May 2019

Author-Topic Modeling of DESIDOC Journal of Library and Information Technology (2008-2017), India

Manika Lamba
Department of Library and Information Science, University of Delhi, lambamanika07@gmail.com

Margam Madhusudhan
Department of Library and Information Science, University of Delhi, madhumargam@gmail.com

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

Part of the Library and Information Science Commons

https://digitalcommons.unl.edu/libphilprac/2593
Author-Topic Modeling of DESIDOC Journal of Library and Information Technology (2008-2017), India

Abstract
This study presents a method to analyze textual data and applying it to the field of Library and Information Science. This paper subsumes a special case of Latent Dirichlet Allocation and Author-Topic models where each article has one unique author and each author has one unique topic. Topic Modeling Toolkit is used to perform the author-topic modeling. The study further which considers topics and their changes over time by taking into account both the word co-occurrence pattern and time. 393 full-text articles were downloaded from DESIDOC Journal of Library and Information Technology and were analyzed accordingly. 16 core topics have been identified throughout the period of ten years. These core topics can be considered as the core area of research in the journal from 2008 to 2017. This paper further identifies top five authors associated with the representative articles for each studied year. These authors can be treated as the subject-experts for the modeled topics as indicated. The results of the study can serve as a platform to determine the research trend; core areas of research; and the subject-experts related to those core areas in the field the Library and Information Science in India.

Keywords Author-Topic Modeling; DESIDOC Journal of Library and Information Technology (DILIT); Latent Dirichlet Allocation (LDA); Information Retrieval; Information Processing and Management; Text Mining

1 Introduction
Topic modeling acts as a text mining tool to process, organize, manage and extract knowledge. It is based on probabilistic modeling and is used to discover hidden structures in large archives of documents on the basis of similar patterns of word usage in each document. It is typically used to determine the underlying “topics” in text documents. A “topic” represents the broader concept shared by a document corpus and this “topic” evolves with time. Modeling topics without considering time will cause major problems. Topic evolution will help to identify topics within the context and how they evolve over time.

Lately, the focus has either on a group of highly cited authors or a sample of journal articles. This leads to skew the results by incorporating highly cited works which are not necessarily representative of the works produced and including a few articles/authors which can heavily influence the results. This defying problem can be solved by performing author-topic modeling where not only the core research topics related to a discipline is identified but also the core authors which may be treated as the subject-experts for the respective modeled topics. Further, understanding the development of the discipline and the changes in topics over time will help to recognize the discipline identity. Therefore, this study provides the groundwork of identifying the identity of Library and Information Science (LIS) discipline in India by finding the main topics in DESIDOC Journal of Library and Information Technology (DILIT) by using Latent Dirichlet Allocation (LDA) modeling technique.

2 Related literature
Few important studies which focus on author-topic modeling are by Caglieri et al. (2018) who “address the discovery of research collaborations among multiple authors on single or multiple topics. Specifically, they exploit an exploratory data mining technique, i.e., weighted association rule mining, to analyze publication data and to discover correlations between ATM topics and combinations of authors”; and Mao et al. (2017) who “propose an approach to detect Topical Scientific Communities (TSCs) with both topology and topic features by applying machine learning techniques and network theory”. This study follows the works of Sugimoto et al. (2011) where they “identify changes in dominant topics in LIS over time, by analyzing the 3,121 doctoral dissertations completed between 1930 and 2009 at North American Library and Information Science programs using author-topic modeling”; and Figueroa et al. (2017) who “offers an overview of the bibliometric study of the domain of Library and Information Science (LIS), with the aim of giving a multidisciplinary perspective of the topical boundaries and the main areas and research tendencies. Based on a retrospective and selective search, they have obtained the bibliographical references (title and abstract) of academic production on LIS in the database LISA in the period 1978–2014. They apply latent Dirichlet allocation, in order to identify the main topics and categories in the corpus of documents analyzed” with major modifications”.

