.
- ❖ To know yearly update of theses and dissertations.
- ❖ Text analysis of the titles to understand the growth of topics, topic modelling of the documents, publication proportion of the topics.

3. Significance

The significance of this study lies in its comprehensive exploration of the electronic theses and dissertations (ETDs) landscape, with a focus on Shodhganga, a prominent digital repository in India. The study's findings carry several noteworthy implications:

- 3.1 Academic Advancement and Collaboration:** By shedding light on Shodhganga's usage patterns, the study showcases the repository's increasing popularity among researchers, students, and academicians. This highlights the importance of digital platforms in facilitating academic progress and fostering collaborations within the scholarly community.
- 3.2 Global Reach and Impact:** Global and regional engagement analysis underscores Shodhganga's international appeal, with scholars from various countries accessing its resources. This reveals the repository's role in connecting researchers worldwide, enhancing the global impact of Indian research, and promoting cross-cultural academic exchange.
- 3.3 Digital Transformation of Research:** The study's exploration of yearly updating trends reflects the ongoing digital transformation of academic research. The continuous increase in uploaded theses and dissertations signifies the growing acceptance and adoption of digital formats, which can streamline research dissemination and accessibility.
- 3.4 Efficient Information Retrieval:** The text analysis using Latent Dirichlet Allocation (LDA) technique contributes to the efficiency of information retrieval within Shodhganga's vast repository. Researchers can now navigate and explore specific topics more effectively, saving time and enhancing the quality of their research.

Thematic Insights: Identifying key thematic areas through LDA, such as women's empowerment, rural development, and technology, provides insights into the prevailing research trends. This information aids researchers in understanding the evolving academic landscape and directing their research endeavours accordingly.

3.5 Plagiarism Detection and Quality Control: Shodhganga's role in plagiarism detection through the "ShodhShuddhi" program demonstrates its commitment to maintaining research integrity and quality. This study highlights the repository's proactive approach to fostering ethical research practices and maintaining the credibility of academic work.

3.6 Contribution to Academic Literature: The study adds to the academic discourse by comprehensively analysing Shodhganga's impact, usage, and trends. Its findings can serve as a reference for scholars, institutions, and policymakers interested in understanding the evolving landscape of digital repositories and their implications for education and research.

3.7 Future Research and Development: The study sets the stage for further research into digital repositories, text mining, and the dynamics of ETDs. It encourages scholars to delve deeper into understanding user behaviour, technological advancements, and the evolving role of such platforms in academia.

Overall, this study's significance lies in its ability to offer insights into the transformative role of digital repositories like Shodhganga in shaping the present and future of research, education, collaboration, and knowledge dissemination.

4. Scope

The scope of this study is multifaceted and encompasses various dimensions of electronic theses and dissertations (ETDs), with a specific focus on the Shodhganga digital repository. The study's scope can be outlined as follows:

4.1 Usage Patterns Analysis: The study delves into understanding the usage patterns of Shodhganga since its inception. It aims to uncover trends in user engagement, exploring how the platform's popularity has evolved. This analysis provides insights into the changing landscape of academic research and the increasing adoption of digital repositories.

4.2 Global and Regional Engagement: By examining the global and regional interest in Shodhganga, the study broadens its scope to assess the repository's impact beyond India. It identifies the countries and regions where the platform has gained prominence, contributing to cross-border collaborations and information dissemination.

4.3 Yearly Update Trends: The study investigates the yearly update trends of theses and dissertations on Shodhganga. This scope reflects the ongoing transition from print to digital formats in academic work and reveals insights into the growth trajectory of ETDs in the digital age.

4.4 Text Analysis and Topic Modeling: Employing the Latent Dirichlet Allocation (LDA) technique, the study delves into the content of ETD titles to identify prevalent topics. This scope allows for a deeper understanding of the thematic landscape of Shodhganga, enabling researchers to navigate its content more efficiently.

4.5 Implications for Research and Education: The study's scope extends to exploring the implications of Shodhganga's role in research and education. It highlights how digital repositories contribute to

research collaboration, knowledge dissemination, and plagiarism detection, thereby impacting the quality and credibility of academic work.

