

Analysis of Latent Themes and Sentiments Regarding Outsourcing in Libraries and Information Centres using LDA Topic Modelling and Polarity Distribution: A Study based on Scopus Data

* Ashish Kumar[#]

** Sanjiv Kadyan

* Research Scholar, Department of Library and Information Science, Maharshi Dayanand University, Rohtak (Haryana) - 124001, INDIA; Email: ashish.rs.lisc@mdurohtak.ac.in; ORCID: [https:// 0000-0002-1910-012X](https://orcid.org/0000-0002-1910-012X)

** Professor, Department of Library and Information Science, Maharshi Dayanand University, Rohtak (Haryana) - 124001, INDIA; Email: sanjivkadyan3@gmail.com; ORCID: [https://0009-0005-4194-0386](https://orcid.org/0009-0005-4194-0386)

[#] Corresponding author.

Received: 25 September 2025

Accepted: 29 December 2025

Abstract

The ever-present budgetary constraints in libraries and information centres have created a need for some strategic response to the issue, and outsourcing seems to be one of the possible solutions. Administrators and managers of libraries have adopted this solution to varying extents in various setups. This study provides an analysis of 152 Scopus-indexed publications relating to the topic. The findings reveal a limited output, often attributed to single-time authors. Keyword co-occurrence, along with the LDA topic modelling, reveals the focal areas of research, which include technical services and digital curation, while topics like staff transformation, technology adoption, and research-focused services, and the role of librarians, remained latent. Sentiment analysis shows a major positive outlook (63.1%) towards outsourcing, particularly for its role in improving efficiency and service quality, although concerns about complexity and risks persist. The study will not only be beneficial for researchers interested in the area but also for administrators and managers in making informed decisions.

Keywords: LDA; library outsourcing; polarity distribution; sentiment analysis; text analysis; topic modelling.

1. Introduction

Constantly evolving service demands and rapid technological advancements, along with long-sustained issues like budgetary constraints in libraries and information centres, have created a need for a strategic response in the form of outsourcing. As libraries evolved with these advancements from traditional models to more complex and resource-intensive systems, outsourcing some specific functions like cataloguing and digitization of IT infrastructure has become a common practice with an aim to optimise performance, along with a focus on core activities. Much research has been published focusing on outsourcing in libraries and information centres. Works such as those by Ma and Lee (2016) employed bibliometric methods to identify the trends through publication analysis, while Sawant (2021) conducted a survey-based study to assess the extent and identify the types of services being outsourced in the libraries of academic institutions. The author applied the core competencies

theory, suggesting that libraries outsource non-core tasks like shelving and binding to redirect the professional staff toward core intellectual services like research support. Similarly, Gupta and Sharma (2012) argue that outsourcing is a strategic tool for cost optimization, allowing libraries to manage shrinking budgets while maintaining service standards. All these research studies have been conducted, but most of the published literature has focused either on the quantitative aspect or institutional case studies with a limited integration of bibliometric and content-based analyses. Also, there is a lack of any recent study to map the literature of the field. Key gaps are observed in the understanding of latent themes as well as the sentiment analysis of outsourcing on the basis of scholarly literature.

There is a lack of holistic studies analysing the bibliometric patterns along with qualitative trends, especially the thematic development, topic modelling, and sentiment orientations. The current study addresses these gaps through the conduct of an analysis of the Scopus-indexed publications on outsourcing in libraries and information centres. The study is complemented with thematic and sentiment analyses to uncover the prevalent themes along with the emotional tone of the discourse over time. Topic modelling using Latent Dirichlet Allocation (LDA) will further uncover the latent topics to work upon. By integrating these approaches, the current study aims to present the latent themes along with the polarity-based sentiments from the literature relating to outsourcing in libraries and information centres. Through all these, the paper contributes to the understanding of how outsourcing is conceptualized, debated, and implemented in the scholarly domain through a more advanced way of content-based text mining and sentiment analysis, rather than just mapping the volume and authorship patterns of LIS literature relying on the metadata analysis as done in studies like one by Garg and Tripathi in 2018.

