



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 8, Issue 8, August 2020

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.488

9940 572 462

6381 907 438

ijircce@gmail.com

www.ijircce.com



Personalised User Service in Digital Library in the Era of Big Data - Problem, Changes and Solutions

Reshu¹, Arun Mittal²

M. Tech Student, Department of CSE, Greater Noida Institute of Technology, Greater Noida, Dr. APJ Abdul Kalam Technical University, Lucknow, India¹

Asst. Professor, Department of CSE Greater Noida Institute of Technology, Greater Noida, Dr. APJ Abdul Kalam Technical University, Lucknow, India²

ABSTRACT: Digital Library is a place which consists of very useful information for any academic institution. It is full of knowledge and data resources. Traditionally, libraries were used as a source of information providers to faculties and researchers. But, with the advancement in the technology and increase in the knowledge, library users started demanding individual specific relevant data instead of generic data. This requirement can only be fulfilled by analysing the data which is produced by the user within the library. In this paper, we have discussed many issues which library is facing due to the transformation in the library system.

In order to fulfil the demand of the users, it becomes important to implement latest technologies within the digital library. We have discussed about some of the latest Big Data Technologies which has been used for user's data management within the library and about the experiments which was done in British Digital Library by using Microsoft Azure.

This study mainly focuses on understand the perspective of a user for getting the relevant information according to his/her personal interest from the digital library instead of literature based general search. In order to achieve this personalised information, we have studied about the sources from where we can get this information about an individual.

KEYWORDS: Digital Library, Big Data, User Service, NoSQL, Hadoop

I.INTRODUCTION

Libraries are most important department which help people to gather and access all the useful information related to various subjects and concepts. After digitalization and with continuous development in the field of information technology, libraries are also adapting themselves with the new technology. Librarians are now more technical and are constantly upgrading themselves according to the new digital Era.

Digital Libraries which are also known as digital collection or digital repositories are an online database of major digital objects like audio, video, text, images, digital media, digital documents etc. Digital libraries are categorized in three types:

- a) **Institutional Repository**
- b) **National Library Collection**
- c) **Digital Archive**

Institutional repositories contains the institution's digital data like these, papers, books and other electronic work which can be digitized for reference purpose. National Library collections contain the legal deposits and these legal deposits are covered by copyright legislation. These national libraries contain one or more copies of all the data or material published in a country for the purpose of national level preservation. Evolution in information technology and discovery of digital library was revolutionary and beneficial for archives because in digital archives, primary data of the individual can be recorded instead of group level storage as in case of physical archives in the library and the content can be stored in two or three copies because it was easily to reproduce the digital data for storage purpose.



Due to modernisation in digital libraries with huge number of available content it becomes important to manage this data by latest technologies like Big Data and Artificial Intelligence. Many researches have been going on in this particular area to identify if the library data is a big data or not? According the report published by IFLA (The International Federation of Library Associations and Institutions), Artificial intelligence is among the top four major technologies because it has three major features which can used to implement in digital libraries:

- Advancement in the browsing pattern which will not include keyboard search anymore in upcoming decades.
- Real time translations will be used with the help of latest technologies speech reorganisation systems with the help of machine learning.
- Multiple complex data will be analysed from the cloud [1].

Libraries of current generations with latest technologies like Big Data, data Science and Artificial Intelligence are facing challenges because of the transformation of user's requirement from general data to personalised data. Now they are more interested in the search with are relevant to their research. Now libraries have to improve the data content delivery with exact and relevant searches [2].

The objective of this study is to contribute in the world of Big data, which focuses in transforming the digital library from user's perspective. It mainly enlightens the demand of researchers for personalized model of the library. They want to get the details of the research or the articles which are of their interest instead of general data.

Some of the research questions are mentioned below:

1. How big data technology can be used for transforming the traditional general data digital library into personalized data service provider?
2. How Big Data has helped the libraries?
3. Is Library data is a big data? If yes, how?
4. Factors which has been considered while analysing the data and approach which need to be implemented as per user's perspective.
5. What can be the further open research which needs to be addressed?

According to the above mentioned research questions, we have provided following contribution via advance research and latest technology:

- 1) Use big data technology to adapt the personal user requirement in order to deliver the appropriate content, automatically tag the related content to the user, data sharing facility should also be implemented so that they can interact with different users through online portals.
- 2) Digital libraries have accepted big data because it has resolved their problem of overloading of information which was produced by web. It has also helped in archiving of the database within the library. Big data provide the exact facts and figure of the data which helps in auditing, talks and showcasing the current condition of a library.
- 3) Traditional libraries were based on the general data and was working on the transitional database model i.e. RDMS. As compared to other fields in which big data research is going on, research of big data in digital library is quite limited because of scattered data, budget constraints in the library and less emphasis in this domain for implantation.
- 4) While analysing the data we have considered the perspective of user for interest in personalized knowledge and data. Users are more interested in time saving tasks i.e. they want to search the relevant topic in less time, they want to share and discuss about the content with other users. They need the facility of automatically tagging, recording and management.
- 5) Further open research which can be addressed will be implantation of Big Data with Artificial Intelligence and Data Science in order to get the log of the individual user which can result in getting the idea of the interest as well helps in tracking the usage of the journals or databases subscribed by the library.