Few important studies have been reviewed with regard to the application of LDA, Topic Modeling, and Text Mining. A pioneering research classic paper on LDA by Blei et al. (2003) showed the “efficient approximate inference using Markov Chain Monte Carlo methods based on vibrational methods and an EM algorithm for empirical Bayes parameter estimation. They report results in document modeling, text classification, and collaborative filtering, comparing to a mixture of unigram models and the probabilistic LSI model”. Some of the selected articles which show the implementation of the LDA in their studies are written by Monttazi (2018); Kim and Kang (2018); Zhao et al. (2016); Guo et al. (2017); Zhang et al. (2017); Chen (2017); Lu et al. (2017); Liu et al. (2016); Zhang et al. (2018); Ma et al. (2018); Wolttmann and Alkaoesig (2018); Chen et al. (2016); Wang et al.(2013); Koltsouva and Kolteov (2013); Nichols (2014);Yan et al. (2017); Katsurai et al. (2016); and Huet al. (2014).

Few studies which applied topic modeling in the field of Library/Museum under various sub-fields are as follows: Mehler and Waltinger (2009) applied topic modeling in library classification by presenting a topic classification model using the “Dewey Decimal Classification” (DDC); whereas Bae et al. (2014) presented their “study in threefold: First, they suggested an alternative approach to real-time big data analysis, which has become an extremely important issue. Second, they applied a topic modeling technique that is used in various research areas, including Library and Information Science (LIS). Based on this, they can confirm the utility of storytelling and time series analysis. Third, they developed a web-based system, and make the system available for the real-time discovery of topics”. Lu and Wolfram (2012) “present static and dynamic word-based
approaches using vector space modeling, as well as a topic-based approach based on latent Dirichlet allocation for mapping author research relatedness. Outcomes for the two word-based approaches and a topic-based approach for 50 prolific authors in Library and Information Science are compared with more traditional author co-citation analysis using multidimensional scaling and hierarchical cluster analysis” in contrast to Efron, Organisciak, and Fenlon (2011) who applied topic modeling in Museum in which “they propose a way to improve topic modeling in large collections by identifying documents that convey only weak topical information using the corpus from the Institute of Museum and Library Services Digital Collections and Content aggregation”.

3 Research Objectives

The following are the main research objectives of the study which are both historical and methodological in nature:

(a) To identify the core topics of DESIDOC Journal of Library and Information Technology diachronically; and
(b) To identify the authors associated with the representative articles.

4 Methodology

A total of 393 full-text articles were downloaded from DJLIT journal for the period 2008-2017. The articles were analyzed according to the author-topic model and were divided into the ten-time slices. The author-topic modeling technique was used to determine the top five topics, top five words, top five representative articles and the top five authors associated with those articles. In this method, for each year, 5 topics were identified. Each topic contained a probability value, that is, the likelihood that the topic identified should be associated with the year. These topics were ranked by descending probability values and the top five were selected as being most representative of that year. Similarly, a probability for each word was calculated to represent the association between a word and the given topic and the top five words were chosen as most representative of the topic. Lastly, the authors were assigned probability values for each topic and these too were ranked. The top five authors were chosen as highly representative authors for the given topic. In this approach, the hyperparameter α (prior parameter determining the topic distribution per document) and hyper-parameter β (prior parameter determining the word distribution per topic) are fixed at 5/T (where the T= number of topics taken in the study, i.e. 5 for the present study) and 0.01 respectively for each year as default. Fixing the hyper-parameters maintain the uniformity and consistency of results for each year. The author-topic modeling approach taken in this study is similar to that of Sugimoto et al. (2011) with major modifications, where they did topic modeling over 3,121 North American LIS doctoral Dissertations completed between 1930 and 2009 but this paper is restricted to full-text articles published in DJLIT journal from the year 2008 to 2017 excluding the Guest editorials.

