A Study on the Research Trends of Big Data at Public Libraries : with a Focus on the Journal "Public Library Quarterly"

Seongsin Lee∗, Hyunsook Kim∗∗, Sumin Baeck∗∗∗, Subin Yoon∗∗∗∗

ARTICLE INFO

ABSTRACT

This study aims to analyze the Big Data-related research trends in thefield of public library by using the social network analytical method. One hundred seventeen articles published in the journal 'Public Library Quarterly' were analyzed with author keywords and the frequency, degree centrality, while the betweenness centrality of the keywords were examined. The keywords "programs" and "community development" demonstrated the highest degree of centrality. The keywords "programs", "strategic planning", "community development", "future of libraries", "outreach", and "evaluation" posted a high degree of centrality and betweenness centrality. The keywords "measurement", "survey", "community partnerships", and "community engagement" demonstrated a high degree of centrality but not a high betweenness centrality. Meanwhile, the keywords "planning", "marketing", "community needs", and "community building" demonstrated a high betweenness centrality but not a high degree centrality. Based on the results of this study, public libraries should focus on the following directions of research when focusing on the Big Data. First, as mentioned in the above, a more in-depth discussion is needed regarding COVID-19 and social media. Second, the academic interest in the Big Data related education and training for the public librarians and educational contents is needed. Third, public libraries should think about the ways to efficiently perform their roles as a local data center, including their cooperation with other organizations in their local community.

∗   Seongsin Lee, Professor, Kyungpook National University, Department of Library & Information Science, (leess@knu.ac.kr) (First and Correspondence author)
∗∗   Hyunsook Kim, Doctorate student, Kyungpook National University, Department of Library & Information Science, (ahiran01@naver.com) (Co-Author)
∗∗∗ Sumin Beack, Master student, Kyungpook National University, Department of Library & Information Science, (qsm555@naver.com) (Co-Author)
∗∗∗∗ Subin Yoon, Master student, Kyungpook National University, Department of Library & Information Science, (subinprague@gmail.com) (Co-Author)

1. Introduction

The Fourth Industrial Revolution is a buzzword across all different sectors of the Korean society. The term was first coined in 2016 by Klaus Schwab. Since then, all academic fields, including the LIS, have paid heightened attention to the Fourth Industrial Revolution. The Fourth Industrial Revolution refers to technological developments such as hyper-connectivity and super-intelligence based on smart information technology, namely IoT, Big Data, Cloud Computing, and AI, respectively. However, the term does not only refer to the technological innovation. The intelligent information technology mentioned in the above is integrated into the economic and social fields. The term is also used to indicate a wide range of changes taking place across all areas of our lives. Such innovative changes are accelerating as contactless activities are now common since the outbreak of COVID-19.

This study focuses on the Big Data, one of the technological pillars of the Fourth Industrial Revolution. Big Data is defined differently according to many scholars and academic fields. Hoy (2014) stated that Big Data is "the idea that computers can gather trillions of pieces of information about billions of different things and find useful patterns in that information" (321-322). Furthermore, Song (2014) stated that Big Data is "a large set of data or various types of the real-time data collected utilizing data collection, storage, management, and analytical tools which are more sophisticate d than any of the existing DB management tools" and "the amount of data produced and distributed has increased exponentially since the advent of various information channels such as social media and smart devices" (74). Traditionally, a library is an institution that manages and provides information. Big Data collected and generated internally and externally can help develop innovative library services (Lee et al., 2015). As Hoy (2014) mentioned, libraries and librarians are very suitable for working and dealing with Big Data. In this context, the research and application cases related to Big Data in the field of library have been steadily increasing over the past ten years. Domestic and international examples of Big Data utilization in the field of library are as follows. First, in Korea, there is the 'Data for Library' platform developed and operated by the National Library of Korea. The Data for Library platform is a Big Data Platform that provides various data to 1,333 public and small libraries as of January 2022. The examples of the information provided are the information relating to library collections/loan data, library information, popular books which the users borrowed, detailed book inquiries, and usage rate for each book. Another example would be the Big Data Platform called 'Solomon', which provides support services for librarians. Solomon collects and analyzes bibliographic data inside the library, online bookstore data provided outside the library, and public data to support librarians. Only library managers can sign up to use it (Lee & Lee, 2021).

