Tracking the Evolution of Research Themes in CALIBER Conferences: A Topic Modelling Approach Using LDA

Subhayan Chakraborty **Manoj Kumar Verma**

Abstract

The Convention on Automation of Libraries in Education and Research Institutions (CALIBER) has made a significant contribution to Indian Library and Information Science research and dissemination. The present study aims to analyze the evolving landscape of LIS research presented at five CALIBER conferences (2013, 2015, 2017, 2019, and 2022). We apply LDA (Latent Dirichlet Allocation) to identify key latent research topics from full-text papers published in these conferences. The study also uncovers trends and thematic evolution in those research topics over time. The results reveal a steady research focus on areas such as 'Open-access and Academic Publishing' and 'Use of Library Resources'. It also highlights some emerging topics, e.g., 'Technology-Enhanced Library Services'. The study highlights underexplored areas and suggests directions for future LIS research in India. The paper will benefit future researchers and LIS professionals in understanding the LIS research landscape in India.

Keywords: CALIBER, Latent Dirichlet Allocation, Research Trends, Scholarly Communication, Text Mining, Topic Modelling

1. Introduction

Library and Information Science (LIS) is a dynamic academic field. It continuously evolves to meet the information needs of society. Research in Library and Information Science (LIS) contributes to the development of society, which is driven by information. LIS as a research domain has been growing exponentially in recent years. With the advancements in the digital age, LIS research focus has also shifted from traditional library-focused topics towards areas related to digital technology, social media, and data science (Elia & Mhando, 2024). Academic Conferences are important in sharing innovations, best practices, and academic research among LIS professionals and researchers. Academic conferences provide significant scientific and social benefits for researchers (Hauss, 2021). LIS conferences assist in shaping the research landscape and the emergence and evolution of research topics in LIS (Harrison, 2010). In India, the CALIBER conference, organized by the INFLIBNET Centre, has been a major event in the LIS community.

Corresponding Author
Subhayan Chakraborty
Email: subhayan@drtc.isibang.ac.in

1.1 About the Selected Conference

Convention on Automation of Libraries in Education and Research Institutions (CALIBER) is organised by the INFLIBNET Centre, India. It is a biennial convention that was started in 1994. Since then, INFLIBNET has been organising CALIBER every two years in different parts of the country. It is a highly reputable platform that provides maximum opportunities to Indian LIS researchers to share their research work. So far, 13 conventions have been organized in collaboration with various universities (Gandhinagar, n.d.).

Table 1 :Selected CALIBER Conferences

CALIBER Year	Conference Theme	Location	Number of Published Papers
2013	"Library Vision 2020: MovingTowards the Future"	Gandhinagar, Gujarat	51
2015	"Innovative Librarianship: Adapting to Digital Realities."	Shimla, HP	57
2017	"Re-Envisioning Role of Libraries: Transforming Scholarly Communication"	Chennai, Tamil Nadu	53
2019	"Library 2030: Moving Towards Smart Technologies, Services and Resources"	Bhubaneswar, Odisha	46
2022	"Envisioning Digital Transformation in Libraries for NextGen Academic Landscape"	Varanasi, UP	45

This study examines the topics discussed in the five latest CALIBER conference papers from 2013 to 2022. Table 1 lists five selected CALIBER conferences with their respective themes and the number of papers published. The purpose is to identify the major research topics of these conferences and to analyse how the research focus has changed over the years. The present study employs topic modelling to evaluate the major research topics. It is classified as a probabilistic text mining technique, and it seeks to uncover latent and acute topics in a collection of textual documents or a corpus (Kochedykov et al., 2017).

1.2 Objectives of the Study

- To identify the top twenty research topics presented in the five CALIBER conferences (2013, 2015, 2017, 2019, and 2022) using topic modelling.
- To analyze the year-wise evolution and frequency of topics in those conferences using topic modelling.
- To visualise the topic-wise trends in the CALIBER conference during the selected time.

