International Journal of Information Science and Management
Vol. 23, No. 2, 2025, 195-207
DOI: https://doi.org/10.22034/ijism.2025.2027507.1469 / DOR: https://dorl.net/dor/20.1001.1.20088302.2025.23.2.10.6

Original Research

Thematic Analysis of Iranian Medical Library and Information Theses and Dissertations: Applying Text Mining Techniques

Shadi Asadzandi
Assistant Prof., Health Management and Economics Research Center, Health Management Research Institute, Iran University of Medical Sciences, Tehran, Iran.
asadzandi.sh@iums.ac.ir
ORCID iD: https://orcid.org/0000-0002-1350-9629

Atieh Sadat Hosseini Ghahjavarestani
Master's student in Medical Library and Information Science, Department of Medical Library and Information Science, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran.
Corresponding Author: Atiye.ho3eini@gmail.com
ORCID iD: https://orcid.org/0000-0003-1337-5488

Received: 28 April 2024
Accepted: 09 September 2024

Abstract

With the development of Internet technology, the amount of data has increased astonishingly. This article aims to determine the thematic trend of theses and dissertations in librarianship of postgraduate courses in the universities of medical sciences. The study analyzed 110 theses from five Iranian universities in Medical Library and Information Science (MLIS) between 2017 and 2021. A Python script processed Persian text to create a word cloud and extract significant keywords using TF-IDF. Data was preprocessed, normalized, and filtered, and keywords were grouped by year and university for analysis. Results show key research topics in MLIS from 2017 to 2021, including information-seeking behavior, Altmetrics, health, and information services, as shown in the word cloud. Common themes included information-seeking behavior, digital literacy, and health literacy, indicating stable research trends in medical library science topics. Each university emphasized different areas: Isfahan focused on clinical aspects, Tehran on foundational issues, Shahid Beheshti on scientometrics, while Hamadan addressed information retrieval. Trends showed limited variation in keywords over the years. The study highlights evolving Medical Library and Information Science (MLIS) topics, emphasizing a shift towards patient information and health literacy. However, a gap exists in addressing issues like AI and big data. Limited collaboration and data hindered the research scope, suggesting that future studies should explore faculty and student topic selection interests.

Keywords: Data Mining, Text Mining, Dissertation, Theses, Library and Information Sciences (LIS), Library Subjects, Medical Library and Information Sciences (MLIS), Thematic Analysis.

Introduction

Modern information systems have made obtaining a vast amount of data possible. However, among these, unstructured data cannot be quickly processed and made accessible. These data are usually part of textual data. Traditional systems for organizing and retrieving information are not responsive in this field, so new methods, such as text mining, are used (Hwang, Wei, Lee & Chen, 2017). Conceptual data mining means analytical methods to extract meaningful

knowledge from massive amounts of data. Data can be presented as text, graphs, video, audio, and images. (Cheng & Lui, 2021) This relatively new concept emerged in the mid-1990s as a new approach to data analysis and to eliminate the knowledge gap. The first conference related to data mining was held in the United States in 1995 and registered in MeSH (Medical Subject Headings) in 2009 (Yoo, Alafaireet, Marinov, Pena-Hernandez, Gopidi & Chang, 2012).

The approach of automatically assigning topics to texts based on text mining methods and their specific algorithms has made information science specialists and librarians able to identify topics and interpret resources and their functions in optimal resource retrieval (Bittermann & Fischer, 2018; Fang, Yang, Gao & Li, 2018; Sanandres, Abello & Madariaga, 2020). Text mining not only analyzes large texts such as books and journal articles, but also covers the texts of E-mails, tweets, and various comments on social networks text mining methods, which are one of the special fields of data mining, were developed to automate the process of analyzing this type of data (Mansouri, Zarmehr & Karshenas, 2020).

Today, with the development of computer and internet technology, the amount of data has increased at an incredible speed; therefore, research in any field shows its importance for further development and shows each discipline's exact direction of movement. Research plays an essential role in the growth of any profession and allows researchers to share their findings with their wider professional communities through publication. In LIS, various qualitative and quantitative research methods were used and studied to test hypotheses and answer various research questions. (Armann-Keown & Patterson, 2020; Islam, Islam & Mondal, 2018; Wu, Li, Feng, Li, Huang, Xu & Lyu, 2021)

Graduate education, as a component of the national educational system, plays a significant role in enriching culture, enhancing competitiveness with other countries, and advancing science and technology. A well-developed scholarly work can contribute to civilization's scientific, economic, and social development. Theses from master's and doctoral programs are prime examples of such academic contributions. These theses typically serve as mandatory coursework for degree completion. They encompass various sections, including an introduction, a review of prior studies, a critical analysis of the research background, the presentation of research hypotheses, and ultimately, a report of the findings and their comparison with other studies (Putri & Saputra, 2021). Dissertations help the growth process of knowledge, showing significant information and providing evidence of significant scientific achievements. These materials also reveal the educational and research priorities and centers of research institutes and universities, and create a unique insight into the field of scientific promotion of each discipline (Zong, Shen, Yuan, Hu, Hou & Deng, 2013). Studying the trend of publications in LIS has shown social, cultural, and technological changes and developments in recent decades. Studies have shown that in recent decades, the use of librar* words and their derivative terms, such as library, libraries, librarian, librarians, as dominant keywords have decreased (Finlay, Sugimoto, Li & Russell, 2012; Sugimoto, Li, Russell, Finlay & Ding, 2011).