The following example illustrates how the libraries overseas utilize the Big Data related technology. First, the British Library conducted an ancient document digitization project with Google to digitize 250,000 books out of copy books published between 1700 and 1870 (Song, 2014). Another example would be the fact where the OCLC established an integrated digital library system by implementing a cloud computing-based library system. Third, DPLA (Digital Public Library of America) has launched a new platform to integrate and provide digital materials for the public libraries and museums

in the United States.

As we have seen thus far, the cases in the field of library that utilize Big Data are steadily increasing, and academic interest is growing accordingly. Therefore, this study intends to examine the Big Data research achievements in libraries, emphasizing public libraries. To this end, this study aims to observe what areas of interest public library staffs have regarding the Big Data. Unfortunately, the information science related research results concerning the data science published by the librarians employed by domestic libraries are inadequate. Therefore, this study will focus on the published articles through the journal 'Public Library Quarterly'. The public library staffs mainly write the articles published in the PLQ. The relevant analysis will help identify the main areas of interest which the international public libraries have concerning the Big Data. A future comparison with domestic cases will be available if adequate domestic research is conducted.

2. Methodology

This study uses the social network analytical method to examine the Big Data related research trends for the articles published in the Public Library Quarterly. The social network analytical method is commonly used to analyze the research trends for a specific topic. In particular, this study identifies characteristics for each network by constructing a network which targets the author keywords and examines the frequency, degree centrality, and betweenness centrality of the keywords. According to Lee (2016), "social network analysis can be used to discover what the main topics of a specific research community have been with a text-mining tool" and "example of this is the keyword network analysis" (p. 265). The node with the highest degree centrality demonstrates the most critical node in the network. High degree centrality means that many direct connections with other nodes in the network are made. In this study, a keyword with high degree centrality is essential in the relevant research field. It also means that it is a keyword with high direct relevance to other keywords.

Next, the betweenness centrality proves which node is best positioned in the network. The betweennes s centrality analysis mainly measures as to which node can best control the relationship between other nodes in the network. That is, a high betweenness centrality means that "it can act as a broker in the network" (Lee, 2016, p. 265). As Lee and Sohn (2016) mentioned, "an actor with ha igh betweenness centrality can choose to pass on information or not to the other actors within the network" (p. 84). In this case, the word actor means a node, and high betweenness centrality implies that there will be a disruption in network connection when the said node is removed (Lee & Sohn, 2016).

Next, the data collection process is as follows.

First, the author keywords for the journal Public Library Quarterly that were searched for using Big Data as a keyword were analyzed. The Public Library Quarterly is a peer-reviewed journal. Furthermore, "every issue of Public Library Quarterly contains informative articles written by the

directors and staffs of leading public libraries, news of current public library events, and book reviews covering issues of interest to those in public library work" (Public Library Quarterly, 2022).

As mentioned in the journal's website above, the Public Library Quarterly publishes papers and book reviews written by public library staffs. However, book reviews were excluded from the analysis in this study.

Second, the collected data were used for the analysis after undergoing the following pre-processing processes.

1) *This study aims to analyze research trends related to Big Data in the field of public library through analyzing author keywords. Generally, search terms are excluded from the network analysis process. Therefore, the keyword Big Data was excluded from the analysis. In the case of the keyword "public library", the word "public" was excluded, whereas the word "library" was used for analysis. This decision was made because if the word "public library" were deleted from searches about public library management, only the term "management" would remain, and search results would not be so different to search results for general management. Another example would be search results for "public library research". Again, if the term "public library" were excluded, only the word "research" would remain, and the search results would be no different from those about research in the general sense.*

2) *Words connected by the conjunction "and" were separated for analysis. For example, in the case of "marketing and promotion", the word "marketing" and "promotion" have different meanings, so they were treated as two separate keywords.*

3) *Keywords in abbreviated forms were used for analysis, and words spelled out were abbreviated. For example, the keyword for library services and technology is "LSTA", and the keyword for young adults is "YA".*

4) *Proper nouns were deleted.*

5) *Uppercase letters were converted to lowercase letters for analysis. For example, the keyword "Readers advisory" was changed to "reader advisory" and "Library services" were changed to "library services".*

Third, frequency, degree centrality, and betweenness centrality results were derived from the preprocessed keywords, and were analyzed according to the criteria mentioned in the above by using the NetMiner4 software.