4.1 DESIDOC Journal of Library and Information Technology

“DESIDOC Journal of Library and Information Technology (DJLIT) was started in the year 1981. It is a peer-reviewed open access bi-monthly journal that publishes original research and review papers related to library science and IT applied to library activities, services, and products. It is meant for librarians, documentation and information professionals, researchers, students and others interested in the field. DJLIT is a prestigious journal of LIS studies in India with 0.364 SJR (SCImago Journal Rank); 0.600 SNIP (Source Normalized Impact per Publication), and 0.47 Cite Score. It is abstracted and indexed in Scopus, LISA, LISTA, EBSCO Abstracts/Full-text, Library Literature, and Information Science Index/Full-text, The Informed Librarian Online, Open J-Gate, Indian Science Abstracts, Indian Citation Index, Full-text Sources Online, WorldCat, ProQuest, Google Scholar, Ulrich's International Periodical Directory, Index Copernicus, and OCLC” (DESIDOC Journal of Library and Information Technology 2018).

4.2 Latent Dirichlet Allocation (LDA)

This paper focuses on the use of Latent Dirichlet Allocation (Blei et al, 2003), which is based on Dirichlet distribution to model the topics from the corpus of LIS articles. In this study, each article gets represented as a pattern of LDA topics making every article appear. LDA automatically infers the topic discussed in a collection of articles and these topics can be used to summarize and organize the articles. LDA is based on probabilistic modeling and the observed variables are the bags of words per article whereas hidden random variables are the topic distribution per article. “The main goal of LDA is to compute the posterior of the hidden variables given the value of the observed variables” (Allahyari et al., 2017). The assumptions of LDA for the study are (i) articles with similar topics will use similar groups of words, (ii) articles are a probability distribution over latent topics, and (iii) Topics are probability distributions over words (Fig.1).
Fig. 1: Graphical model representation of Latent Dirichlet Allocation
(Source: Blei et al., 2003)

Fig. 1 demonstrates the functioning of LDA where the outer box represents documents, while the inner box represents the repeated choice of topics and words within a document. The variables shown in the figure are defined as follows (Blei et al., 2003):

\( \alpha \) – parameter of Dirichlet prior on the per-document topic distribution
\( \beta \) – parameter of Dirichlet prior on per-topic word distribution
\( \theta \) – topic distribution for the document, \( d \)
\( z \) – topic for the \( n \)th word in the document, \( d \)
\( w \) – is the specific word
\( N \) – total number of words in the corpus
\( M \) – total number of documents in the corpus

4.3 Author-Topic Model

The LDA model mentioned by Blei et al. (2003) is “extended to what is called the author-topic model” (Rosen-Zvi et al. 2010). In this model, each document can be viewed as a mixture of probabilistic topics and authors. In this study, each author is associated with a single article. The author-topic model allows not only examine which topics are most salient across the various time period but also which authors are most associated with these topics.

Fig. 2: Author-topic model based on Latent Dirichlet Allocation
(Source: Rosen-Zvi et al., 2010)
Fig. 2 explains the functioning of the model, “when author model as a group of authors, \( \alpha \), decide to write the document \( D \). For each word in the document an author is chosen uniformly at random. Then, as in the topic model, a topic is chosen from a distribution over topics specific to that author, and the word is generated from the chosen topic. As in the author model, \( x \) indicates the author responsible for a given word, chosen from \( \alpha \). Each author is associated with a distribution over topics, \( \theta \), chosen from a symmetric Dirichlet (\( \alpha \)) prior. The mixture weights corresponding to the chosen authors are used to select a topic \( z \). and a word is generated according to the distribution \( \phi \) corresponding to that topic, drawn from a symmetric Dirichlet (\( \beta \)) prior. The author-topic model subsumes the two models described above as special cases: topic models like LDA correspond to the case where each document has one unique topic, and the author model corresponds to the case, where each author has one unique topic” (Rosen-Zvi et al., 2010).

4.4 Topic-Modeling-Toolkit (TMT)

Topic-Modeling-Toolkit (Google code archive, 2011a) is powered by Java, a graphical interface tool for LDA topic modeling. “It is a simple GUI-based application for topic modeling that uses the popular MALLET toolkit for the back-end” (Abinaya and Winster, 2014). “Topic models provide a simple way to analyze large volumes of unlabeled text. A ‘topic’ consists of a cluster of words that frequently occur together. Using contextual clues, topic models can connect words with similar meanings and distinguish between uses of words with multiple meanings. The GUI has two main windows - Basic and Advanced” (Google code archive, 2011b). All the 393 full-text articles are first converted into text format and processed using Topic-Modeling-Toolkit. A total of 10 hours were spent on the aforesaid process. In the toolkit, following parameters were being fixed for the study: (i) Number of topics: 5, (ii) Number of iterations: 200, (iii) Number of topic words printed: 5, and (iv) Topic proportion threshold: 0.05.

