




# Knowledge Discovery from the Digital Library's Contents: Bangladesh Perspective

Md. Habibur Rahman<sup>1,2</sup> , Sohaimi Zakaria<sup>3</sup>, and Azree Ahmad<sup>4</sup>

<sup>1</sup> Chattogram Veterinary and Animal Sciences University, Chattogram, Bangladesh  
librarian@cvasu.ac.bd

<sup>2</sup> Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia

<sup>3</sup> Faculty of Information Management, Universiti Teknologi MARA (UiTM),  
Shah Alam, Malaysia  
sohaimiz@uitm.edu.my

<sup>4</sup> Faculty of Information Management, Universiti Teknologi MARA (UiTM),  
Kedah Branch, Merbok, Malaysia  
azree@uitm.edu.my

**Abstract.** The purpose of this study is to explore the present trends of knowledge discovery (KD) from digital library (DL) systems in Bangladesh. The main obstacles of KD from the contents of DL and ways to overcome the barriers are also described. This study uses both qualitative and quantitative approaches along with the review of related literature. The present scenario of the KD from the contents of DL in Bangladesh is presented by a survey with a structured questionnaire and reviewing related literature. This study identifies the challenges of KD from the contents of DL in Bangladesh, which are inaccurate bibliographic metadata, a lack of accurate holdings data, a lack of synchronized bibliographic metadata, etc. This study also suggests some suitable ways to overcome the existing challenges of KD from DL contents, such as providing high-quality data, preserving complete bibliographic metadata, preserving accurate holdings data, synchronizing bibliographic metadata and holdings data, and using consistent data formats.

**Keyword:** Knowledge discovery · Digital library · Contents of digital library · Information retrieval · Library of Bangladesh · Discovery tools · Metadata

## 1 Introduction

The contents of the digital library (DL) are critical for any twenty-first-century library user. Especially it is very important for students, faculty members, researchers, and scientists who want to get information in the quickest possible time. As a result of the advancement of ICT and the demand for current users, DL has been established all over the world to effectively and efficiently serve those communities. With the advent of modern science and technology, information management services that enable the creation of digital libraries have significantly improved [1]. As a consequence, the authorities of that DL would purchase a significant number of digital resources each year based on their ability to meet patrons' demands. However, due to a lack of good searching tools, many

great library materials are overlooked. Libraries spend a significant amount of money each year growing their collections through various information resources or platforms, yet most of these resources are underutilized owing to a lack of a robust search engine with a single interface [2].

Knowledge discovery (KD) or retrieval of information from sources is very critical for any library user. KD tools are essential for any library to make its materials visible to users so that they can make appropriate use of them. KD tools use many approaches to search for and extract usable knowledge from data, databases, and documents [3]. However, KD tools can work effectively and efficiently if the library metadata and holdings data are properly managed. Discovery and access can be difficult or impossible without accurate bibliographic metadata and holdings data [4]. Nowadays, the most crucial responsibility of library professionals is to ensure KD from DL content is effectively and efficiently done in order to provide good library services to their users. As a consequence, they must rely on advanced technology, cutting-edge tools, and approaches that are critically needed in the DL to help make use of the contents of crucial information sources and locate the knowledge [5].

The purpose of this study is to explore the present trends of KD from DL systems in Bangladesh. It tried to identify the advantages and disadvantages, current barriers and challenges, and the tools and technologies required for KD from DL contents. It also attempted to offer effective strategies for overcoming the existing barriers and obstacles associated with KD. As a result, this research will be highly useful to all Bangladeshi and worldwide information scientists in understanding KD from the content of DL. It will serve as a guideline for library academicians, practitioners, and patrons.

The rest of the paper is structured as follows. The second section describes the conceptual map by reviewing related literature. The third section explains the aims and objectives of this study. The fourth section presents the research questions. The fifth section presents the research methodology. The sixth section describes the analysis of data, findings and discussion. The seventh section presents the technique of discovering knowledge as per the library professionals in Bangladesh. The eighth section mentions the limitations and direction for future research and the final and ninth section concludes the paper.

