Topic and Trend Analysis of LIS Research Articles in the UGC-CARE Group I Accredited LIS Journals Using Voyant Tools

Vinay Anand and Kumar Gaurav

This study aims to identify specific topics and trends of academic journals, based on 435 articles published for the limited period of 2021 to 2022 in the UGC-CARE Group I approved LIS journals through Voyant tools. Using this tool, topic modelling, term frequency analysis, trend analysis and keyword context analysis was conducted. The results found ten topics for research articles that were published in the various UGC-CARE listed Group I LIS journals in the limited period of the time. The analysis reveals that most of the research published in UGC-CARE Group I LIS, deals with the topics of diverse library applications and various case studies practices. The results also reveals that the most of the articles are written in the Indian context. It indicates that the researchers write their articles in the Indian prespective, is the most prevanlent in UGC-CARE listed Group I journals.

Introduction

A massive volume of literature is produced by research that is expanding quickly. As time goes on, new issues arise and outmoded ones are replaced. It is crucial for researchers to be aware of emerging subjects, ideas, or disciplines that are relevant to their particular area of study, so that they may gradually recognize new ideas and disciplines. Thus, bibliometric investigations are carried out in the field of library and information science. For the purpose of sharing information finding and decision-making, platforms like social networks, websites, blogs, and academic journals produce a large volume of unstructured text data. It is crucial to evaluate, synthesize, and process this data. In order to manage such massive amounts of data and extract the necessary knowledge for decision-making, this presents a difficulty. New tools and methods are needed to manage this boom of electronic documents properly. In the past ten years, machine learning and statistics have created new method known as topic modeling—for identifying word trends in big collections of texts. At this point**,** analyzing the research articles in the academic journals from the perspective of text mining,may offer logical information to researchers community.

Therefore, The purpose of this study is to identify the topics of research articles published in UGC-CARE Listed LIS Group I academic journals by using topic modelling technique. Using topic modeling techniques, we can organize, comprehend, and summarize massive volumes of textual data. It helps in identifying topical patternsacross the collection, adding theme annotations to documents andmaking use of themes to arrange, search, and sum up texts.

2. Literature Review

Topic modeling is a well-known technique in the Data mining field. Data analysts use various tools and algorithms to analyze the text or data. Here we are trying to include a few background studies to properly understand how topic moedellingor text mining technique is beneficial.

A study by Kaila (2020) focused on the information flow on Twitter during the coronavirus outbreak. They extracted the tweets related to coronavirus and studied the information flow using the sentiment analysis and topic modeling technique. Of equal importance, authors evaluated two online social networks Twitter and Reddit, with a topic modeling technique for demonstrating a method for interpreting the clusters with topwords (Curiskis et al., 2020). The same technique was used by Moro et al. (2019) to enhance the automated approach for establishing the dimensions of ethnic marketing research. Furthermore, the study Zafari and Ekin (2019) proposed using topic models to group drugs with respect to the billing patterns and exhibit the potential aberrant behaviours while using medical specialties as a covariate.

Few studies on the library and information science domain are found regarding topic modelling.

Lamba and Madhusudhan (2019) analyzed full text research article (retrieved from DESIDOC Journal of Library Information Technology), tagged with the modelled topics using a topic mining algorithm Latent Dirichlet Allocation (LDA). Again, major modification, Manika Lamba and Madhusudhan (2019) described the importance and usage of metadata tagging and prediction modelling tools for researchers and librarians. They use Topic Modelling Toolkit and RapidMiner toolbox for employed prediction for topic mining. Mazumder and Barui (2021) analyzed the titles of the Indian LIS theses for discovering the topics using topic modelling LDA. They discovered ten topics where the topic 'Library use' was the mjor topic.

The study by Katsurai (2021) has explored the recent trends of adopting data minning method in library and information science. Their study found that the recent popularity of machine learning techniques in LIS research is increasing. The result of this study has been proved b.y Pawde, (2021) to study the identifying emerging information needs of library users using the data mining method. The author proposed a single data mining technique that analyses the past user queries and accurately identifies under-provisioned and under-stocked information needs.

As a result of integrating these studies, three points were clarified. First, the topic modelling was well known method in information communication technology domain. Second, in library and information science domain data for research could be collected from both journals and databases. Third, LDA algorithm was mostly prefer for topic modelling. Based on the three points and to achieve the objectives of the study we used an open source text analysis tool for topic modelling named Voyant tools, and term frequency analysis, trend analysis and keyword context analysis were performed.