5 Results and Analysis

On the basis of the output files generated by the TMT, a comprehensive analysis has been performed in a chronological order in a yearly manner for the studied period. The files were available in both CSV and HTML format.

5.1 Topic Analysis

Tables 1 and 2 summarize the LDA result of the DJLIT articles. Table 1 shows the labeling of the topics, \( a \) through \( e \) for each year, and are organized in descending order according to their probability values (where \( a \) having the highest probability value). It summarizes the core topics for the corresponding years. Table 2 lists the word co-occurrence pattern over the time and summarizes the top 5 words or the high loading keywords, ranked by the probability value for each year in the descending order.

<table>
<thead>
<tr>
<th>Table 1: Extended Latent Dirichlet Allocation Topic Result for Corresponding Year (2008-2017)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic b</td>
</tr>
<tr>
<td>Topic c</td>
</tr>
<tr>
<td>Topic d</td>
</tr>
<tr>
<td>Topic e</td>
</tr>
</tbody>
</table>
50 articles for the year 2008 were processed. The evidence from high-loading keywords and most representative articles (Appendix-A) show that Topic a is about *bibliometrics* with a special emphasis to scientometrics in the geographical region particular to India. **Topic b** is about *open access* emphasizing that the research resources like journals be accessible to all. High-loading articles in **Topic c** show a focus on *information literacy* in libraries. **Topic d** displays a focus on *digital libraries* with a focus on virtual and online exhibitions. The representative articles from **Topic e** is on *knowledge management* using the web architecture for its users.

47 articles for the year 2009 were processed and reveals the dominance of *academic libraries in Topic a* (Appendix-A) which includes collecting data and information about the students/universities. **Topic b** indicates an interest in *bibliometrics* with an emphasis on papers published in India. **Topic c** is on *digital libraries*, with a focus on the digital access management. Research articles and keywords for **Topic d** indicate a focus on *e-learning* in libraries. **Topic e** is focused on *user studies*, particularly as a case study with respect to certain libraries.

36 articles for the year 2010 were processed and presents the keywords for **Topic a** reveals an interest in *library consortia* in particular reference to resource sharing among the libraries and institutions. A review of the representative research articles (Appendix-A) exposed **Topic b** as *bibliometrics* which is dominated by words that reflected research output in particular to Indian universities. **Topic c** has an interesting split between the keywords and representative research articles. The keywords focused on *user studies*, however, the articles share a focus on a mixture of both *bibliometrics* and *user studies* topic. Interestingly, *information literacy* dominates in **Topic d**. **Topic e** is focused on *library services*, with a particular emphasis on users, services, and technology of the libraries.

35 articles for the year 2011 were processed and shows *user studies* topic with a focus on students and the internet in **Topic a**. Representative research articles in **Topic b** (Appendix-A) suggest *knowledge management* in libraries. **Topic c** reveals an interest in *bibliometrics*. The keywords for **Topic d** indicate *OPAC* with an emphasis on searching documents. **Topic e** indicates the *library services* topic with emphasis on technology in libraries.