II.LITERATURE SURVEY

In this section we will be discussing about the definition, features and technology used in Big Data which will justify its usage in digital libraries. We will also discuss about the past research which were made in the field of digital libraries and how big data can be considered in digital library for integration as per user's requirements. This section will help us in understanding the basics of Big Data and its functional structure as well.

2.1 Data related to Digital Library

Information systems which are used currently cannot process, analyse and store wide range of data. The data is too vast that even if we process that on regular basis it cannot cope up with the current need. For example, as the hits of searches are continuously rising on daily basis over the internet, it is resulting in big data [3]. Due to increase interaction of users over the internet for searching the articles in the library, the amount of data which is daily produced in the library has been increased [4].

There is abundant number of increase in the scholarly data. In the year 2014, it has been estimated that 114 million English scholars data was produced which was approximately 27% of the data which was freely available [5] as the number of scholarly data is increasing it has become quite important to manage such data, organize the data, finding out the relationship between the entities of the documents or we can say we need to understand the science of the science [6]. In recent times, digital libraries should learn to handle data from external sources along with resources which are internal because even the internal data has started showing the attributes of Big Data (Li & Zhang, 2013). Within the digital library, primary data is electronic documents; user's browsing history, data of users information and some of the data in different formats and data types (images, audio, video, etc.) [7].

As per statistics, the largest document repository in digital format is with National Library of China. They have more than 1000 terabytes of data as digital resource and the hike rate is around 100 terabyte per year [1]. In 2007, National Library of China had 200TB of digital data but till the end of the year 2011 they had 561.3TB of data [8]. This is showing a huge increase in the data which is satisfying the characteristics of Big Data. Even Harvard University Library announced the big data in the year 2011 which includes more than 12 million books which were provided by their 73 library branches. Each and every collection was offering 100 different data for each attribute like maps, manuscripts, data, etc. Users had the facility to access the data of U.S. Digital Public Library [1].

With the advancement of big data, now most of the digital libraries have started using tools for data analysis, performance analysis of the research as compared to the traditional methods of data analysis.

2.2 Big Data Definition

In this section, we will list and understand some widely used definition of Big Data. It is quite difficult to define big data in a simple definition. The idea of Big data was first introduced by Laney in 2001 during his research [9]. We can begin with some basic definition Big Data is a collection of huge amount of complex data, which is received from sources which generate excess data in fraction of seconds or produce real time data. These data are mostly structured, unstructured and semi structured. Such data are difficult to be handled by traditional databases[29]. Big data has been characterised in 4 V's : Volume, Variety, Velocity and Veracity [10].

Detailed description of four V of Big Data has been given below:

1. Volume – It refers to the amount of data which is need for processing in big data. Volume of the data should be quite high and should be continuous in order to be considered as big data. For example, in social media (Facebook, Twitter, etc.), GPS, healthcare data, etc., huge amount of data has been produced in a fraction of second which results in big data.
2. Variety – It refers to the different types of data in different formats. Big data has the capability to handle variety of data i.e. structured, unstructured and semi-structured data. Now days, data is flooded in different formats which includes, audio, text, video, web data, streamed data, tweets, log files etc. Therefore, it is important to deal with such data for getting the accurate analysis.
3. Velocity – It refers that at how much speed the data is creating, processing and analysing for processing. In recent time, the data is accelerated with very high speed because the data is coming from real-time applications and live streaming. It is difficult for traditional data management system to deal with such a huge and complex data. Most of the organisations are finding big data reliable because the accurate data is received within specified time frame.
4. Veracity – It refers to uncertainty of the data. It also indicates the reliability which is associated with certain type of data. For example, it is quite unpredicted to calculate the data related to weather or related to economy because it may or may not provide the accurate results of the predictions but the data and analytical report received after utilizing the big data techniques are accurate as compared to other technologies.