3. Aim and Objectives of Study

- To identify the topics of UGC-CARE-listed Group I LIS Journals
- To findout the trends of research articles in UGC-CARE-listed Group I Library and Information science journals from 2021 to 2022.
- To find an efficient text analysis tool for topic modeling.

4. Metohdology

Four methodologies were used in this study. Topic modelling was performed on the title of 435 articles published from 2021 to 2022 in the UGC-CARE listed LIS Group I journals to find the topics. Term frequency analysis was used to identify prominent theme of the topics. Trend analysis was conducted to find out the trends of topics. At the end, Keyword Contexts analysis was conducted to find out the left and right context of the terms.

4.1. Data Extraction and collection

The data was collected from the UGC-CARE Group I listed LIS journals to search for articles. A web scrapping tool was used for data extraction from various LIS journals of UGC-CARE Group I. The retrieve was targeted on the title and conducted on the limited period from 2021 to 2022. As a result 435 articles were extracted. The title of articles were selected as data for text analysis because title of the article is the short reflection of the content of the research article.

4.2. Loading of Data

Voyant Tools is a web-based reading and analysis environment for digital texts, supporting various data formats, viz Corpus, Text, XML, HTML, JSON, Tables, and URLs. We imported our datasets into XML format and uploaded them into Voyant tools. A few seconds of processing voyant tools revealed the results.

4.3. Data Pre-processing

During the process of data extraction, other than title of the article, we got authors, publication, year, abstract, and keywords data. In this study our main focus was to analyse the title of the article, so we went through the data preprocessing phase to remove this unwanted data from the dataset. Although in this study we were using an open-source text analysis tool 'Voyant tools' for data analysis or text analysis, so we did not performed tokenization, lemmatization and stemming of words, because the Voyant tools has in built Stanford Core Natural Language Processing package, that automatically tokenized, lemmatized and stemmed the words.

4.4. Voyant Tools

Voyant Tools is an open-source, web-based application for performing text and data mining. It was developed by Stefan Sinclair at McGill University and Geoffrey Rockwell at the University of Alberta. It was a scholarly

project designed to facilitate reading and interpretive practices for digital humanities students. In Voyant tools, we used topic modelling, term frequency analysis, trends analysis and keyword context analysis to achieve our objectives of the study (Alhudithi, 2021).

5. UGC -CARE : A Quality Mandate for Indian Academia

The University Grants Commission (UGC) seeks to support and empower the Indian academic community through its "Quality Mandate" in order to bring all academic subjects within its purview up to par with international standards of high-quality research. On November 28, 2018, the UGC made a public announcement announcing the creation of a specialised Consortium for Academic and Research Ethics (CARE) to serve this goal.

5.1. Objectives of the UGC-CARE LIST

- To support academic honesty, high standards for research, and ethical publishing practises at Indian institutions.
- To encourage reputable journals to publish high-caliber work that will help it rise in the rankings internationally.
- To provide a strategy and process for locating high calibre journals.
- To stop academics from publishing work in journals that are predatory, questionable, or subpar, which reflects poorly on Indian academia and tarnishes its reputation.
- To promote quality research, academic integrity, and publication ethics
- To create and maintain a "UGC-CARE Reference List of Quality Journals" (UGC-CARE List) for all academic purposes.

5.2. Need for UGC-CARE List

- Research publications' reliability is crucial since it reflects the academic reputation of the institution, the country, and not simply the author.
- Predatory, questionable, and Non-standard journals have grown to be a major global concern.
- Publishing in questionable or Non-standard journals negatively affects academic performance over the long run and tarnishes a person's reputation as well as that of their institute and their country.
- According to reports, India has a high number of research publications that are published in low-quality journals, which has a negative impact on the country's reputation.

5.3. UGC-CARE LIST

The UGC-CARE List includes only TWO groups, making the search procedure easier. These are not ordered or hiearachical categories.

- **UGC-CARE List Group I:** Journals approved under UGC-CARE guidelines.
- **UGC-CARE List Group II:** Journals indexed in globally renowned international databases.

5.4. UGC-CARE Listed Group I LIS JOURNALS

The UGC-CARE List has been divided into four categories; science, social sciences, arts and humanities, and multidisciplinary based on Scopus's All Science Journal Classification (ASJC) classifications (Elsevier Science). Journals indexed in Web of Science and Scopus or discontinued are not included in the count.

