Application of Big Data in Libraries

Washington Kamupunga
Honors Degree in Information systems
Zhejiang University of Science and Technology,
318 Liuhe road, Xihu district, Hangzhou, China

Yang Chunting, PhD
Zhejiang University of Science and Technology,
318 Liuhe road, Xihu district, Hangzhou,
Hangzhou, China

ABSTRACT
In the dynamic environment, records and management systems are independently maintained by education institutions, libraries and books whilst data are not readily accessible in a centralized position. Big data is being created due to digitalization of libraries and this has imposed limitations to researchers, educationists, scholars and policy maker's efforts in improving the quality and efficiency. As a result, serving the users with books and articles that are in line with their interests is a great challenge. This paper addresses the issues of bringing various sources of information from different sources and institutions into one place in real time which can be time saving. The primary objective is to decrease the time that lapses between searching the reading material and the actual reading. Thus, a mechanism by which this bridge can be gapped is of paramount importance as access to information is costly especially to those with limited internet access. The research focuses on the development of a strategy that reduces time of finding reading material and this is in line with current recommendation system. Through this system there is great analyses of book descriptions to identify books that are in line with users' interests. Within time huge amount of data is collected from the researchers, educationists, scholars and policy makers and this big data will be used to train machines to automate the tasks to some extent. As a result, the valuable information gained from analyzing massive amounts of aggregated libraries data can provide key insights in improving information accessibility. This makes researchers, educationists, scholars and policy maker's reach out for research solutions easily and cheaply and also makes information more accessible to the underprivileged and marginalized.

Keywords
big data, digitization, dynamic, analyzing, aggregated libraries, recommendation system

1. INTRODUCTION
The application of Big data as a resource has been outspoken in recent years due to its use in educational analysis and data-driven decision making. Jim. (2018). Highlighted that Big Data comprised of high-velocity, high-variety and high-volume information that is costly and can be used for innovation to enhance insight and decision making. Digitization is conversion of images, text, video and audio into digital form that be understood and processed by the computer (bits). Pramod (2019). outlined that recommendation systems plays a pivotal role in recommending relevant reading materials and driving customer conversion. According to Reinhalter & Wittmann (2014), each sector has developed a fascination with the ostensibly new discovery of Big Data and its extraordinary capabilities to fuel analytical breakthroughs since 2012. Big data analytics have affected academic libraries in two ways: first, due to the massive volume, selection, and speed of the knowledge concerned, so the storage and process needs of the system are rather overwhelming. Secondly, the analytics techniques and algorithms are complicated, which makes Big Data analytics a computing-intensive task.

2. SCENARIO
Through globalization, many people worldwide have fallen in love with social media. Readers and researchers are now approaching the online platforms such as Google search, LinkedIn, Twitter, Facebook etc. to research, purchase books and find new library services. Some libraries in the in this information age have been adopting technology well. These libraries and the librarians involved are uniquely suited to utilize Big data. Big data gives inherent challenges of getting out through and keeping up-to-date Librarians and users. This challenge has been orchestrated by the inability of many algorithms to perform well in search and filtering settings hence they instigated their failure in forecasting user preference, interest and change in taste. Big data in this scenario should be no exception hence the need for big data applications in libraries that offers predictive analytics of user reading habits, online services and applications that better understands the user's needs and requirements. Thus, big data better forecasts for patterns in habits, usage of systems and resources and future library planning.

3. LITERATURE REVIEW
Rul'& Ryabov (2018) noted that there are four recommender systems that can be used and hybrid recommender is an amalgamation of two recommender system. One recommender systems cover the weakness of another and this managed to respond positively to changes consumer taste. The same methodology was by them and it was successful. According to Daniel (2017), Big Data is the growth in volume and variety of data which can no longer be managed using the traditional database due to its size and different data formats. Big Data can mainly be described using three Vs which are volume, velocity and variety**.** Most big data-processing techniques emphasize high-speed and efficient use of large amounts of data. In his study Qian (2015), emphasized that big data has some associations can lead to the and these associations lead to the discovery of new knowledge. In their research, Wu, Su, and Deng (2013), highlighted that big data can be used to make decision making as analysis of reader behavior. In 2016, Zhang highlighted in his research how easier it has become for libraries to make use of big data to implement knowledge and as well as improving the services. As highlighted by Coelho (2011), libraries face internal and external pressure to change, they have to respond positively to user requirements and content delivery. Digital libraries have shifted from literature to used-based and also from general to personalized services (Li, 2012; L.X. Wang, 2015). The issue of personalized services has been of great concern to most digital libraries. In his study, Gu, (2010), highlighted that personalized digital libraries help to reduce wastage of time for users since information can be accessed easily. He also

