Impact of big data on library services: prospect and challenges

Abid Hussain and Ramsha Shahid

Introduction

Big data is has become a major phenomenon in the age of digitalization, which profoundly impacts people living around the globe. It has changed the traditional way of thinking regarding data usage in different fields. The application of big data can be observed in various domains, as its use has been rapidly increasing since the past decade. Big data aims to advance the existing technologies by collecting, storing, preserving, managing and analyzing a massive amount of data. In the age of data abundance, optimal use of data has accelerated technological advancement, resulting in greater exploration in science and technology by transforming old learning techniques. The major characteristics of big data can be defined by the 5Vs model, which includes volume, variety, velocity, value and veracity. Volume is referred to the capacity of a system to store data; variety is referred to different forms of data that can be processed; velocity defines the processing speed of data gathered through multiple sources; and veracity is linked with the conformity and accuracy of data, while value is related to the worth or benefits of compiled information (Anna and Mannan, 2020).

With increased number of virtual classrooms, a massive volume of data is produced by online courses, academic sessions and teaching activities. Currently, with e-libraries, millions of people worldwide access resources and services over the Web (Hussain, 2020a, 2020b). Over the past 50 years, libraries have compiled an enormous amount of data in electronic form related to library collections. The data is only limited to that particular library and cannot be reused by any other library community or data store. The data can play a vital role in enhancing library services by giving students and library members personalized library services. With the application of big data, libraries will be directed to tryout different user-friendly services. However, several challenges exist which involve a lack of skilled human resources and technology to process the data effectively.

Doug Laney first introduced the term big data in 2001. In his research notes, he argued that big data defines the new and innovative method of managing and analyzing large data sets. Traditional data management tools cannot process big data effectively, a distinguishing feature of big data (Xu *et al*., 2017). Big data includes information to aid in intelligent decision-making and detailed evaluation. These information resources consist of several private electronic and computing devices generating massive amounts of data every second through online shopping, research, GPS navigation and social media posts (Reinhalter and Wittmann, 2014).

Libraries have been managing and sorting intellectual assets for decades, widely used by different universities, research institutes, the public and government. Big data can be used by researchers to improve user centered services. Librarians also need to understand how this transformation of available data will facilitate knowledge creation (Xu *et al*., 2017).

According to Khan *et al*. (2014), as many opportunities are linked with big data for users, it is also confronted with many challenges. These hurdles lie in capturing, storing, sharing, analyzing and visualizing the data. Some other challenges related to big data include scalability, security, inconsistency and incompleteness of data. Data must be well structured for analysis, and unstructured and semi-instructed data need to be reviewed as well. As data sets are often enormous and compiled from several resources, the pre-processing of data can prove beneficial in this regard. It will also improve the quality of data and results. Currently, libraries' databases contain noisy and inconsistent data, which needs to be cleaned, integrated and transformed before being used as big data.

According to Zhan and Widén (2018), librarians could not make a pragmatic and comprehensive implementation of big data in public and academic libraries. To ensure the provision of big data services, librarians must hold system-oriented and service-oriented roles. Libraries have a massive amount of data, which can improve the quality of library services. As the data of modem digital library holds all the characteristics of big data, the library can use big data methods for innovation and resource utilization and transferring (Hamad *et al*., 2020a).

Challenges and opportunities in the application of big data in libraries

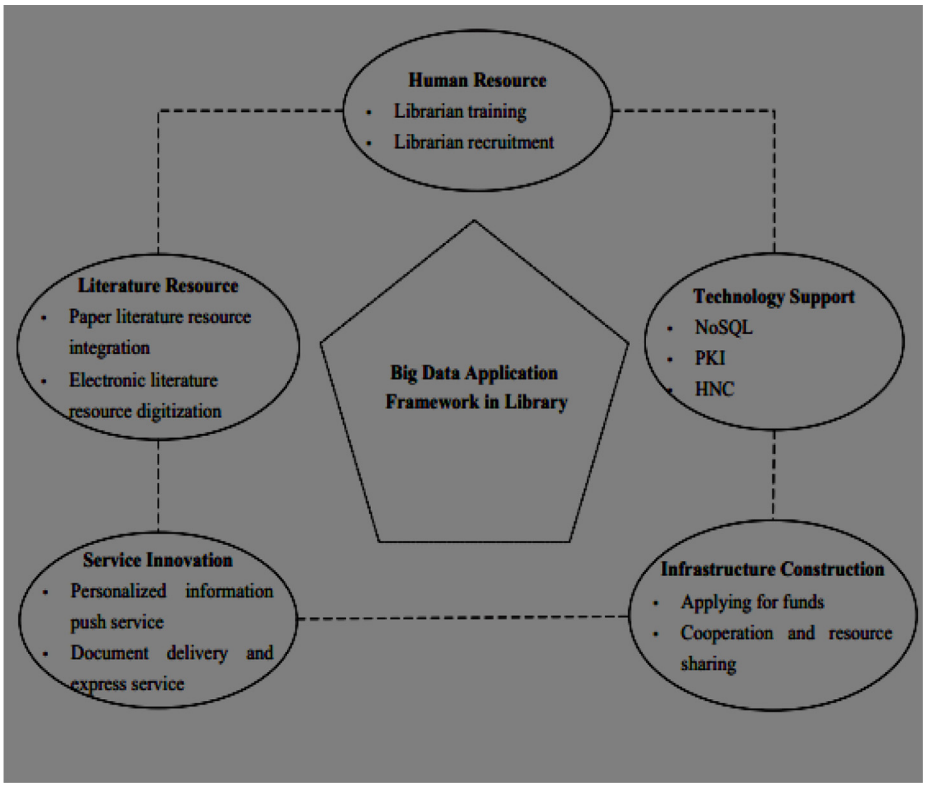
The application of big data in a library involves a complex framework as different modules interact and work together. The following diagram illustrates the big data application framework from the library. It involves library staff, resource collection, infrastructure and technological support (Figure 1).

