

Chapter 10

The Library Big Data Research: Status and Directions

Shaochun Xu

Department of Computer Science, Algoma University, Sault Ste Marie, Canada

Wencai Du

Hainan University, Hainan, China & City University of Macao, Macao, China

Chunning Wang

National Geological Library of China, Beijing, China

Dapeng Liu

Gradient X, Santa Monica, CA, USA

ABSTRACT

Libraries are widely used by government, universities, research institutes, and the public since they are storing and managing intellectual assets. The library information directly stored in libraries and about the people interaction with libraries can be transformed into accessible data which then will be used by researchers to help library better serve users. Librarians need to understand how to transform, analyze, and present data in order to facilitate such knowledge creation. For example, the challenges they face include how to make big datasets more useful, visible and accessible. Fortunately, with new and powerful analytics of big data, such as information visualization tools, researchers/users can look at data in new ways and mine it for information they intend to have. Moreover, interaction of users and stored information has been taken into librarian's consideration to improve library service quality. In this work, the authors discuss the characteristics of datasets in library and argue against a popular confusion that data involved in library research is not big enough, conduct a review for the research work on library big data and then summarize the applications and research directions in this field. The status of big data research in library in China is discussed. The challenges associated with it are also discussed and explored.

DOI: 10.4018/978-1-7998-8051-6.ch010

1 INTRODUCTION

Big data is well received with the three common characteristics: high-volume, high velocity, and/or big variety (Press, 2013). In spite of a few advancing researches who had linked library data into big data, some researchers raised questions about that since there is no clear characteristics of velocity (Hessman, 2013); one such argument was that because the database management systems is enough for storing and processing library data, it seems that it does not require big data technology such as distributed systems for analysis or processing. As a fact there is no much general review works on the research for library “big data”.

This paper will review the researches that have been done in the usage of big data technology in library, highlight some challenges of adoption of such technique that had not been well perceived, in order to provide a basis for further related work. While in many fields big data has demonstrated the capacity to improve predictions, save money, boost efficiency and enhance decision-making such as weather forecasting, disaster prevention, finance, business transaction, national security, education, and health care, our take is that library big data could also more effectively serve researchers or ordinary users better. Our observation is that library big data research just started. While the future is promising, there are some unique challenges that are specific to this field. Librarians are not sure how to integrate it into library data: facing new technology, they need to possess a deep understanding of how to transform, analyze, and present data to facilitate knowledge creation. As an outcome, the ability to leverage library big data will also make it possible to better understand how users perform research and serve them better with new ways.

The rest of the paper is organized as follows: Section 2 provides related work. Section 3 describes the characteristics of library datasets. The issues associated with library big data research is discussed in Section 4. Section 5 explores the opportunities or research directions in library data with big data technology. The status of big data research in library in China is summarized in Section 6. The conclusions and the future work are present in Section 7.

2 RELATED WORK

“Big data” describes innovative techniques and technologies to capture, store, distribute, manage and analyze datasets that traditional data management methods are normally unable to handle. The concept of “Big data” was first defined by Laney in his research note (Laney, 2001). According to the definition, big data is mainly characterized by three Vs: Volume, Velocity, and Variety (Zikopoulos et al., 2012). The first V, refers to the data volume. General speaking, the size of the data sets of big data is huge compared to regular data. However, it seems that there is no fixed definition for the size, i.e. how big of data could be classified as big data. Therefore, the size might vary based on the disciplines. Traditional software usually can handle megabyte and even gigabyte sized data sets, while big data tools should be able to handle terabyte and petabyte sized data sets. The second V, velocity, refers to the situation where data is created dynamically and accessed in a fast way; The data come in frequently, such as every second or so, and data access often has to be finished in a fraction of a second. Sometimes, data processing has to be done in real time therefore the software system has large throughput. The third V, referring to variety, indicates data heterogeneity which makes big data sets harder to organize and analyze. The regular type of data collected by researchers or businesses is strictly structured, such as data entered into

12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/the-library-big-data-research-status-and-directions/274751

Related Content

A Survey of Users' Satisfaction in the University Library by Using a Pareto Analysis and the Automatic Classification Methods

Jaouad Oukrichand Belaid Bouikhalene (2017). *International Journal of Library and Information Services* (pp. 17-36).

www.irma-international.org/article/a-survey-of-users-satisfaction-in-the-university-library-by-using-a-pareto-analysis-and-the-automatic-classification-methods/181685

The Research Process and Indigenous Epistemologies

Kgomotso H. Moahi (2020). *Handbook of Research on Connecting Research Methods for Information Science Research* (pp. 245-265).

www.irma-international.org/chapter/the-research-process-and-indigenous-epistemologies/240529

The Voortrekker Monument as Memory Institution: Mediating Collective Memory, Tourism and Educational Programming for a Local and Global Audience

Annie R. Antonitesand Johan Nel (2019). *Handbook of Research on Advocacy, Promotion, and Public Programming for Memory Institutions* (pp. 348-365).

www.irma-international.org/chapter/the-voortrekker-monument-as-memory-institution/219913

Doctor's Awareness and Perception of Medical Library Resources and Services: A Case Study of Delta State University Teaching Hospital (Delsuth), Nigeria

Violet E. Ikolo (2020). *International Journal of Library and Information Services* (pp. 58-71).

www.irma-international.org/article/doctors-awareness-and-perception-of-medical-library-resources-and-services/253104

Becoming Trusted Research Partners Through InfoExpress at The University of Toronto: (Re)Connecting Researchers to the Library

Manda Vrkljanand Adrienne Findley-Jones (2021). *Cases on Research Support Services in Academic Libraries* (pp. 187-214).

www.irma-international.org/chapter/becoming-trusted-research-partners-through-infoexpress-at-the-university-of-toronto/260638