30 articles for the year 2012 were processed and indicates a focus on *library resources* with a special emphasis on web technology and their application in universities in **Topic a**. Research articles in **Topic b** focus primarily on *library websites* (Appendix-A). **Topic c** is an interesting split between the keywords and representative articles: the keywords focused on *digital libraries* topic and research articles share a focus on both *information literacy* and *online information*. The focus of **Topic d** is on *bibliometrics* with an emphasis on scientometrics. **Topic e** is on *library services* which focus on the application of cloud computing in libraries.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Research, Cent, Science, India, Journals</td>
<td>University, Information, Students, Data, Universities</td>
<td>Resources, Consortium, University, Institutions</td>
<td>Resources, Students, Web, Internet</td>
<td>Library, Resources, Technology, Web, Universities</td>
<td>Articles, University, Journal, Subject, Study, Total</td>
<td>Journals, Library, Users, Information, User, NIT, Law</td>
<td>Study, Library, Search, Web, Citation, Information, Online</td>
<td>Papers, Research, Science, University, Journals, Books</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
33 articles for the year 2013 were processed and reveals that Topic a is about bibliometrics with a focus on university libraries. Topic b is about citation analysis which is a method of bibliometrics. High-loading articles in Topic c show a focus on digital libraries with an emphasis on the library search, digital access and users (Appendix-A). Topic d displays a focus on library resources with an emphasis on journals, receiving information and school libraries. The representative articles from Topic e is on library services and their impact on libraries and research.

40 articles for the year 2014 were processed and shows that journal evaluation dominates in Topic a (Appendix-A). Topic b indicates an interest in library resources on Internet related to university students. Topic c is on library services with a focus on students as users and the services provided by the university libraries. Research articles and keywords for Topic d indicate bibliometrics with a focus on research and academic libraries, particularly to India. Topic e is focused on knowledge management in libraries.

36 articles for the year 2015 were processed and reveals user studies as Topic a. A review of the representative research articles (Appendix-A) shows that Topic b is dominated by words reflecting on MOOCs which is an open platform based on HTTP and is used for e-learning. Topic c is about bibliometrics with a focus on universities. Digital libraries dominate Topic d with a focus on metadata. Topic e is focused on library resources in particular to academic libraries.

36 articles for the year 2016 were processed and presents that Topic a is related to digital libraries based on the keywords and research articles (Appendix-A). Representative research articles in Topic b suggest library software as a topic with a focus on their quality and their usage in academic libraries. Topic c reveals an interest in library services, with an emphasis on patents in libraries. Keywords for Topic d indicates knowledge management in libraries as a topic with an emphasis on ontologies restricted to Indian region. Topic e indicates library resources topic with an emphasis on books and journals for the students.

50 articles for the year 2017 were processed and the emphasis that the research articles in Topic a indicate bibliometrics with an emphasis on scientometrics. Research articles in Topic b focus primarily on user studies (Appendix-A), particularly students being as respondents when querying about the library resources. Topic c is about academic libraries. The focus of Topic d is on library services. Topic e is on mobile technology in libraries.

5.2 Automated Author Analysis

Table-III summarizes the LDA result generated by TMT for each year from 2008 to 2017. It presents the top 5 authors associated with the corresponding articles in Appendix-A ranked by the probability value for each year in the descending order. It is important to consider that if same authors are coming twice under the same topic, then those authors have contributed different research articles whereas if same authors are coming under two or more topics then same research article is composed of two or more topics. This observation demonstrates that a research article can be composed of a mixture of topics.

Table 3: Extended Latent Dirichlet Allocation Author Result for Corresponding Topic a-e (2008-2017)

|--------|------|------|------|------|------|------|------|------|------|------|

Table-III summarizes the LDA result generated by TMT for each year from 2008 to 2017. It presents the top 5 authors associated with the corresponding articles in Appendix-A ranked by the probability value for each year in the descending order. It is important to consider that if same authors are coming twice under the same topic, then those authors have contributed different research articles whereas if same authors are coming under two or more topics then same research article is composed of two or more topics. This observation demonstrates that a research article can be composed of a mixture of topics.
6 Discussion

16 core topics have been identified in the study and these core topics for DJLIT journal from the year 2008 to 2017 are summarized in Table IV. Further, Table V ranks the topics from the highest number of occurrences in the journal to the lowest over the epoch. As it can be observed from Tables IV and V, Bibliometrics is the most researched topic in DJLIT which occurred in almost all the years except in 2016 followed by User Studies and the emphasis of the bibliometric method was more towards scientometrics studies in the years 2008, 2012 and 2017. The least common research topics in DJLIT are found to be e-learning, journal evaluation, library consortia, OPAC and open access which occurred only once over the 10 year period. Moreover, topics like Library software; MOOCs; and mobile technology in libraries are some of the novel areas of research in the context of India which are being introduced in the DJLIT in the past three years. Lastly, the modeled topics have been compared with the bibliometrics study conducted by Garg and Sharma (2017) in LIS discipline in India during 2004-2015. The findings from the topic analysis (Tables IV and V) in the present study somehow match the sub-disciplines in LIS founded by Garg and Sharma (2017).