## **2 Literature Review**

### **2.1 Concept of Knowledge Discovery**

Knowledge discovery has been a hot topic not only in data mining and artificial intelligence but also in many other disciplines throughout the history of humanity in general and in particular over the past decades in the digital age [7]. Knowledge discovery is responsible for the quality of data entering discovery systems [8]. The high-level process of extracting effective, undiscovered, possibly valuable, and ultimately understood patterns from enormous amounts of data is referred to as knowledge discovery [9]. Web software that searches journal articles and library catalog metadata in a unified index and shows search results in a single interface are known as discovery tools [2]. The knowledge discovery process has some distinct characteristics, particularly when dealing with data of high velocity, variety, and volume [10]. Characteristics such as large data volume,

knowledge discovery efficiency, the accuracy of discovered knowledge, and discovery automation are linked and must be addressed together for reliable knowledge discovery [10, 11]. Knowledge Discovery in databases is the process of extracting usable information from a huge data set using a data mining algorithm with specified metrics and thresholds and in more recent times, artificial intelligence approaches have been utilized to achieve this goal [11]. Because humans' intelligent actions include discovering knowledge, extracting knowledge, and mining knowledge, knowledge discovery is a type of intelligent behavior [7, 12].

## 2.2 Tools of Knowledge Discovery

A knowledge discovery tool is defined as a search engine that uses unified indexes of licensed scholarly information to search across multiple library databases provided by various vendors and can be customized for size, range, and comprehensiveness of data inclusion for targeted solutions [6]. Discovery tools that ingest metadata into a single index employ a single set of search algorithms to retrieve and rank results [2]. Although it is evident that a tool like Summon is required and value to library users, libraries must take care to position such devices in such a way that they can be a successful addition to the resource discovery dynamic [13]. Discovery tools, such as EBSCO Discovery Service, summon service (Serial Solution), Encore Discovery (Innovative interface), and Primo Central (Ex Libris Group), provides both opportunities and challenges for library instruction, depending on the academic discipline, users' knowledge, and information-seeking need [2]. A metasearch engine, sometimes known as a search aggregator, is an online information retrieval tool that generates its own results using the data of a web search engine. Metasearch engines take user input and query search engines for results right away. The users are supplied with enough data that has been acquired, ranked, and presented to them [14]. Federated search is a key component of an Information Portal, which serves as a gateway to a variety of information sources. When a user types a search query into the Information Portal's search box, the system uses federated search technology to send the search string to each resource included in the Portal [15].

## 2.3 The Current State of Digital Library Practice in Bangladesh

We have become comfortable with the DL, which is paperless, borderless, and always accessible from anywhere in the globe, due to the use of information and communication technology. Digital libraries and Institutional Repositories have grown increasingly popular as a means of gaining quick access to electronic information, however digital library activities in Bangladesh are still in their infancy [16]. In Bangladesh, the expansion of IRs is modest and it has a poor rate of IR development as compared to other Asian countries [17]. According to ROAR (Registry of Open Access Repositories), the number of repositories in Bangladesh is quite low, with only twelve [18]. They suggested that Bangladeshi librarians may require assistance in developing repositories in their own organizations. Islam & Naznin pointed out that the implementation of DL in Bangladesh is difficult, and specified library users should be knowledgeable of how to utilize it. It also necessitates an educated workforce, enhanced infrastructural facilities, and government attention [19]. Islam and Naznin; Rahman et al. noted significant progress in

library digitization efforts in Bangladesh over the last decade, as well as issues like capacity building, shrinking funding, insufficient facilities, and traditional duties being supplanted by modern technologies. They also promoted joint digitalization efforts and provided helpful recommendations [19, 20].

### 3 Objectives of the Study

The main objective of this study is to explore the present trend of KD from digital library systems in Bangladesh. The more particular objectives are:

- a) To determine the pros and cons of knowledge discovery from DL contents
- b) To identify the existing barriers and challenges of DL
- c) To identify the tools and technologies that are required for KD from DL contents
- d) To offer effective ways to overcome the existing barriers and challenges of KD.

### 4 Research Questions (RQs)

Based on the above objectives, one major research question (MRQ) and three subsidiary research questions (SRQs) have been formulated that will guide the study.

MRQ1. What is the present trend of KD from the DL system in Bangladesh?

SRQ2. What difficulties and challenges are being faced in knowledge discovery from the DL content in Bangladesh?

SRQ3. What types of tools and technologies are required for knowledge discovery from DL contents?

SRQ4. How could the existing barriers and challenges of KD be overcome?

### 5 Research Methodology

Both quantitative and qualitative methods have been applied in this study. A structured questionnaire was designed and sent to the library professionals through emails for collecting data. The questionnaire was sent to 150 library professionals of different organizations, including universities, research institutes, etc. in Bangladesh. We received 85 filled questionnaires. Among 85 respondents 78 respondents, filled questionnaires appropriately and the remaining 7 respondents did not fill properly. Therefore, we used 78 responses in this study. Due to the current pandemic circumstances, many professionals are staying in remote areas where internet connectivity does not work properly. Hence, they could not participate in this study. We conducted telephone interviews with key professionals who are actively involved in directing DL. The quantitative data were analyzed using SPSS (version 20.0) software and qualitative data were analyzed thematically.

