Journal of Scientometric Research, 2024; 13(1):161-175.
https://www.jscires.org

Research Article

Visualising Knowledge, Research Hotspots and Trends of Literacy Studies in the Context of Library, 1969-2021

Anupta Jana*, Rosalien Rout

School of Library and Information Science, Gangadhar Meher University, Sambalpur, Odisha, INDIA.

ABSTRACT

In this study, we conducted an in-depth analysis spanning 53 years, from 1969 to 2021, focusing on the field of literacy studies within the context of libraries. Our exploration involved a dataset of 4,986 articles retrieved from the Scopus database. Our primary objective was to visualize knowledge by identifying and exploring prominent trends and hotspots in literacy studies. To achieve this, we adopted a comprehensive approach. The methodology employed in this study combined traditional approaches with contemporary tools. The dataset was analyzed using the R software for conventional methodologies, while MATLAB was utilized for cutting-edge techniques. The multifaceted approach allowed us to uncover patterns of continuous growth, identify key contributors, and employ the Latent Dirichlet Allocation (LDA) model to recognize emerging and significant topics. The study revealed a consistent pattern of continuous growth in the field of literacy studies, indicating the acquisition of new knowledge over time. Key contributors, including productive authors, influential journals, and active countries, were identified. The application of the LDA model enabled us to recognize newly emerged, developed, and important topics. The significance of this research lies in its contribution to understanding the dynamic landscape of literacy studies within library contexts, offering valuable insights for future research and practical applications in the field.

Keywords: Literacy, Information Literacy, Library, Bibliometrics, Topic model, MATLAB, Library and Information Science.

Correspondence:
Anupta Jana
School of Library and Information Science, Gangadhar Meher University, Sambalpur, Odisha, INDIA.
Email: anuptajana@gmail.com
ORCID: 0000-0002-2863-0691

Received: 11-03-2023;
Revised: 11-10-2023;
Accepted: 30-01-2024.

INTRODUCTION

Literacy involves the ability to read, write, speak, and listen effectively, enabling individuals to communicate with the world. A lack of essential literacy skills can hinder a person's progress at every stage of life.[1] Over time, different literacy skills have evolved, with the concept of an information-based society emerging in the 1950s after World War II.[2] The term 'information society' was coined by economist Fritz Machlup.[3] Since then, information has played a crucial role in economic and cultural activities in terms of creation, distribution, and manipulation. However, with the continuous growth of information, concerns about misinformation and disinformation have also risen, such as fake news, propaganda, clickbait, hoaxes, and rumours. To address these issues, individuals need to adopt various literacy programs, and library and information science professionals can play a vital role in providing guidance. Libraries and professionals have resorted to various literacy methods to prevent misinformation, including library instructions, user and library orientation, OPAC instruction, information literacy instruction, media literacy, digital literacy instruction, and more. Studies have been conducted to analyse and explore the roles of libraries and professionals in promoting literacy among user communities. This study attempts to conduct a bibliometric analysis using a variety of statistical models to recognize the state of literature and explore the structure of knowledge, assuming that "data speak for themselves".[4,5]

LITERATURE REVIEW

Despite the brief history of bibliometric studies on literacy in the Library and Information Science field, they have been conducted for several reasons, including:

In their bibliometric study, Pinto *et al.*[6] used statistical, mathematical, and content analyses to investigate international scientific productivity in Information Literacy (IL) literature published between 1974 and 2011, as indexed in the Web of Science (WoS) and Scopus databases. The authors extracted results about authors' productivity, distribution of disciplines, co-authorship, affiliations, and frequently used journals. The study's findings showed continuous growth in literature and the prevalence of Anglo-Saxon authors. The authors' affiliation demonstrated significant dispersion, and the study confirmed the

DOI: 10.5530/jscires.13.1.14

fit of Lotka's law and Bradford's law. To investigate research trends on IL, Shaheen *et al*.[7] conducted a bibliometric study on the data published between 2003 to 2012 in the Scopus database. The study examined the annual growth rate of preferred journals for article publication, authors' productivity, countries' contribution, and subject-wise distribution of publications. Analysis of the data revealed a steady increase in literature over the past decade, with authors from North America and the United Kingdom showing higher productivity compared to other countries. A similar study was also conducted by Shankar Reddy Kolle[8] in a publication related to information literacy literature between 2005 and 2014. The study found that annual growth was higher in 2007, 2008, and 2011. The most productive author and institute were identified as "Pinto, M." and the "University of Granada, Spain," respectively. The "Journal of Academic Librarianship" was the most commonly preferred. The study also found that the contribution of the USA was high. The research topics of "digital divide," "media literacy," "pedagogy," "higher education," and "critical thinking" were identified as prominent in the field of Information Literacy. In the study by Parlina, Ramli, and Murfi,[9] a decade of big data research (2009-2018) is scrutinized using bibliometrics and text mining. Results reveal a surge in publications since 2014. Core journals, prolific authors, and influential entities are identified. A unique thematic analysis uncovers dominant themes, with big data analytics standing out, covering methods, tools, infrastructure, and applications. Security, privacy, social networks, and the Internet of Things emerge as critical aspects. To assess the scientific productivity of mobile IL literature in higher education Pinto *et al*.[10] conducted a bibliometric study. The study analysed papers covered by Library and Information Science Abstracts (LISA), Library and Information Science and Technology Abstracts (LISTA), Web of Science (WoS), Scopus, and Education Resources Information Centre (ERIC) published between 2006 and 2017. The study's findings provided an overview of mobile information literacy and identified the most relevant journals, average authors' productivity, productive authors, and research trends in this field. To analyse the evaluation of IL in literature indexed by the Scopus database between 1975 and 2018, Omwoyo[11] conducted a bibliometric. The study showed that the concept of IL has expanded beyond the field of social science and into various disciplines. It also revealed the emergence of different types of literacy such as content literacy, media literacy, health literacy, metaliteracy, and business-related literacy after 2000. Additionally, the paper proposed a collaborative and multidisciplinary approach to research activities in the field of IL. Apart from all these traditional studies, with a modern approach, Li *et al*.[12] conducted a study using topic modelling to evaluate subject characteristics of IL topics published between 2005 and 2019. Latent Dirichlet Allocation (LDA), introduced by Blei *et al*.[13] in 2003, serves as the foundation for topic modelling in this study. LDA operates as a probabilistic model, representing each document within a corpus as a random mixture across latent topics. These latent topics, in turn, are defined by distributions over words, encapsulating the probabilistic relationships between documents and the underlying thematic content. The study found that core topic words remained relatively stable, while also identifying new research directions published in recent years.

Based on the above discussion, it is evident that numerous studies have already been conducted, particularly in the field of IL, to identify the productivity of authors, institutions, countries, and keyword frequencies. However, this area of study has not adequately explored the utilization of topic modelling techniques to gain insights into library-related aspects within the context of Information Literacy. Furthermore, comprehensive studies that delve into historical developments and focal points in literacy within library settings are noticeably lacking. The study aims to fill this gap by providing a holistic analysis of the literature from 1969 to 2021, shedding light on the knowledge and research trends in this specific area. Specifically, the study seeks to answer:

How has scholarly productivity evolved in the field of literacy in libraries during the specified period?

What is the knowledge structure within this field, and how do different research themes relate to each other?

Objectives of the study

This study aims to assess the academic literature on literacy that was implemented in libraries from 1969 to 2021. To achieve this goal, the study has the following objectives:

To assess the scholarly productivity and evolution of literacy in library research during these 52 years through a bibliometric analysis.