Finally, a keyword network was constructed.

3. Results and Discussion

First, the total number of analyzed papers and the frequency of each keyword are as follows. The Public Library Quarterly is a journal currently published six times a year and has been accessible through the journal's website since the 1979 papers. The reports published from 2008 that relate to the keyword "Big Data" were included for the search. Table 1 below demonstrates the number of articles that were searched for from the year 2008 through 2021 (up to December). The total number of articles to be analyzed is 117, excluding two articles without author keywords.

Table 1. Number of Articles

Publication Year	Number of Articles
2008	5
2009	6
2010	19
2011	5
2012	7
2013	5
2014	4
2015	5
2016	9
2017	10
2018	7
2019	16
2020	11
2021	10

Keywords that appeared more than three times are illustrated in **Table 2** below.

Table 2. Important keywords

Rank	Keyword	Frequency
1	programs	4
1	collection development	4
1	community development	4
1	strategic planning	4
1	survey	4
1	community building	4
2	opioids	3

2	community engagement	3
2	library services	3
2	future of libraries	3
2	evaluation	3
2	PLA	3
2	future	3
2	content analysis	3
2	technology	3

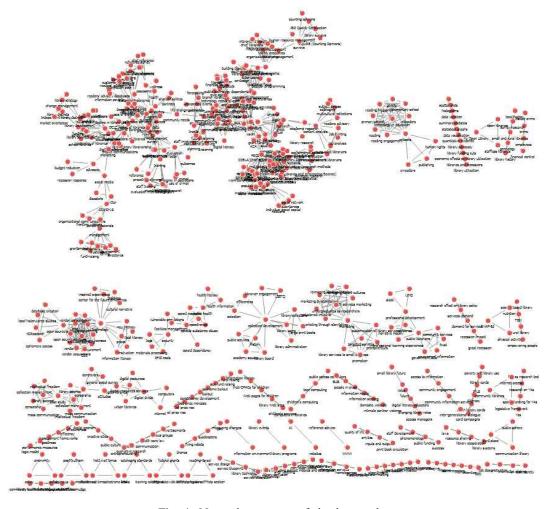Next, the network structure of all keywords is shown in **Figure 1** below.

Fig. 1. Network structure of the keywords

Figure 1 demonstrates that a significant portion of the network structure is largely separated, meaning Big Data-related studies are not focused on any one topic but are related to the various topics of public library.

The analytical results for the top 40 keywords with a high degree centrality are illustrated in **Figure 2** and **Figure 3** to examine the above network structure more clearly.
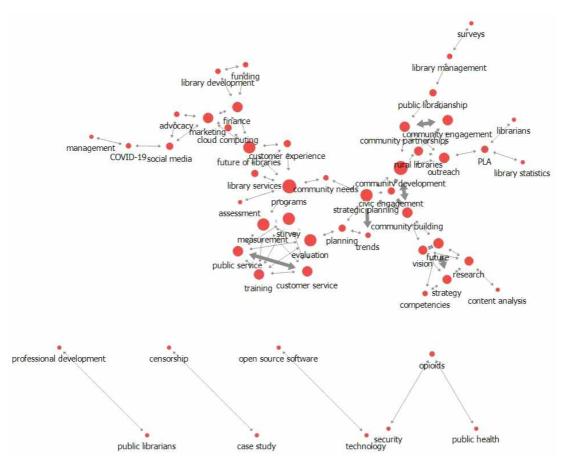

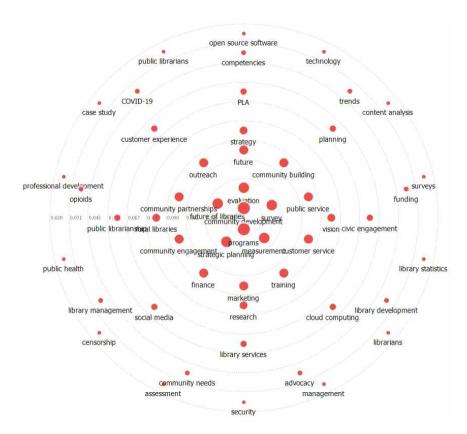**Fig. 2.** Degree centrality of the top 40 keywords

Fig. 3. Degree centrality of the top 40 keywords

The detailed results of **Figure 2** and **Figure 3** are as follows:

1) The keywords "programs" and "community development" had the highest degree centrality. The keyword "programs" was connected with the keywords "community needs", "library services", "measurement", and "assessment". Furthermore, the keyword "community development" was connected with the keywords "community engagement", "community building", "community partnerships", and "civic engagement".