III.EXISTING TECHNIQUE

According to the definition of big data provided by Manyika, the size of the datasets of big data is beyond the capability of traditional database software for storing, managing, capturing and analysing the data.[11]

Big Data Technologies

Big data technologies works upon the below mentioned three fundamental factors:-

1. Performance and Capacity of the hardware.
2. By reducing the data size before processing. This can be achieved by data compression and using Colum-oriented database (NoSQL) like Cassandra, HBase etc.
3. Parallel processing of the data by distributing the data. This is achieved by distributing the data in multiple disks for parallel processing and assigning the job to different computers so that they can work parallel in order to maximize the throughput and fault tolerance. This is the concept which is used by Map Reduce.

Above mentioned factors can be achieved by following technologies of big data : Massively Parallel Processing (MPP), NoSQL (Not only SQL), Hadoop and Mapreduce.

3.1.1 Massively Parallel Processing – MPP

In Massively Parallel processing, the data is spread over the independent nodes or servers which are visible to the users who are accessing the database [12]. In Big Data, the concept of MPP is usually called shared-nothing database [13], because the cluster is created by independent nodes, these nodes does not share disk or memory but they communicate by network.

After the invention of multicore CPU, these MPP databases were configured in such a manner that they considered each core as a node and within the single server they were able to perform parallel processing.

3.1.2NoSQL (Column Oriented Databases)

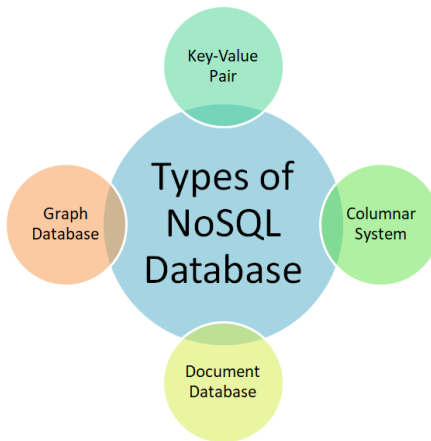
Traditional Relational Database Management System (RDMS) [13] introduced the concept of accessing the structured data by using SQL. RDMS was capable of accessing, categorising and normalizing the queries which was used in multiple IT industries as well as many other organisations. The record which was processed by RDMS was structured and had fixed length.

To overcome the logic and concept of traditional database system (RDMS), column based database evolved [14] in which the data is stored in column instead of row-column format. The data is compressed and stored in a single column so whenever the required data need to be fetched, the processing become quite fast. Once the data is compressed in the column it helps in query optimization as well. In 90s, the first example of column based SQL was introduced which were available like (e.g. Sybase IQ [MacNicol 2004]) but now it has established as NoSQL[15], this data model is beyond the relational model. Initially, NoSQL stands for No SQL but now it has been modified as Not Only SQL. Some of the examples of NoSQL are Cassandra, HBase, MongoDB etc.

Extremely large datasets can be processed by this database. In this database, there is no fixed schema to store the data because it adapted itself as per the data. Due to this it holds the feature of scalability which makes the physical data representation easier.

The type of databases which will fall in the category of NoSQL will possess following features :-

- **Key-Value Pair System**
- **Columnar System**
- **Document Databases**
- **Graph Databases**



Key-Value Pair System

It uses a hash table which has unique key and the value associated to that key which will be associated by the key using a pointer [16]. These key value database does not requires any schema as they do not follow the properties of traditional databases (RDBMS) like ACID or the concept of data replication, data placement etc., due to which these key-value databases provide high scalability and flexibility while accessing the data. The data retrieved from key value pair systems are mostly stored in a string.

Columnar System

This database is used to processes the massive data which is distributed among multiple machines [17]. In traditional Database Management System RDMS, these data are stored in tuples and column format. In column oriented database, it is quite easy to add up the column for a given key. This functionality provides great scalability, flexibility and optimisation for accessing the data in a faster way. Whenever required, columns can be easily added. Most common example of Columnar Database System is Google’s Big Table, in that the rows are identified by row key with the associated data with it. Some of the examples of columnar NoSQL databases are Cassandra, Hypertable, HBase etc.

Document Databases

Document databases are also similar to key value database but in this case these key-values are documented in a document and the structure of the document is provided by JSON (JavaScript Object Notation) or Binary JSON (BJSON). These databases are mostly used for documentation purpose and showcased for presenting any report or data analytics. Example of document databases are MongoDB, CouchDB, etc.