To reveal the knowledge structure within this field, utilizing probabilistic topic modelling.

METHODOLOGY

In a nutshell, our methodology blends bibliometrics and probabilistic topic modelling to gain insights into the academic literature on library-based literacy over the specified time frame.

Data source

To conduct our analysis, we utilized a bibliographic dataset, extracted from the SCOPUS database that spans from 1969 to 2021. We chose this database due to its comprehensive features, widespread usage as one of the leading bibliographic databases, and its inclusion of only peer-reviewed publications.

Data inclusion criteria

To narrow down our focus on literacy in the library context, we conducted an advanced search using specific queries (mentioned in the textbox below) in the SCOPUS database, resulting in the retrieval of 4986 relevant documents.

TITLE-ABS-KEY ("literac*" AND "librar*") AND PUBYEAR > 1968 AND PUBYEAR < 2022 AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English"))

Data summary

On December 31st, 2022, we extracted the bibliographic data, for this study. We restricted our bibliographic dataset to include only sources published in the article format and written in the English language from 1969 to 2021.

Bibliometric analysis

To assess the status of the literature based on the bibliographic data (such as authors, journals, countries, and author keywords) of the retrieved documents, we employed the bibliometrix package of R, a statistical computing language.[14] Biblioshiny, Java software, was created by Massimo Aria at the University of Naples Federico. It seamlessly merges the capabilities of the bibliometrix package with the user-friendly interface of web applications.[15,16] Descriptive statistics were used to evaluate the productivity of authors, journal sources, and countries, as well as the occurrence of keywords related to the themes of the studies in this field.

The keywords used in the articles can provide an accurate explanation of the topics covered. Therefore, analyzing the networks of keyword co-occurrence can reveal the connections among the topics covered in the articles. Implementing this method helped to identify the research themes and establish relationships among them.

Probabilistic topic modelling

Additionally, to reveal the underlying structure of knowledge from the titles, author's keywords, and abstracts of the documents in the field of interest, we utilized a probabilistic topic model called Latent Dirichlet Allocation (LDA). LDA is a technique that enables automatic topic mining by identifying clusters of word co-occurrences to uncover hidden topics or thematic structures within documents.[12,13] In this model, documents are viewed as a combination of topics and these topics are viewed as a combination of words. These topics generate words through probability distributions. Documents are therefore considered as a distribution of topics and topics are considered as a distribution of words. This model is advantageous due to its exchangeability of words and documents among the topics, leading to improvement. The LDA model is performed in two phases. In the first phase, the distribution of topics is arbitrarily selected. In the second phase, for each word in the documents, a topic is randomly assigned from the distribution of topics.[17,18] This model can be expressed mathematically as follows,

Assuming we have a set of documents D, where, each of the documents is a square of N words, $=(w_{d,1}, w_{d,2}, \ldots, w_{d,N})$ and each word is an item from vocabulary indexed by {1, …, V}. Let there be K latent topics $\beta_{1:K}$, representing the distribution of the vocabulary. Based on this model, the posterior distribution will be

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{k=1}^{K} p(\beta_K) \prod_{d=1}^{D} p(\theta_d) \left(\prod_{n=1}^{N} p(z_{d,n}|\theta_d) p(w_{d,n}|\beta_{1:K}, \beta_{d,n})\right)$$

Where,

θ_d = Topic proportions of the d^{th} document,

z_d = Topic assingments for d^{th} documents,

w_d = set of observed words.

A set of documents was prepared for analysis with the LDA model by combining the corresponding articles' titles, abstracts, and authors' keywords. Articles without abstracts were removed. Additionally, the documents underwent several cleaning and processing steps, such as lowercasing, tokenization, removing stop words, normalizing words, erasing punctuation, and removing short words. The '*Text Analytics Toolbox*' of MATLAB 2018 software was used to perform these steps.[19]

Limitation

The dataset extracted for this study from the Scopus database, which is comprehensive, but may not capture all pertinent articles and could be subject to inclusion and coverage biases. It purely concentrates on English-written articles, potentially introducing language bias by omitting crucial research conducted in other languages. Given its bibliometric nature, the study might not consider other factors such as research impact on practice, societal relevance, etc.

Literacy studies in the Library Science domain

Research trends

Literacy studies in the Library Science domain show notable research trends. The dataset, covering the period from 1969 to 2021, indicates that 8,121 authors contributed to 4,986 publications from 733 distinct sources. The dataset has been growing at an average annual rate of 12.2% (Figure 1). This suggests that the number of documents or sources has been increasing steadily over the years. On average, each document in the dataset has been cited 12.03 times. This is a measure of the impact or influence of the documents. On average, each document is cited once per year. This is another way of assessing the frequency of citations. There are a total of 7,600 author-generated keywords associated with the documents. These keywords are often used to describe the main topics or themes of the research. 8,121 unique authors have contributed to the documents in the dataset. Out of the total authors, 1,660 have authored documents on their own without co-authors. There are 2,048 documents in the dataset that have been authored by a single author without collaboration. On average, each document has 2.11 co-authors. This indicates that

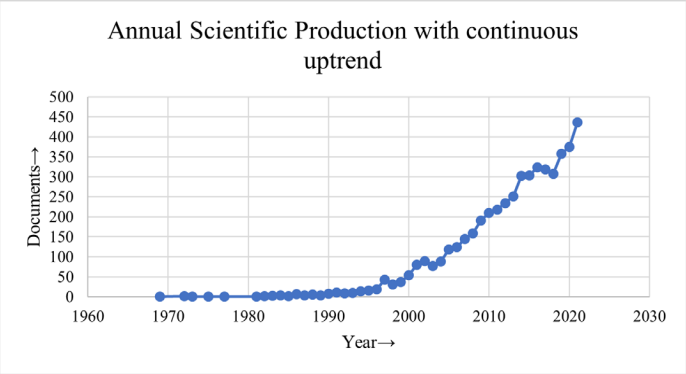

Figure 1: Annual scientific production of the documents with continuous growth.

most documents are the result of collaboration between multiple authors.

Aparicio *et al.*[20] proposed that each scientific production undergoes three stages of growth. Using their framework, we identified that during the *initial stage* (1969-1996), an average of 4.5 articles per year were published, resulting in a total of 126 documents. In the *development stage* (1997-2004), the publication rate increased to 62.25 articles per year, resulting in a total of 498 documents. Finally, in the *expansive stage* (2005-2021), the publication rate further increased to 256.58 articles per year, resulting in a total of 4362 documents.

Influential authors

Table 1 provides data on the research productivity of various authors, considering the number of articles they have published, the total citations their work has received, the year of their first publication, and their years' contribution to this field as of 2021. This information provides valuable insights into the authors' contributions and impact within their respective domains.

Julien, H., with 28 articles and 562 citations since 1997, appears to be a highly productive and influential researcher, showcasing a consistently impactful track record in the academic community. Despite publishing fewer articles (19), *Pinto, M.*, has accrued a notable 314 citations since 2008, indicating substantial recognition within a relatively short timeframe. *Lloyd, A., Gross, M.,* and *Latham, D.* exhibit comparable numbers of articles and citations, signifying that their work since 2003 and 2007, respectively, has consistently left an impact on their respective fields. *Saunders, L., Spring, H., Ameen, K.,* and *Merga, M.K.* all having commenced their research journey in the mid to late 2000s, showcase variability in their article numbers, with their total citation counts reflecting diverse levels of influence within the academic community. *Hicks, A.*, started publishing in 2010 with 10 articles and 39 total citations, and is in the early stages of the research career, with the potential for further growth in impact.