Table 4: Summary of Topic Analysis

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Academic Libraries</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2.</td>
<td>Bibliometrics</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>3.</td>
<td>Digital Libraries</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>4.</td>
<td>E-Learning</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>5.</td>
<td>Information Literacy</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>6.</td>
<td>Journal Evaluation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>7.</td>
<td>Knowledge Management</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>8.</td>
<td>Library Consortia</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>S.No.</td>
<td>Modeled Topic</td>
<td>Number of Occurrences during 2008-2017</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------------------------------------------</td>
<td>----------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Bibliometrics</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>User Studies</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Digital Libraries</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Knowledge Management</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Library Services</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Academic Libraries</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Information Literacy</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Library Resources</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>E-Learning</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Journal Evaluation</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Library Consortia</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Library Software</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Mobile Technologies in Libraries</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>MOOCs</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>OPAC</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Open Access</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The present study determines not only the major research areas of DJLIT articles for each year but also the authors contributing to the work. As pointed by Lin et al. (2013), "expert finding is of vital importance for exploring scientific collaborations to increase productivity by sharing and transferring knowledge within and across different research areas". Therefore, Table-6 indicates the highly-cited authors from the bibliometrics study of DJLIT journal for the period of 2011-2015 which has been compared with Tables I and III of our study to determine if the highly-cited authors of DJLIT journal are domain-experts or not. As it can be observed, out of the 17 highly-cited authors, 7 authors from the Bapte (2017) study are not present as the subject-experts in our study. This shows how the results of highly-cited authors and works can overshadow the real domain-experts in a field.

Table 6: Identification of Domain for Highly-Cited Authors in 2011-2015 by Bapte (2017)

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Highly-Cited Author</th>
<th>Domain Expert for Modeled Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gupta, B.M.</td>
<td>Bibliometrics</td>
</tr>
<tr>
<td>2</td>
<td>Garg, K.C</td>
<td>Citation Analysis</td>
</tr>
<tr>
<td>3</td>
<td>Kademani, B.S.</td>
<td>Bibliometrics</td>
</tr>
<tr>
<td>4</td>
<td>Rousseau, R.</td>
<td>*</td>
</tr>
<tr>
<td>5</td>
<td>Ramaiah, C.K.</td>
<td>Digital Libraries and Library Resources</td>
</tr>
<tr>
<td>6</td>
<td>Prathap, G.</td>
<td>*</td>
</tr>
<tr>
<td>7</td>
<td>Egghe, L.</td>
<td>*</td>
</tr>
<tr>
<td>8</td>
<td>Sen, B.K.</td>
<td>Digital Libraries</td>
</tr>
<tr>
<td>9</td>
<td>Dhawan, S.M.</td>
<td>Bibliometrics</td>
</tr>
<tr>
<td>10</td>
<td>Kalyane, V.L.</td>
<td>*</td>
</tr>
<tr>
<td>11</td>
<td>Satija, M.P.</td>
<td>Digital Libraries</td>
</tr>
<tr>
<td>12</td>
<td>Padhi, P</td>
<td>Library Services</td>
</tr>
<tr>
<td>13</td>
<td>Glanzel, W.</td>
<td>*</td>
</tr>
<tr>
<td>14</td>
<td>Schubert, A.</td>
<td>*</td>
</tr>
<tr>
<td>15</td>
<td>Foo, S.</td>
<td>Knowledge Management</td>
</tr>
<tr>
<td>16</td>
<td>Lynch, C.</td>
<td>*</td>
</tr>
<tr>
<td>17</td>
<td>Sagar, Anil</td>
<td>Bibliometrics</td>
</tr>
</tbody>
</table>