Graph Databases

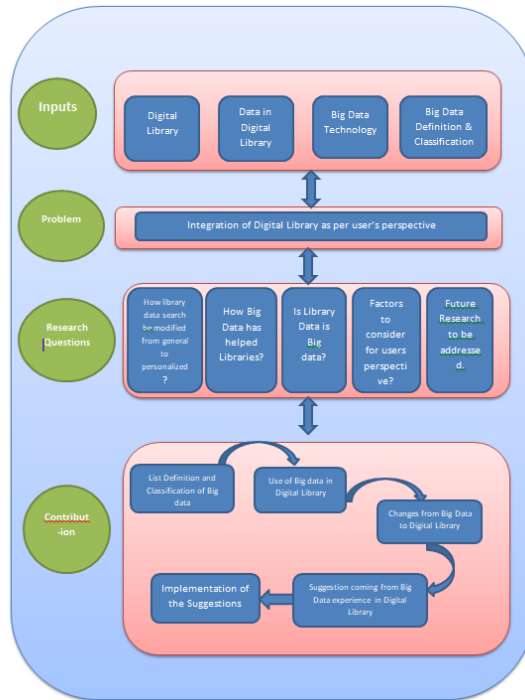
Fundamental structures of Graph databases are called node- relationship [18]. As the name is suggesting, this data base is based on the graphical view of nodes which shows the connection or relationship with other node. Data within this database is also stored in key value pair unlike traditional RDBMS because they had the rigid schemas which need to be followed. In these databases we reach to the data by navigating through multiple nodes which have relations among them. Example of Graph Database is Neo4J, Virtuoso.

3.1.3 Hadoop and MapReduce

Hadoop, framework developed in Java for running the application in parallel across the large clusters was first developed by Doug Cutting, who was a developer of Lucene (search engine) and Nutch (Web Crawler) [19]. Cutting was working on a project where he tried to implement Google File System with different requirements, at that time he studied GFS and got influenced to work upon Hadoop project [20]. Doug was hired by Yahoo! To work upon a project which was similar to Google File System in order to provide the distributed processing with MapReduce platform. As a result, Hadoop was declared the basic MapReduce framework which was used outside the Google. Hadoop is now taken care by Apache Software Foundation [21], which is helps in execution of data centric applications. There are multiple open source Apache projects which ran on Hadoop, they are Zookeeper, Hive, Cassandra, Pig, HBase.

Hadoop architecture is consisting of three parts:-

IV.FLOW DIAGRAM OF METHODOLOGY



V.BIG DATA IN DIGITAL LIBRARY FROM USER'S PERSPECTIVE – CHALLENGES

This section discuss about the challenges which digital libraries are facing while integrating in the latest technology like Big Data and also to meet the demands of the users as per their personalized search.

5.1 Is Library data is considered Big data?

Contrasted with different areas, an investigation for libraries is limited in Big Data. For instance, the general exploration intensity is commonly frail, research efforts are dissipated, and there is an absence of observational examination. Specifically, there is insufficient accentuation on the functional use of large information in libraries [22]. There are numerous explanations behind this. For instance, a few researchers accept that advanced libraries will in general act naturally sufficient comparative with hierarchical administration and frequently do not think about rising expertise [23]. A few researchers contend that the reasons might be related with spending capital [24]. Despite the fact that hardware costs are declining quickly and programming convenience is expanding, it remains difficult to decide if such changes will help ease the difficulties related with insufficient assets for digital library development construction [25]. A few researchers believe that data security and security assurance issues are inevitable while thinking about Big Data applications. For instance, private data may disclosure if client data is utilized to distinguish client interests. Such dangers will likewise affect the broad utilization of Big data information innovation in digital libraries [26].

Indeed, even the benefits of customary information assets have created some shortcomings comparative with huge scope investigation into and the utilization of Big data in the field of libraries. For simple use, such information must be changed, e.g., digitization of paper records, and incorporated with other information. Nonetheless, this information has encountered different improvement stages and different data handling advancements have been embraced.

In this manner, unlike rising fields, changing and incorporating conventional information assets regularly causes significant handling and transformation costs, which speaks to another difficulty for digital libraries comparative with the utilization of new data technologies.



5.2 Big Data Impact of Digital Library

The rise of Big Data fuels the issue of data over-burden that has been related with the advancement of the web. The expenses and difficulties related with efficient utilization of data assets are expanding. For instance, from a client point of view, Big Data can create a developing prerequisite for feeble data among clients [27], which is regularly portrayed by fluffy structures, hazy information scope, an absence of clear and orderly recovery and revelation steps, and the need to powerfully investigate an incredible number of documents to accomplish just fractional fulfillment. Nevertheless, even with unpredictable and dynamic exploration issues, the necessity for powerless data is getting progressively significant and normal.

As far as technological ability in the big data age, there is developing requirement for data specialist, for example, data engineers with big data handling information and expertise, data analysts who can display big data and perform application examinations, and data stewards who can oversee and find important information and guarantee information accessibility. Data stewards include data archivists, subject librarians, and other comparable experts [28]. Librarians have helped analysts gather and break down scientific information for quite a while. In the period of big data, subject librarians are turning out to be increasingly specific information librarians. Ability related information for data librarians incorporates open data licensing agreements, protected innovation rights, rights the executives, data management plans, asset usage, information investigation and application, big data deployment and related designs, the board of institutional stores, information reference, and information distribution (Journal Center of the School of Information Management, Wuhan University, 2016).