## 6 Findings and Discussion

### 6.1 Demographic Profile of the Respondents

This section focuses on the male and female ratio as well as the respondents' highest levels of education. Among the 78 respondents, 70.5% (55) respondents were male, and 29.5% (23) were female. The educational qualification of the respondents is seen that the highest 77% of respondents have a Master of Arts (M.A.) degree, while the lowest 4% have the Masters of Social Science (M.S.S.) degree. The second highest 13% has a master's of Philosophy (M.Phil) degree and 6% of respondents have a Doctor of Philosophy (Ph.D.) degree.

### 6.2 Designation of the Respondents

According to the survey, it is seen that the highest percentage of responses (36%) from the "Assistant librarian", followed by "Deputy Librarian" (32%), "Librarian" (17%), and "Library Officer" (4.5%) "Additional Librarian" (3%) Deputy Director (3%). We received only one response (1%) from "Chief librarian", "Principal Scientific Officer", "Assistant Library officer" and "Cataloguer" respectively.

### 6.3 Competencies of Respondents in Digital Library Management

This section summarizes the results of the survey on respondents' digital library management skills. The highest 44.9% of respondents have an intermediate level of competencies in the management of the digital library, on the other hand, 30.8% of respondents are beginners in the management of the digital library. 14.1% of respondents stated that they have excellent knowledge in the management of digital libraries, while 10.3% mentioned there have no setup of digital libraries.

### 6.4 The Tools and Technologies Needed to Discover Knowledge from the DL Contents

The respondents were asked to indicate their level of agreement with some specific tools and technologies which are required for Knowledge discovery from the content of digital libraries. The level of agreement of the respondent is analyzed on 1–5 Likert scales in the Sect. 6.4, 6.5 and 6.6. The mean and standard deviation of the responses were calculated according to the following scores: strongly disagree = 1.00, disagree = 2.00, neutral = 3.00, agree = 4.00, and strongly agree = 5.00 using the descriptive analysis techniques of SPSS.

The highest mean score was 4.49 for the statement of "OPAC," while the lowest mean score was 4.03 for the statement of "Apps." The second highest mean score was 4.42 for the statement of "Library Catalogue," followed by 4.35, 4.26, 4.21, 4.18, 4.15 and 4.14 for the statements of "Search engine," "Index," "Discovery software", "Document Object Identifier (DOI)", "Information of contents page", and "Metadata" respectively.

## 6.5 The Difficulties and Obstacles of Discovering Knowledge

The opinion of the respondents regarding the difficulties and obstacles of discovering knowledge from the content of a digital library is stated below.

### **Inaccurate Bibliographic Metadata**

Without the proper bibliographic metadata, users won't be able to find the information they need [4]. The respondents agreed with the statement (with a mean score of 3.91) that content could not be traced due to inaccurate bibliographic metadata.

### **Lack of Accurate Holdings Data**

The content from the digital library is unable to discover due to a lack of accurate holdings data. The respondents agreed (with a mean score of 3.83) that one of the most significant barriers to discovering knowledge from the digital library's contents is a lack of accurate holdings data.

### **Lack of Synchronized Bibliographic Metadata and Holdings Data**

The respondents agreed with the statement (with a mean score of 4.24) that unsynchronized metadata and holdings data creating hindrance in discovering knowledge.

### **Libraries Receive Data in Multiple Formats**

The respondents agreed with the statement (with a mean score of 4.21) that the libraries have to receive data in multiple formats, which causes difficulties discovering knowledge from the digital library's content.

## 6.6 Ways of Overcoming the Barriers and Challenges of KD from Contents of DL

The respondents were asked to specify how to overcome the barriers and challenges of KD from the digital library's content. The responses of the respondents are presented below.

The highest mean score was 4.60 for the statement of "Using Knowledge Bases And Related Tools (KBART) and MARC standards," while the lowest mean score was 4.12 for the statement of "Preserving complete bibliographic metadata." The second highest mean score was 4.37 for the statement of "Providing high-quality data," followed by the score of 4.32, 4.28, 4.24 for the following statements of "Using consistent data formats", "Preserving accurate holdings data" and "Synchronizing bibliographic metadata and holdings data" respectively.