Numerous studies have been conducted across various subject areas utilizing text mining technology. This tool provides a valuable opportunity for automatically analyzing a collection of documents with unknown research structures. It also facilitates the identification of the latest research trends and the delineation of domains and trends across different fields. Therefore, examining research trends within a specific domain over various periods can offer a better understanding for researchers and policymakers, aiding in the effective planning of future research and allocation of research resources (Gholami, Sals Faal, Behzadi & Jampour, 2023;

Kim & Delen, 2018; Thakur & Kumar, 2022). Consequently, this study aims to utilize text mining methods to analyze master's and doctoral theses in the Medical Library and Information Science (MLIS) in Iran.

Literature Review

A review of past studies has shown that text mining is an effective method for identifying changes in the thematic trends of research and emerging keywords within them. Numerous studies in various thematic areas have been conducted using this method. For example, studies have been carried out in the fields of medical informatics (Kim & Delen, 2018), mobile learning (Salloum, Al-Emran, Monem & Shaalan, 2018), sleep disorders (Lam, Lai, Wang, Lai, Hsu & Chung, 2016), and electronic records (Shokouhian, Asemi, Shabani & Cheshme-Sohrabi, 2019). Another area of interest is MLIS. Dastani Chelak, Ziaei, and Delghandi (2021) evaluated the texts of library articles up to 2020 using data mining methods. Additionally, another study analyzed medical library articles in specialized journals up to 2019 using text mining techniques (Dastani, Mousavi Chelak, Ziaei & Delghand, 2022). In another study, the same researcher examined published topics concerning health information technology in global journals using text mining methods (Dastani, Ehtesham, Javanmard, Sabahi & Bahador, 2022). Gholami Sals Faal (2023) also utilized text mining to investigate the thematic trends of articles in knowledge and information science.

However, examining texts in this field has not been limited to text mining; other methods such as co-word analysis, bibliometric and scientometric methods, and the description of citation trends of articles and theses in this area have also been utilized. Therefore, no study has investigated these in MLIS using text mining methods and techniques. Furthermore, these are significant studies in any discipline and are crucial in determining the research trends of various universities and their topics of interest. Thus, examining them would greatly assist in creating a clear map of research trends in universities and developing issues in new areas according to community needs.

Materials and Methods

The research process is shown in Figure 1, and the steps are explained as follows.

Data collection

This study considered seven universities across the country offering graduate programs in MLIS. These universities included Tehran University of Medical Sciences, Shahid Beheshti

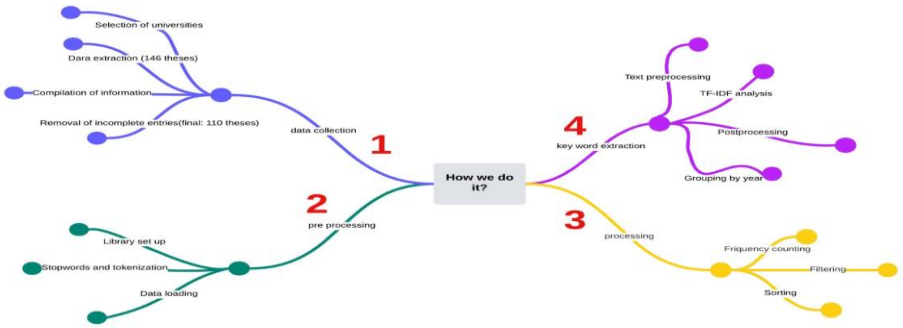

Figure 1: The methodology is presented sequentially, detailing the tasks performed

University of Medical Sciences, Iran University of Medical Sciences, Isfahan University of Medical Sciences, Tabriz University of Medical Sciences, Hamadan University of Medical Sciences, and Kerman University of Medical Sciences. Due to a lack of response and access to data from Tabriz and Kerman universities, these institutions were excluded from the study's statistical population. In the remaining five universities, a total of 146 theses from master's and doctoral programs were collected through a census method between the years 2017 and 2021. All information regarding the theses, including titles, abstracts, keywords, subject headings, publishers, and degree levels, was compiled and categorized in an Excel file. Initially, the data were reviewed, and incomplete entries were removed from the study, resulting in a final analysis of 110 thesis and dissertation abstracts.