5.3 Problem in Innovation of Digital Library in Big Data Era

Since forever, libraries have been continually advancing and altering themselves to adjust to changes in human culture and innovation [29] and, in the big data time, the current elements of digital libraries are confronting new changes. A few issues have arrived at a point that requests the hypothesis of eradication of the library proposed by American librarian Lancaster must be thought of [30]. With the far-reaching utilization of Internet innovation and cell phones, the library is no longer solely resource management.

1. **Delivery of Resources:** The digital library has become more advanced. Previously, the exchange of resources was performed by librarians. In cutting-edge technology, the main part of this work is performed by the end user themselves. Getting paper reports can be completely automated. Getting digital assets has become a fundamental learning expertise, and most clients have learned the aptitudes and innovation required to find and access different data resources. Nowadays, libraries are no longer termed as information resource hub as it was treated before. For example, according to OCLC's Library Awareness, 2010, only 1% people used library portal for searching and gathering required information.
2. **Utilization of Resources:** The benefit of digital library is to provide functions and form of digital resource services, which a traditional library system can never provide. But, with steady and continuous advancement in academic big data on internet, present users intent to utilize digital library's resources are also being weakened constantly and users give preferences to use other digital material on internet. Due to increase number of web searches by scholars now they are more comfortable in fetching the information from internet resulting in least interest in library data [31].
3. **Social Reorganisation:** It has been observed that the value of the library and the librarians are decreasing in the new world of technologies. For example, a former research assistant who is now designated as subject librarian does not have any importance in new discoveries, analysis and discussions because of the change in user's requirements. One of the survey which was done by Ithaka Institute of the United States indicated that from last three years, the importance of library is decreasing [32]. Libraries are still considered as place of data storage and prevention. Although, the interest is declining but the libraries are still purchasing the new resources for them. Due to all these factors some of the researchers has pessimistic thinking for the development of the library [33].
4. **Change in thinking:** Libraries are also going through different change of thinking because as per traditional methods some researchers still wants to explore the literature, want to identify the resources but some of the researchers find the new trends of the technology. This different requirement of the users resulting in difficult decision making. After discussing about such situation by multiple librarians, OCLC is believing that future library technology will face the crisis which will have a huge impact in future [34].



VI.RESULT AND ANALYSIS

6.1 Introduction

This chapter discuss about the analysis which we have done during our research regarding the new techniques which can be used in digital library and integration of the library as per user's requirements. It also included the answer of the research questions which were raised in the starting of the research. Transformation of the digital library for user service is an important domain where more research and implementations are still needed. We have tried to gather all the relevant information to understand the need of library up gradation as per users service and the technologies and methods which and be used for achieving this goal.

6.2 Analysis on the importance of user services in Digital Library

New technologies like Big Data has resulted in many challenges for digital libraries and also provided multiple reasons for transformation. Demand of the researchers from the library is rapidly increasing, now they are more interested in personalized data in comparison to the data-intensive research. Library is facing many internal and external pressures due to change in system therefore, it has become important for the library to adapt with the latest technology in order to deliver the exact content [35].

Traditional digital libraries always played a role of service providers to the researchers and never gave any special interest to the individual requirements for some specific data. With the development in the technologies, libraries felt the drastic change in the users which now demand for the personalised service instead of traditional mass service model because of that, content within the library got transformation from traditional literature based to modern user based. After this transition, users started using the resourced of the digital library in an optimized manner [36].

There are basic three reasons which explain that why big data era has resulted a personalized service in digital library:-

- 1) Due to production of the huge amount of data within the library it has become very difficult to understand and separate the useful data among that. Overloading of the search within the library has produced combination of irrelevant and relevant data in different data types. Filtering out irrelevant data among such a huge data is the main task to provide the user centric data in order to provide user satisfaction within the library.
- 2) Continuous increasing in data also results in continuous increase in data connections. These connections help us to understand the raw data in an efficient and effective manner so that research can be done on such data for understanding the core interest of an individual user. For example, the social data of a user helps in understanding the correction information by which library can provide the relevant search to the user. Digital library is a reservoir of user information like user information, personality information (this can be gathered by keeping the track on the browsing history, retrieval and history of the downloads). By analysing such data of an individual user, we will get more comprehensive result for the user and we will be able to understand the habit, behaviour and interest of a particular user.
- 3) In order to obtain the knowledge of a particular application, a user searches for the data. The understanding of that particular data is totally dependent upon the platform, environment and the information which is currently available.