2) The keywords "future of libraries", "strategic planning", "measurement", "survey", and "evaluation" also had a high degree centrality. The keyword "future of libraries" was connected with the keywords "finance", "cloud computing", "marketing", and "customer experience". The keyword "strategic planning" was connected with the keywords "planning", "community needs", "civic engagement", "community development", "planning", "trends", and "community building". Furthermore, the keyword "measurement" was connected with the keywords "survey", "customer service", "evaluation", "public service", and "training".

3) The degree of the keywords "outreach", "community partnerships", "community engagement", "finance", "marketing", "training", "customer service", "public service", "community build-

ing", and "future" was high. The keyword "marketing" was connected with the keywords "cloud computing", "social media", "advocacy", and "future of libraries". The keyword "community partnerships" was connected with the keywords "community engagement", "outreach", "public librarianship", "rural libraries", "community development", and "library management".

4) The keyword "training" was connected with the keywords "measurement", "public service", "evaluation", "customer service", and "survey".

5) The degree of the keywords "rural libraries", "research", "vision", and "strategy" was also high. The keyword "rural libraries" was connected with the keywords "community engagement", "community partnerships", and "outreach". The keyword "vision" was connected with the keywords "community building", "future", and "strategy".

Table 3. Degree centrality of the top 21 keywords

Rank	Keyword	Degree centrality
1	programs	0.125
1	community development	0.125
3	future of libraries	0.107143
3	strategic planning	0.107143
3	mesurement	0.107143
3	survey	0.107143
3	evaluation	0.107143
8	outreach	0.089286
8	community partnerships	0.089286
8	community engagement	0.089286
8	finance	0.089286
8	marketing	0.089286
8	training	0.089286
8	customer service	0.089286
8	public service	0.089286
8	community building	0.089286
8	future	0.089286
18	rural libraries	0.071429
18	research	0.071429
18	vision	0.071429
18	strategy	0.071429

Next, the result of analysis of betweenness centrality is the following **Table 4**. As Lee and Sohn (2016) mentioned, "a keyword showing high betweenness centrality score connects sub research areas" in the subject of the study and "these keywords leads to convergence of the research" (180). **Table 4** demonstrates the keywords that have a high betweenness centrality.

Table 4. Betweenness centrality of the top 21 keywords

Rank	Keyword	Value
1	strategic planning	0.297403
2	programs	0.269264
3	community development	0.207792
4	community needs	0.206926
5	future of libraries	0.192857
6	community building	0.14026
7	marketing	0.098701
8	outreach	0.075974
9	planning	0.069048
10	evaluation	0.064502
11	future	0.061039
12	PLA	0.052597
13	public librarianship	0.051948
13	social media	0.051948
13	finance	0.051948
16	community partnerships	0.037013
16	community engagement	0.037013
18	vision	0.036039
19	COVID-19	0.026623
19	library management	0.026623
19	research	0.026623

Table 4 demonstrates that the keywords "programs", "strategic planning", "community development", "future of libraries", "outreach", and "evaluation" post a high degree of centrality and betweenness centrality. The keywords "measurement", "survey", "community partnerships", and "community engagement" demonstrated a high degree of centrality but not high betweenness centrality. Meanwhile, the keywords "planning", "marketing", "community needs", and "community building" demonstrated high betweenness centrality but not high degree centrality, meaning that the keywords "planning", "marketing", "community needs", and "community building" are important topics that connect Big Data related sub-topics related in thefield of public library.

The main analytical results are the following.

First, "programs", which was a keyword with a high degree of centrality, was connected with the keywords "community needs", "library services", "measurement", and "assessment". Such results imply that an understanding of community needs is essential when designing and providing user-customized library services, and, to this end, Big Data could be a helpful tool. Furthermore, the keyword "programs" was linked to other keywords such as "measurement" and "assessment". Such results show heightened interest in data produced through the evaluation of library programs or data for evaluating library programs.