Global Research Productivity and Collaboration

Table 2 provides a comprehensive overview of research productivity and collaboration patterns among countries, considering both the number of articles produced and the citation impact.

In the first part of Table 2, countries were ranked based on the number of articles they produced. The United States, with 1985 articles, is the clear leader, signifying its substantial research output, and is followed by the UK, Canada, Australia, and other countries. It indicates countries' contribution to global research. The Table 2 reveals the percentage of Single-Country Publications (SCP) versus Multiple-Country Publications (MCP) for each country. A higher percentage of SCP suggests a more independent research landscape, while a higher percentage of MCP implies greater international collaboration. In this context, Nigeria, with 2.42% SCP indicates a relatively higher degree of international collaboration, whereas countries like the USA (57.40% SCP) and China (1.85% SCP) show different patterns. The number of MCP reflects the diversity and strength of international research collaborations for each country. For example, the UK has 12 MCP, indicating a moderate level of research collaboration with other nations, while China has 13 MCP, illustrating a growing international presence in research collaborations. Countries like India, Pakistan, and South Africa, with lower overall article counts, show relatively high percentages of MCP. This suggests that these countries are actively engaging in international research collaborations, potentially serving as regional research hubs.

The second part of Table 2, focused on research impact, ranking countries by the number of citations received and the average number of citations per article. While the USA has the highest citation count, its average number of citations per article (11.09) is lower than countries like the UK (12.98) and Australia (15.34). This suggests that the USA's research output is extensive, but individual articles may receive fewer citations on average compared to the UK and Australia. The wide range of average citations per article among countries, from Sweden's 23.35 to Nigeria's 4.74, reflects the variability in research impact and quality. Sweden's high average citation count suggests a strong emphasis on producing highly influential research. This data underscores the importance of considering both the quantity and quality of research output. While high citation counts signify research influence, a high average citation per article implies that a country's research is consistently impactful.

Journals' productivity

Journals' productivity has been assessed through their rankings (Table 3). This ranking of journals is based on their *h*-index and total citations. The *Journal of Academic Librarianship* tops the list with an *h*-index of 35 and 4404 total citations. *Reference Services Review* follows closely with an *h*-index of 32 and 4161 citations. *College and Research Libraries* take 3rd place with an *h*-index

Table 1: Most influential authors in the said domain.

Author	Articles	Total citation	Year of 1st publication	Years of contribution (as of 2021)
Julien, H.	28	562	1997	24
Pinto, M.	19	314	2008	13
Lloyd, A.	10	190	2003	18
Gross, M.	11	182	2007	14
Latham, D.	11	182	2007	14
Saunders, L.	13	127	2007	14
Spring, H.	14	85	2010	11
Ameen, K.	13	83	2009	12
Merga, M.K.	13	82	2015	6
Hicks, A.	10	39	2010	11

Table 2: Most productive countries in the defined domain (top 10).

Rank by articles	Country	Articles	Freq.	SCP*	MCP#	Rank by citation	Country	Citation	Avg. article
1.	USA	1985	57.40%	1941	44	1.	USA	22008	11.09
2.	UK	232	6.70%	220	12	2.	UK	3011	12.98
3.	Canada	195	5.63%	181	14	3.	Australia	2239	15.34
4.	Australia	146	4.22%	132	14	4.	Canada	2206	11.31
5.	Nigeria	84	2.42%	79	5	5.	Spain	569	16.74
6.	China	64	1.85%	51	13	6.	Pakistan	473	8.30
7.	India	63	1.82%	60	3	7.	Nigeria	398	4.74
8.	Pakistan	57	1.64%	50	7	8.	New Zealand	397	17.26
9.	South Africa	50	1.44%	44	6	9.	Sweden	397	23.35
10.	Spain	34	0.98%	29	5	10.	South Africa	349	6.98

[*SCP → Single country publication, #MCP → Multiple country publications]

of 26. For journals with similar *h*-indices, total citations play a conclusive role, as seen with the *Journal of Documentation* and *Portal* sharing the 4th rank. Further down, Communications in *Information Literacy, Journal of the Medical Library Association*, and *New Library World* tie with an *h*-index of 19, but total citations vary. *Health Information and Libraries Journal* and *College and Undergraduate Libraries* share the 9th position, with *Health Information and Libraries Journal* having more total citations. This ranking highlights the importance of considering both the *h*-index and total citations when evaluating a journal's impact and significance within a field.

Most globally cited documents

Globally cited literacy research reflects a dynamic and diverse field, addressing digital evolution, social nuances, practical applications, geographic considerations, interdisciplinary collaboration, enduring impact, paediatric literacy, and emerging concepts, collectively contributing to a comprehensive understanding of literacy in our changing world. The rankings of the studies in Table 4 are determined based on the total number of citations received, reflecting their impact on shaping a holistic comprehension of literacy within the evolving global context.

Highly Cited Articles

Articles with higher citation counts, such as "Information and digital literacies: a review of concepts" and "The role of home literacy environment in the development of language ability," have likely made significant contributions to their respective fields. Higher citation counts may indicate that these articles are influential and have been widely referenced by other researchers.

Diversity of Topics

The articles cover a diverse range of topics, including information literacy, healthcare decision-making, literacy development in children, and digital competence. The variety of topics suggests that these articles address important issues in different domains, reflecting the multidisciplinary nature of research in information science, education, and healthcare.

Table 3: Journal's productivity.

Rank	Sources	*h*-index	Total citation	Impact factor
1.	Journal of Academic Librarianship	35	4404	2.9
2.	Reference Services Review	32	4161	1.9
3.	College and Research Libraries	26	2122	3.9
4.	Journal of Documentation	21	1930	3.1
5.	Portal	21	1550	0.2
6.	Communications in Information Literacy	19	1523	1.5
7.	Journal of the Medical Library Association	19	1377	4.2
8.	New Library World	19	973	-
9.	Health Information and Libraries Journal	18	1331	4.8
10.	College and Undergraduate Libraries	18	1209	1.6

Table 4: Global Ranking of Cited Literacy Studies.

Sl. No.	Paper	Author(s)	Name of journal	Year of publication	Total citation
1.	Information and digital literacies: a review of concepts.	David Bawden	Journal of Documentation	2001	516
2.	The role of home literacy environment in the development of language ability in preschool children from low-income families.	Adam C. Payne, & *et al.*	Early Childhood Research Quarterly	1994	374
3.	Access to Print in Low-Income and Middle-Income Communities: An Ecological Study of Four Neighborhoods.	Susan B. Neuman, Donna Celano	Reading Research Quarterly	2001	357
4.	Critical Information Literacy: Implications for Instructional Practice.	James Elmborg	The Journal of Academic Librarianship	2006	301
5.	Adopting evidence-based practice in clinical decision making: nurses' perceptions, knowledge, and barriers.	Shaheen Majid, & *et al.*	Journal of the Medical Library Association	2011	249
6.	Digital competence-an emergent boundary concept for policy and educational research.	Liisa Ilomäki, & *et al.*	Education and Information Technologies	2016	238
7.	Risk Factors and Screening Instruments to Predict Adverse Outcomes for Undifferentiated Older Emergency Department Patients: A Systematic Review and Meta-analysis.	Christopher R. Carpenter, & *et al.*	Academic Emergency Medicine	2015	227
8.	The Readability of Pediatric Patient Education Materials on the World Wide Web.	Donna M. D'Alessandro & *et al.*	Archives of Pediatrics and Adolescent Medicine	2001	226
9.	Information grounds and the use of need-based services by immigrants in Queens, New York: A context-based, outcome evaluation approach.	Karen E. Fisher & *et al.*	Journal of the American Society for Information Science and Technology	2004	222
10.	Digital literacy and informal learning environments: An introduction.	Eric M. Meyers & *et al.*	Learning, Media and Technology	2013	194

Figure 2: Conceptual framework.