The value added-feature of this study is that various librarians, stakeholders, publishers, and institutions from all around the world can recognize the Indian authors from Table-3 as subject-experts and domain-creators for their respective domain of research in LIS in India. These subject-experts can also be considered for numerous national or international collaborative projects by potential collaborators and institutions. Further, editors can acknowledge these subject-experts as potential reviewers in the field of LIS for their respective research topic. Recognizing experts along a mutual research domain can lead to better networking between the experts and researchers all around the world. Now, new professionals; current graduate students; and research scholars working in the LIS field can reach out to these subject-experts to receive guidance in their career. Lastly, the recognition of top-ranking authors can further help the experts to do well in their own professional career by getting more citations for their representative articles (Appendix-A) and making a strong stance in their professional society.

One of the major limitations of this study would be its small sample size but this small data size only made it possible to closely review the top core research areas, and the subject-experts related to those core areas for DJLIT on a yearly basis for the studied period. The methodological limitations of the study will include the prior identification of an appropriate number of topics for the articles before performing the Latent Dirichlet Allocation and the incompetence of the Dirichlet topic distribution to correlate among topics and manual interpretation and labeling of ‘topics’. Although some topics were fairly straightforward to label (e.g., Topic b of 2008, the top three loading words of which were (a) access, (b) OA, and (c) open), others proved more difficult to ascertain the content or methodological relationship that connected the words and the research articles. This work has broad application to those who want to know the research areas in LIS education which are highly researched and which are under-researched in India. From a methodological standpoint, this work may hold interest for biometricians/ scientometricians interested in new techniques for evaluating journals and articles in India.

7 Conclusion

In this study, 393 full-text articles from DESIDOC Journal of Library and Information Technology were analyzed according to topic-author modeling to identify the 16 core topics throughout the period of ten years. We also recognized the highly researched and under-researched areas published in DJLIT over the epoch. Thus, the current study provides a complementary lens to previous metric studies popular in India by exploring a form of analysis which can overcome the problem of skewing of results by highly cited works. Further, the top five subject-experts for the modeled topics have been identified with their representative articles. Such ranking can
encourage experts to compete with each other to do better quality research to maintain their rank in the top five author list and ultimately help to improve the quality of research in LIS in India. This study can be extended to perform prediction analysis by labeling the articles with the modeled topics and classify them to predict the topics of unlabeled articles in DJLIT.

REFERENCES


| Topic | 1. In Search of Information Literacy  
Programmes and Practices  
Survey of Selected Institutions at Bangalore  
2. Information Literacy—  
Competency Standards and Performance Indicators: An Overview  
3. Information Literacy Standards, Guidelines and their Implementation—An Analysis  
4. Information Literacy—Essential Skills for the Information Age  
5. Improving Information Skills of the Postgraduate Students of NILIS  
6. University of Colombo: A Case Study | 1. Building Subject Gateway in a Shifting Digital World  
2. Electronic Medical Records Management Systems: An Overview  
3. Access Management for Digital Repository  
4. Safeguarding the Digital Contents: Authentication and Access Control  
5. Secure Authentication Key Exchange Protocol for Credential Services  
2. Information Needs and Use Pattern of District Court Lawyers of Salem and Erode in Tamilnadu  
3. Information and Communication Technology Literacy among Library Professionals at Calcutta University, Kolkata  
4. Application of ICT and Related Manpower Problems in the College Libraries of Bhubaneshwar  
2. Information Needs and Use Pattern of District Court Lawyers of Salem and Erode in Tamilnadu  
3. Information and Communication Technology Literacy among Library Professionals at Calcutta University, Kolkata  
4. Application of ICT and Related Manpower Problems in the College Libraries of Bhubaneshwar  
2. Information Needs and Use Pattern of District Court Lawyers of Salem and Erode in Tamilnadu  
3. Information and Communication Technology Literacy among Library Professionals at Calcutta University, Kolkata  
4. Application of ICT and Related Manpower Problems in the College Libraries of Bhubaneshwar  
5. Scientometric Analysis of Literature Output of Prof. G.N. Ramachandran in the Subjects of Biophysics and Crystallography |
4. Intranet Sharing through Image Background System with Authoring Online Classification Application Considerations