2. Literature Review

2.1 Application of Topic Modelling in Research Areas

Several studies have contributed to tracking the evolution of research themes in different domains by applying topic modelling techniques. Murakami et al. (2017) demonstrated the usefulness of topic modelling

for uncovering meaningful patterns and thematic structures in academic papers. It highlighted the advantages of topic modelling over traditional corpus linguistics methods. Clare & Hickey (2019) investigated the evolution of community forestry (CF) research from 1990 to 2017 using bibliometrics and topic modelling. The analysis identifies four major research areas and 20 key topics. Lee & Kang (2018) established topic modeling as a more effective approach to discover hidden topics in a large document collection than traditional ways, with subjective judgments and predetermined categories. They focused on 11,693 articles from 11 journals of technology and innovation management (TIM) research and applied LDA modeling to uncover trending topics. The result showed 'hot and cold' topics in TIM research. Corado Simões et al. (2025) applied topic modelling to 5,296 papers presented at EIBA conferences. They identified 24 key topics. The study highlighted the evolution of international business research.

2.2 Application of Topic Modelling in LIS

Figuerola et al. (2017) used topic modelling on over 92,000 LIS records from 1978–2014. The purpose was to map the domain's structure. They identified 19 key topics. These topics were further grouped into four main areas: processes, information technology, libraries, and applied information fields. Lamba & Madhusudhan (2019) examined the hidden research topics in articles published in the DESIDOC Journal of Library and Information Technology from 1981 to 2018. They included 928 full-text articles and used LDA topic modelling to tag each article with the modelled topics. The results indicate that bibliometrics, ICT, and user studies are the key areas of Indian LIS research. Han (2020) used LDA to trace the evolution of LIS research from 1996 to 2019. The results showed that library-focused topics are declining. Bibliometrics and information retrieval are consistent as research topics. It also analysed a shift in technological focus. Miyata et al. (2020) used LDA on 1,648 full-text articles from five LIS journals across two periods (2000–2002 and 2015–2017). The study concluded that the Internet became a central topic in LIS. It was observed that overall topic diversity declined over time. Thakuria & Deka (2024) investigated trends in scholarly publication in open-access library and information science journals between 2013 and 2022. The study identified the most prevalent topics and research themes and stated that traditional research topics are declining.

2.3 Research Gap

Despite the growing number of studies using Latent Dirichlet Allocation (LDA) to examine trends in LIS studies, a knowledge gap remains for analysing the key research topics and research trends in papers published in LIS, especially in the context of India. This present study aims to fill this gap up by investigating key research topics in CALIBER conference papers.

3. Methodology

3.1 Dataset

The present study includes full-text research papers presented at the five CALIBER (Convention on Automation of Libraries in Education and Research Institutions) conferences held in 2013, 2015, 2017, 2019,

and 2022. The full-text papers were collected from the Information and Library Network (INFLIBNET) Centre's publicly available Institutional Repository (https://ir.inflibnet.ac.in/home). The total number of papers included in the study is 251. Initially, each paper was downloaded and saved in PDF format. Later, Python (version 3.11.8) was incorporated to extract textual data from these full-text PDFs and stored into a CSV file for further preprocessing and analysis.

3.2 Data Preprocessing

Python (version 3.11.8) was implemented to preprocess the extracted textual data to prepare the CALIBER conference papers for topic modelling. The preprocessing steps were as follows:

3.2.1 Stopword removal

Stopwords add noise and do not contribute meaningfully to topic modelling. It is a pre-processing step in text mining applications that can improve efficiency and accuracy (Ladani & Desai, 2020). Stopwords were removed using the NLTK (Natural Language Toolkit) library of Python. The aim was to focus on content-rich terms.

3.2.2 Lemmatization

Lemmatization treats different grammatical forms of the same word as a single token (e.g., "libraries" '! "library"). It improves topic modeling by reducing vocabulary size and increasing semantic coherence of topics (Chauhan et al., 2023). Present Textual data was lemmatized using the NLTK library of Python.