The unique skills of librarians are essential to make use of big data. Librarians' jobs have evolved from traditional setup into data mining, consultant, data management, librarian data services, data research scientist, architecture librarians, librarian data designs, coordinator, scientific data curation, librarian specialist/metadata, data curator, etc. Librarians should change their traditional role and learn the new skills needed in the 21st century.

Large data sets can improve the quality of library services. Data resources of digital libraries can be used as big data by using techniques to bring innovation and introduce important digital changes. Big data provides insights into the

  0741-9058, DOI 10.1108/LHTN-02-2022-0022

Figure 1. *Big data application framework in library*



utilization of resources, decision-making, and library user needs more accessible. Academic libraries are already using high-level services to collect, evaluate and manage library resources. Changes in technology and new ways of communication are demanding the redefined academic library.

A significant milestone was announced in 2010, which was the collaboration between the world's most extensive library – Library of Congress and Twitter. The purpose of the partnership was to analyze all the tweets on Twitter to retain and archive users' data. In 2012, the Harvard University library also published its metadata, including books, videos, audio recordings, manuscripts and other contents. Ensuring data privacy and security are also concerns when implementing big data analytics in library services (Bhat, 2018).

Recommendations and concluding remarks

The application of big data in libraries is constantly evolving and includes the following challenges some of which have already been noted. Digital libraries have data in different forms, including literature, digital resource collections, database resources, etc. Libraries need to improve the storage size for big data resources and integrating newly incorporated resources of varying data formats.

Technology is an inseparable part of big data application. New advanced technological methods must be implemented in existing systems to innovate changes from traditional library services. Data mining, acquisition and analysis techniques and new solutions such as parallel computing and distributed framework are also needed.

Big data brings and requires technically competent librarians. Many librarians currently lack the essential skillsets for managing data needed in a world of digital libraries. These libraries require knowledgeable staff who can coordinate data processing tools and technologies. While recruiting new staff, the technical demands of the job must be taken into consideration.

Big data in libraries is more related to the enhancement of user service. To understand the user's requirements, trained data scientists are required who can compile and analyze the interests of users through available data from different resources. There must be allocations of funds in the budget to address the financial constraints of training and retooling librarians so that the data can be appropriately used in libraries for decision-making. There are some universities worldwide that have experienced the implementation of big data to bring innovation in traditional library services with funds for technical support and training. Data security and privacy is also a significant concern in this regard. Also, to ensure big data implementation in library services, a big data management policy should be devised to address all predicted challenges and ensure sustainability.

REFERENCES

Anna, N.E.V. and Mannan, E.F. (2020), "Big data adoption in academic libraries: a

literature review", *Library Hi Tech News*, Vol. 37 No. 4, pp. 1-5, doi: 10.1108/LHTN-11-2019-0079.

Bhat, W.A. (2018), "Long-term preservation of big data: prospects of current storage technologies in digital libraries", *Library Hi-Tech*.

Hamad, F., Fakhuri, H. and Abdel Jabbar, S. (2020), "Big data opportunities and challenges for analytics strategies in Jordanian academic libraries", *New Review of Academic Librarianship*, pp. 1-24.

Hussain, A. (2020a), "Cutting edge: technology's impact on library services", in Jesubright, J. and Saravanan, P. (Eds), *Innovations in the Designing and Marketing of Information Services*, IGI Global, pp. 16-27, doi: 10.4018/978-1-7998-1482-5.ch002

Hussain, A. (2020b), "Big data in library services, daily times, Pakistan", available at: https://dailytimes.com.pk/533352/big-data-in-library-services-daily-times/ (accessed 15 September 2021).

Khan, N., Yaqoob, I., Hashem, I.A.T., Inayat, Z., Mahmoud Ali, W.K., Alam, M., Shiraz, M. and Gani, A. (2014), "Big data: survey, technologies, opportunities, and challenges", *The Scientific World Journal*, Vol. 2014.