4.6 Technological Advancements: The study indirectly touches on the technological advancements driving the digital transformation of academic resources. It acknowledges how platforms like Shodhganga utilize advanced techniques such as text mining and data analysis to enhance the accessibility and utility of their content.

4.7 Future Research Avenues: The study's scope extends to stimulating further research in areas related to digital repositories, text mining, and the analysis of scholarly content. It allows researchers to explore this field's user behaviour, content preferences, and technological innovations.

4.8 Policy Implications: The study's findings could affect educational and research institutions' policies. Insights into trends and preferences can inform decisions about resource allocation, technology integration, and content management.

In summary, the scope of this study encompasses a broad range of topics within the realm of ETDs and digital repositories, with a central focus on Shodhganga. It provides insights into user behaviour, trends, and the implications of digital resources for research and education while paving the way for future research endeavours in related domains.

5. Methodology

Google Trends has been used to evaluate the trends of Shodhganga use patterns, local and global use, and related topics and queries. The data from google trends are available online from the google trends website, which is a part of Google. The data for yearly website updates was extracted from Shodhganga with an advanced search option. In the search option, we typed all of Shodhganga; in the "for" option "theses" was the keyword we put, and the filter used was "date completed 2000 till 2023." The result showed 38427 within 0.11 seconds. To perform the text analysis, the Latent Dirichlet Allocation (LDA) method of text mining is applied. Preprocessing is done before applying LDA on the titles of the ETD, i.e. tokenisation, filtering and text cleanup (like text normalisation, stop words removal, etc.) to eliminate some irrelevant textual information in the text mining process. The diagrams have been generated using R software.

6. Analysis and Results

First objective was to know the shodhganga uses pattern since its inception. From google trends analysis, we can see that the term Shodhganga has been in Shodhganga has been growing popular in recent years with the search term shodhganga on the rise since 2010. This is likely due to the increased popularity of online research, the growing number of Indian students pursuing higher education and the increasing availability of high-speed internet in India. Figure 1 shows the term shodhganga was searching in google through google trends in graphical manner.

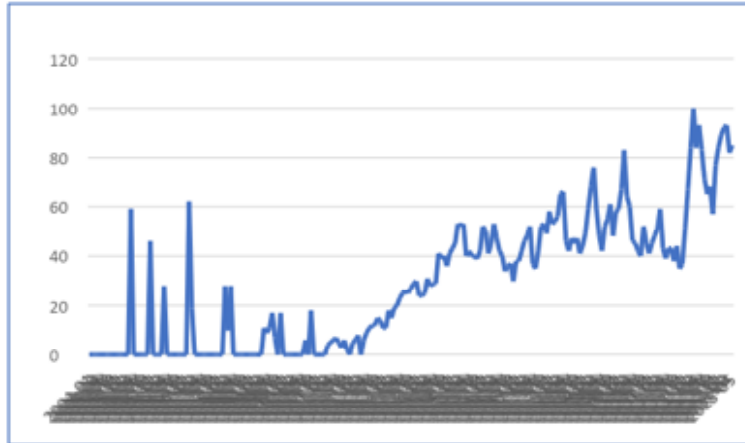


Figure 1: Popularity of Shodhganga since its inception

Second objective of the study was to know what the regions are involved worldwide and within India. Google Trends showed that the Shodhganga is most popular in India, followed by Nepal and Pakistan. Research scholars from the United States, Spain, and the United Kingdom visited shodhganga more frequently than scholars from other countries. China was the country with the least number of visitors. Figure 2 shows the global interest, and Figure 3 shows the regional interest of shodhganga.