highlighted that the cognitive and behavioral design processes are involved in the personal customization of the digital library. In their paper, Ferran, Mor, & Minguillón (2005), highlighted that libraries play a pivotal role for public social service systems. In order to continue satisfying users of the library, digital libraries must continue to reflect changes in technology thereby using big data. In 2013, Li & Zhang in their research titled 'The problem of large data in the construction of digital library' noted that digital libraries should focus more on their data rather than external data since it has characteristics of big data and this would increase effective use of resources. Shuqing Li, Fusen Jiao, Yong Zhang and Xia Xu (2019) outlined that a lot of benefits can be derived from the use of big data in digital libraries which include knowing the user reading habits and thereby effectively utilizing resources. According to Audrey (2012), OverDrive (Ohio, USA) claimed in its first Big Data Report in April 2012, using data mining technologies and data analyses they found out that circulation of books has great influence to publishers. Huang, Lu, Cheng, & Gui, (2016) heighted that most digital libraries are using third part to process the big data. For example, in November 2017 another scholar for Baidu announced on its official website that it had collected 1.29 million academic sites and built 400 million academic documents. According to Zhang, 2001 libraries have been trying come up with new ideas so as to meet the needs of its clients.

4. METHODOLOGY

Data were collected using an online survey. A total of 116 responses were from the online survey that was administered. The aim of the survey was to collect opinions from different users that will help the researcher to pinpoint the algorithms to implement based on the change in user taste, preference and interest. Taking into consideration that the combination of Big Data and public libraries is in its infancy, there is no mature questionnaire that could be used as reference and this resulted in the questionnaire being designed. The online questionnaires were targeted to those who uses social medial platforms and digital content in searching books and librarians who have already worked with challenges stemming from Big data. The data collection process seeks to address the following questions:

- What are the Big data analyzing tools and techniques suitable for academic libraries?
- What are the definitions and approaches to Big Data in academic libraries?
- What are the benefits of Big Data in academic libraries?
- What are the gaps in Big Data studies related to academic libraries and research trends of the future?

The respondents' experiences were assumed to be vital in the provision of additional interpretations of the application of Big Data in libraries

5. RESULTS

5.1 Big data tools

In order to determine the Big data tools and techniques suitable for academic libraries relevant questions were asked to respondents through an online survey. Respondents were asked about the different tools suitable for analyzing data in the libraries. Table 4.5 below shows the responses.

Table 1: Big Data tools

	Summary of Counter		
Big_Data_Tools	Mean	Std. Dev.	Freq.
Batch analysis	45.861111	43.746773	36
Interactive analy..	57.023256	31.22155	43
Stream analysis	72.513514	15.838809	37
Total	58.5	33.630343	116

Source: Research data 2019

The table above shows the items that were used in the survey to determine the Big Data analyzing tools and techniques suitable for academic libraries. It can be noted that the responses had frequencies ranging from 36-43 which indicates that the respondents were in agreement with the dimensions that were used in determining the Big data analyzing tools and techniques suitable for academic libraries. The dimension used that has the highest response is "Interactive analysis" with mean=57 and standard deviation of 15.8. This is followed by Stream analysis and Batch analysis with frequencies of 37 and 36 respectively.

The figure below shows the summary of big data tools which academic libraries are recommended in order to benefit from Big Data opportunities.

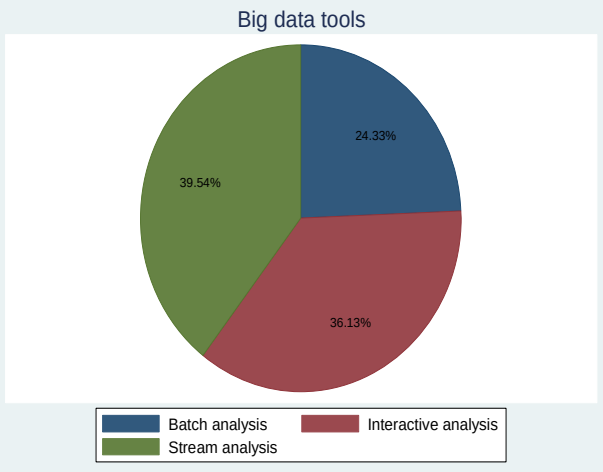

Figure 1: Big data tools

Source: Research data 2019

From the figure above it can be denoted that 36.13 % of the Big Data tools used in academic libraries relate to Interactive analysis. About 39.54 % are attributed to stream analysis. Bansal (2019) stream refers to the continuously arriving of unbounded data. Analytics of such real-time data has become an utmost necessity. It mainly requires technology of efficiency for computing data from different clusters. Academic libraries can use high technology efficiency to analyze data. According to Nikos (2018) stream data analysis promotes on the fly processing over large sets of data and libraries can use it. The data will be processed and analyzed on arrival. Fig also shows that 24.33 % of the Big tools used in academic libraries was also ascribed to Batch analysis.