3.2.3 Tokenization and Basic Cleaning

The Gensim library of Python was used to tokenize the texts, remove punctuations, and convert the texts to lowercase. Tokenization is an important phase in Natural Language Processing, and it involves the segmentation of text into meaningful units (Webster & Kit, 1992).

3.3 Topic Modelling

Topic Modelling was conducted on the preprocessed text by using the Latent Dirichlet Allocation (LDA) algorithm. LDA is a popular unsupervised machine learning technique (Jelodar et al., 2019). Initially, a dictionary was created from the list of tokenized documents. Then the corpus of CALIBER conference papers was created. Each document was converted into a bag-of-words format using the dictionary. Furthermore, LDA was trained. Words were assigned probabilistically to topics and topics to documents. The top 10 keywords for each of the 20 topics were extracted and saved in a CSV file. For each document, the model outputs a distribution over all 20 topics. The number of topics was decided by calculating the coherence score. The purpose was to assess the interpretability of topics. The model with 20 topics achieved the best balance between thematic clarity and coherence score.

3.4 Visualization

The present study employed both tabular and graphical visualizations. pyLDAvis was used to generate an interactive HTML visualization. Bar Charts were generated using matplotlib to display the overall frequency of dominant topics across all documents. Heatmaps created using seaborn present the frequency of topics by publication year.

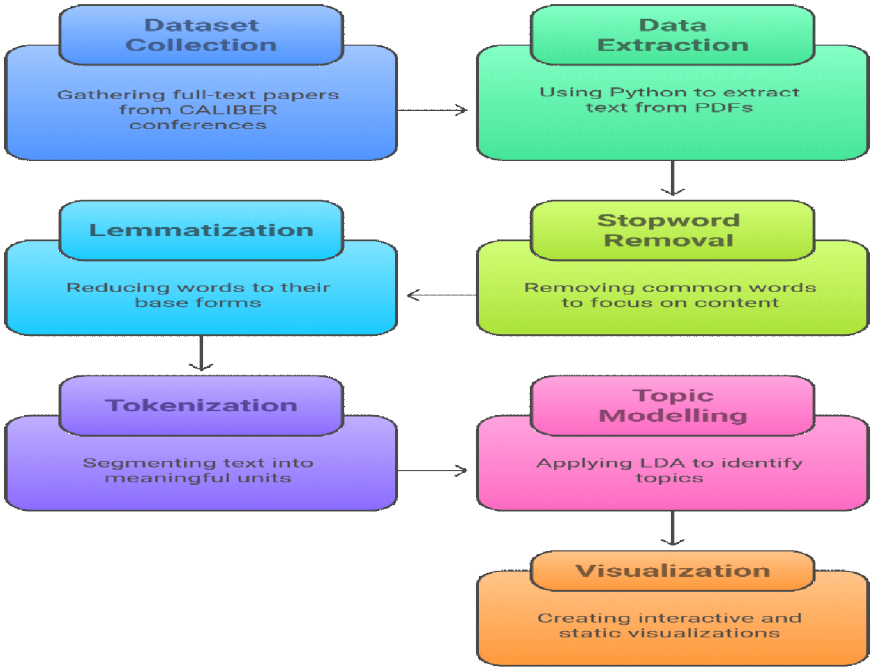

Figure 1: Research Methodology Workflow

4. Data Analysis and Interpretation

4.1 Topic Distribution Overview

The present study applies LDA topic modelling to the corpus of CALIBER conferences held in 2013, 2015, 2017, 2019, and 2022. This unsupervised, probabilistic model identifies 20 distinct topics. Each topic is represented by its top ten keywords (See Table 2). Topics are labelled based on the keywords they are represented by. Assigning labels to the topics ensures a reasonable connection of the topics to actual research areas in Library and Information Science (LIS). For example, Topic 1 is represented by the keywords 'research', 'journal', 'open', 'university', 'access', 'science', 'India', 'international', 'article', and 'cultural'. Therefore, it is labelled as 'Open-access and Academic Publishing'.