Temporal Trends

The articles span a range of publication years, from 1994 to 2016. This indicates that influential articles can emerge at different points in time. It may be interesting to explore whether there are trends or shifts in the topics and methodologies of these articles over the years.

Impact on Practice

Articles like "Critical Information Literacy: Implications for Instructional Practice" may have implications for teaching and practice in academic librarianship. "Adopting evidence-based practice in clinical decision making" and "Risk Factors and Screening Instruments" may have influenced healthcare practices and policies.

Potential for Future Research

The Table 4 may guide researchers in identifying key areas of interest based on citation counts and topics with fewer citations. It could be an indicator of gaps or emerging trends in the literature.

Conceptual framework

Figure 2 depicts the co-occurrence network of the top 50 authors' keywords, and the '*Louvain community detection algorithm*'[21] was used for this purpose. Based on the algorithm's analysis, three major research themes were identified and are described below.

The theme of information literacy heterogeneity (red) encompasses the various forms of information literacy that are adopted by libraries to educate users on how to deal with the different types of information provided by the library.

The roles of librarians in promoting IL (blue) refer to how professionals in the library and information science field promote literacy within user communities through various teaching and learning methods.

The theme of health literacy and the library (green) involves the promotion of health literacy among library users through various channels, including the Internet, social media, and outreach programs.

The themes were categorized using a thematic map to better understand the structure of the knowledge[4] (Figure 3). Based on their density and centrality, the themes were divided into four sections. The *Motor themes*, which had higher centrality and density, were found in the upper right quadrant and played a vital role in shaping the conceptual structure of the scientific field. This means they are not only at the core of the knowledge structure but also well-developed and interconnected. These themes, including health literacy, information skills, and outreach programs, play a pivotal role in shaping the conceptual framework of the scientific field. Their high centrality suggests that they are frequently cited and form the basis for much of the research in the field. So, researchers should pay close attention to these Motor themes as they represent the foundational concepts and areas where further exploration and innovation can lead to advancements in the field. The *Niche themes*, located in the upper left quadrant, had low centrality but high density. These themes were already developed, specialized, and were in a marginal stage. Studies related to professional development and literature fall into this quadrant. The *Emerging or Declining themes*, located in the lower left quadrant, had lower centrality and density and were modern and evolving. Themes are relevant because they signify areas that are in a state of flux. Emerging themes, such as digital literacy, social media, and fake news, are indicative of new directions in the field.

Researchers may find opportunities for pioneering work in these areas. Lastly, the *Basic themes* were found in the lower right quadrant and were characterized by higher centrality but lower density. These themes were essential for the scientific structure and were not yet well-developed. This quadrant included themes related to literacy, libraries, librarians, information literacy, academic libraries, and library instruction. Analysing the thematic map, it is clear that digital literacy and health literacy are highly relevant in today's world. Health literacy is critical for understanding healthcare information and making informed decisions, especially during health crises like the COVID-19 pandemic. Digital literacy is essential in an increasingly connected society, impacting education, employment, cybersecurity, and access to government services. Both literacies play pivotal roles in individuals' well-being, opportunities, and participation in the digital age.

Topic modelling

Performing topic modelling using the LDA technique Figure 4 provides an insightful structure of a given data set (corpus) which is a random mixture of latent topics. However, there is no specific rule for determining the exact number of latent topics.

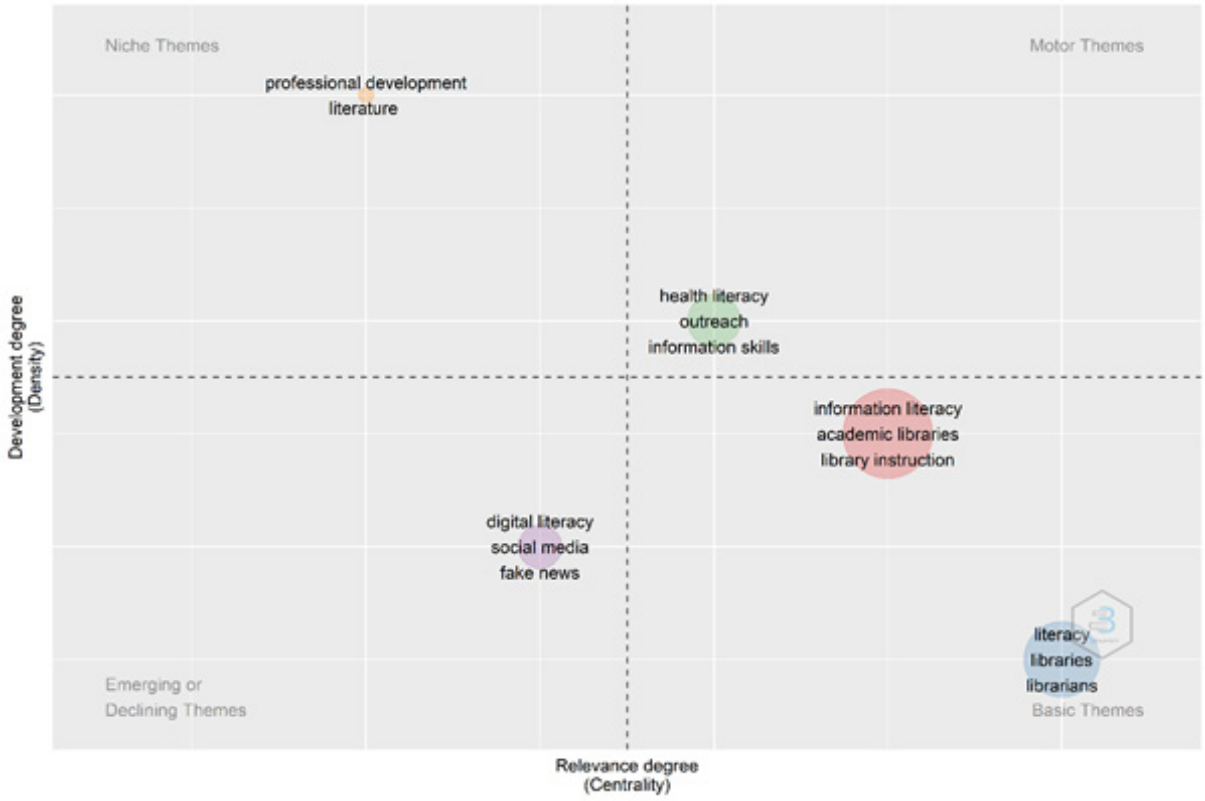

Figure 3: Thematic map.

```
repairNotes = jana.notes;
repairNotes = regexprep(repairNotes,'[^A-Za-z\'']',' ');
repairNotes = lower(repairNotes);
repairNotes = strtrim(repairNotes);
documents = tokenizedDocument(repairNotes);
documents = removeStopWords(documents)
```

```
131 tokens: critical literacy practices problem based learning projects science digital technologies part many children s everyday lives entertainment info
228 tokens: assessing effect virtual education information literacy competency evidence based practice among undergraduate nursing students background info
267 tokens: digital competencies teachers transformation educational environment purpose study determine main directions development digital competencies t
167 tokens: information needs breast cancer patients educational status influence information needs ghana health information literacy plays critical role s
131 tokens: malaysian research support librarians self directed learning traits examining demographic differences relationship competencies research suppor
122 tokens: effect librarians digital skills technology acceptance academic libraries jordan digital skills necessary work manage electronic library infras
125 tokens: deconstructing information literacy discourse peeling back layers higher education discourses information literacy practice create epistemologi
133 tokens: role libraries misinformation programming research agenda misinformation fake news exploded social media platforms communities past few years s
```

Figure 4: Preparing dataset for perforing LDA.