Researchers can search for journals in the UGC-CARE portal using the criteria of Subject, Title, ISSN, Publisher, and Language of Publication. For this study we searched UGC-CARE Listed Group I by Subject "Library and Information Science" and found total 20 journals where 1 journal now discountinued since April 2022 (see Table 1).

Table 1: UGC-CARE Listed LIS Journal (Source :https://ugccare.unipune.ac.in/)

Sr. No.	Journal Title	Publisher	ISSN
1	Art Libraries Journal	Cambridge University Press	0307-4722
2	Catholic Library World	Catholic Library Association	0008-820X
3	Citaliste: the Scientific Journal on Theory and Practice of Librarianship	The City Library of Panevo and Faculty of Philosophy of the University of Novi Sad	2217-5563
4	College Libraries	West Bengal College Librarians Association	0972-1975
5	Georgia Library Quarterly	Georgia Library Association	2157-0396
6	IASLIC Bulletin	Indian Association of Special Libraries and Information Centres	0018-8441
7	International Journal of Information Retrieval Research	IGI Global	2155-6377
8	Journal of Indian Library Association	Indian Library Association	2277-5145
9	Journal of Library and Information Studies	Department of Library and Information Science, National Taiwan University	1606-7509

10	Journal of the Canadian Health Libraries Association	Canadian Health Libraries Association	1708-6892
11	Journal of University Librarians Association of Sri Lanka	University Librarians Association of Sri Lanka	NA
12	KELPRO Bulletin	Kerala Library Professionals Organisation	0975-4911
13	Knowledge Quest	American Library Association	1094-9046
14	Library Herald	Delhi Library Association	0024-2292
15	New Review of Children's Literature and Librarianship	Taylor and Francis	1361-4541
16	RBU Journal of Library and Information Science	Department of Library and Information Science, Rabindra Bharati University	0972-2750
17	South African Journal of Libraries and Information Science	Library and Information Association of South Africa	0256-8861
18	SRELS Journal of Information Management	Informatics Publishing Limited and Sarada Ranganathan Endowment for Library Science	0972-2467
19	Virginia Libraries(discontinued)	Virginia Libraries Association	NA
20	World Digital Libraries: An International Journal	Energy and Resources Institute	0974-567X

6. Results

6.1. Topic Modelling

Topic modelling was performed using MALLET (MAching Learning LanguagE Toolkit), an inbuilt package in Voyant tools. After processing of the dataset in Voyant tools, a corpus of one document was created, having a total of 5,697 words and 1,599 unique word forms. The tool was calculated vocabulary density as 0.281, readability index as 17.429, and average words per sentence as 271.3. As a result of MALLET topic modelling in voyanttool, we found 10 topics and extracted representative words(Table 2).

Table. 2: Topics of articles in UGC-CARE Listed Group I journals

Topic 1	Library
Topic 2	Study
Topic 3	research
Topic 4	information
Topic 5	analysis
Topic 6	university
Topic 7	libraries
Topic 8	india
Topic 9	science
Topic 10	scientometirc

6.2. Term Frequency Analysis

The top 20 words in the topic and the term frequency (TF) were extracted. The frequency of words was calculated by inbuilt Latent Dirichlet Allocation (LDA) algorithm available in Voyant tools. Relative count was also calculated by the LDA. The results shows that the term 'library' has the highest frequency and the highest relative count(Table 3). If we see top ten terms and its frequency count then we may understand that most used terms in the UGC-CARE listed LIS journals'research articles. The most used terms are 'library'. 'study', 'research', 'information', 'analysis', 'university', 'libraries', 'india', 'scinece', and 'scientometric'.