Table 2: Top Twenty Research Topics

Topic	Top Keywords	Label
Topic 1	research, journal, open, university, access, science, india, international, article, cultural	Open-access and Academic Publishing
Topic 2	data, research, library, author, repository, study, science, researcher, indian, li	Research Data and Repositories
Topic 3	library, social, service, use, medium, user, mobile, sm, respondent, using	Library User Services & Social Media
Topic 4	information, university, institute, researcher, ranking, use, link, search, science, term	Information Retrieval
Topic 5	university, yes, information, security, model, management, cgpa, library, ranking, system	Information Security & Management
Topic 6	research, university, library, analysis, law, information, science, sentiment, legal, social	Legal Information
Topic 7	library, information, service, institution, college, digital, practice, technology, rural, new	Digital Library and Technology Services
Topic 8	library, student, information, study, use, table, respondent, user, resource, service	Use of Library Resources
Topic 9	college, library, woman, naac, study, family, table, work, accreditation, automated	College Libraries & Accreditation
Topic 10	course, moocs, mooc, student, education, university, open, online, video, content	E-learning and MOOCs
Topic 11	library, service, technology, user, information, web, system, university, resource, book	Technology-Enhanced Library Services
Topic 12	open, library, learning, science, journal, user, information, online, student, content	Open Learning Resources and Services
Topic 13	data, publication, author, research, study, table, paper, analysis, number, science	Bibliometric and Scientometric Analysis
Topic 14	information, data, knowledge, metadata, library, ontology, web, international, literacy, th	Metadata, Ontology, and Information Literacy
Topic 15	lis, library, professional, information, skilled, year, manpower, science, skill, financial	Library Professionals, Skills & Workforce
Topic 16	ra, lis, readersa, service, syllabus, education, advisory, library, reading, research	Library and Information Science Education
Topic 17	plagiarism, research, academic, respondent, citation, impact, work, article, altmetrics, scholar	Plagiarism, Citation and Impact Analysis
Topic 18	digital, repository, learning, library, content, research, access, user, information, science	Digital Repository
Topic 19	library, resource, research, service, consortium, user, information, access, study, university	Library Resources and Consortium
Topic 20	university, education, learning, library, college, india, platform, higher, educational, teacher	Higher Education and Teaching Platforms

Table 3 shows the total distribution of top topics according to their frequency in CALIBER conference papers. Topic 8 (Use of Library Resources) has the highest number of papers published in the selected conferences (42). Topic 16 (Library and Information Science Education) has the fewest papers published (1). Topic 11 (Technology-Enhanced Library Services), Topic 1 (Open-access and Academic Publishing), Topic 13 (Bibliometric and Scientometric Analysis), and Topic 19 (Library Resources and Consortium) also have significant numbers of papers published (34, 24, 17, and 17, respectively). Figure 2 depicts the distribution of top topics in papers with a bar chart.

Table 3: Total Distribution of Top Topics

Topic	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Number of Papers	24	15	7	5	2	8	9	42	2	6	34	8	17	16	8	1	3	16	17	11