2. Objectives

The current study is particularly based on the text analysis of the Scopus-indexed documents to analyse the concept of outsourcing in the field of libraries and information centres. For the conduct of the study, the following objectives were taken:

- i. To map a brief overview of the related publications in the domain.
- ii. To identify the latent topics in the domain using LDA (Latent Dirichlet Allocation) topic modelling.
- iii. To analyse the sentiments expressed in the documents using lexicon-based polarity analysis.

3. Methodology

This is a secondary data-based study conducted on publication data extracted from the Scopus database on the 5th of July 2025, using title, abstract, and keywords searching. The search string is combined with various relevant refinement criteria to get the best results out of the database that relate to outsourcing in libraries and information centres. The complete search string is, “(TITLE-ABS-KEY (outsourc* OR "contracting out" OR "third-party services" OR "external service provider*" OR "vendor services")) AND (TITLE-ABS-KEY (librar* OR "information center*")) AND (LIMIT-TO (SUBJAREA , "SOC") OR LIMIT-TO (SUBJAREA , "ARTS") OR LIMIT-TO (SUBJAREA , "DECI") OR EXCLUDE (SUBJAREA , "ENER") OR EXCLUDE (SUBJAREA , "ENVI") OR EXCLUDE (SUBJAREA , "PSYC") OR EXCLUDE (SUBJAREA , "AGRI") OR EXCLUDE (SUBJAREA , "CENG") OR EXCLUDE (SUBJAREA , "CHEM") OR EXCLUDE (SUBJAREA , "MEDI") OR EXCLUDE (SUBJAREA , "MATH") OR

EXCLUDE (SUBJAREA , "ENGI") OR EXCLUDE (SUBJAREA , "COMP") OR EXCLUDE (SUBJAREA , "ECON")) AND (EXCLUDE (DOCTYPE , "le") OR EXCLUDE (DOCTYPE , "ed")) AND (EXCLUDE (PUBYEAR , 2025))". The search string gave a recall of 213 documents, which were extracted in CSV format. After removing the duplicates, 195 documents were screened for their suitability, and 161 were found relevant. After the removal of 9 documents that were originally published in a language other than English, 152 documents were selected for the study. These were refined and harmonized using OpenRefine (Delpeuch et al. 2025). For the topic modelling and sentiment analysis of the data and visualization of the results, Biblioshiny (Aria and Cuccurullo 2017), VOSviewer (Van Eck and Waltman, 2009), TALL (Text Analysis for All) shiny app (Aria et al. 2025), R Studio (Posit team 2025), and Microsoft Excel were used during the study.

For the identification of latent themes in the research related to outsourcing in libraries, Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003) topics are identified using the abstracts of the selected documents. The number of topics ($k=5$) is validated through a qualitative review, where 5 number of topics provided an optimal balance between the semantic distinctiveness as well as the unique nature of the topics. To ensure the linguistic accuracy through preprocessing and tagging, that are performed using the Universal Dependencies model that is pretrained using the GUM corpus with a tree bank data of 208,446 tokens, 212,035 words, and 12,146 sentences was used, which is proposed to be a significantly accurate model (words- 99.57%, Lemma- 96.55%, sentences- 95.07% and PoS- 95.77% accurate) (Zeldes, 2017) however the actual analysis and modelling were performed strictly on the abstracts of the selected publications from the WoS with Scopus datasets. Sentiment polarity is calculated using Hu and Liu's (2004) Bing lexicon within TALL through binary polarity aggregated per document to classify the sentiments among very positive (≥ 2), positive (1), neutral (0), negative (-1) and very negative (≤ -2) based on the net score. Limiting the study to a single database and a single publication language might have left behind some of the important studies of the field.