Table 3: Top 20 frequent terms

S.No.	Term	Count	Relative
1.	library	107	18781.81
2.	study	97	17026.5
3.	research	66	11585.05
4.	information	65	11409.51
5.	analysis	62	10882.92
6.	university	57	10005.27
7.	libraries	55	9654.204
8.	india	38	6670.177
9.	science	34	5968.053
10.	scientometric	34	5968.053

11.	case	30	5265.929
12.	covid	30	5265.929
13.	students	29	5090.398
14.	services	24	4212.744
15.	academic	22	3861.682
16.	assessment	22	3861.682
17.	librarians	22	3861.682
18.	social	21	3686.151
19.	using	21	3686.151
20.	global	20	3510.62

A wordcloud was formed of 50 terms, helped to visualize the terms. In Voyant tools wordcloud is known as 'Cirrus'. The wordcloud postions the word such that the terms that occur the most frequently are bigger in size. In Fig. 1 the terms 'library' 'study' and 'research' are the top three highest frequent terms, so that the letter of these terms are bigger in size. As the algorithm goes through the list and continues to attempt to draw words as close as possible to the center of the visualization it will also include small words such as citation, data, art, impact, and health within space left by larger words that do not fit together snugly .

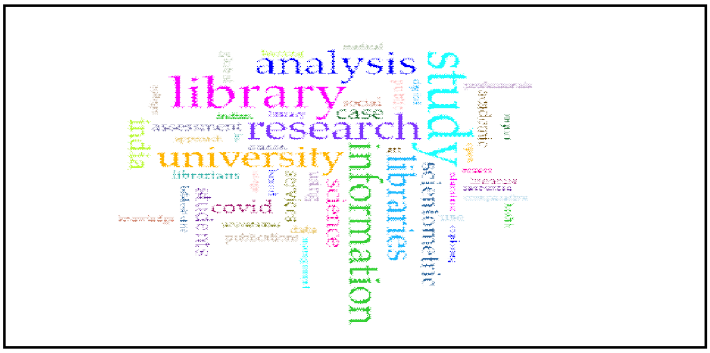

Fig 1: Visualization of Terms into Wordcloud

6.3. Trend Analysis

We conducted a trend analysis, shows a line graph of the most frequent words used in a corpus. Each series in the graph is coloured according to the word it represnets. At the top of the graph a gegend displays which words are associated with certain colours. By default, the trends tool shows the relative frequencies of words in corpus. The line graph shows the term library and the term study in the upward direction and the terms 'analysis', 'information and research are colliding several times and in downword direction, that may show, the case studies based practices related to library, is the current research trends of UGC-CARE listed LIS journal (Fig. 2).

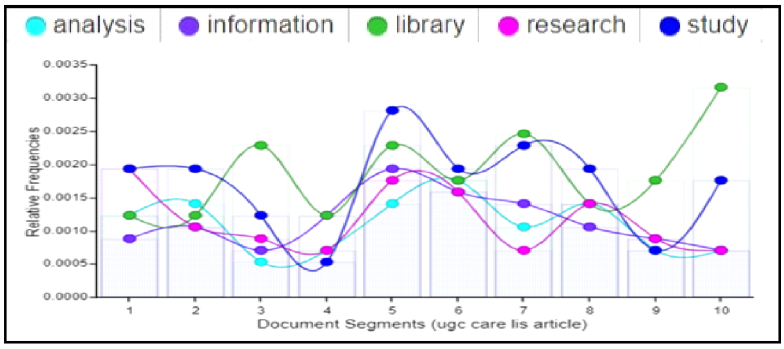

Fig 2. Trends of Terms and Relative Frequencies

6.4. Keyword Contexts Analysis

The Context Analysis was performed to show each occurrence of a keyword with a bit of surrounding text (the context). It can be useful for studying more closely how terms are used in different contexts. Table 4 shows the left and right context of the top ten terms. Services and Facilities related to library and library website content analysis are the two context of the term 'library'. Whereas for the term 'study' we found that the study related to scholars data collection and the study related to faculty of social science are the left and right context.

Table 4: Keyword Context of the terms

Left Context	terms	RightContext
Services and Facilities Available in	library	Websites: A Content Analysis of
scholars in Data collection: A	study	of the faculty of Social
Aligarh Muslim University Big Data	research	publications in Library and Information
Performance among Library Personnel Health	information	Literacy among Rural women of
in Library Websites: A Content	analysis	of Indian Institutes of Management
Information Science students of Panjab	university	Chandigarh, India: A Study Information
towards Adopting Artificial Intelligence in	libraries	Impact of MOOC on Agricultural
The National Digital library of	india	(NDLI):- A single window platform
performance of library and information	science	(LIS) academia from central universities
Covid-19 and psychology: A	scientometric	assessment Scientometric analysis of image