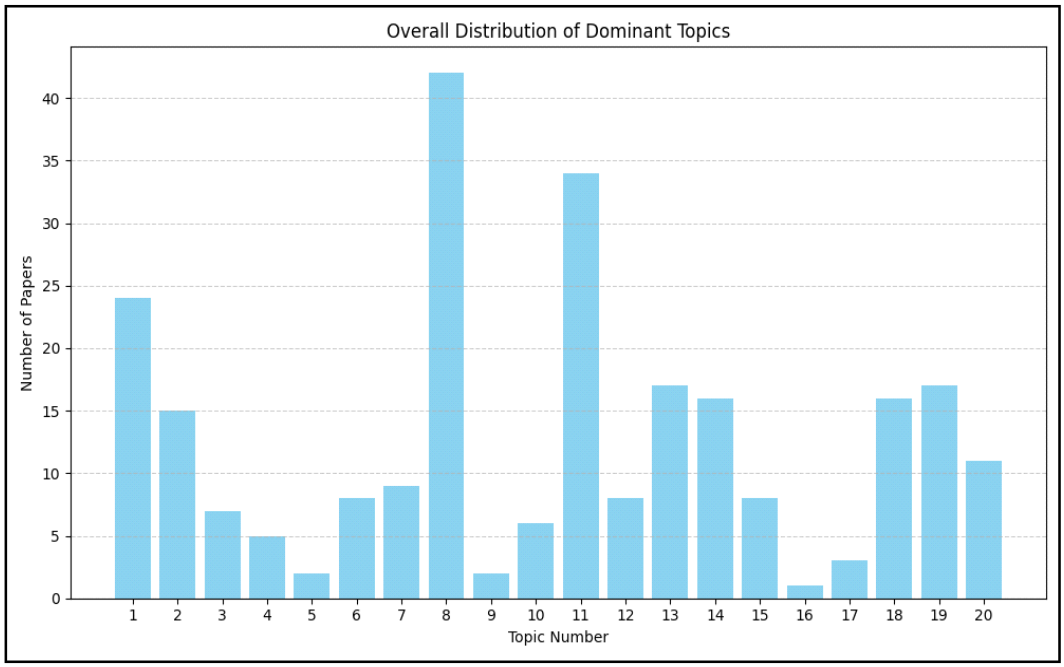

Figure 2 Frequency Distribution of Top Topics

4.2 Year-wise Topic Trend

Table 4 depicts a topic-year distribution matrix. The purpose was to explore how topics evolved over time. Each document's dominant topic was mapped to its publication year, and the cumulative topic counts per year were calculated. Furthermore, a year-wise topic distribution is plotted on a two-dimensional grid of mini-line charts (see Figure 3). Each mini plot shows the trend of one topic across years.

Table 4: Topic-Year Distribution Matrix

Year		2013	2015	2017	2019	2022
Frequency	Topic 1	4	9	4	2	5
	Topic 2	1	5	2	2	5
	Topic 3	0	1	2	4	0
	Topic 4	1	2	1	0	1
	Topic 5	0	1	1	0	0
	Topic 6	1	1	4	1	1
	Topic 7	1	1	2	4	1
	Topic 8	8	7	15	9	3
	Topic 9	0	1	0	0	1
	Topic 10	0	2	3	1	0
	Topic 11	13	4	3	8	6
	Topic 12	2	5	0	0	1
	Topic 13	0	10	5	0	2
	Topic 14	8	1	4	2	1
	Topic 15	1	1	1	4	1
	Topic 16	0	0	0	1	0
	Topic 17	0	1	0	1	1
	Topic 18	6	2	1	3	4
	Topic 19	2	2	3	3	7
	Topic 20	3	0	2	1	5

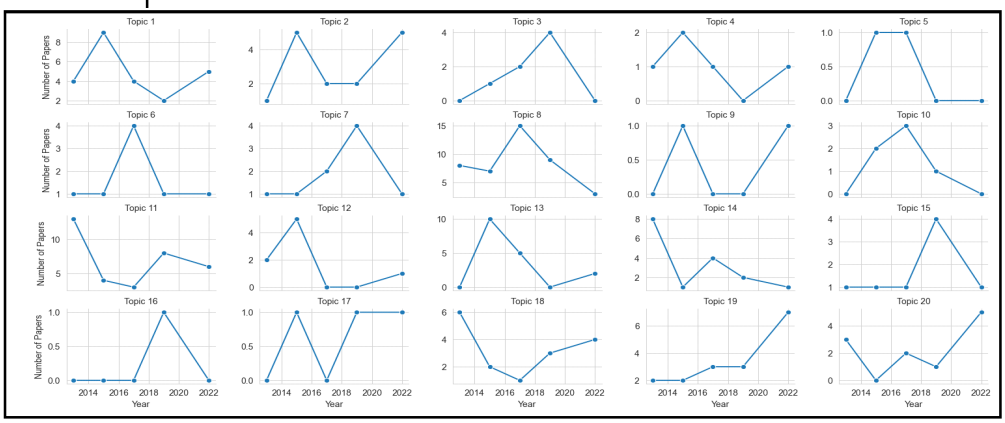

Figure 3 Year-wise Distribution of Research Topics

Table 4 and Figure 3 show patterns in the year-wise distribution of research topics in CALIBER conference papers.