## 6.7 The Most Commonly Used Discovery Tool/software in the Libraries of Bangladesh

The most commonly used discovery tool/software in Bangladeshi libraries. The highest 55 (66.7%) respondents agreed that "VuFind" is used to discover knowledge in Bangladeshi libraries, while the lowest 2 (2.6%) respondents believed that "Encore" and "Blacklight" are used for the same purpose. The second highest 37 (47.4%) respondents

agreed that “EBSCO’s Discovery Service” are being used in libraries in Bangladesh followed by 16 (20.5%), 15 (19.2%), 9 (11.5%), 8 (10.3%), 6 (7.7%) for the tools/software “eXtensible Catalog”, “WorldCatLocal”, “BiblioCommons”, “ProQuest AquaBrowser”, “Ex Libris” respectively.

## **7 The Technique of Discovering Knowledge as Per the Library Professionals in Bangladesh**

An open-ended question, “Would you please share your thoughts on how to effectively discover knowledge from the digital library’s content?” was asked through the questionnaire. The responses to this question were analyzed and categorized into themes as discussed below.

Creating systematic index and keywords: IP-1, IP-5, and IP-7 (Interview Participant) mentioned that the knowledge could be discovered from the content of the digital library by creating a systematic index of resources and searching knowledge applying appropriate keywords of desired information. Powerful search engine: IP-2 and IP-10 stated that a powerful search engine should be developed for discovering knowledge. A search engine can find information from any organized digital source. It is a popularly used tool all over the world for discovering knowledge. The resources should be collected from authentic sources: IP-50 and IP-72 emphasized the collection of resources from authentic, renowned, and standard organizations. They revealed that a world-famous organization generally maintains proper quality and standard system from creation to delivery of information. Hence the information provided by world-class organizations that information can be easily discovered using searching tools. Use of Artificial Intelligent and Big Data technology: IP-22, IP-66, and IP-71 stated that knowledge might be discovered effectively and efficiently from the contents of the digital libraries utilizing Artificial Intelligence (AI) and Big Data technologies. DL can leverage AI and Big Data technologies to discover knowledge for its databases, as they are used in almost every industry. Following standard metadata: IP-33, IP-77 and IP 80 stated that every document should be uploaded in the digital library with the following standard metadata and setting all the possible keywords and subjects should be added as much as possible. The database should be relational; the author, subject, place, publisher, supplier, etc. should be hyperlinked. Search can be filtered within the search result. The interface should be user-friendly.

As per the study respondents, KD from the contents of DL would be simple to discover using the approaches indicated.

## **8 Limitations and Direction for Future Research**

In this pandemic circumstance, educational institutes and most research organizations in Bangladesh and the rest of the world are physically closed. The information and library professionals from those educational institutes and research organizations make up the research population for this study. We were unable to communicate with many professionals since they were not available at their workstations; as a result, we could

not reflect their views in the article. Future research could involve a larger sample size and a look at the demographic information of the diverse responders. As a result of the limitations mentioned above, other scholars may be inspired to conduct further empirical research in this field.

## 9 Conclusion

The aim of the study was to explore the present trend of KD from digital library systems in Bangladesh. The findings showed that 70 respondents (out of 78) stated that they are successfully capable of running the DL and some of them are experts in DL management and only 8 respondents mentioned they have not existed the DL. Moreover, the study exposed that “VuFind” and “EBSCO’s Discovery Service” software are being popularly used as KD tools in the libraries in Bangladesh. As a result, Bangladesh has a healthier overall scenario in terms of DL setup, skilled manpower, management, and KD from DL’s content. The findings also revealed that the DL faces numerous challenges in discovering knowledge from its content, including inaccurate bibliographic metadata, a lack of accurate holdings data, a lack of synchronized bibliographic metadata and holdings data, and data in a variety of formats, among others. However, this study recommended that DL should maintain high-quality data, preserving complete bibliographic metadata and accurate holdings data, synchronizing bibliographic metadata and holdings data, using consistent data formats and using Knowledge Bases And Related Tools (KBART) and MARC standards. Every year, almost every library acquires valuable resources for library clients based on its financial capabilities. Nevertheless, library users may not be able to read all of these materials since adequate discovery techniques are not used to retrieve information from the sources. Therefore, discovery tools are playing a crucial role to solve that problem. We may conclude that KD from DL’s content could be a success if data suppliers, service providers, and competent professionals work together to update and synchronize bibliographic and holdings data, as well as standardize data formats.