7. Discussion

Table 3 shows the terms that appear most frequently in the titles of the articles and also the relative value, which shows how a term relates to the different titles of the articles. See Table 4 for a clearer understanding

of how the terms are related to the article. In this study, the articles' trend was mined based on the terms for the years 2020 and 2021. The analysis reveals that most of the research published in UGC -CARE Group I LIS publications, deals with the topic of diverse library applications. Additionally, case studies based practices of various kinds are highly common in these publications. If we see topics in (Table 2), the frequency of the term 'India' is under the top 10, indicates that researchers are keenly engaged in case studiesbased research. It also indicates that the Indian perspective of the research article are the most prevalent in the UGC-CARE listed Group journals. Additionally, another well-known and well-liked field in the LIS is 'scientometrics'. Unexpectedly, 'COVID' has emerged as a brand-new category of the study topic. Since we selected the papers during the period of the pandemic, some study has allowed us to evaluate COVID through other LIS eyes.

At the end we found an efficienct open source tool 'Voyant tools' for text mining and topic modelling. This tool empowers the researchers for performing text mining and data visulization without any coding.

8. Conclusion

Machine learning is a powerful technique for text analysis since it enables us to examine the text, identify the emotions, and model the topics. The results above may be used to identify the most popular study areas in the UGC-CARE -listed Group I LIS journals. Through this study, Indian LIS researchers may learn more about the popular research areas in these journals. Additionally, it expedites the discovery of study trends by researchers and the effective tool for text mining and topic modelling. The fourth law of library and information science, "To save the time of users", is therefore satisfied by this method, which also serves as a quick information retrieval tool for researchers.

References

1. About - Voyant Tools Help. (2022). https://voyant-tools.org/docs/#!/guide/about

2. Alhudithi, E. (2021). Review of Voyant Tools: See through your text. Language Learning & Technology. 25(3), 43–50.

3. Curiskis, S. A., Drake, B., Osborn, T. R., & Kennedy, P. J. (2020). An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. Information Processing & Management, 57(2), 102034. https://doi.org/10.1016/j.ipm.2019.04.002

4. Kaila, D. R. P. (2020). INFORMATIONAL FLOW ON TWITTER - CORONA VIRUS OUTBREAK – TOPIC MODELLING APPROACH. 7. http://files/300/Kaila - INFORMATIONAL FLOW ON TWITTER - CORONA VIRUS OUTBR.pdf

5. Katsurai, M. (2021). Adoption of Data Mining Methods in the Discipline of Library and Information Science. 17. http://files/306/Katsurai - 2021 - Adoption of Data Mining Methods in the Discipline .pdf

6. Lamba, M, & Madhusudhan, M. (2019). Metadata Tagging and Prediction Modeling: Case Study of DESIDOC Journal of Library and Information Technology (2008–17). 57. http://files/316/Lamba and Madhusudhan - Metadata Tagging and Prediction Modeling Case Stu.pdf

7. Lamba, Manika, & 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

8. Mazumder, S., & Barui, T. (2021). Discovering Topics from the Titles of the Indian LIS Theses Discovering Topics from the Titles of the Indian LIS Theses Discovering Topics from the Titles of the Indian LIS Theses. https://digitalcommons.unl.edu/libphilprac

9. Moro, S., Pires, G., Rita, P., & Cortez, P. (2019). A text mining and topic modelling perspective of ethnic marketing research. Journal of Business Research, 103, 275–285. https://doi.org/10.1016/j.jbusres.2019.01.053

10. Pawde, A. (2021). IDENTIFYING EMERGING INFORMATION NEEDS OF LIBRARY USERS USING DATA MINING. 57, 10. http://files/318/Pawde - 2021 - IDENTIFYING EMERGING INFORMATION NEEDS OF LIBRARY .pdf

11. Zafari, B., & Ekin, T. (2019). Topic modelling for medical prescription fraud and abuse detection. Journal of the Royal Statistical Society: Series C (Applied Statistics), 68(3), 751–769. https://doi.org/10.1111/rssc.12332

Keywords: LIS Research; UGC-CARE List; LIS Journals; Topic Modelling; Research Trends; Text Analysis

About Authors

Mr. Vinay Anand
Ph.D. Research Scholar
Department of Library and Information Sciencet
University of North Bengal, WB, India
Email: rs_vinay@nbu.ac.in

Mr. Kumar Gaurav
Library Assosication of Bihar
Email: kgaurav525@gmail.com