Preprocessing

The first part of the project involved a Python script designed to create a word cloud for visualizing frequently used keywords in theses from Persian textual data. Initially, the necessary libraries were installed. These included the `hazm` library, which is used for text processing in Persian, the `persian` library for handling Persian characters, and `wordcloud_fa`, a library for generating word clouds from Persian text. Next, the libraries were imported into the program. The script imports essential Python libraries and functions, including numerical computation (`numpy`), text data handling (`pandas`), text processing (`hazm`), file management (`os`), and word cloud generation (`wordcloud_fa`). Subsequently, stopwords and tokenization were set up. The function `merge_stopwords` was defined to compile all stopwords from various files within a directory into a single list. This list was then expanded with a list of stopwords obtained from the `hazm` library. A tokenizer was also created using the `hazm` library. The next step involved loading data from an Excel file and removing any row containing '***' in the 'abstract' column. Following this, text processing was performed to prepare the data for analysis.

Processing

With the initiation of the data processing phase, frequency counting of subcategories was conducted. A frequency count was performed on the 'Subject Identifier' column, which was stored in the `subcategory_frequency` dictionary. Subsequently, the keywords in the 'Keywords' column were counted and divided into individual keywords, which were similarly counted as the subcategories. Next, normalization of the keys was carried out. The keys in the `subcategory_frequency` dictionary (which included both subcategories and keywords) were normalized using the `transform_key` function, which employs the `Normalizer` class from the `hazm` library.

In the following step, the keys were filtered. The `transformed_dictionary` keys underwent several filtering stages, where initially, any keys that did not contain Persian characters were removed, followed by the elimination of any leading colons. Finally, any key equal to '***' was also removed. In the last stage, the `final_dictionary` was sorted in descending order based on its values. This final dictionary represents the frequency of keywords and is utilized to create the word cloud. The `WordCloudFa` class from the `wordcloud_fa` library was then employed to generate a word cloud from the `final_dictionary`. The word cloud was subsequently saved as an image file and displayed.

Furthermore, Python executed text preprocessing and keyword extraction from the Persian text dataset. The objective was to identify the most significant keywords for each year, creating a DataFrame where each row corresponded to a specific year and contained a list of keywords for that year. In subsequent sections of the code, the following tasks were completed:

1. Renaming Columns: The program renamed the 'Publication Year' and 'Abstract' columns in the DataFrame to 'Year' and 'Abstract'.

2. Text Preprocessing: A function named `preprocess_text` was defined to normalize the text, separate tokens, remove stopwords, filter out short tokens, and ultimately lemmatize the text.

3. Applying Text Preprocessing: The `preprocess_text` function was applied to the 'abstract' column of the DataFrame, converting the result into a list of documents.

4. Keyword Extraction Using TF-IDF: The program utilized `TfidfVectorizer` from `sklearn.feature_extraction.text` to convert documents and the list of keywords into TF-IDF feature matrices. It then calculated the cosine similarity between documents and keywords, assigning each document the keyword with the highest similarity score.

5. Postprocessing Keywords: A function named `postprocess keyword` was defined to remove stopwords and lemmatize the keywords.

6. Grouping Keywords by Year: The DataFrame was grouped by 'Year', and the `postprocess_keyword` function was applied to the 'Keyword' column. The results are reported in Table 1.

7. Grouping Keywords by Universities: The DataFrame was grouped by 'University', and the `postprocess keyword` function was applied to the 'Keyword' column. The results are reported in Table 2.

TF-IDF (Term Frequency-Inverse Document Frequency) is a numerical statistical method that reflects the importance of a word to a document within a collection or corpus. The value of TF-IDF increases proportionately to the number of times a word appears in a document and is balanced by the number of documents in the corpus containing that word. This helps to adjust because some words may occur more frequently across the entire corpus. In this context, TF-IDF converts both documents and the list of keywords into a matrix of TF-IDF features. Subsequently, cosine similarity is calculated between the documents and the keywords to determine how closely each document resembles each keyword. The keyword with the highest similarity score is then assigned to each document, providing an effective method for extracting the most relevant keyword from each document (Dastani et al., 2021). Although this method has limitations such as low quality and semantic differentiation, it offers statistically better performance compared to other approaches in text mining. (Zhang, Yoshida & Tang, 2011) Given the limited dataset of fewer than 1,000 cases, other methods did not provide accurate statistics or results regarding the keywords.

Results

The data for this study comprised theses from master's and doctoral programs at Iranian medical universities, specifically those that included at least one of these degree levels. Initially, 146 titles were extracted from the library management system; however, due to incomplete information in various fields, only 110 titles remained. These titles included 10 doctoral theses and 100 master's theses. The number of theses, categorized by year from 2017 to 2021, was as follows: 17, 16, 36, 22, and 19 titles, respectively.