4.2.1 Emerging Topics

Research areas like 'Technology-Enhanced Library Services' and 'Library Resources and Consortium' have emerged as significant research areas in recent years (particularly in 2019 and 2022).

4.2.2 Consistent Topics

Topics like 'Open-access and Academic Publishing' and 'Use of Library Resources' remained consistent throughout the years.

4.2.3 Declining Topics

Research topics like 'Open Learning Resources and Services' have faded out in recent years.

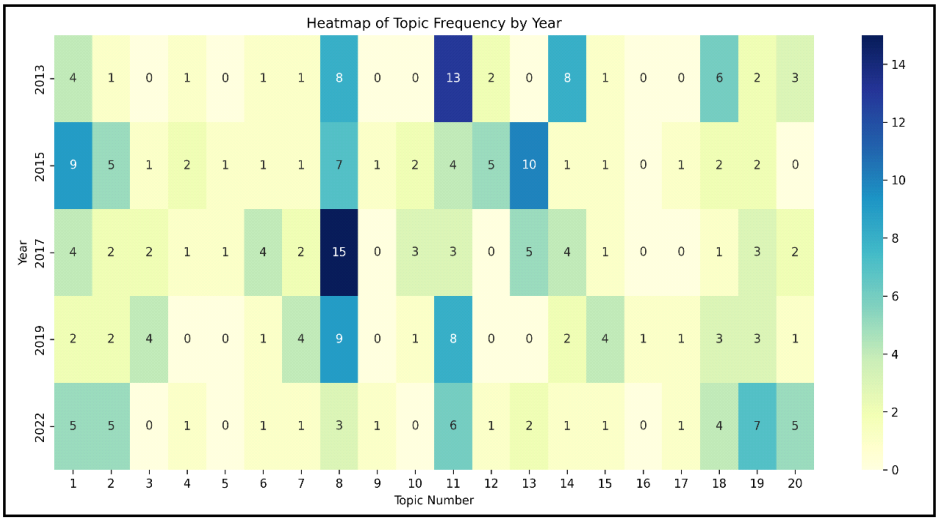

Figure 4: Heatmap of Topic Frequency by Year

Figure 4 visually presents a heatmap. It shows the frequency with which each discovered topic appeared in papers. Each cell in the heatmap represents the average weight of a particular topic. The degree of the existence of the topic corresponds to the colour intensity. Darker shades indicate higher frequencies, while lighter shades reflect lower relevance.

Additionally, pyLDAvis is used to explore the topics in an interactive HTML page (see Figure 5). Each bubble represents one topic. The frequency of the topics is represented by the size of the bubbles. The position of the circles shows how different the topics are from each other. For each topic, the contribution of the top keywords can be explored. Although in this section, only a static view of Topic 1 on the page is shown.