Figure 2: Global interest of Shodhganga

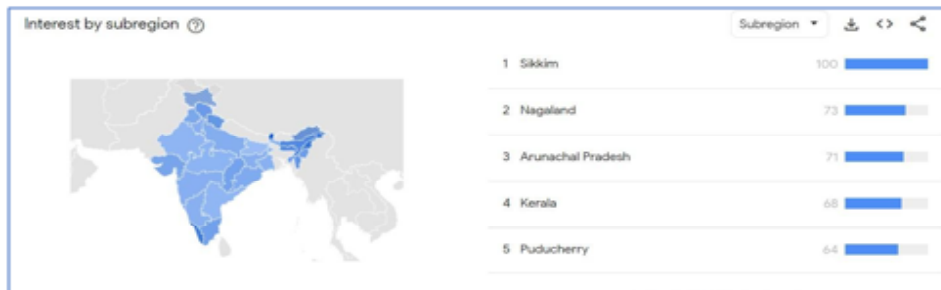


Figure 3: Regional interest of Shodhganga

The most popular topic searched on shodhganga are education, science, and technology. and most popular searched topic searched on shodhganga are engineering, medicine and management

Third objective was to know the yearly updating of theses and dissertations. This was achieved through shodhganga advanced search, which it showed 38427 theses had been updated till 2023. The number of uploaded theses has been growing since 2010, peaking in 2019. we have calculated the yearly updating of the thesis in shodhganga as it was searched from the shodhganga website. Figure 4 shows the increased year wise theses updates peaking at 2019.

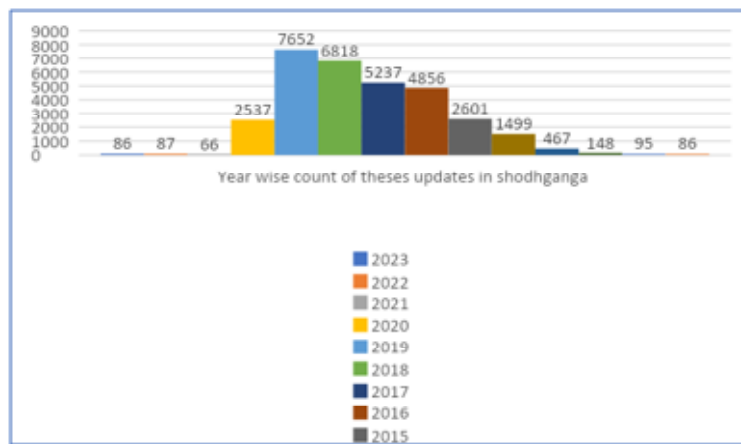


Figure 4: Year wise theses updates in Shodhganga

Forth objective was text analysis of the titles to understand the growth of the topics.

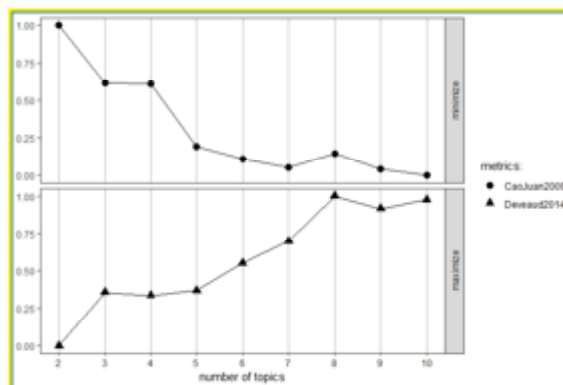


Figure 5: Number of topics

LDA uses statistical associations of words in a text to generate latent topics (Hannigan et al., 2019), and the number of topics (K) is different for different corpus of text or documents based on the LDA metrics. At least one of the measurements, either the best k or a k that is a decent compromise between the measurements,

can be taken. In Figure 1, two measurements reliably improve and most likely have their optimal point at $k > 10$ (CaoJuan, 2009; Deveaud, 2014).

LDA consider a document as a mixture of the topic and a topic as a mixture of words based on their presence and weightage in the content. In the mining process, all the texts are filters and trace how words and expressions co-happen in them and naturally “learn” gatherings. Such words address a coherent theme or topic. These words are attached to each topic based on hyperparameters alpha and beta values of each term in the corpus. Alpha means the density of the document’s topic, and Beta means the density of words with the topic. The higher the alpha value, the documents consist of more topics, and the lower the alpha value, the fewer topics. On the other hand, the higher the Beta, the more corpus items are composed of many words, and the lower the beta value, it consists of several terms (Srinivasa-Desikan, 2018, Ravikumar et al., 2023). Figure 2 shows how the LDA identifies topics from the corpus of text formed by the titles of the ETD.