Active research topics in these MLIS

Figure 2 illustrates the topics of interest in MLIS theses from 2017 to 2021. Notably, these terms were extracted and organized from subject headings assigned to the theses during the indexing process. In the word cloud, more significant words indicate greater significance than others. Additionally, the variation in color serves solely for visual differentiation and does not convey any specific meaning. The most frequently researched topics during these years include information-seeking behavior, Altmetrics, Informational need, Health, and Information services.

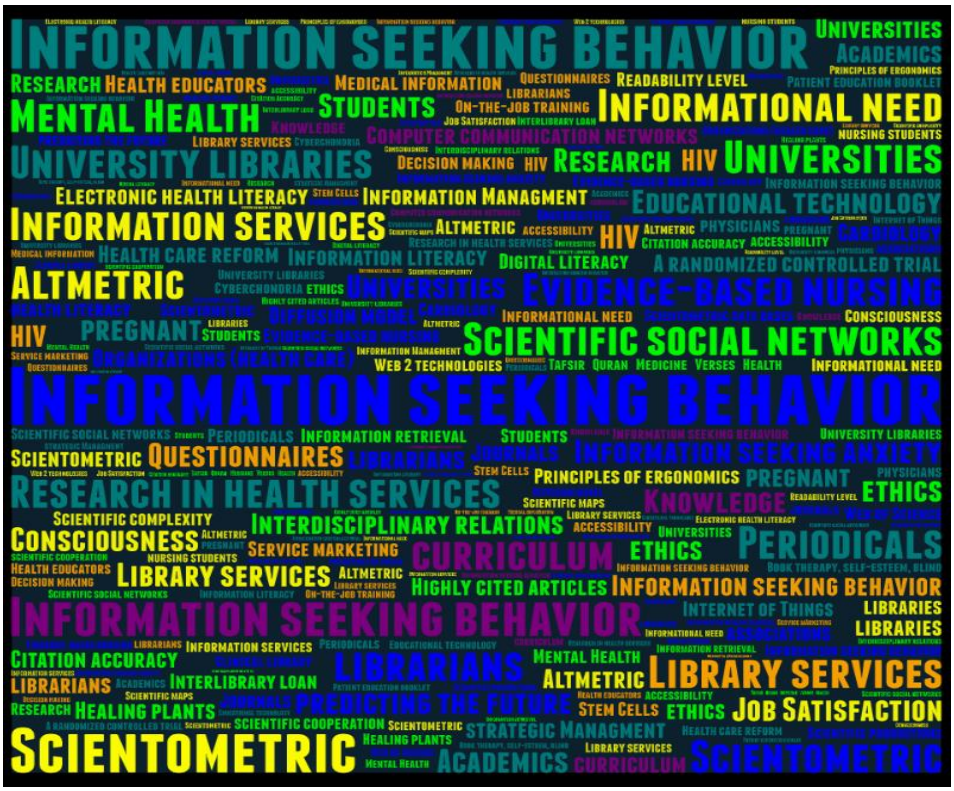

Figure 2: Active subjects researched in theses of postgraduate studies in MLIS of the country (2017-2021)

Most essential keywords in these on medical librarianship and medical information by year

We categorized the keywords used in the theses by year to provide a clear view of the frequency of topics for each year and to illustrate trends in topic changes. In this section, we employed the TF-IDF algorithm. By analyzing the abstracts of the theses and creating a word matrix, we classified them by year. As shown in Table 1, from 2017 to 2021, the most commonly used keywords included digital literacy, health literacy, altmetrics, ResearchGate, scientometrics, journals, cancer and patient education. This suggests that there was no significant change in the trends of MLIS studies during these years, as most keywords were repetitive and confined to specific subject areas. Based on Table 1, the main topics from 2017 to 2021 included bibliometrics, information literacy, health literacy, patient education, and ethics, with interdisciplinary keywords and the clinical aspects of library science receiving

renewed attention in 2021.

Table 1
Keywords researched in theses of postgraduate studies in MLIS of the country (2017-2021)