In the LDA algorithm, it is necessary to specify the number of topics, and this can be done by estimating perplexity Figure 5. Measuring perplexity helps in choosing the appropriate number of latent topics from a given corpus, and these latent topics are identified by the distribution of words.[17] After analysing the perplexity values over time, we selected 32 topics for our study.

Topic interpretation

In the LDA topic model, meaningful interpretation of the latent variables is necessary, and this is done by examining the top keywords of each topic.[22] Our study identified thirty-two topics, and their corresponding top keywords are illustrated in Figures 6, 7, 8, and 9. Below, we systematically explain each of the identified topics.

Topic 1 is about the significance of information literacy in today's global society. It discusses various models, emphasizes their role in lifelong learning, and recommends alternative approaches. It advocates for a holistic view of information literacy, highlighting its importance beyond education in enhancing well-being and societal progress.

Topic 2 focuses on the integration of information literacy into university curricula. It highlights the importance of collaboration between students, faculty, and librarians in developing programs and courses that embed information literacy as a core competency. It discusses various approaches and models for curriculum development, emphasizing the role of information literacy in undergraduate education across disciplines. It also outlines strategies for designing assignments and learning outcomes that promote information literacy skills among students.

```
rng(3)
cv = cvpartition(bag.NumDocuments,'HoldOut',0.1);

dataTrain = bag.Counts(cv.training,:);
numTopicsRange = 2.^(1:5);
ppl = zeros(numel(numTopicsRange),1);
for i = 1:numel(numTopicsRange)
        numTopics = numTopicsRange(i);
        mdl = fitlda(dataTrain,numTopics,'Verbose',0);
        dataTest = bag.Counts(cv.test,:);
        [~,ppl(i)] = logp(mdl,dataTest);
end

figure
plot(numTopicsRange,ppl,'-*')
xlabel("Number of Topic")
ylabel("Perplexity")
```

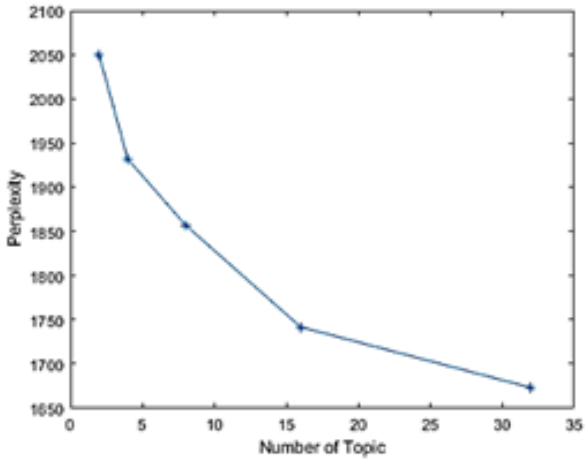

Figure 5: Measuring perplexity.

Topic 3 is on examining information literacy skills in graduate and undergraduate students using surveys, tests, and questionnaires. It explores factors impacting skill levels, such as academic orientation and prior behaviours. It informs curriculum design and targeted literacy programs, emphasizing the importance of enhancing students' skills for academic success.

Topic 4 is about investigating diverse research methods, including qualitative and quantitative approaches, to gain insights into topics such as engineering, international perspectives, and attitudes. It identifies barriers, attitudes, and the impact of training and consultation in research.

Topic 5 is about the changing roles of librarians in academic institutions, emphasizing collaboration, education, and adaptability. It discusses shifting perceptions, the importance of information literacy, and challenges faced by librarians.

Topic 6 highlights the crucial role of libraries in rural and local communities. It explores how libraries provide services, create unique spaces, and foster development. It discusses the impact of library initiatives, collaborations, and leadership in addressing challenges and benefiting communities. It emphasizes the need for sustainable funding and technology access to ensure libraries continue to serve as vital community hubs.

Topic 7 explores how academic libraries play a dynamic role within universities, emphasizing their impact on student success and campus partnerships. It highlights the importance of collaboration, sustainability, and addressing poverty, drawing from case studies and innovative models. It also encourages continued adaptation and revision to ensure libraries remain integral to the modern university environment.

Topic 8 deals with the evolving landscape of library instruction and assessment, highlighting librarians' role in teaching and improving student learning. It explores various methods and tools, such as rubrics and program evaluations, to measure outcomes. It also discusses practical approaches in subjects like chemistry and English composition and emphasizes the impact of instruction on student success. It encourages continuous improvement, collaboration, and reinforcement of library literacy in academia.

Topic 9 is on how public libraries contribute to societal development, emphasizing their role in promoting culture, inclusion, and sustainability. It highlights their significance as information hubs, fostering diversity and community engagement. It also indicates that public libraries are fundamental in shaping societal values and driving positive change.

Topic 10 is about the evolving concept of critical literacy, focusing on its relevance in the 21st century. It examines new approaches and perspectives, considering technology's impact on traditional analysis. It introduces innovative tools and calls for multidisciplinary discussions, emphasizing its significance in understanding complex narratives and identities in the modern world.

Topic 11 investigates how social media influences young individuals, focusing on factors that affect their preferences and behaviours. It explores key issues, such as ethical considerations and legal implications, with a focus on the Indian context. It highlights both the positive and negative effects of social media engagement and emphasizes the importance of credible sources, ethics, and policies in guiding responsible use among adolescents.

Topic 12 explores strategies to improve access to digital resources and information outreach. It highlights the importance of information literacy and progress in meeting diverse information needs, particularly in Canada and other regions. Collaboration and communication are essential for effective solutions.

Topic 13 is on creative projects and workshops as effective tools for engagement and learning. Librarians play a key role in supporting students through virtual workshops and peer teaching. Interactive activities like games and video creation enhance learning experiences, with a focus on feedback and reflection for improvement.

Topic 14 explores how evidence-based research informs nursing practice. It emphasizes the practitioner's role in adopting and applying research findings, bridging gaps between theory and