4. Data analysis

4.1 Brief overview

The study analyses a total of 152 documents published between 1988 and 2024 to identify the latent topic through LDA topic modelling and sentiment analysis through polarity distribution. Over a span of 38 years, these publications have attained an average of 6.965 citations per cited publication. A total of 228 authors contributed to the domain across 60 different sources, achieving an average of 3.45 citations per author during the period. The publication list includes 121 articles, 13 reviews, 10 book chapters, 8 books, and 1 conference paper. Of these 152 documents, 125 are published in open access, while the remaining 27 are available through a subscription model.

4.2 Topic modelling using Latent Dirichlet Allocation (LDA)

The topic model is estimated using the 100 most frequent words identified through a frequency measure. Figure 1 presents the five topics identified with the beta probability of each word belonging to specific topics. The visualization shows the ten representative words of each topic selected on the basis of the probability of that particular topic.

Topic 1: Project-based outsourcing for digital resources

This topic is particularly related to the outsourcing activities for the digital library collections and academic library resources. The project-based outsourcing is the basis of the development of the topic.

Topic 2: Information management and the role of staff

This particular topic is related to the activities of information management, the role of organizational staff, and the availability of resources for the same. The topic also covers the terms relating to the technological transformation.

Topic 3: Outsourcing of public services of libraries and information centres

This particular topic includes the outsourcing culture in the libraries for the public or user-oriented services like study spaces, and the management of resources in the government institutions.

Topic 4: Vendor management and the outsourcing culture in cataloguing activities

The topic covers three major areas, namely the vendor relationship management within academic libraries, catalogue management services, and the issues relating to the rights and access management for the outsourced library services.

Topic 5: Research-oriented library services

This topic is different from the other four as it discusses the research support services and the role of a librarian for the same. The presence of the words “cost” and “law” suggests the study relating to the budgetary provisions and the legal provisions. The word “law” might be there relating to special libraries in the field of Law.

The presence of the words “library” and “outsource” in multiple topics, like topics one and three, is attributed to the selection of the studies that particularly relate to the outsourcing of library activities.

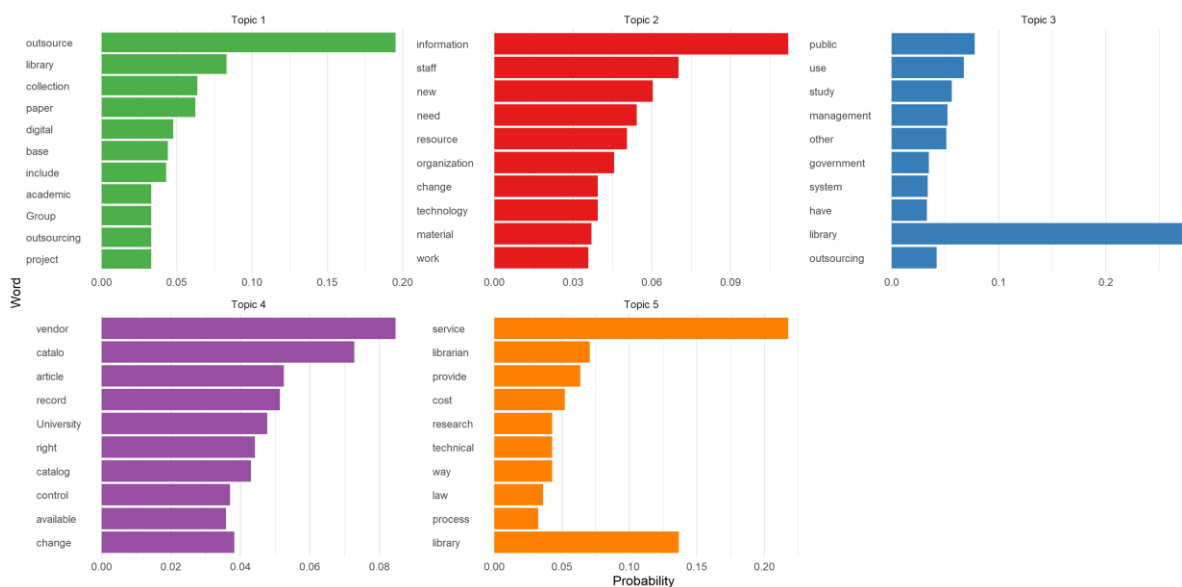


Figure 1: Topics identified through Latent Dirichlet Allocation (LDA) with word results