5.2 Big data analyzing techniques

Table 2: Big Data analyzing techniques

Analying_tools	Summary of Counter Mean	Std. Dev.	Freq.
Data mining	69.133333	32.310694	15
Machine learning	45.714286	41.27244	21
Signal process..	58.193548	32.511966	31
Statistics	48.72	26.494528	25
Visualisation	73.625	29.347709	24
Total	58.5	33.630343	116

Source: Research data 2019

From Table 2 above it can be denoted that most of the frequencies range from 15 up to 24 which indicates that respondents were in total agreement with tools that researcher used to determine the analyzing techniques suitable for academic libraries. The technique that received the most responses is item 5 labeled "signal processing" with frequency of 31 and mean =58.2. The other techniques that has relatively higher mean values are "Statistics" and "Visualization". These have mean values of 48 and 73.6 respectively. However, the technique that has the least frequency is "Data mining" with mean=69 and standard deviation=32.3.

The figure below shows the results on data analyzing techniques that were recommended for use in public libraries.

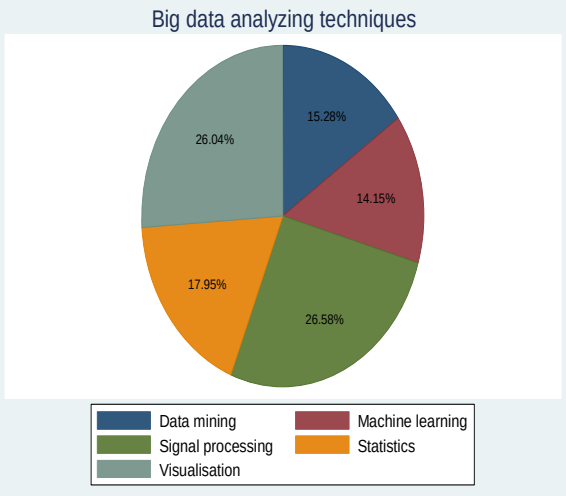

Figure 2: Big Data analyzing techniques

Source: Research data 2019

The figure above shows summary of the Big Data analyzing techniques recommended for use in academic libraries. Most of the respondents (26.58%) recommended signal processing applications in academic libraries for quick access. About 26.04% recommended visualization techniques. These results were consistent with Nikos findings (2018) who outlined that data visualization provides users with intuitive means to interactively explore and analyze data, enabling them to effectively identify interesting patterns, infer correlations and causalities, and supports sense-making activities.

Other Big Data analyzing techniques that were recommended for use in academic libraries are data mining with thresholds of (15.28%). Data mining is also a technique used to analyze large information repositories and discover implicit, but potentially valuable information. Evidently, analysis of large amounts of information from Big Data leads to the emergence of new techniques such as data mining. It is also known as knowledge discovery in databases (Han, Kamber, & Pei, 2011). The purpose of data mining is to reveal hidden relationships and unknown patterns and trends by mining into giant amounts of data (Sumathi & Sivanandam, 2006). Three techniques are used during data mining analysis, namely classical statistics, artificial intelligence, and machine learning (Girija & Srivatsa, 2006). Data mining in academic libraries is called the bibliomining. This concept is used to track patterns, behavioral changes, and trends in library system contacts (Siguenza-Guzman, Saquicela, Avila-Ordóñez, Vandewalle, & Cattrysse, 2015)."

As sees in 2 Figure 2, about 14.15% recommended machine learning for use in academic libraries. Bu et al. (2012) outlined that the purpose of this technique is to transfer the observational data in a way that can help predict any hidden data. Nowadays, machine learning is widely used as a method that drives the market and sales of online shops, keeps out spam emails, organizes advertising systems, and builds a content recommender system to suggest targeted users. It also supports scientists and researchers in different fields of science to interpret Big Data to acquire applicable knowledge, especially in certain fields such as high energy physics and biology.

According to Figure 2, 17.95% recommended statistics for use in academic libraries. Manyika et al. (2011) suggested that applied mathematics techniques, which are often used in decision-making about the type of relationships existing between variables, might have occurred inadvertently. These relationships might result from certain underlying causative relationships. However, these techniques are used to reduce the probability of type I errors (false positive errors) and type II errors (false negative errors).