If the library starts providing the personalized service to the researches then libraries can provide the wide range of diversity and serviced to its users resulting in their time saving [37]. Due to lack of personalised service within the library a user need to spend its most of the time in searching the data which is relevant to him. These hit and trial of data search generated frustration in the users and pessimistic thinking about the library.

As traditional library need advancement due to increase demand of user centric data, it has become very important to deal with the latest technology like Big Data. Although digital library is facing slow research in this field by experts as well as by individuals, but it is the area where advancement is much needed.

Many issue of the traditional digital library can be resolved by using technologies of Big Data.

Characteristic of Academic Data which need to be handled by Big Data

- 1) Authors name and the format to mention them.
- 2) Citations of the Research article
- 3) Subject Area
- 4) Same name in literature as in publications
- 5) Different Formats

Issues within the diversity of the citation have recently been resolved because the formats were different for the different platform [38].



6.3 Analysis on the process and data of Digital Library

Implementation of advanced technology in digital library is always a point of discussion for scholars. As per technical concept, scholars has suggested to use advanced technologies like cloud computing and service oriented architecture because the new libraries will have media data as well which will need management of the resources, management of the business capabilities [39].

According to some other scholars, digital library should consider three aspects before transforming by Big Data technologies:

- 1) Resource Collection
- 2) Application of the Technology
- 3) Services of the Library

According to more scholars, the change in digital library is dependent upon five aspects:-

- 1) Library to manage big data due to huge amount of data flow.
- 2) Library should also deal with data centric environment due to scientific research method.
- 3) Library should also take care of the Business development model.
- 4) Libraries are transforming from data-centric approach to personalised approach, so library should have all the knowledge resources.
- 5) Library should be capable of dealing with change in information technology [40].

There are basic three reasons which explain that why big data era has resulted a personalized service in digital library:-

- 1) Due to production of the huge amount of data within the library it has become very difficult to understand and separate the useful data among that. Overloading of the search within the library has produced combination of irrelevant and relevant data in different data types. Filtering out irrelevant data among such a huge data is the main task to provide the user centric data in order to provide user satisfaction within the library.
- 2) Continuous increasing in data also results in continuous increase in data connections. These connections help us to understand the raw data in an efficient and effective manner so that research can be done on such data for understanding the core interest of an individual user. For example, the social data of a user helps in understanding the correction information by which library can provide the relevant search to the user. Digital library is a reservoir of user information like user information, personality information (this can be gathered by keeping the track on the browsing history, retrieval and history of the downloads). By analysing such data of an individual user, we will get more comprehensive result for the user and we will be able to understand the habit, behaviour and interest of a particular user.
- 3) In order to obtain the knowledge of a particular application, a user searches for the data. The understanding of that particular data is totally dependent upon the platform, environment and the information which is currently available.

If the library starts providing the personalized service to the researches then libraries can provide the wide range of diversity and serviced to its users resulting in their time saving [41]. Due to lack of personalised service within the library a user need to spend its most of the time in searching the data which is relevant to him. These hit and trial of data search generated frustration in the users and pessimistic thinking about the library.

As traditional library need advancement due to increase demand of user centric data, it has become very important to deal with the latest technology like Big Data. Although digital library is facing slow research in this field by experts as well as by individuals, but it is the area where advancement is much needed.

Many issue of the traditional digital library can be resolved by using technologies of Big Data.

Characteristic of Academic Data which need to be handled by Big Data

- 1) Authors name and the format to mention them.
- 2) Citations of the Research article
- 3) Subject Area
- 4) Same name in literature as in publications
- 5) Different Formats
- 6) Copyright Protection
- 7) Property Rights

6.4 Analysis on the process and data of Digital Library

Implementation of advanced technology in digital library is always a point of discussion for scholars. As per technical concept, scholars has suggested to use advanced technologies like cloud computing and service oriented architecture



because the new libraries will have media data as well which will need management of the resources, management of the business capabilities [42].

According to some other scholars, digital library should consider three aspects before transforming by Big Data technologies:

- 1) Resource Collection
- 2) Application of the Technology
- 3) Services of the Library

According to more scholars, the change in digital library is dependent upon five aspects:-

- 1) Library to manage big data due to huge amount of data flow.
- 2) Library should also deal with data centric environment due to scientific research method.
- 3) Library should also take care of the Business development model.
- 4) Libraries are transforming from data-centric approach to personalised approach, so library should have all the knowledge resources.
- 5) Library should be capable of dealing with change in information technology [43].

VII.CONCLUSION AND FUTURE ACTION

7.1 Conclusion

In Digital Libraries, demand of the services which are user specific is increasing day by day. Researches are mostly interested in the data which is relevant to their area of interest and they do not want to waste their time to dig into the general data. Due to increase in such demands of the users it has become very necessary to transform the library with latest technologies and tools in order to keep the existence of digital library.

