THINK INDIA JOURNAL

ISSN:0971-1260

Vol-22-Special-Issue-27-December-2019

Big Data and Libraries: Looking Deeper into Technologies

Kulveen Kaur*

****Bhai Kahn Singh Nabha Library, Punjabi University, Patiala***

Abstract

Advances in technology have already influenced libraries to embrace ICT applications in daily housekeeping operations in order to provide effective and efficient services to the library users. With the onset of big data tools, plethora of information can be extracted from large and massive datasets to yield useful and visible results. Big data gives helps to visualize search patterns and trends in human behavior. Also, it helps in strengthening and optimizing various management operations and enhance user services in accordance with the user behavior.

Keywords: Big Data, Data Mining, Library big data, Visualization, Evaluation, Services

Introduction

Twenty first century libraries have seen tremendous change in terms of growth and development. An unprecedented growth of information on three fronts of volume, velocity and variety results in advancement of data storage technologies and emergence of big data. Availability of data in various forms, various sizes (huge) coming from any source, any media, etc. produces massive data sets that yields myriad of surprising results on analysis. However, the biggest challenge for the libraries coming their way is to manage and process an ever-increasing amount of data (raw information).

The fascination for big data is increasing worldwide encompassing every sector and libraries are not far behind where opportunities for data services are expanding rapidly. Managing and processing large data sets of log files, blogs, online catalogue record, etc. has now becoming beyond the ability of traditional relation databases. Remarkable developments in technology creates new opportunities to extract more value from big data sets for better and faster decisions. Some of the prominent definitions of big data has been studied to know the nomenclature of big datasets. According to International Data Corporation (IDC) "Big data technologies describe a new generation of technologies and architecture designed to economically extract value from very large volumes of a wide variety of data enabling high velocity capture, discovery and/or analysis." Hoy's (2014) referred big data to the ability of computers to "gather trillions of pieces of information about billions of different things and find useful patterns in that information". Likewise, Sherman (2014) defined it as a collection of large and complex datasets that required new technologies and applications for processing. Libraries are the center of knowledge and learning and there is no excuse for libraries not being involved with data mining. Zhan (2016) stated that "Therefore, libraries and data mining are naturally connected from the knowledge point of view." Kim & Cooke (2017) performed big data analysis of operation and services of public library in two different cities of two countries i.e. London and Seoul. The services offered by the libraries in these two countries were compared and it was found that the library in London was performing better than that in Seoul. The analysis was performed by using Chernoff face method. Reinhalter & Wittmann (2014) concluded that the use of Big Data as an information resource will continue to become more prevalent in academics and research

Blummer and Kenton (2019) explicitly discussed free data sources and data analytic tools used for improvement of services in the libraries. Many tools and techniques are available for handling big data operations and some of the most commonly used tools are:

- **Hadoop:** It offers features such as robust ecosystem flexibility and faster data processing to meet the analytical needs for data analysis.
- **HPCC:** It offers features such as complex data processing using lesser code. The graphical IDE simplifies development, testing and debugging.
- **Storm:** It is an open source system capable of performing real data computations. The process of analysis using this system is considered one of the easiest.
- **Qubole:** It provides a single platform for end to end big data processing.
- **Cassandra:** Its versatility lies in the fact that it is used for applications where the users can't afford to lose data.
- **Statwing:** As the name suggests, it is used for statistical analysis such as creating charts, plots and heatmaps etc.

121

ISSN:0971-1260

Vol-22-Special-Issue-27-December-2019

- **CouchDB:** It uses JavaScript for accessing data. It allows accessing data by defining Couch Republication Protocol

Some other systems are Flink, Cloudera, Openrefine, Rapidminer, DataCleaner and Kaggle etc.

It is clear from above discussion that there is no specific tool of analysis which will fit every library. The library professionals will have to use different tools for analysis of different services but the application of Big Data analysis technology is surely going to help improve the infrastructure in two libraries. The use of high level languages and artificial intelligence tools is going to increase manifolds in the coming days and it is the right time for library professionals to have hands on experience on languages such as R, Python C, etc.