Figure 6: Topic modeling of the documents

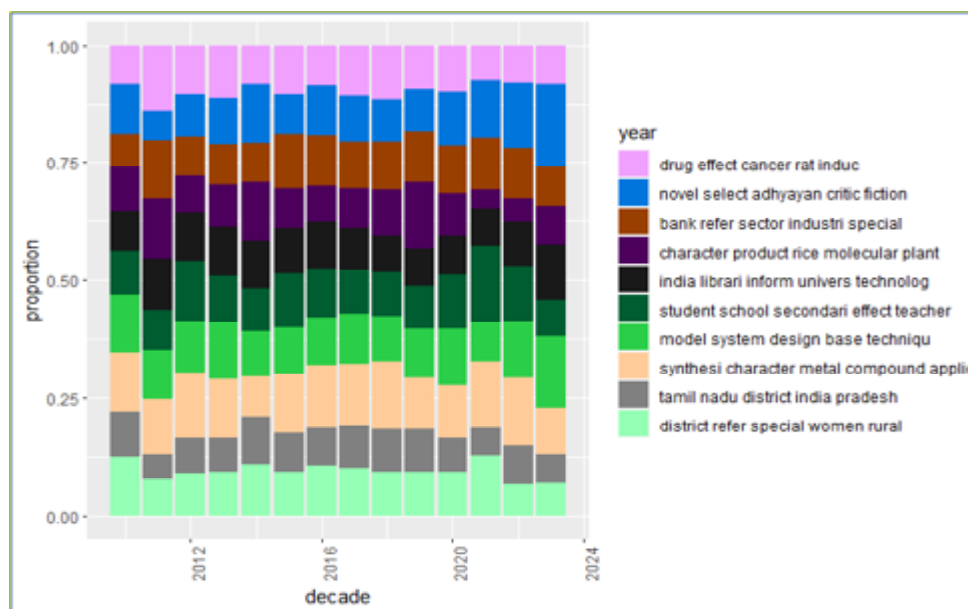


Figure 7: Publication proportion of topics from 2010 to 2023

Ten prominent topics identified from the text analysis of the ETD titles are- ‘Women empowerment’ and ‘rural development’ under the broad topic of Economics, study on South Indian states, Inorganic chemistry; ‘Education’; ‘Technology’; ‘Library and Information Science’ and ‘Indian Information System’; ‘Agriculture’ and ‘Botany’; ‘Literature’ and Medical Science. In figure 3, mean topic proportions per decade for the ETD titles are aggregated and presented through a bar plot. The figure shows the topics that dominate the publication proportion in the interval. From 2010 to 2023, publications on ‘inorganic chemistry’ slightly dominated over the other topics; however, the publication proportion of the topics in the time interval under study does not differ much. From 2020 onwards, there is a significant growth in the publication of Literature related ETD.

7. Discussions

7.1 Usage Pattern and Online Research: The increasing popularity of “Shodhganga” since 2010 reflects the changing landscape of academic research. The surge in online research can be attributed to factors such as the proliferation of digital resources, ease of access, and the availability of high-speed internet. This trend underscores the shift towards digital platforms for academic information seeking, which offers convenience and broader access to scholarly content.

7.2 Global and Regional Engagement: The study’s revelation that Shodhganga is most prevalent in India, followed by Nepal and Pakistan, aligns with geographic proximity and shared linguistic and cultural ties. However, the interest from countries like the United States, Spain, and the United Kingdom

demonstrates the repository's international relevance. This global engagement opens opportunities for cross-border collaborations and the exchange of research ideas.