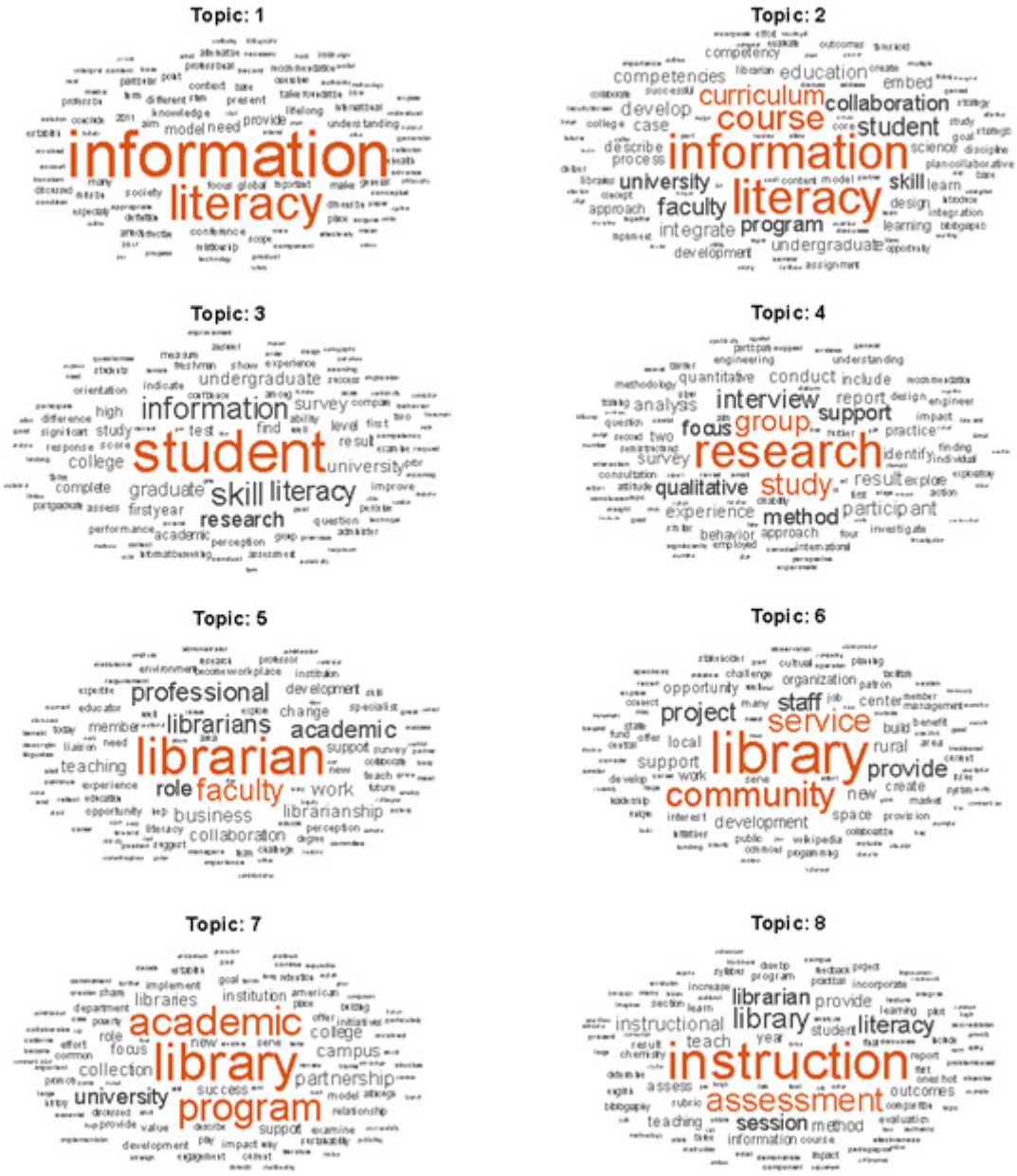

Figure 6: Word cloud of topics 1-8.

practice. Reflective learning and affective understanding play vital roles in nursing education.

Topic 15 examines development initiatives in Africa, emphasizing interdisciplinary efforts and key challenges. It highlights the importance of support and engagement across various sectors. Case studies from South Africa and other African nations demonstrate potential for improvement and outcomes. Technical involvement, organizational initiatives, and interdisciplinary approaches are crucial for addressing complex challenges and promoting growth.

Topic 16 explores strategies to enhance university library services, especially in the digital age. It emphasizes the importance of training, technology, and user satisfaction. It highlights challenges and the need for innovation, recognizing the influence of technology and user feedback.

Topic 17 focuses on enhancing student engagement, with librarians and instructors playing active roles. It emphasizes interactive assignments, critical thinking, and providing relevant sources. It addresses challenges, encourages adaptability, and promotes personal engagement while tackling issues like privacy and misinformation.

Topic 18 explores the challenges and opportunities in the digital landscape, emphasizing the importance of technology skill development and education. It discusses the role of libraries,

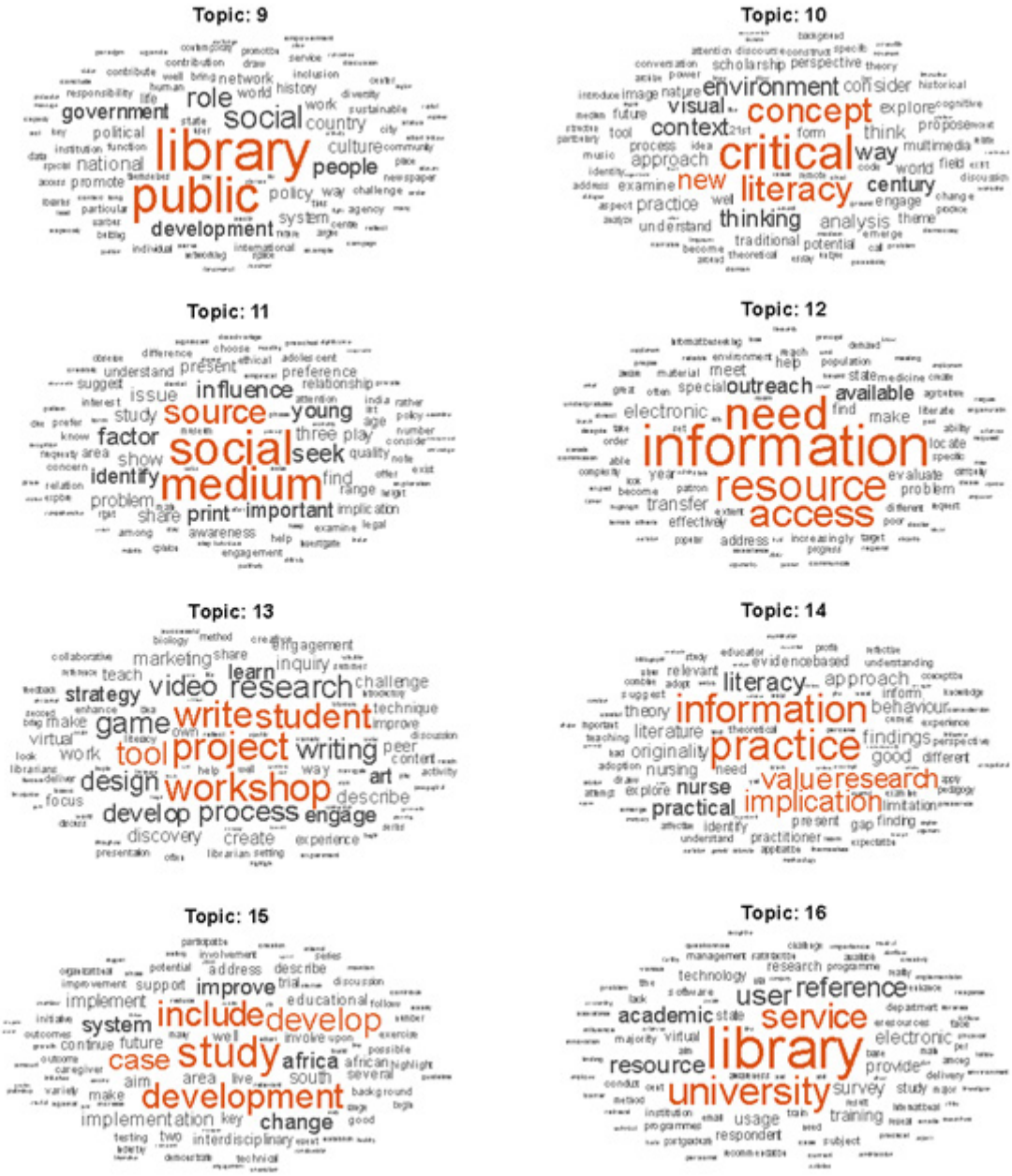

Figure 7: Word cloud of topics 9-16.

institutions, and universities in enhancing digital competence and accessibility, particularly in the context of the COVID-19 pandemic.