5. Universal Taiwan (ROC) Archives Technology: A Case Study of a Virtual Implementation Effective than


1. Availability, Use and Barriers to ICT for Libraries: A Case Study of the Jodhpur City Public Library

1. Quality Assessment of Libraries

1. Web Interface in Library Management Software Systems


2. OAID: Designing a Web-based Information System for Aerospace Grey Literature: System Design


2. Current Trends in Information Technology in Academic Libraries of India

2. Information Seeking Strategies for University Libraries in India

1. Training Needs of School Librarians in India


1. Metadata Diversity: Interoperability and Resource Discovery

1. Digitization of Library Collections: A Case Study of the Jodhpur City Public Library

1. A Critical Look at Online Exhibitions and Online Collections: When Creating One Resource is more Effective than the Other

1. Implementation of a Virtual Archive: Developing Virtual Reality Technology: A Case Study of the National Archives Administration Taiwan (ROC)

1. Universal Decimal Classification: Past and Present

1. Semantic Web Services: A Study of Existing Technologies, Tools and Projects

1. Online Virtual Exhibitions: Concepts and Designs

1. Application Domain and Annotational Classification Recommendation—A Survey

1. Exhibition Authoring System with Intelligent Affective Background Image Composition

1. Knowledge Sharing through Internet

Topic 2

1. An Overview of Online Archives

2. Voices: A Virtual Exhibition as an Awareness Model to the Forum 2004

3. A Critical Look at Online Exhibitions and Online Collections: When Creating One Resource is more Effective than the Other

4. Implementation of a Virtual Archive: Developing Virtual Reality Technology: A Case Study of the National Archives Administration Taiwan (ROC)

5. Universal Decimal Classification: Past and Present

Topic 3

1. Semantic Web Services: A Study of Existing Technologies, Tools and Projects

2. Online Virtual Exhibitions: Concepts and Designs

3. Application Domain and Annotational Classification Recommendation—A Survey

4. Exhibition Authoring System with Intelligent Affective Background Image Composition

5. Knowledge Sharing through Internet

1. Availability, Use and Barriers to ICT for Libraries: A Case Study of the Jodhpur City Public Library

2. Current Trends in Information Technology in Academic Libraries of India

3. Websites of Indian Institutes of Technology Libraries


5. Implication of ICT in Libraries of Indian Universities

1. Library Management Software Systems

2. Current Trends in Information Technology in Academic Libraries of India

3. Websites of Indian Institutes of Technology Libraries


5. Implication of ICT in Libraries of Indian Universities

1. Web Interface in Library Management Software Systems

2. Current Trends in Information Technology in Academic Libraries of India

3. Websites of Indian Institutes of Technology Libraries


5. Implication of ICT in Libraries of Indian Universities

1. Library Management Software Systems

2. Current Trends in Information Technology in Academic Libraries of India

3. Websites of Indian Institutes of Technology Libraries


5. Implication of ICT in Libraries of Indian Universities

1. Web Interface in Library Management Software Systems

2. Current Trends in Information Technology in Academic Libraries of India

3. Websites of Indian Institutes of Technology Libraries


5. Implication of ICT in Libraries of Indian Universities

1. Library Management Software Systems

2. Current Trends in Information Technology in Academic Libraries of India

3. Websites of Indian Institutes of Technology Libraries


5. Implication of ICT in Libraries of Indian Universities

1. Web Interface in Library Management Software Systems

2. Current Trends in Information Technology in Academic Libraries of India

3. Websites of Indian Institutes of Technology Libraries


5. Implication of ICT in Libraries of Indian Universities

1. Library Management Software Systems

2. Current Trends in Information Technology in Academic Libraries of India

3. Websites of Indian Institutes of Technology Libraries


5. Implication of ICT in Libraries of Indian Universities