Years	Keywords
2017	['Service Marketing', 'Digital Literacy', 'Academia', 'Cancer', 'Research Gate', 'Individual Factors', 'University', 'Rehabilitation Sciences', 'Information Anxiety', 'Scientific Productions', 'Iran, Malaysia, Turkey', 'Pregnant women', 'Strategic management', 'Randomized controlled trial', 'Cited articles', 'Students', 'Quality.', 'Cancer]
2018	['Computer Communication Networks', 'Ministry of Health', 'Interpretation of the Qur'an, Medicine, Verses, Health', 'Patient Education', 'Knowledge Sharing', 'Health Literacy', 'Evidence-Based Nursing', 'Infrastructure, ', 'journals', 'university libraries', 'interdisciplinary relations', 'health information', 'online information']
2019	[Access to information, 'Information', 'Financial gratitude', 'Cancer', 'Surveys and questionnaires', 'Altometrics', 'Microbiology', 'Musculoskeletal disorders', 'Patient education', 'Knowledge', 'Health literacy', 'Digital literacy', 'Information storage and retrieval', 'In-service training', 'Patient education', 'Clinical librarian', 'Research Gate', 'Health websites', 'Quality of life-related to Health', 'Endocrine and Metabolism', 'Bibliographic databases', 'Book therapy, self-esteem, blind', 'Diffusion model', 'Cancer', 'Scientometric system', 'Information seeking behavior', 'Qualification' professional', 'information-seeking behavior', 'health information', 'health information']
2020	['transgenic', 'university libraries', 'health system priority', 'knowledge', 'ethics', 'occupational therapy', 'knowledge', 'ethics', 'health information', 'health literacy', 'heart diseases' and vessels', 'knowledge', 'citation', 'health websites', 'information-seeking behavior', 'body image']
2021	['Patient Education', 'Digestion and Liver', 'Scientometric System', 'Pregnant Women', Information Science', 'Patient Education', 'Citation', 'Evaluation Studies', 'Libraries', 'Cyberchondria', 'Open access journals', 'Coronavirus-related health literacy', 'Health literacy']

Popular research topics in theses on MLIS among universities

This study also examined the popular topics of interest among faculty and students at each university. As shown in Table 2, the subjects explored at MUI were primarily focused on the clinical aspects of the field, including cancer studies, health literacy, and information-seeking behaviors. In contrast, the TUMS concentrated more on the foundational elements of LIS, such as library services, information dissemination, academic libraries, educational programs, and health literacy, while addressing some clinical aspects. At SBMU, the topics predominantly revolved around scientometrics and bibliometrics, with researchers investigating journals, articles, library websites, academic social networks, scientific outputs, and bibliographic databases. Due to the absence of a Ph.D. program in this field and the limited number of theses, UMSHA encompassed fewer topics, primarily focusing on information retrieval and related behaviors and discussions on health knowledge and information. The IUMS, similar to the MUI, has placed a greater emphasis on patient education and evidence-based medicine.

Table 2
The keywords researched in the theses of graduate studies in MLIS of the country (2017-202) by the university

Universities	Keywords
IUMS[1]	['Individual factors', 'Health websites', 'Rehabilitation sciences', 'Knowledge sharing', 'MS,' 'Evidence-based nursing', 'Knowledge', 'Health literacy', 'Digital literacy', 'Information', 'Information storage and retrieval', 'In-service training', 'Patient education', 'Clinical librarian', 'Groups', 'Knowledge', 'Ethics', 'Conceptual model', 'Knowledge', 'Ethics', Library services', 'mental health', 'health literacy', 'cardiovascular diseases', 'knowledge', 'pregnant women', 'information science', 'patient education', 'citation', 'personal factors' ', 'evaluation studies', 'library services']
TUMS[2]	['Library Services', 'Digital Literacy', 'Academia', 'Cancer', 'Research Gate', 'Computer Communication Networks', 'Ministry of Health', 'Tafseer، Qur'an، Medicine ، Verses، Health', 'Patient Education', 'Information', 'Educational Program', 'Financial Gratitude', 'Cancer', 'Physician Assistants', 'Altometrics', 'Microbiology', 'Musculoskeletal Disorders', 'Education' Patient', 'transgenic', 'patient education', 'university libraries', 'priority of health system', 'digestive and liver', 'scientometric system']
UMSHA[3]	['information need', 'information seeking behavior', 'gynecology and obstetrics', 'bibliometrics', 'organizational ethics', 'library services', 'information seeking behavior', 'information centers', 'knowledge', 'individual factors', 'health information', 'physical fitness', 'information-seeking behavior', 'health-related quality of life', 'information-seeking anxiety', 'knowledge', 'online assessment tools']
MUI[4]	['Students', 'Library services', 'Breast cancers', 'Health information', 'Online information', 'Information science', 'Cancer', 'Scientometric system', 'Individual factors', 'Qualification' professional', 'information-seeking behavior', 'health information', 'health information', 'information-seeking behavior', 'body image', 'health literacy']
SBMU[5]	['Information anxiety', 'Scientific productions, Iran, Malaysia, Turkey', 'Pregnant women', 'Infrastructure', 'Strategic management', 'Consort checklist', 'Journals', 'Referenced articles', 'University Libraries', 'Research Gate', 'Interdisciplinary Relations', 'Health Websites', 'Quality of Life-Related to Health', 'Endocrine and Metabolism', 'Bibliographic Databases', ', Citation', 'Book of therapy, self-esteem, blind', 'Academic website', 'Cyberchondria', 'Open access journals', 'Coronavirus-related health literacy']