- 7.3 Digital Transformation and Updating Trends:** The consistent growth in yearly updates of theses and dissertations on Shodhganga signifies the ongoing digital transformation of academic content. This trend reflects academia's increasing embrace of digital formats, which enhances content accessibility, preservation, and dissemination. The peak in updates in 2019 suggests a pivotal moment in the repository's expansion, likely due to a confluence of factors, including increased awareness and institutional support.
- 7.4 Thematic Evolution and Topic Modelling:** The successful application of the Latent Dirichlet Allocation (LDA) technique in understanding thematic growth is noteworthy. Identifying key topics like women's empowerment, rural development, education, and technology underscores the repository's relevance to diverse academic disciplines. This analysis aids researchers in identifying prevalent research trends and informs them about topics gaining traction over time.
- 7.5 Research Focus and Emerging Topics:** The prominence of "inorganic chemistry" in publication proportion emphasizes the sustained interest in specific disciplines. Additionally, the growth in Literature-related publications from 2020 onwards points to emerging research areas. This insight can guide researchers in making informed decisions about their research directions and topic choices.
- 7.6 Academic Collaboration and Plagiarism Detection:** Shodhganga's broad accessibility and focus on plagiarism detection contribute to maintaining research integrity. The repository's role as a platform for scholars worldwide encourages collaboration, knowledge sharing, and cross-pollination of ideas.
- 7.7 Educational Policy and Digital Literacy:** The increasing reliance on digital repositories like Shodhganga underscores the importance of digital literacy in academia. Educational institutions and policymakers can use these insights to develop strategies that support the effective utilization of digital resources in teaching and research.
- 7.8 Future Research and Technological Integration:** The study's success in employing LDA for topic analysis opens avenues for future research in text mining and natural language processing techniques. As digital repositories evolve, further research can explore advanced methods for enhancing content discoverability and usability.

The study's results demonstrate the evolving landscape of academic research driven by digital transformation, accessibility, and international collaboration. Shodhganga's role as a facilitator of knowledge dissemination, research collaboration, and content preservation highlights the transformative power of digital repositories in shaping the academic ecosystem. The insights garnered from this study provide valuable guidance for researchers, institutions, and policymakers as they navigate the digital age of scholarship.

Limitations

Its heavy reliance on data from Google Trends and Shodhganga's database could introduce biases by omitting individuals less engaged online or offering an incomplete representation of user behaviour. Moreover, while Google Trends provides relative interest scores, it lacks exact figures and could be influenced by external events or seasonal variations. Using LDA for topic modeling, the application of text analysis is subject to data quality and parameter selection, potentially leading to subjective topic definitions. Technological advancements in text mining beyond those utilized might impact the accuracy of the study's findings.

8. Conclusion

In conclusion, the evolution of electronic theses and dissertations (ETDs) has brought a transformative shift in the academic landscape. The emergence of platforms like Shodhganga, a digital repository of theses and dissertations in India, has further catalyzed this change by providing an invaluable resource for researchers, students, and academicians alike. The significance of Shodhganga lies in its role as a repository and as a tool for promoting research collaboration, knowledge dissemination, and plagiarism detection.

The objectives of this paper are effectively met through a meticulous analysis using various methodologies. The first objective, understanding the usage pattern of Shodhganga, revealed its growing popularity since its inception, driven by factors such as the surge in online research, increased Indian student enrollment in higher education, and improved internet accessibility. The second objective highlighted the global and regional interest in Shodhganga, with India leading the pack, followed by Nepal, Pakistan, and international interest from countries like the United States, Spain, and the United Kingdom.

Furthermore, the study successfully examined the yearly updating trends of theses and dissertations on Shodhganga. It showcased a continuous increase in uploads, peaking in 2019, correlating with the ongoing digitalization of academic work and the repository's growing prominence. The fourth objective, focusing on text analysis through Latent Dirichlet Allocation (LDA), enabled a deeper understanding of the growth of topics within the repository. This technique illuminated key thematic areas like women's empowerment, rural development, inorganic chemistry, education, technology, etc.