Second, the keyword "community development" posted a high frequency and degree centrality. In particular, the keyword "community development" was closely related to the keywords such as "community engagement", "civic engagement", "community partnerships", and "community building". Such findings are pertinent that the authors of the journal 'Public Library Quarterly' are mainly librarians or library staffs. That is, such findings show that Big Data is a topic of interest in areas such as data availability in library management or the library's role as a data center. That is, Such results emphasize that the public library should play a central role as a local data center and can contribute to community development by fulfilling this role. Such findings also underscore that community engagement, civic engagement, and community partnerships are crucial for community development.

Third, the keyword "future of libraries" demonstrates that the library staffs are attentive to the relationship between the future of libraries and finances. In general, any organization, including library, must secure the proper financial investment and resources to achieve its goals. The resources that the organizations must ensure include both the human and physical resources. Physical resources pertain to information, equipment, and budget, with budgetary resources being the most important. Adequate financial support is essential above all for the library to fulfill its goals and ultimately make sure that the users recognize it as an important community institution. Such results demonstrate that library officials are aware of how library marketing is related to the future of the library. Most importantly, the future of libraries and their correlation to the Big Data surfaced as a keyword with a high frequency and degree centrality. This may be so because it is recognized that the interest in Big Data is essential for the future of public libraries.
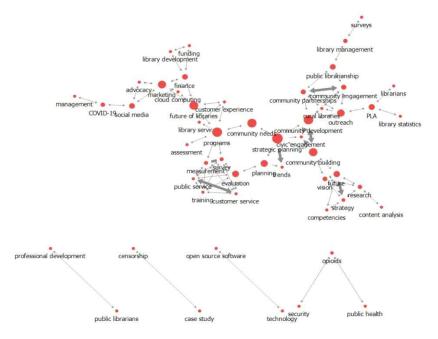
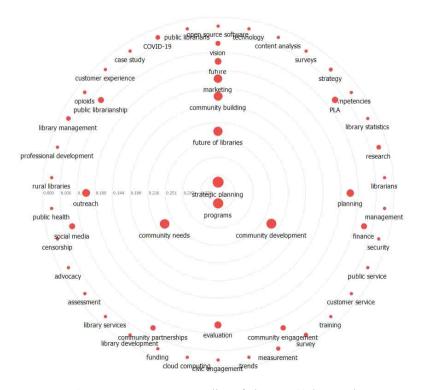**Fig. 4.** Betweenness centrality of the top 40 keywords

Fig. 5. Betweenness centrality of the top 40 keywords

Fourth, as illustrated in **Figure 4** and **Figure 5**, the keywords "planning", "marketing", "community needs", and "community building" demonstrated a high betweenness centrality yet a moderate degree centrality. Furthermore, the keywords "COVID-19" and "social media" did not make it to the list of top 21 keywords with high degree centrality but were included among the top 21 keywords with a high betweenness centrality. Such results imply that, while not many studies related to the six keywords were mentioned in the above, they are the important topics connecting the sub-topics of Big Data at the public libraries. Moreover, the keywords mentioned in the above are the scalable topics related to the public libraries' Big Data research. Finally, the public libraries can integrate the Big Data-related topics of research with other issues by enhancing accessibility through keywords such as "planning", "marketing", "community needs", "community building", "COVID-19", and "social media".

Fifth, the keywords "opioids", "PLA", "content analysis", and "technology" posted high frequency but low degree centrality and betweenness centrality, which implies that few scholars are leading studies related to the keywords mentioned in the above.

Sixth, the keyword "training" was connected with the keywords "measurement", "public service", "evaluation", "customer service", and "survey". Again, such results emphasize that public librarians need training to handle Big Data, especially for evaluation and performance measurement.

Seventh, the keyword "collection development" posted high frequency but demonstrated no significant analytical results for degree centrality and betweenness centrality, which demonstrates that collection development is not a meaningful keyword related to Big Data.