Discussion

This study aimed to examine master's and doctoral theses in MLIS to determine their thematic trends among medical universities in Iran. The findings revealed that the most active topics during the years under review included information-seeking behavior, scientometrics, information needs, health, and information services. Additionally, the results indicated that the topics of greatest interest to faculty and students were knowledge sharing, digital literacy, library services, information-seeking behavior, and bibliometrics. Dastani, Mousavi Chelak, et al. (2022) identified information services, scientometrics, bibliometrics, web-based treatment, and information literacy as six main topics in MLIS articles up to 2019, aligning with the findings of this study. Gholami, Sals Faal, 2023) concluded that the primary issues in knowledge science included organizational culture, social networks, and information literacy, with scientific, social networks, and information literacy appearing in this study, albeit not classified as popular or primary topics.

In the study by Kurata, Miyata, Ishita, Yamamoto, Yang, and Iwase (2018) based on a

thematic model approach, the topics of information literacy, citation analysis, and information sharing were included among the thirty main topics from 2000 to 2017. This study also analyzed the keywords used in the abstracts, which were categorized by year. The evaluations revealed little variation, with only a few new keywords emerging in specific years. For instance, in 2017, the keywords service marketing, rehabilitation science, and information anxiety were not addressed in subsequent years. Keywords consistently focused on by authors included knowledge, cancer, health literacy, patient education, libraries, and information. The analysis showed that in recent years, students have increasingly focused on treatment and health services within the realm of library science and interdisciplinary topics. However, there were no significant differences across the years.

Dastani et al. (2021) highlighted the importance of keywords such as patient, librarian, intervention, inform review, social, care, student, and online from 2015 to 2020. They also concluded that "patient" and "care" emerged as essential keywords in library articles post-2000. Since medical librarians serve as providers of health information services for patients and physicians, they must adapt to the evolving needs of society. Ashrafi, Hodhodinezhad, Shahrzadi, and Soleymani (2017) concluded that the activities of medical librarians and information professionals extend beyond libraries and that they can play significant roles in areas such as media literacy, health literacy, health information, and clinical services. This study similarly indicated an increased focus on these topics in recent years.

The study also addressed the popularity of various topics among medical universities in the country, revealing that specific themes, such as scientometrics and altmetrics, were common across all institutions. As demonstrated in the study by Sahoo, Mohanty, Biswal, Dash, and Sahu (2020), research impact measurement and citation-based studies constituted 71% of the topics in articles published in 2019. Additionally, Sahu and Parabhoi (2020) noted that bibliometrics and scientometrics were the most studied topics in India from 2014 to 2018. However, some universities, such as IUMS and MUI, emphasized the field's clinical aspects and topics related to patients and health literacy. In contrast, the University of Tehran focused more on library services, academic libraries, and educational programs.

Myers (2020) reviewed the abstracts of the Medical Library Association from 2001 to 2019, identified themes such as academic infrastructure, librarian roles, publication, literature searching, and library services, which were also present in the IUMS and Shahid Beheshti University studies. Ghanadi Nezhad, Osareh, and Ghane (2022) showed that the topics studied in library and information science up to 2020 included information technology, web studies, libraries, information marketing, and information and digital literacy, which were notably absent from medical library theses. The keywords used in the theses from 2019 and 2020 indicated that information literacy and evidence-based medicine, along with a focus on the clinical aspect of library science, have gained more attention in recent years. The results of Akbari, Soleymanpour, and Sedghi's (2023) study were consistent with this study, indicating that this area has become a research priority for faculty members and students in library science, compared to other topics.

Conclusion

The topics published across various fields continuously evolve, with new subjects frequently brought to discussion and less significant or older topics becoming obsolete. The MLIS is not exempt from this trend. While numerous studies have been conducted on published

articles in MLIS in Iran and other countries, research on library theses provides a comprehensive perspective for faculty and students in selecting non-repetitive topics, thereby saving time and resources. Additionally, it serves policymakers in clarifying pathways for utilizing findings in educational discussions and identifying research gaps within the field. The current study indicates that there has been an increasing focus on new areas such as patient information, health information, health literacy, scientometrics, and library services in recent years. However, the pace of change has not kept up with the advancements in science. Recent technological revolutions, including the advent of artificial intelligence and big data, have led many topics in this field to shift towards these areas. Yet, there remains a noticeable absence of such subjects in medical library theses.

Furthermore, given the importance of the thesis as a course requirement and a prerequisite for graduation for doctoral and master's students, the findings suggest that students and faculty are not as invested in selecting thesis topics as they are in publishing articles in this domain. Consequently, a gap exists regarding current and challenging issues relevant to the information age. Due to the lack of collaboration with two other universities in Iran and the incomplete data recorded by library systems during indexing, along with the slow pace of research, the timeframe for this study was limited to 2020, resulting in a smaller sample size of data collected. Future research could yield more comprehensive analytical results by considering a broader statistical population. Additionally, it is recommended that a qualitative study be conducted to explore the research interests of students and faculty and the significance they place on selecting thesis topics.