## References

1. Pan, Z.: Optimization of information retrieval algorithm of digital library based on semantic search engine. In: 2020 International Conference on Computer Engineering and Application (ICCEA). IEEE (2020)
2. Karadia, A., Pati, S.: Discovery Tools and Services for Academic Libraries (2015)
3. Shi, H., He, W., Xu, G.: Workshop proposal on knowledge discovery from digital libraries. In: Proceedings of the 18th ACM/IEEE Joint Conference on Digital Libraries, JCDL 2018, Fort Worth, Texas USA, 3–7 June 2018, vol. 2 (2018)
4. Kemperman, et al.: Success Strategies for Electronic Content Discovery and Access: A Cross-Industry White Paper. OCLC, Dublin (2014). <http://www.oclc.org/content/dam/oclc/reports/data-quality/215233-SuccessStrategies.pdf>
5. Viet, N.T., Kravets, A.G.: Analyzing recent research trends of computer science from academic open-access digital library. In: 8th International Conference on System Modeling & Advancement in Research Trends, 22nd–23rd November 2019, Proceedings of the SMART–2019. IEEE (2019)



6. Shi, X., Levy, S.: An empirical review of library discovery tools. *J. Serv. Sci. Manag.* **8**, 716–725 (2015). <https://doi.org/10.4236/jssm.2015.85073>
7. Sun, Z., Stranieri, A.: A knowledge discovery in the digital age. *PNG UoT BAIS* **5**(1), 1–11 (2020)
8. Sharma, G., Tripathi, V.: Effective knowledge discovery using data mining algorithm. In: Fong, S., Dey, N., Joshi, A. (eds.) *ICT Analysis and Applications*. LNNS, vol. 154, pp. 145–153. Springer, Singapore (2021). [https://doi.org/10.1007/978-981-15-8354-4\\_15](https://doi.org/10.1007/978-981-15-8354-4_15)
9. Zhu, H., Li, X.: Research on the application of knowledge discovery in digital library service. In: 7th International Conference on Social Network, Communication and Education (SNCE 2017), *Advances in Computer Science Research*, vol. 82 (2017)
10. Misra, S., Mukherjee, A., Roy, A.: Knowledge Discovery for Enabling Smart Internet of Things: A Survey. Wiley Periodicals, Inc. (2018)
11. Soundararajan, E., Joseph, J.V.M., Jayakumar, C., Somasekharan, M.: Knowledge discovery tools and techniques. In: Proceedings of the Conference on Recent Advances in Information Technology, Kalapakkam, India, vol. 141 (2005)
12. Russell, S., Norvig, P.: *Artificial Intelligence: A Modern Approach*, 3rd edn. Prentice Hall (2010)
13. Boyer, G.M., Besaw, M.: A study of librarians' perceptions and use of the summon discovery tool. *J. Electron. Resour. Med. Libr.* **9**(3), 173–183 (2012). <https://doi.org/10.1080/15424065.2012.707056>
14. Wikipedia: Metasearch Engine (2021). [https://en.wikipedia.org/wiki/Metasearch\\_engine](https://en.wikipedia.org/wiki/Metasearch_engine). Accessed 14 Sept 2021
15. Lingam, A.S.: Federated search and discovery solutions. *IP Indian J. Libr. Sci. Inf. Technol.* January-June **5**(1), 39–42 (2020)
16. Rahman, M.S.: Challenges and initiatives of digital library system and institutional repository: Bangladesh scenario. *Eastern Libr.* **25**(1), 1–23 (2020)
17. Elahi, M.H., Mezbah-ul-Islam, M.: Open access repositories of Bangladesh: an analysis of the present status. *IFLA J.* **44**(2), 132–142 (2018). <https://doi.org/10.1177/0340035218763952>
18. Chowdhury, M.H.H., Mannan, S.M.: Identifying the possible contents for university repositories of Bangladesh. *Eastern Libr.* **25**(2), 1–12 (2020)
19. Islam, M.S., Naznin, S.: Present status of digital library initiatives in Bangladesh. In: Proceedings of the 6th International Conference on Asia-Pacific Library and Information Education and Practice. *Asia-Pacific LIS: Exploring Unity Amid Diversity*, Philippine International Convention Center, Manila, Philippines, 28–30 October (2015)
20. Rahman, A.I.M.J., Rahman, M.M., Chowdhury, M.H.H.: Digital resources management in libraries: step towards digital Bangladesh. In: Proceedings of the National Seminar on Cross-Talk of Digital Resources Management: Step Towards Digital Bangladesh. Bangladesh Association of Librarians, Information Scientists, and Documentalists (BALID), Dhaka, pp. 1–24 (2015)