4. Conclusion

One hundred seventeen articles published in the journal "Public Library Quarterly" were analyzed with author keywords to examine the Big Data-related research trends in the field of public library. As a result, the keywords "programs", "collection development", "community development", "strategic planning", "survey", and "community building" recorded high frequency. Furthermore, the keywords "programs", "future of libraries", "strategic planning", "measurement", "survey", and "evaluation" demonstrated a high degree of centrality. The keywords "strategic planning", "community development", "community needs", "future of libraries", "community building", and "marketing" posted a high betweenness centrality. The keywords "planning", "marketing", "community needs", and "community building" demonstrated a high betweenness centrality yet a moderate degree centrality. Furthermore, the keywords "COVID-19" and "social media" did not make it to the list of top 21 keywords with a high degree centrality yet were included among the top 21 keywords with a high betweenness centrality. Such results imply that, while not many studies were related to the six keywords mentioned in the above, they are important topics connecting the sub-topics of Big Data for the public libraries. Moreover, the keywords mentioned in the above are the scalable topics related to the public libraries' Big Data research. Finally, the public libraries can integrate the Big Data-related research topics with the other issues by enhancing accessibility through keywords such as "planning", "marketing", "community needs", "community building", "COVID-19", and "social media".

Anna and Mannan (2020) stated that, in comparison to a significantly large number of Big Data studies published at colleges and the academic field of library, the Big Data studies published in the field of public library are inadequate. The research on the library and information science in Korea demonstrates that the Big Data studies have been published since the public library's Big Data platform service was launched in 2016, yet the number is insignificant. Furthermore, most of the articles are the case studies such as the Big Data usage related studies. Based on the results of this study, public libraries should focus on the following directions of research when focusing on the Big Data.

First, as mentioned in the above, a more in-depth discussion is needed regarding COVID-19 and social media.

Second, the academic interest in the Big Data related education and training for the public librarians and educational contents is needed.

Third, public libraries should think about the ways to efficiently perform their role as a local data center, including their cooperation with the other organizations in their local community.

In this study, the authors analyzed the author keywords of articles published in the journal "Public Library Quarterly", written mainly by the public library staffs. Therefore, it was possible to discern as to which fields librarians and staffs were interested in concerning the Big Data. However, this may not be consistent with the areas of academic interest. This study could not analyze the academic research related to Big Data because there is not enough LIS academic research conducted on this topic thus far. Furthermore, this study was not able to derive an implication through comparison with the other cases overseas because, in Korea, there are not many articles that measured the interest towards Big Data in the related academic areas and in the field.

References

Anna, N. E. V., & Mannan, E. F. (2020). Big data adoption in academic libraries: a literature review. *Library Hi Tech News, 37*(4), 1-5. doi:10.1108/LHTN-11-2019-0079

Hoy, M. B. (2014). Big data: An introduction for librarians. *Medical reference services quarterly, 33*(3), 320-326. doi:10.1080/02763869.2014.925709

Lee, E. J., Kim, W. J., & Kim, H. S. (2015). Development and management of library services based on big data analysis: A case study on "study on big data analysis and adoption in library". *Digital Library, 80*(0), 17-30.

Lee, H., & Sohn, I. (2016). *Fundamentals of big data network analysis for research and industry.* John Wiley & Sons.

Lee, S. (2016). A study on research trends in public library research in Korea using keyword networks. *Libri, 66*(4), 263-274. doi:10.1515/libri-2016-0052

Lee, S. Y., & Lee, S. (2021). A study on Big Data Analysis of Public Library in Busan: Based on the Library Collection/Circulation Data. *Journal of the Korean Library and Information Science, 55*(4), 89-114. http://journal.kci.go.kr/kslis/archive/articleView?artiId=ART002778030

Public Library Quarterly (2022). *Public Library Quarterly, Volume 41, Issue 1 (2022)* https://www.tandf

online.com/toc/wplq20/41/1?nav=tocList

Song, M. (2014). Big Data and Libraries. *Survey Research, 26*, 72-101.

[About the authors]

Seongsin Lee is a professor in library and information science at Kyungpook National University. Her research interests include library management, library marketing, reading education, and information ethics. Her research appeared in Libri, as well as diverse journals of the Republic of Korea. She is a first author.

Hyunsook Kim is a doctoral student in library and information science at Kyungpook National University. Her research interests include reading education and information literacy. Her research appeared in the Journal of Korean Library and Information Science Society. She is a co-author.

Sumin Beack is a master student in library and information science at Kyungpook National University. Her research interests include library marketing, knowledge sharing, and cooperative networks of libraries. She is a co-author.

Subin Yoon is a master student in library and information science at Kyungpook National University. Her research interests include library cooperation, social recognition, and user perception. She is a co-author.