Acknowledgment

This study is part of a Student Grant for the first score in the master's entrance exam in the medical library, and information sciences approved (code: IUMS_1401-3-75-24605) at Iran University of Medical Sciences, Iran.

Conflict of interest

The authors declare that there is no conflict of interest.

Funding sources

This work was supported by the Vice-Chancellor of Research and Technology, Iran University of Medical Sciences (Grant Number: 140233726930)

Endnotes

1. Iran University of Medical Sciences
2. Tehran University of Medical Sciences
3. Hamadan University of Medical Sciences
4. Isfahan University of Medical Sciences
5. Beheshti University of Medical Sciences

References

Akbari, Z., Soleymanpour, S. & Sedghi, S. (2023). Research priorities of Iranian faculty members and PhD students in medical library and information science. *Journal of Modern Medical Information Sciences*, 8(4), 350–361. https://doi.org/10.32598/jmis.8.4.4 [in Persian]

Armann-Keown, V. & Patterson, L. (2020). Content analysis in library and information research: An analysis of trends. *Library & Information Science Research*, 42(4), 101048. https://doi.org/10.1016/j.lisr.2020.101048

Ashrafi-Rizi, H., Hodhodinezhad, N., Shahrzadi, L. & Soleymani, M. (2017). A study on the novel services of medical librarians in health information services: A narrative review. *Health Information Management*, 13(6), 438-444. Retrieved from https://him.mui.ac.ir/article_11550_f3f3eb3f0cf67c429b059ba8d3d245a5.pdf?lang=en [in Persian]

Bittermann, A. & Fischer, A. (2018). How to identify hot topics in psychology using topic modeling. *Zeitschrift für Psychologie*, 226(1), 3-13. https://doi.org/10.1027/2151-2604/a000318

Cheng, Q. & Lui, C. S. (2021). Applying text mining methods to suicide research. *Suicide and Life-Threatening Behavior*, 51(1), 137-147. https://doi.org/10.1111/sltb.12680

Dastani, M., Chelak, A. M., Ziaei, S. & Delghandi, F. (2021). Identifying emerging trends in scientific texts using TF-IDF algorithm: A case study of medical librarianship and information articles. *Health Technology Assessment in Action*. https://doi.org/10.18502/htaa.v4i2.6231Vol (Issue), pages [in Persian]

Dastani, M., Ehtesham, H., Javanmard, Z., Sabahi, A. & Bahador, F. (2022). Identifying the trends of global publications in health information technology using text-mining techniques. *Shiraz E-Medical Journal*, 23(11), e123803. https://doi.org/10.5812/semj-123803

Dastani, M., Mousavi Chelak, A., Ziaei, S. & Delghandi, F. (2022). Discovering research topics from medical librarianship and information using text mining. *International Journal of Information Science and Management (IJISM)*, 20(2), 201-216. Retrieved from https://ijism.isc.ac/article_698388_cd6d711c740fe816aa92070b5f40b00e.pdf

Fang, D., Yang, H., Gao, B. & Li, X. (2018). Discovering research topics from library electronic references using latent Dirichlet allocation. *Library Hi Tech*, 36(3), 400-410. https://doi.org/10.1108/lht-06-2017-0132

Finlay, C. S., Sugimoto, C. R., Li, D. & Russell, T. G. (2012). LIS dissertation titles and abstracts (1930–2009): Where have all the librar* gone?. *The Library Quarterly*, 82(1), 29-46. https://doi.org/10.1086/662945

Ghanadi Nezhad, F., Osareh, F. & Ghane, M. R. (2022). Forecasting the subject trend of international library and information science research by 2030 using the deep learning approach. *International Journal of Information Science and Management (IJISM)*, 20(1), 471-492. Retrieved from https://ijism.isc.ac/article_698379_4c1260976bb764e8a20fcb26d45721e1.pdf

Gholami Sals Faal, H., Behzadi, H., & Jampour, M. (2023). Topic trends in knowledge and information science in domestic prestigious Iranian journals based on the LDA model. *Sciences and Techniques of Information Management*, 9(2), 31-58. https://doi.org/10.22091/stim.2022.8532.1853 [in Persian]

Hwang, S.-Y., Wei, C.-P., Lee, C.-H. & Chen, Y.-S. (2017). Coauthorship network-based literature recommendation with topic model. *Online Information Review*, 41(3), 318-336. htpps://doi.org./ 10.1108/OIR-06-2016-0166

Islam, M. S., Islam, M. N. & Mondal, M. (2018). Research trends in library and information science in Bangladesh: An analytical study. *Journal of Information Science Theory and Practice*, 6(2), 36-45. https://doi.org/10.1633/JISTaP.2018.6.2.3