It is also noted that 3 out of 5 topics identified through topic modelling have already been explored, as they are found to be similar to the six major themes identified through keyword co-occurrence of authors' keywords (outsourcing in specialized & international library contexts, technical and cataloguing functions outsourced, digital outsourcing & vendor management, academic library outsourcing practices, cost efficiency & outsourcing in specialized libraries, and collaborative outsourcing in public & university libraries). Topic 1 resembles themes 3 and 4, topic 3 resembles theme 6, and topic 4 explores similar areas as identified in themes 2 and 3.

Topics 2 and 5, which discuss information management, technology adoption, staff transformation, research-oriented services, and the role of librarians, are found to be partially latent, and hence, studies can be conducted to analyse the adoption and efficiency of outsourcing in these areas of libraries.

4.3 Sentiment analysis using polarity distribution

Polarity detection is a sentiment analysis technique widely used to determine whether the set of documents expresses a positive, negative, or neutral sentiment. The current analysis is done using the lexicon-based approach, incorporating contextual adjustments to refine the sentiment scoring. Figure 2 presents the plot of the polarity distribution among the 152 documents. The figure consists of two bar charts showing the top positive (top left) and negative (bottom left) words, and the pie chart (right) showcasing the polarity distribution among the selected documents. The interpretation shows that the corpus has a general positive sentiment (49.3% very positive, reinforced by 13.8% positive sentiment) towards the outsourcing, but a significant negative sentiment is present too, with 19.7% negative and 7.89% very negative sentiment. 9.21% of the documents fall under the neutral categorization that might be attributed to either their ambiguous nature or their being factual publications. The presence of negative views makes it even more important to examine the specific drivers behind this negativity.

The positive words can be categorised into two major categories, one with a focus on getting the work done efficiently (right, work, quality, available, well), while the second one relates to the value and promotion of new ideas in the field (improve, important, support, benefit, originality). These collectively make the themes relating to success, effectiveness, innovation, and the importance of the services and activities.

In contrast, the negative words are related to problems or obstacles faced, risks associated, detection of failures, complexity of the processes, and the uncertainty about the impact indicated by words like issue, limit, implication, concern, failure, error, loss, obstacle, critical, critic, disadvantage, complex, etc.