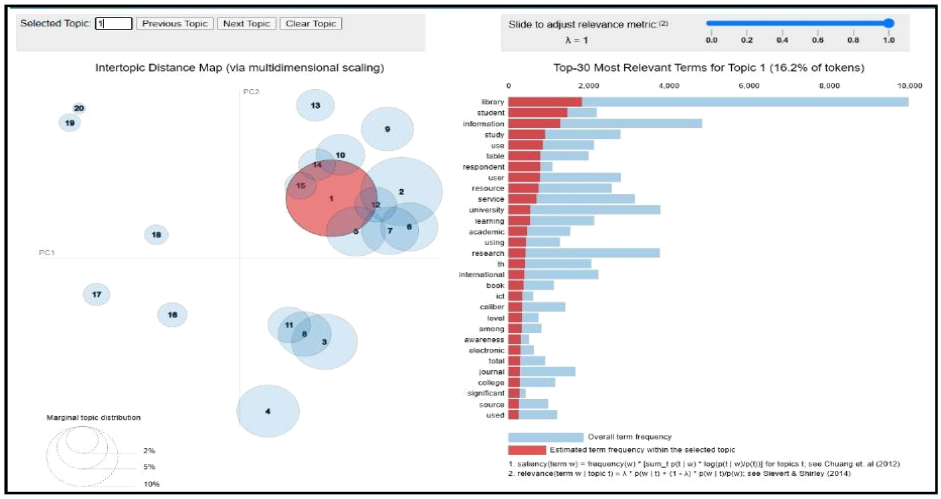

Figure 5: Interactive HTML Page of Topic Distribution

5. Discussion

The present study analyses the research topics and their trends in five CALIBER conferences (2013, 2015, 2017, 2019, and 2022). It applied Latent Dirichlet Allocation (LDA) to perform topic modelling on full-text CALIBER papers. The interpretation of data analysis shows a variety of topics such as Open-access publishing, library resources, social media, higher education, consortium, bibliometric and scientometrics studies, technology-based services, research data management, etc. Furthermore, these topics are to some extent interrelated and overlapping. The study confirms that topics like 'Use of Library Resources', 'Technology-Enhanced Library Services', 'Open-access and Academic Publishing', 'Bibliometric and Scientometric Analysis', and 'Library Resources and Consortium' have the most of the researchers' attention. Year-wise distribution of topics states that 'Technology-Enhanced Library Services' and 'Library Resources and Consortium' are major emerging topics in recent years. While 'Open-access and Academic Publishing' and 'Use of Library Resources' are the most consistent among other topics. 'Open Learning Resources and Services' is declining as a research topic. These findings can help library professionals, researchers, and policymakers identify underexplored areas and prioritize resource allocation. The results provide the LIS researchers with insights into the LIS research landscape in India.

6. Conclusion

This study has elaborated the present research status and emerging trends in research topics in the context of the CALIBER conference. It applied Latent Dirichlet Allocation (LDA) to analyze the thematic structure of research papers. The study identified 20 major research areas where LIS researchers are focusing on in CALIBER conferences. The findings show that LIS research in India has gradually shifted from foundational topics like Learning Resources and Services to more advanced and emerging areas such as digital libraries, open access, and technology-enhanced library services. The year-wise topic analysis and heatmap visualization highlighted changing trends. The study provides an overview of the research landscape in

CALIBER conferences. Also, it serves as a guide for future research directions. It demonstrates how text mining techniques, such as topic modeling, can be an effective tool for analysis in library science. Future researchers may aim to apply other topic modelling techniques, such as BERTopic, for more robust mapping of topics. Furthermore, they can include other Indian (such as PLANNER) and international conferences (such as COLLNET, WLIC) under their scope of study.

In conclusion, this research contributes to a better understanding of the intellectual structure of LIS in India. It encourages further use of text mining techniques to analyze scholarly communications in other national and international conferences. It will help LIS researchers to stay updated on emerging and underexplored research areas. Also, the study will be beneficial for policymakers in forming their funding priorities. Furthermore, evolving research topics may reinforce LIS curricula and professional development.

References

1. Chauhan, U., Shah, S., Shiroya, D., Solanki, D., Patel, Z., Bhatia, J., Tanwar, S., Sharma, R., Marina, V., & Raboaca, M. S. (2023). Modeling Topics in DFA-Based Lemmatized Gujarati Text. Sensors, 23(5), 2708. https://doi.org/10.3390/s23052708

2. Clare, S. M., & Hickey, G. M. (2019). Modelling Research Topic Trends in Community Forestry. Small-Scale Forestry, 18(2), 149–163. https://doi.org/10.1007/s11842-018-9411-8

3. Corado Simões, V., Rocha, J. P., Piepenbrink, A., Cantwell, J., & Gugler, P. (2025). The evolution of IB research: A topic model analysis of EIBA conference papers (1975-2020). Competitiveness Review: An International Business Journal, 35(1), 1–31. https://doi.org/10.1108/CR-07-2024-0132