. **Opportunities and Challenges**

Information industry is quickly moving to digitize their records and images to optimize operations for immediate opportunities along with an increase in statistical logistics, video surveillance (IPTV), world cat, RFID tag readers, etc. at an infinite rate and subsequently the size of the datasets. Big data research has just started and library professionals are trying to integrate this technology into library data. However, it is unclear that how big of data could be classified as big data. It is assumed that big data includes data with terabyte and petabyte of size. In addition, Big data enables deeper and more valuable insights from the data in number of ways:

- Library is a service providing institution and the data mining operations in services such as membership record, issue and return of books, web OPAC, utilization of e-resources such as e-books, e-journals, e-databases such as Proquest, etc, can help know the use pattern of popular and most sought library resources which further can be used by the institution management to focus on strengthening the most used resources in library and improve upon the lagging services through investment and staff strengthening. It helps the institution in optimal management of library resources.
- The user pattern would be helpful for libraries as well as for its users in managing inter library document delivery services. Mining of the user data will also help the business enterprises dealing with libraries in managing their stocks. Chang and Chen (2006) used data mining technique to classify the users into five clusters according to the library resources they were interested in and found that the graduates and associate researchers preferred digital information over traditional resources. Bansal & Kaur (2013) analyzed data of UIET library using Association Rule Mining (ARM) technique and compared it with Structured Query Language (SQL) mining. On analysis of book transaction data, it was found that the book Chemical Reaction Engineering was the most commonly issued book for a particular stream of engineering.
- Cataloguing and classification of documents involves a lot of manual intervention. However, there is a big scope of performing classification of books and journals by using artificial intelligence techniques. For that, however, the data needs to be mined to know which types of books/ journals have been classified into which particular category. The library professional would have to enter different attributes of the book such as its name, author's background, field or contents of the book and the cataloguing would be done automatically. This process would save manhours of the staff. However, developing such a mining and decision-making system is not only challenging, but as well it needs the interdisciplinary faculties to work together.

The data mining has many advantages but require large storage systems and expertise for analysis. The data requires modeling and remodeling as per the applications for which it is being acquired. The cards used by the users require to be changed and the library systems must be connected through proper networking system giving access to the user data and library resources. The big data analysis requires special software and large investment. Getting investment for studying user pattern is a challenge for libraries already reeling under cash crunch.

Conclusion

Integrating technology in libraries will simplify the library housekeeping operations like daily classification, acquisition, circulation and referencing. Standardization of resources (e.g. books, journals, magazines) can be setup to extract useful information which can help in better relativeness with the knowledgeable information. The library professionals are known for their enduring skills and supporting research in a digital world. The amalgamation of technology associated with Big Data Analysis would as well test the programming abilities of library professionals and will make this noble profession interdisciplinary. Libraries have always been at the forefront when it comes to integrating technology with education.

122

ISSN:0971-1260

Vol-22-Special-Issue-27-December-2019

REFERENCES

- Bansal, M. & Kaur, M. (2013). Analysis and Comparison of Data Mining Tools Using Case Study of Library Management System International Journal of Information and Electronics Engineering, 3(5).
- Blummer, B. & Kenton, J. M. (2018). Big data and libraries: Identifying themes in the literature. Internet Reference Service Quarterly, 23(1-2), 15-40.
- Chang, C. & Chen, R. (2006). Using data mining technology to solve classification problems A case study of campus digital library. *The Electronic Library 24(*3) 307-321
- Hoy, M. B. (2014). Big data: An introduction for librarians. *Medical Reference Services Quarterly*, *33*(3), 320–326.
- Kim, Y. & Cooke, L. (2017). Big data analysis of public library operations and services by using the Chernoff face method. Journal of Documentation, 73(3), 466-480
- Reinhalter, L. & Wittmann, R. J. (2014). The Library: Big Data's Boomtown. The Serials Librarian, 67(4), 363-372
- Sherman, C. (2014). "What's the big deal about big data?". *Online Searcher*, *38*(2), 10–16.
- Zhan, M. (2016). Exploring the feasibility of applying data mining for library reference service improvement: A case study of Turku Main Library. Master thesis, Åbo Akademi University, Åbo.