In a time of information overload, effective information retrieval and analysis are paramount. This study has demonstrated the power of text mining techniques, like LDA, in unravelling the underlying themes within vast volumes of textual data. By identifying significant topics, the study aids researchers and scholars in navigating the repository's content efficiently and extracting meaningful insights.

Shodhganga's role as a hub for academic exploration, knowledge sharing, and plagiarism prevention remains vital in pursuing quality research. As the repository continues to grow, it has the potential to contribute significantly to the advancement of various fields of study, serving as a central repository of Indian research output. This article sheds light on the importance of such platforms in fostering academic progress, facilitating

interdisciplinary research, and fostering global collaborations. As the academic landscape evolves, Shodhganga stands as a testament to the power of digital repositories in shaping the future of research and education.

Reference

- Asmussen, C. B., & Møller, C. (2019). Smart literature review: a practical topic modeling approach to exploratory literature review. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0255-7>
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72(7–9), 1775–1781. <https://doi.org/10.1016/j.neucom.2008.06.011>
- Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective Latent Concept Modeling for ad hoc information retrieval. *Document Numerique*, 17(1), 61–84. <https://doi.org/10.3166/dn.17.1.61-84>
- Electronic Theses & Dissertations (ETDs). (2023). Sheridan Libraries. Retrieved April 1, 2023, from <https://www.library.jhu.edu/library-services/electronic-theses-dissertations/>
- Electronic Theses & Dissertations (ETDs)—Sheridan Libraries. (2023). Retrieved April 1, 2023, from <https://www.library.jhu.edu/library-services/electronic-theses-dissertations/>
- Electronic Theses and Dissertations | The Graduate School. (2023). Retrieved April 1, 2023, from <https://grad.ncsu.edu/students/etd/>
- Gohil, L. (2015). Text Mining: Process and Techniques. *International Journal of Innovative Research in Computer Science and Technology*, 3, 70–72.
- Hannigan, T., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V., Wang, M. & Jennings, P. D. (2019). Topic modeling in management research: Rendering new theory from textual data. *Academy of Management Annals*. <https://doi.org/10.5465/annals.2017.0099>
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. In *Multimedia Tools and Applications* (Vol. 78, Issue 11). <https://doi.org/10.1007/s11042-018-6894-4>
- Kaur, M., & Verma, N. (2022). Citation analysis of doctoral thesis related to citation analysis from the Maharashtra region available at Shodhganga. *International Journal of Information Dissemination and Technology*, 12(1), 37–42. <https://doi.org/10.5958/2249-5576.2022.00008.5>
- Ravikumar, S., Boruah, B. B., & Gayang, F. L. (2023). Latent topics identification from the articles of Sri Lankan authors using LDA. *Global Knowledge, Memory and Communication*. <https://doi.org/10.1108/GKMC-08-2022-0206>

Srinivasa-Desikan, B. (2018) *Natural Language Processing and Computational Linguistics: A Practical Guide to Text Analysis with Python, Gensim, SpaCy and Keras*. Packt Publishing Ltd., Birmingham.

Suresh, M., Head, &, Ali, M. A. A., & Rao, M. A. S. (2014). A Study on Information Retrieval Methods in Text Mining. *International Journal of Engineering Research and Technology*, 184–190.

Shodhganga: A reservoir of Indian theses @ INFLIBNET. (2023). Retrieved April 1, 2023, from <https://shodhganga.inflibnet.ac.in/>

Shodhganga—Wikiwand. (2023). Retrieved April 1, 2023, from <https://www.wikiwand.com/en/Shodhganga>

Sinha, M. K., & Purkayastha, N. (2018). Awareness and Use of Electronic Theses and Dissertations (ETD) with Special Reference to Shodhganga and Shodhgangotri of INFLIBNET: A Study of Scientific Community Library Users of Assam University, Silchar. 2018 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (ETTLIS), 240–247. <https://doi.org/10.1109/ETTLIS.2018.8485256>

What is Shodhganga? A Comprehensive Guide (2023)—Worthleaf.com. (2023). Retrieved April 1, 2023, from <https://worthleaf.com/shodhganga/>