Topic 19 focuses on improving search and evaluation skills to measure quality and impact. It identifies factors influencing self-efficacy and emphasizes the effectiveness of search engines like Google. Overall, enhanced skills lead to improved search results and behaviour.

Topic 20 is about exploring research trends and data analysis, focusing on citation patterns, financial openness, and performance over time. It addresses plagiarism concerns and highlights the impact of senior researchers. It utilizes tools like Google Scholar for comprehensive data analysis, contributing to a better understanding of research patterns and their influence on academic institutions.

Topic 21 explores scientific research, literature review, and journal articles, emphasizing copyright and ethical considerations. It discusses key criteria, relevance in various fields, and the role of keywords. It highlights ethical aspects and the importance of close examination in the research process.

Topic 22 explores data collection techniques in Nigeria, revealing significant findings about population demographics, resource utilization, and distribution methods. It analyses the impact of gender, age, and satisfaction in various districts, including prison

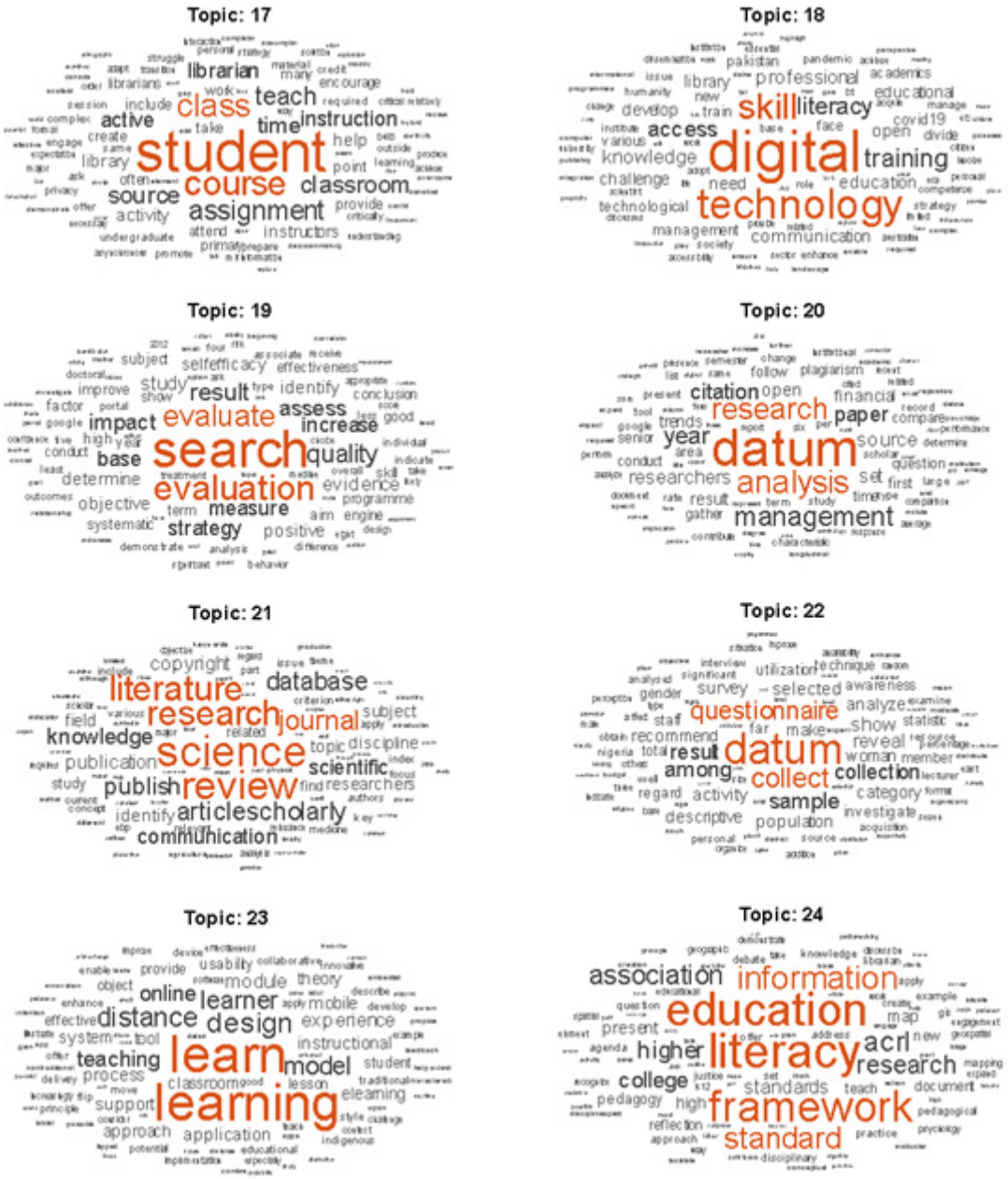

Figure 8: Word cloud of topics 17-24.

settings, enhancing our understanding of data collection in the region.

Topic 23 explores online learning, emphasizing distance design, instructional models, and mobile tools. It discusses effective teaching styles, innovative delivery methods, and challenges in online education.

Topic 24 focuses on higher education's information literacy framework, standards set by ACRL, and pedagogical approaches. It explores the use of GIS and interdisciplinary engagement, emphasizing the role of librarians. It outlines an action plan for educators, promoting participatory models and principles of pedagogy in various disciplines.

Topic 25 focuses on literacy development in primary and secondary education, involving teachers and librarians. It addresses curriculum development and cultural aspects in education policies worldwide. It highlights the transition from primary to secondary education and emphasizes the importance of functional literacy, especially in urban and underprivileged areas.

Topic 26 evaluates computer and internet literacy in Nigeria, addressing factors influencing competence. It uses questionnaires and readability tools to assess ICT adoption and performance. It aims to improve computer and internet literacy, particularly in community-based and educational settings.

Figure 9: Word cloud of topics 25-32.

Topic 27 aims to improve online information retrieval and user experience by discussing web tutorials, interactive content, and new features. It emphasizes user convenience and redesign strategies for better online information access, particularly in libraries and education.

Topic 28 investigates anxiety reduction interventions for adults, including youth and older individuals. Findings indicate significant effects on anxiety levels, with positive outcomes for decision-making and well-being. It also identifies potential barriers and associations, offering insights for future programs.

Topic 29 explores the impact of fake news on information literacy in the United States. Findings reveal an increase in fake news consumption, with a significant impact on comprehension and emotional response. Librarians play a crucial role in addressing this issue, offering mentorship and resources to enhance critical thinking and information literacy skills among the public.

Topic 30 emphasizes the importance of promoting reading and literacy in elementary education. It highlights the role of diverse reading materials, e-books, and classroom activities in motivating students to develop a reading habit. The research also discusses the impact of reading on vocabulary, language, and cultural knowledge, advocating for literacy advocacy programs to empower young learners and improve their overall literacy skills.

Topic 31 highlights the significance of family engagement in promoting early childhood literacy. It emphasizes the impact of a literacy-rich home environment, family reading activities, and community programs in supporting children's literacy development. The study discusses the positive influence of storytelling, reading aloud, and parental involvement on a child's readiness for school. It underscores the importance of promoting these practices to bridge literacy gaps, especially in low-income communities, and foster a lifelong love for reading.

Topic 32 focuses on strategies to improve health literacy, especially in clinical and medical settings. It emphasizes training, awareness, and coordinated care, especially for chronic conditions. It aims to promote healthcare autonomy and wellness in diverse populations. It also discusses the utilization of technology and electronic databases like PubMed to facilitate evidence-based medicine.