4. Elia, E. F., & Mhando, J. A. (2024). Global Library and Information Science Research Patterns: A Twenty-Five-Year Reflection. University of Dar Es Salaam Library Journal, 18(2). https://doi.org/10.4314/udslj.v18i2.12

5. Figuerola, C. G., García Marco, F. J., & Pinto, M. (2017). Mapping the evolution of library and information science (1978–2014) using topic modeling on LISA. Scientometrics, 112(3), 1507–1535. https://doi.org/10.1007/s11192-017-2432-9

6. Gandhinagar, I. C. (n.d.). INFLIBNET Centre Gandhinagar.https://www.inflibnet.ac.in/convention/caliber.php

7. Han, X. (2020). Evolution of research topics in LIS between 1996 and 2019: An analysis based on latent Dirichlet allocation topic model. Scientometrics, 125(3), 2561–2595. https://doi.org/10.1007/s11192-020-03721-0

8. Harrison, R. (2010). Unique Benefits of Conference Attendance as a Method of Professional Development for LIS Professionals. The Serials Librarian, 59(3–4), 263–270. https://doi.org/10.1080/0361526X.2010.489353

9. Hauss, K. (2021). What are the social and scientific benefits of participating at academic conferences? Insights from a survey among doctoral students and postdocs in Germany. Research Evaluation, 30(1), 1–12. https://doi.org/10.1093/reseval/rvaa018

10. 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. Multimedia Tools and Applications, 78(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

11. Kochedykov, D., Apishev, M., Golitsyn, L., & Vorontsov, K. (2017). Fast and modular regularized topic modelling. 2017 21st Conference of Open Innovations Association (FRUCT), 182–193. https://doi.org/10.23919/FRUCT.2017.8250181

12. Ladani, D. J., & Desai, N. P. (2020). Stopword Identification and Removal Techniques on TC and IR applications: A Survey. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 466–472. https://doi.org/10.1109/ICACCS48705.2020.9074166

13. Lamba, M., & Madhusudhan, M. (2019). Mapping of topics in DESIDOC Journal of Library and Information Technology, India: A study. Scientometrics, 120(2), 477–505. https://doi.org/10.1007/s11192-019-03137-5

14. Lee, H., & Kang, P. (2018). Identifying core topics in technology and innovation management studies: A topic model approach. The Journal of Technology Transfer, 43(5), 1291–1317. https://doi.org/10.1007/s10961-017-9561-4

15. Miyata, Y., Ishita, E., Yang, F., Yamamoto, M., Iwase, A., & Kurata, K. (2020). Knowledge structure transition in library and information science: Topic modeling and visualization. Scientometrics, 125(1), 665–687. https://doi.org/10.1007/s11192-020-03657-5

16. Murakami, A., Thompson, P., Hunston, S., & Vajn, D. (2017). 'What is this corpus about?': Using topic modelling to explore a specialised corpus. Corpora, 12(2), 243–277. https://doi.org/10.3366/cor.2017.0118

17. Thakuria, A., & Deka, D. (2024). A decadal study on identifying latent topics and research trends in open access LIS journals using topic modeling approach. Scientometrics, 129(7), 3841–3869. https://doi.org/10.1007/s11192-024-05058-4

18. Webster, J. J., & Kit, C. (1992). Tokenization as the initial phase in NLP. Proceedings of the 14th Conference on Computational Linguistics -, 4, 1106. https://doi.org/10.3115/992424.992434

About Authors

Subhayan Chakraborty, Research Scholar, Department of Library and Information Science, Mizoram University, Aizawl, India
Email : subhayan@drtc.isibang.ac.in
ORCID: https://orcid.org/0009-0004-6161-5102

Prof. Manoj Kumar Verma, Professor, Department of Library and Information Science, Mizoram University, Aizawl, India
Email : manojdlis@mzu.edu.in
ORCID: https://orcid.org/0000-0002-3009-3258