5.3 Benefits of Big Data in academic libraries

Table 3: Benefits of Big Data in academic libraries

Benefits	Summary of Counter Mean	Std. Dev.	Freq.
Management plans	35.647059	38.097866	34
Opportunities for Big Data inve..	67.074074	14.667929	27
Supporting researchers	89.521739	20.10624	23
Understanding of Data analytics	53.25	28.353756	32
Total	58.5	33.630343	116

Source: Research data 2019

The table above shows the techniques that were used in the survey to find out the benefits that are brought by Big data in academic libraries. The benefits of Big Data in academic libraries are meant for management plans, supporting researchers, investing in the opportunities of Big Data and text-miming methodology. They also foster conceptual and theoretical understanding of Big Data and analytics with academic libraries. It can be noted that the responses had frequencies ranging from 27-34 for each dimension which indicates that the respondents were in agreement with the

dimensions that were used in finding out the benefits that are brought by Big Data in academic libraries. The dimension used that has the highest response is "Management plans" with mean=35 and standard deviation of 38.

The figure below shows the summary of the benefits of Big Data in academic libraries as

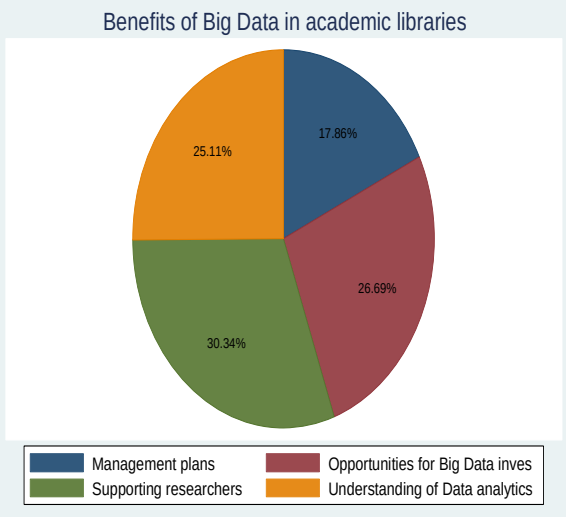

Figure 3: Benefits of Big Data in academic libraries

Source: Research data 2019

The figure above shows summary of the benefits of Big Data in academic libraries. Most of the respondents (30.34%) pointed out that Big data in academic libraries is beneficial in supporting researchers, about 26.69% believed that Big data applications provides opportunities for Big Data investment while 17.86% believed that Big data applications helps in management plans. Most of the decisions made by management these days are data driven. About 25.11% believed that it is beneficial in understanding the concept of data analytics. According to Faisal (2015) big data and data analytics are intertwined. For good results to be obtained big data and analytics must be used together. There are many analytical methods that can be applied in the library and these include Visualization, Association Rule mining, Classifiers and Basic Statistical Procedures. These results indicate that big data is very important in academic libraries. Big data is very important since it is one of the essential tools used for planning. Management relies heavily on big data for planning since they can use past experience to plan for the future. The results also show that big data is mostly important to the researchers since they will make use of it to find problems faced by library used and suggest solutions. According to Gerald & Lindhult (2019), systemic innovations can be made with the use of big data. The researchers have to understand the history of the library in order to understand where it is going. Big data can help researchers most of the information that they need.

6. CONCLUSION

Big data applications in academic libraries are too complex and hence the recent increasing demand in solving needs of users or researchers. The main purpose of this research was to provide an overview of tools and techniques used for analysis of big data in e libraries and in order to meet dynamic needs and wants of clients. can be used. Understanding the hail of Big Data in academic libraries is vital so that corrective measures can be taken in order to improve the current recommendation system as well as improving the efficiency. With the tremendous amount of data being generated in e libraries, techniques and tools used in this research can improve the current recommendation because the results were clear and accurate. Literature review in this study indicates that Big Data applications in libraries results in the creation of new knowledge and libraries that use big data analytics are more productive and efficient than others. To conclude Big Data tools and analyzing tools require special skills and they facilitate for improved performance in this competitive world.

Area for further research.

The future studies should survey the actual platforms or technologies that are used in academic library Big data. The possible arrears to explore above this study may include:

- ✓ A research study that seeks to determine how the academic libraries can benefit from analyzing digital content and social media material for enhancing their information services
- ✓ A research study that determines how Big Data can systematically affect the economic value in academic libraries.

7. REFERENCES

[1] Lindell, Jim. (2018). What are Big Data and Analytics?. 10.1002/9781119512356.ch1.

[2] Singh, Pramod. (2019). Recommender Systems: With Natural Language Processing and Recommender Systems. 10.1007/978-1-4842-4131-8_7.