Findings

In the realm of research trends, a dataset spanning 1969 to 2021 reveals a steady 12.2% annual growth in scholarly publications, with an average of approximately 12 citations per document. Collaborative efforts among authors are commonplace, with most documents involving an average of 2.11 co-authors. Notable authors such as Julien, H., and Pinto, M., have made substantial contributions to their fields, with varying levels of research impact. Global research productivity and collaboration unveil the United States as a research leader, and different countries exhibit diverse collaboration patterns. The citation impact varies, with the USA having a high citation count, while countries like the UK and Australia receive more average citations per article. Journals are ranked by their *h*-index and total citations, underlining their significance within the field. Additionally, a conceptual framework identifies three major research themes, with an emphasis on the importance of digital and health literacy as highly relevant areas in today's world, essential for individual well-being and participation in the digital age. Furthermore, employing topic modelling with LDA based on title, authors' keywords, and abstracts has revealed 32 research hotspots.

DISCUSSION AND IMPLEMENTATION

The descriptive analysis of this study suggests a consistent expansion in literacy research within the framework of library studies, mirroring the trends observed in S.R. Kolle.[8] This study's examination of international scientific collaboration delineates the distinct features of research endeavours across various countries, echoing the findings of Pinto, María and *et al*.[6]

The outcomes of LDA topic modelling yielded an estimation of 32 latent topics within the field of literacy studies related to library research. The diverse array of topics reveals the multifaceted landscape of information literacy. The global significance of information literacy (Topics 1 and 9-16) emphasizes its holistic role beyond education, touching on critical aspects like public libraries, societal development, and digital competence.[23] The integration of information literacy into university curricula (Topics 2 and 17-24) is underscored, with collaboration between students, faculty, and librarians being pivotal.[24] Assessing information literacy skills (Topics 3 and 25-32) emerges as a crucial aspect.[25] The evolving roles of librarians (Topics 5, 13, 16, and 29-31) further contribute to this landscape, recognizing librarians as dynamic contributors in combating fake news, fostering creativity, and championing health literacy.[26]

In summary, examining research trends, this study employed a combination of bibliometrics and LDA topic modelling. The use of computer algorithms for quantifying and analysing textual data eliminates subjectivity, revealing patterns in texts and aiding researchers in extracting meanings from extensive datasets.

CONCLUSION

In conclusion, the study highlights the evolution of research productivity in the area of literacy, which has gone through initial, development and expansive stages. The emergence of various concepts related to literacy, such as digital literacy, media literacy, health literacy, etc., has led to an increase in research productivity. Authors from Western countries have been found to dominate this field in terms of the total number of articles and citations, frequency of publication, research cooperation, single-country publication, and multiple-country publication. The study also identifies the key research themes and their relationships using thematic maps, which suggest that professional development, health literacy, and library outreach programs are the prominent areas of study. The LDA model analysis helped in grouping the literature into 32 topics, which provides a crystal-clear idea of emerging research areas like digital literacy, social media, and fake news. Overall, this study provides valuable insights into the evolution of research productivity and the trends in literacy-related research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

1. Gee JP. What is literacy. Negotiating academic literacies: Teaching and learning across languages and cultures. 1989;171(1):18-25.
2. Bell D. The Coming of Post-industrial Society: A Venture in Social Forecasting. Basic Books; 1976. 602 p.
3. Machlup F. The Production and Distribution of Knowledge in the United States. New Jersey: Princeton University Press; 1962. 440 p.
4. Fontanella S, Cucco A, Custovic A. Machine learning in asthma research: moving toward a more integrated approach. Expert Rev Respir Med. 2021;15(5):609-21.
5. Gould P. Letting the Data Speak for Themselves. Annals of the Association of American Geographers. 1981;71(2):166-76.
6. Pinto M, Escalona-Fernández MI, Pulgarín A. Information literacy in social sciences and health sciences: a bibliometric study (1974-2011). Scientometrics. 2013;95(3):1071-94.
7. Shaheen M, Chang YK, Nu AH, Win KM, San YW, Majid S, Aye HN, Khine MM. Analyzing publishing trends in information literacy literature: A bibliometric study. Malaysian

Journal of Library & Information Science 2015;20(2). Available from: http://ojie.um.edu.my/index.php/MJLIS/article/view/1766

8. Kolle SR. Global research on information literacy: a bibliometric analysis from 2005 to 2014. The Electronic Library. 2017;35(2):283-98.
9. Parlina A, Ramli K, Murfi H. Theme Mapping and Bibliometrics Analysis of One Decade of Big Data Research in the Scopus Database. Information. 2020;11(2):69.
10. Pinto M, Fernández-Pascual R, Caballero-Mariscal D, Sales D, Guerrero D, Uribe A. Scientific production on mobile information literacy in higher education: a bibliometric analysis (2006-2017). Scientometrics. 2019;120(1):57-85.
11. Onyancha OB. Knowledge visualization and mapping of information literacy, 1975-2018. IFLA Journal. 2020;46(2):107-23.
12. Li Y, Chen Y, Wang Q. Evolution and diffusion of information literacy topics. Scientometrics. 2021;126(5):4195-224.
13. Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. Journal of Machine Learning Research. 2003;3:993-1022.
14. Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics. 2017;11(4):959-75.
15. Rusydiana AS. Bibliometric analysis of journals, authors, and topics related to COVID-19 and Islamic finance listed in the Dimensions database by Biblioshiny. Sci Ed. 2021;8(1):72-8.
16. Huang JH, Duan XY, He FF, Wang GJ, Hu XY. A historical review and Bibliometric analysis of research on Weak measurement research over the past decades based on Biblioshiny. arXiv; 2021. Available from: http://arxiv.org/abs/2108.11375
17. Tontodimamma A, Nissi E, Sarra A, Fontanella L. Thirty years of research into hate speech: topics of interest and their evolution. Scientometrics. 2021;126(1):157-79.
18. Blei DM. Probabilistic topic models. Commun ACM. 2012;55(4):77-84.
19. MATLAB. Massachusetts: The MathWorks Inc.; 2018. Available from: https://in.mathworks.com/products/new_products/release2018b.html
20. Aparicio G, Iturralde T, Maseda A. Conceptual structure and perspectives on entrepreneurship education research: A bibliometric review. European Research on Management and Business Economics. 2019;25(3):105-13.
21. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech. 2008; 2008(10):P10008.
22. Steyvers M, Griffiths T. Probabilistic Topic Models. In: Handbook of Latent Semantic Analysis [Internet]. Psychology Press; 2007. DOI: 10.4324/9780203936399-29.
23. Fredrick OA, Veronica OA. The librarys role in developing information literacy and societal growth. Int J Lib Inf Sci. 2015;7(7):124-9.
24. Asplund J, Hakala E, Sallama S, Tapio S. Integrating Information Literacy Education into the Curriculum at the University of Tampere, Finland. NORIL. 2013;5(1):3-10.
25. LeMire S, Xu Z, Hahn D, Balester V, Dorsey LG. Assessing the Information Literacy Skills of First-Generation College Students. College & Research Libraries. 2021;82(5):730-754.
26. Batchelor O. Getting out the truth: the role of libraries in the fight against fake news. Reference Services Review. 2017;45(2):143-8.

Cite this article: Jana A, Rout R. Visualising Knowledge, Research Hotspots and Trends of Literacy Studies in the Context of Library, 1969-2021. J Scientometric Res. 2024;13(1):161-75.