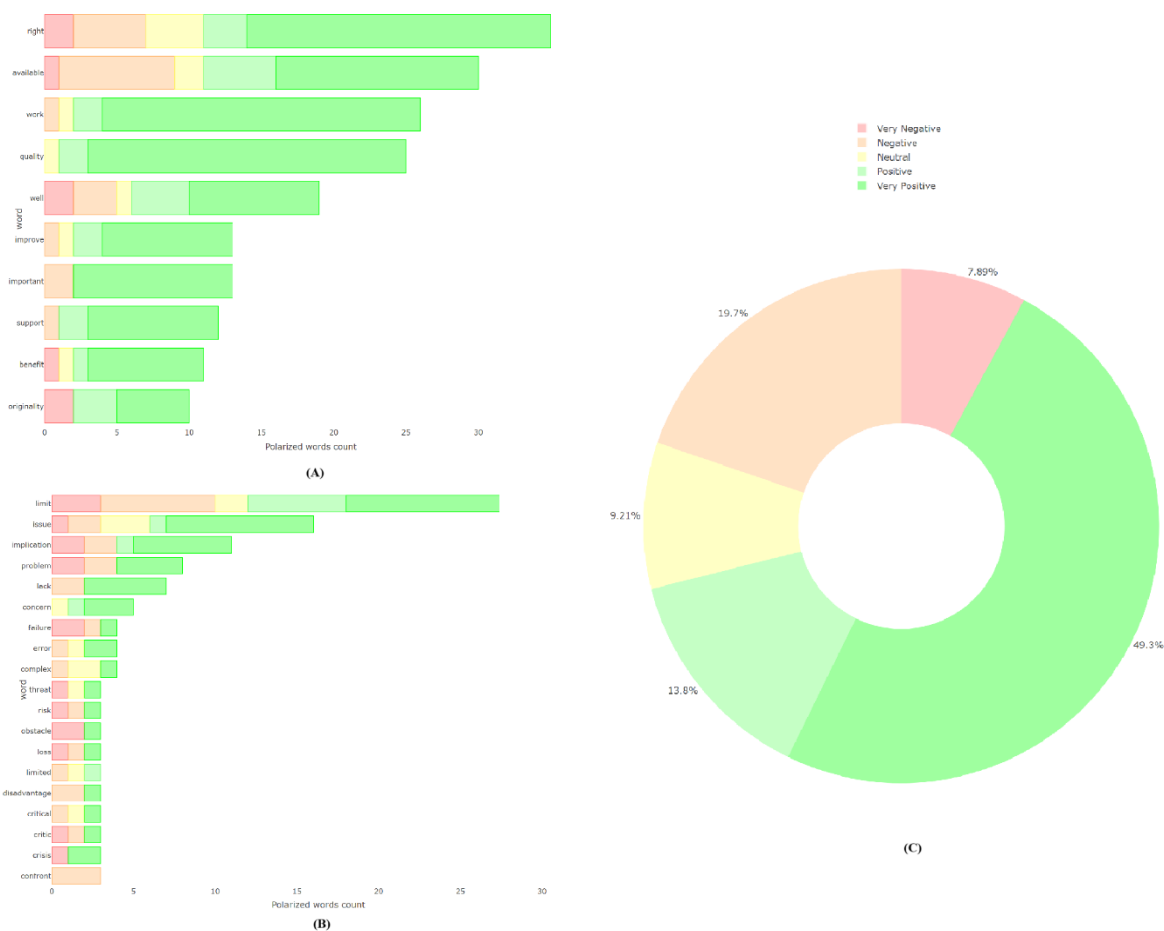


Figure 2: Plots of Polarity Distribution among 152 selected documents. (A) Top positive words, (B) Top negative words, and (C) pie chart showcasing the polarity distribution among selected documents

5. Discussion

The current study, based on 152 Scopus-indexed publications, was conducted to get insights into the latent themes as well as the sentiments regarding the concept of outsourcing in libraries and information centres globally. The study found that a limited number of publications, mainly from single-time authors, are published. The prevalence of “one-off” publications suggests that outsourcing is frequently documented as an institutional case study rather than a sustained theoretical domain.

The major thematic areas identified through the keyword co-occurrence and topics identified using LDA topic modelling resemble, to a greater extent, showing a focused research conduct, but still some of the aspects like management of information in technology-driven libraries, staff transformation, and the implementation of outsourcing in research oriented services along with the role of librarians is still found to be partially latent among the published literature and studies may be conducted to analyse the possibilities and impact of such incorporations. Critics frequently contend that extensive outsourcing can result in the deskilling of professional staff and a diminished institutional memory. The limited attention given to these issues in the analyzed corpus suggests that the LIS community may be prioritizing economic efficiency over long-term workforce stability. Future research should critically examine how librarians' roles are not only redefined but also potentially displaced in highly outsourced environments.