Kim, Y.-M. & Delen, D. (2018). Medical informatics research trend analysis: a text mining approach. *Health Informatics Journal*, 24(4), 432-452. https://doi.org/10.1177/1460458216678443

Kurata, K., Miyata, Y., Ishita, E., Yamamoto, M., Yang, F. & Iwase, A. (2018). Analyzing library and information science full-text articles using a topic modeling approach. *Proceedings of the Association for Information Science and Technology*, 55(1), 847-848. https://doi.org/10.1002/pra2.2018.14505501143

Lam, C., Lai, F.-C., Wang, C.-H., Lai, M.-H., Hsu, N. & Chung, M.-H. (2016). Text mining of journal articles for sleep disorder terminologies. *PloS One*, 11(5), e0156031. https://doi.org/10.1371/journal.pone.0156031

Mansouri, A., Zarmehr, F. & Karshenas, H. (2020). A review of text mining approaches and their function in discovering and extracting a topic. Human Information Interaction, 7(1), 15-26. http://dor.org/20.1001.1.24237418.1399.7.1.5.7 [in Persian]

Myers, B. (2020). What we talk about when we talk about medical librarianship: An analysis of Medical Library Association annual meeting abstracts, 2001–2019. *Journal of the Medical Library Association: JMLA*, 108(3), 364-377. https://doi.org/10.5195/jmla.2020.836

Putri, P. Y. A. & Saputra, K. A. K. (2021). Writing of scientific works in a dissertation context and the difference with thesis. *International Journal of Business, Economics and Law*, 24(4), 1-7. Retrieved from https://ijbel.com/wp-content/uploads/2021/06/IJBEL24-701.pdf

Sahoo, J., Mohanty, B., Biswal, O., Dash, N. K. & Sahu, J. K. (2020). Authorship trend and content analysis: A case study on highly cited articles in library and information science journals. Performance Measurement and Metrics, 21(1), 33-51. https://doi.org/10.1108/PMM-06-2019-0021

Sahu, R. & Parabhoi, L. (2020). Bibliometric study of library and information science journal articles during 2014 2018: LIS research trends in India. *DESIDOC Journal of Library & Information Technology*, 40(6), 390-395. https://doi.org/10.14429/djlit.40.06.15631

Salloum, S. A., Al-Emran, M., Monem, A. A. & Shaalan, K. (2018). Using text mining techniques for extracting information from research articles. *Intelligent Natural Language Processing: Trends and Applications*,740, 373-397. https://doi.org/10.1007/978-3-319-67056-0_18

Sanandres, E., Abello, R. & Madariaga, C. (2020). Topic modeling of Twitter conversations: the case of the National University of Colombia. In *Text Analytics: Advances and Challenges* (pp. 241-251). Springer International Publishing. https://doi.org/10.1007/978-3-030-52680-1_19

Shokouhian, M., Asemi, A., Shabani, A. & Cheshme-Sohrabi, M. (2019). Combined bibliometric and text-mining analysis of scientific productions in PubMed database in the field of electronic health records. *Health Information Management*, 16(4), 190-196. https://doi.org/10.22122/him.v16i4.3953

Sugimoto, C. R., Li, D., Russell, T. G., Finlay, S. C. & Ding, Y. (2011). The shifting sands of disciplinary development: Analyzing North American library and information science dissertations using latent dirichlet allocation. *Journal of the American Society for Information Science and Technology*, 62(1), 185-204. https://doi.org/10.1002/asi.21435

Thakur, K. & Kumar, V. (2022). Application of text mining techniques on scholarly research articles: Methods and tools. *New Review of Academic Librarianship*, 28(3), 279-302. https://doi.org/10.1080/13614533.2021.1918190

Wu, W.-T., Li, Y.-J., Feng, A.-Z., Li, L., Huang, T., Xu, A.-D. & Lyu, J. (2021). Data mining in clinical big data: The frequently used databases, steps, and methodological models. *Military Medical Research*, 8, 44. https://doi.org/10.1186/s40779-021-00338-z

Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J. F., & Hua, L. (2012). Data mining in healthcare and biomedicine: A survey of the literature. *J Med Syst*, 36(4), 2431-2448. https://doi.org/10.1007/s10916-011-9710-5

Zhang, W., Yoshida, T. & Tang, X. (2011). A comparative study of TF* IDF, LSI and multi-words for text classification. *Expert Systems with Applications*, 38(3), 2758-2765. http://doi.org/10.1016/j.eswa.2010.08.066

Zong, Q., Shen, H.-Z., Yuan, Q., Hu, X.-W., Hou, Z.-P. & Deng, S.-G. (2013). Doctoral dissertations of library and information science in china: A co-word analysis. *Scientometrics*, 94, 781-799. https://doi.org/10.1007/s11192-012-0799-1