[3] Yu. Ignat'ev, V & Lemtyuzhnikova, Daria & I. Rul', D & L. Ryabov, I. (2018). Constructing a Hybrid Recommender System. Journal of Computer and Systems Sciences International. 57. 921-926. 10.1134/S1064230718060060.

[4] Lauren Reinhalter, and Rachel J. Wittmann. "The Library: Big Data's Boomtown." Serials Librarian 67.4(2014):363-372.

[5] Daniel K.B. (ed.). (2017). Big Data and Learning Analytics in Higher Education, New Zealand.

[6] DOI: http://arxiv.org/pdf/1203.0160v2.pdf

[7] Qian, Y. H., Cheng, H. H., Liang, X. Y., & Wang, J. X. (2015). Review for variable asso- ciation measures in big data. Journal of Data Acquisition & Processing, 6, 1147–1159.

[8] Wu, K., Su, X. N., & Deng, S. H. (2013). Big data, cloud computing and user behavior analysis. Digital Library Forum. 6. Digital Library Forum (pp. 19–23).

[9] Zhang, H. (2016). Analysis on the present situation of big data research in university libraries in China. Library Work and Study, 7, 46–50

[10] Coelho, H. S. (2011). Web 2.0 in academic libraries in Portuguese public universities: A longitudinal study. Libri, 61(4), 249–257.

[11] Li, L. (2012). Discussion on personalized service mode of digital library in colleges and universities. Lantai World, 23, 91–92.

[12] Gu, L. P. (2010). Research on models of user behavior driven personalized services. New Technology of Library and Information Service, 26(10), 1–9.

[13] Ferran, N., Mor, E., & Minguillón, J. (2005). Towards personalization in digital libraries through ontologies. Library Management, 26(4/5), 206–217 (12).

[14] Li, B. Y., & Zhang, X. Y. (2013). The problem of large data in the construction of digital library. Information Sciences, 11, 26–29.

[15] Li, Shuqing & Jiao, Fusen & Zhang, Yong & Xu, Xia. (2019). Problems and Changes in Digital Libraries in the Age of Big Data From the Perspective of User Services. The Journal of Academic Librarianship. 45. 22-30. 10.1016/j.acalib.2018.11.012.

[16] Audrey, W. (2012). Strata Week: Harvard Library releases big data for its books. Retrieved from http://radar.oreilly.com/2012/04/harvard-book-data-cloudera- hadoop-splunk-ipo.html.

[17] Huang, Y., Lu, W., Cheng, Q. K., & Gui, S. S. (2016). The structure function recognition of academic text—Application in academic search. Journal of the China Society for Scientific and Technical Information, 35(4), 425–431. Information Technology Journal, 5(3), 590–600. DOI: http://dx.doi.org/10.3923/itj.2006.590.600

[18] Zhang, X. L. (2001). Mechanisms of digital library: Evolution of paradigms and its challenges. Journal of the Library Science in China, 27(6), 3–8.

[19] Bansal, Ankita & Jain, Roopal & Modi, Kanika. (2019). Big Data Streaming with Spark. 10.1007/978-981-13-0550-4_2.

[20] Bikakis, Nikos. "Big Data Visualization Tools." (2018).

[21] Han, J., Kamber, M., & Pei, J. (2011). Data mining: Concepts and techniques (3rd ed.). Waltham, MA: Elsevier.

[22] Sumathi, S., & Sivanandam, S. N. (2006). Introduction to data mining and its applications. Berlin: Springer.

[23] Girija, N., & Srivatsa, S. K. (2006). A research study: Using data mining in knowledge base business strategies.

[24] Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J., & Cattrysse, D. (2015). Literature review of data mining applications in academic libraries. The Journal of Academic Librarianship, 41(4), 499–510.

[25] Bu, Y., Brokar, V., Carey, M. J., Rosen, J., Polyzotis, N., Condie, T., … Ramakrishnan R. (2012). Scaling datalog for machine learning on Big Data. Computer research repository (CoRR) (pp. 1–14). Cornell University Library.

[26] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big Data: The next frontier for innovation, competition and productivity. New York, NY: McKinsey Global Institute.

[27] Kalota, Faisal. (2015). Applications of Big Data in Education. International Journal of Social, Education, Economics and Management Engineering. 9. Kalota, F. (2015). 'Applications of Big Data in Education'. World Academy of Science, Engineering and Technology, International Science Index 101, International Journal of Social, Education, Economics and Management Engineering, 9(5), 1501 - 1506.

[28] Midgley, Gerald & Lindhult, Erik. (2019). What is Systemic Innovation?.