The analysis found four major drivers of outsourcing, namely privatization, cost pressures, technological needs, and workflow optimization, that lead to implementation in technical services, digital curation, and strategic functioning across all types of libraries, including academic, public, and special libraries. Vendor management is also in focus in the literature. Outcomes are categorized into economic, operational, and strategic, marking their wider impact that also promotes and guides the libraries for the implementation among varied library services.

Sentiment analysis through polarity detection revealed an overall positive sentiment towards outsourcing in the literature, with 63.1% of the text expressing positive or very positive views heavily weighted towards efficiency, innovation, and service quality. However, there are also some concerns and risks associated with the implementation, such as process complexity and potential failures. The findings suggest a positive intent for outsourcing in the planning phase, but execution challenges arise regarding quality control and vendor dependence. These negative sentiments need further exploration, and studies can be conducted to address the challenges being faced and to enhance the positive views recorded in the corpus. 9.21% of the neutral content reflects the presence of objectivity in sentiments in the literature.

6. Conclusion and recommendations for further research

The study provides valuable insights into the published literature on outsourcing in libraries. Insights from the keyword co-occurrence and LDA topic modelling suggest studying the applicability and efficiency of outsourcing in the latent areas like information management, technology adoption, staff transformation, research-oriented services, and the role of librarians. The negative sentiments associated with the complexity and risks associated with the process also need to be studied to address these challenges and enhance the positivity of the process. The findings of the study can be utilized as a benchmark for future research and also for making informed decisions related to the application of outsourcing in libraries and information centres.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Ethical approval

The authors declare that no animals or human beings were used or involved in this study, and hence, no ethical approval was applicable from any concerned committee. The study adheres to scientific ethics and conduct to ensure transparency.

References

1. Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975. <https://www.sciencedirect.com/science/article/pii/S1751157717300500>
2. Aria, M., Cuccurullo, C., D'Aniello, L., Misuraca, M., & Spano, M. (2025). TALL: A new Shiny app for text analysis. In A. Pollice & P. Mariani (Eds.), *Methodological and applied statistics and demography III. SIS 2024*. Springer. https://doi.org/10.1007/978-3-031-64431-3_11

3. Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent Dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993-1022.
4. Delpeuch, A., Morris, T., Huynh, D., Weblate, Mazzocchi, S., Jacky, Guidry, T., et al. (2025). OpenRefine/OpenRefine: OpenRefine (Version 3.9.2) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.15089184>
5. Garg, Kailash, & Tripathi, Harish. (2018). Bibliometrics and scientometrics in India: An overview of studies during 1995-2014 Part II: Contents of the articles in terms of disciplines and their bibliometric aspects. *Annals of Library and Information Studies*, 65, 7-42.
6. Gupta, D.K., & Sharma, V. (2012). Evidence of outsourcing in science and technology libraries of Delhi. *Library Management*, 33(4/5), 241-252. <https://doi.org/10.1108/01435121211242281>
7. Hu, M., & Liu, B. (2004). Mining and Summarizing Customer Reviews. Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Seattle, 22-25 August 2004, 168-177. <https://doi.org/10.1145/1014052.1014073>
8. Ma, T.J., & Lee, G.G. (2016). Outsourcing library functions: A bibliometric study in SSCI. *The Bottom Line: Managing Library Finances*, 29(4), 251-285. <https://doi.org/10.1108/BL-12-2015-0022>
9. Posit team. (2025). RStudio: Integrated development environment for R (Version 2025.5.1.513) [Computer software]. Posit Software, PBC. <http://www.posit.co/>
10. Sawant, S.S. (2021). Outsourcing of activities and services in academic libraries. *Annals of Library and Information Studies*, 68(1), 39-46.
11. Van Eck, N.J., & Waltman, L. (2009). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523-538. <https://doi.org/10.1007/s11192-009-0146-3>
12. Zeldes, A. (2017). The GUM corpus: Creating multilayer resources in the classroom. *Language Resources and Evaluation*, 51(3), 581-612.