Relevant data collection from the huge amount of data is a difficult task, it need the understanding of the data which has been collected by the library. Traditional database management systems (RDMS) does not has the capability to deal with such huge and diverse data, therefore, it is important to migrate from traditional data management system to databases which can be easily handled by Big Data technologies.

Due to rapid change in the demand from the library and users requirement, it is clear that future is of data and libraries have to adapt themselves to deal with such kind of data as well.

7.2 Future Action

The main focus of this research is to deal with the data which is produced by the users in the digital library and to extract the relevant data from the huge amount of data in order to provide the user interest specific references and materials for their research.

In future, we will work upon gathering the data by using following technology:-

- 1) **Artificial Intelligence** – It will help to search the relevant words used by the users for fetching the similar data over wide range of internet.
 - 2) **Spark Streaming** – It will help to collect the real time data from the users and to convert them in top searches of a particular subject.
 - 3) **Big Data and Cloud** – We can import and export the related data from the cloud quite easily.
- Digital Libraries has huge scope in near future for its integration and modification.

7.3 Limitations

Implementation of new technologies like Big Data in Digital Library is still facing some issue due to following reasons :-

- 1) Digital Library data is not considered as big data by some researchers.
- 2) Library is not in the top list of the researchers for implementing such technologies.
- 3) Libraries are facing some budget issues.
- 4) Librarian also has to upgrade themselves from using traditional way to using a system with latest technologies.
- 5) Information Security is also a constraint according to some researchers. For example, private information can be at a risk if library will start tracking the user's data for analysis purpose.
- 6) Transferring Legacy data in print format into digital format will include huge cost.



REFERENCES

- [1] NMC (2017). NMC horizon report 2017 library edition. Retrieved from <http://cdn.nmc.org/media/2017-nmc-horizon-report-library-EN.pdf>.
- [2] Coelho, H. S. (2011). Web 2.0 in academic libraries in Portuguese public universities: A longitudinal study. *Libri*, 61(4), 249–257.
- [3] Chen, C. F., Qian, O., & Dai, Y. Z. (2014). Study on the construction of digital library in the age of big data. *Library and Information Service*, 58(7), 40–45.
- [4] Brandon, B. (2013). Talk about big data: How the library of congress can index all 170 billion tweets ever posted. Retrieved from <https://www.networkworld.com/article/2162603/uc-voip>.
- [5] M. Khabza and C. L. Giles. The number of scholarly documents on the public web. *PloS one*, 9(5):e93949, 2014.
- [6] The Science of Science Policy: A Federal Research Roadmap. Report on the Science of Science Policy. National Science and Technology Council, 2008.
- [7] Xu, S., Du, W., Wang, C., & Liu, D. (2017). The library big data research: Status and directions. 5(3), 77–88.
- [8] DLPP (2018). About digital library promotion project. Retrieved from http://www.ndlib.cn/szyjs2012/201201/t20120113_57990.htm.
- [9] Laney, D., 2001, 3-D data management: Controlling data volume, velocity and variety, META Research Note, February 6.
- [10] Zikopoulos, P., Eaton, C., DeRoos, D., Deutsch, T. & Lapis, G., 2012, Understanding Big Data. The McGraw Hill.
- [11] J. Manyika et al., Big Data : The next frontier for innovation, competition, and productivity , McKinsey Global Institute 2011 (available at http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation accessed December 2013).
- [12] [DeWitt 1992] D. DeWitt, J. Gray, Parallel database systems: the future of high performance database systems in *Communications of the ACM* n. 35.6 (1992): 85-98 (available at <ftp://ftp.cs.wisc.edu/pub/techreports/1992/TR1079.pdf> accessed December 2013).
- [13] E. Codd, Extending the database relational model to capture more meaning in *ACM Transactions on Database Systems (TODS)* 4.4 (1979): 397 434.
- [14] D. Abadi, P. Boncz, S. Harizopoulos, Column oriented database systems *Proceedings of the VLDB Endowment* 2.2 (2009): 1664 1665.
- [15] R. Cattell, Scalable SQL and NoSQL data stores in *ACM SIGMOD Record* 39.4 (2011): 12 27 (available at http://www.sigmod.org/publications/sigmod_record/1012/pdfs/04_surveys.cattell.pdf accessed December 2013).
- [16] M. Seeger, Ultra Large Sites, Key Value stores: a practical overview in *Computer Science and Media*, 2009 (available at http://blog.marcseeger.de/assets/papers/Ultra_Large_Sites_SS09_Seeger_Key_Value_Stores.pdf accessed December 2013).
- [17] J. Schindler, I/O characteristics of NoSQL databases in *Proceedings of the VLDB Endowment* 5.12 (2012): 2020 2021 (available at http://www.vldb.org/pvldb/vol5/p2020_jirischindler_vldb2012.pdf accessed December 2013).
- [18] R. Angles, C. Gutierrez, Survey of graph database models in *ACM Computing Surveys (CSUR)* 40.1 (2008): 1. (available at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.1072&rep=rep1&type=pdf> accessed December 2013).
- [19] R. Khare, D. Cutting, K. Sitaker, A. Rifkin, Nutch: A flexible and scalable open source web search engine, Oregon State University, 32 (available at http://www.master.netseven.it/files/262_Nutch.pdf accessed December 2013).
- [20] S. Ghemawat, H. Gobioff, S. Leung, The Google file system, *ACM SIGOPS Operating Systems Review* . Vol. 37.No. 5. ACM, 2003 (available at <http://static.googleusercontent.com/media/research.google.com/it/archive/gfssosp2003.pdf> accessed December 2013).
- [21] T. White, Hadoop: the definitive guide, O'Reilly, 2012, ISBN 978 1 44931152-0
- [22] Lu, J. (2014). The evaluation and prospects of big data research in library in China. *Library Journal*, 33(1), 20–25.
- [23] ProQuest (2013). Data mining big data: A strategy for improving library discovery. Retrieved from <http://www.proquest.com/blog/2013/241902501.html>.
- [24] Horstmann, W., & Witt, M. (2017). Libraries tackle the challenge of research data management. *IFLA Journal*, 43(1), 3–4.
- [25] Wang, X. Y. (2015). Analysis of and reflection on the application of big data in university libraries. *Journal of Academic Library and Information Science*, 33(1), 55–60.