Reinhalter, L. and Wittmann, R.J. (2014), "The library: big data's boomtown", in Block, R.J. (Ed.), *The Serials Librarian*, Vol. 67 No. 4, pp. 363-372.

Xu, S., Du, W., Wang, C. and Liu, D. (2017), "The library big data research: status and directions", *International Journal of Software Innovation (IJSI)*, Vol. 5 No. 3, pp. 77-88, doi: 10.4018/IJSI.2017070106.

Zhan, M. and Widén, G. (2018), "Public libraries: roles in big data", *The Electronic Library*, Vol. 36 No. 1.

FURTHER READING

Abid, H. (2021), "Uses of blockchain technologies in library services", *Library Hi Tech News*, Vol. 38 No. 8, pp. 9-11, doi: 10.1108/LHTN-08-2020-0079.

Al-Barashdi, H. and Al-Karousi, R. (2019), "Big data in academic libraries: literature review and future research directions", *Journal of Information Studies & Technology (JIS&T)*, Vol. 2018 No. 2, p. 13.

Ali, I. (2015), "Big data: Apa dan pengaruhnya pada perpustakaan? (What is big data and its influence to library)", *Media Pustakawan*, Vol. 22 No. 4, pp. 15-19, doi: 10.37014/medpus.v22i4.218.

Ball, R. (2019), "Big data and their impact on libraries", *American Journal of Information Science and Technology*, Vol. 3 No. 1, pp. 1-9.

Bertot, J.C., Butler, B.S. and Travis, D.M. (2014), "Local big data: the role of libraries in building community data infrastructures", *Proceedings of the 15th Annual International Conference on Digital Government Research*, pp. 17-23.

Blummer, B. and Kenton, J.M. (2018), "Big data and libraries: identifying themes in the literature", *Internet Reference Services Quarterly*, Vol. 23 Nos 1/2, pp. 15-40.

De Mauro, A., Greco, M. and Grimaldi, M. (2016), "A formal definition of big data based on its essential features", *Library Review*, Vol. 65 No. 3.

Digital Libraries in the Western Region of China: Library & Information Science Book Chapter | IGI Global (2022), available at: www.igi-global.com/chapter/digital-libraries-in-the-western-region-of-china/191584 (accessed 16 September 2021).

Golub, K. and Hansson, J. (2017), "(big) data in library and information science: a brief overview of some important problem areas", *Journal of Universal Computer Science (Online)*, Vol. 23 No. 11, pp. 1098-1108.

Islam, A.A., Ahmad, K., Rafi, M. and JianMing, Z. (2021), "Performance-based evaluation of academic libraries in the big data era", *Journal of Information Science*, Vol. 47 No. 4, pp. 458-471.

Jantti, M. (2016), "Libraries and big data: a new view on impact and effect", *Quality and the Academic Library*, Elsevier, pp. 267-273.

Johnson, V. (2017), "Leveraging technical library expertise for big data management", *Journal of the Australian Library and Information Association*, Vol. 66 No. 3, pp. 271-286.

Kelly, M. (2014), "Applying the tiers of assessment: a holistic and systematic approach to assessing library collections", *The Journal of Academic Librarianship*, Vol. 40 No. 6, pp. 585-591.

Li, J., Lu, M., Dou, G. and Wang, S. (2017), "Big data application framework and its feasibility analysis in library", *Information Discovery and Delivery*, Vol. 45 No. 4.

Li, S., Jiao, F., Zhang, Y. and Xu, X. (2019), "Problems and changes in digital libraries in the age of big data from the perspective of user services", *The Journal of Academic Librarianship*, Vol. 45 No. 1, pp. 22-30, doi: 10.1016/j.acalib.2018.11.012.

Perrin, J.M., Yang, L., Barba, S. and Winkler, H. (2017), All that glitters isn't gold: The complexities of use statistics as an assessment tool for digital libraries – *the Electronic Library*.

Stančin, I. and Jović, A. (2019), "An overview and comparison of free python libraries for data mining and big data analysis", *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 977-982.

The Library Big Data Research: Status and Directions: Computer Science & IT Journal Article | IGI Global (2022), available at: www.igi-global.com/article/the-library-big-data-research-status-and-directions/182538 (accessed 20 September 2021).

ABOUT THE AUTHORS

Abid Hussain is a Library Officer at the Institute of Strategic Studies Islamabad. He holds a MPhil in Library and Information Sciences from the University of Peshawar and writes regularly for national and international journals. So far, he has contributed more than 35 research articles.

Ramsha Shahid is a Graduate from the Department of Defence and Strategic Studies Quaid-i-Azam University Islamabad and also holds Master's degree in Computer Science from the University of Gujarat, Punjab, Pakistan.

Abid Hussain (*abidmardan@gmail.com*) is based at the Department of Library and Information Sciences, Institute of Strategic Studies Islamabad, University of Peshawar, Peshawar, Pakistan.

Ramsha Shahid (*ramshashahid734@gmail.com*) is based at the Department of Defence and Strategic Studies, Quaid-I-Azam University, Islamabad, Pakistan.