- [26] Lee, J. M. (2013). Understanding big data and utilizing its analysis into library and information services. *Journal of the Korean BIBLIA Society for Library and Information Science*, 24(4), 53–73.
- [27] Chen, C. F., Qian, O., & Dai, Y. Z. (2014). Study on the construction of digital library in the age of big data. *Library and Information Service*, 58(7), 40–45.
- [28] Boulton, G. (2014, December). Open data and the future of science. *Septentrio conference series* (no. 1).
- [29] 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.
- [30] Lancaster, F. W. (1982). *Libraries and librarians in an age of electronics*. Information Resources Press, 229.
- [31] Chua, A. Y. K., & Goh, D. H. (2010). A study of web 2.0 applications in library websites. *Library & Information Science Research*, 32(3), 203–211.
- [32] Long, M. P., & Schonfeld, R. C. (2018). The Ithaka S+R library survey 2010: insights from U.S. academic library directors. Retrieved from http://wcm-dev-web1.cul.columbia.edu/content/dam/librarywebsecure/behind_the_scenes/assessment/forums/ithakapresentation_assessmentforum_04122011.pdf.
- [33] Li, Z. X. (2002). Discussion on the disappearance of library in the digital age. *Information and Documentation Services*, (1), 41–44.
- [34] Michalko, J., Malpas, C., & Arcolio, A. (2010). Research libraries, risk and systemic change. *OCLC Research*.
- [35] Coelho, H. S. (2011). Web 2.0 in academic libraries in Portuguese public universities: A longitudinal study. *Libri*, 61(4), 249–257.
- [36] Wang, X. Y. (2015). Analysis of and reflection on the application of big data in university libraries. *Journal of Academic Library and Information Science*, 33(1), 55–60.
- [37] Gu, L. P. (2010). Research on models of user behavior driven personalized services. *New Technology of Library and Information Service*, 26(10), 1–9.
- [38] Williams, K., Wu, J., Choudhury, S. R., Khabsa, M., & Giles, C. L. (2014). Scholarly big data information extraction and integration in the CiteSeerX digital library. *IEEE, International Conference on Data Engineering Workshops* (pp. 68–73). IEEE.
- [39] Yin, H., & Liu, W. (2013). On the new generation of library service system: Functions, visions and prospects. *Journal of the Library Science in China*, 39(5), 26–33.
- [40] Chen, C. F., Qian, O., & Dai, Y. Z. (2014). Study on the construction of digital library in the age of big data. *Library and Information Service*, 58(7), 40–45.
- [41] Gu, L. P. (2010). Research on models of user behavior driven personalized services. *New Technology of Library and Information Service*, 26(10), 1–9.
- [42] Yin, H., & Liu, W. (2013). On the new generation of library service system: Functions, visions and prospects. *Journal of the Library Science in China*, 39(5), 26–33.
- [43] Chen, C. F., Qian, O., & Dai, Y. Z. (2014). Study on the construction of digital library in the age of big data. *Library and Information Service*, 58(7), 40–45.



INNO  SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details