

DATA MINING IS A PERPETUAL CONCEPT FOR LIBRARY AND INFORMATION SCIENCE: AN ESTIMATED OVERVIEW

Dr. Bedadyuti Sahoo

Library In-Charge

School of Computer Engineering, KIIT University, Bhubaneswar-24, Odisha

E-mail: beda_sahoo@yahoo.co.in

Dr. B. S. P. Mishra

Associate Professor

School of Computer Engineering, KIIT University, Bhubaneswar-24, Odisha

E-mail:mishra.bsp@gmail.com

Abstract: Library and information services are not confined in the present day information scenario. The expectations of the users are increasing taking to the research and development of different disciplines. So the work of the present library professionals is increasing in satisfying the needs of the users and for better utilisation of resources available in the library justifying the huge expenditure on online resources. The combination of the artificial intelligence, data management and the statistical implications will bring the concept into full phased reality. So for this, the real assumption of this topic is how the data mining will be useful to extract knowledge from the data ware houses in the synthesized manner. In addition to this, how it will provide better services and good management to the library is aim of this paper.

Purpose: The purpose of writing this paper is to find out the extraction of data from the data mining and its utilisation in the analytic-synthesis manner. It is required for the benefit of the users and to provide better services providing accurate data in a limited period saving the time of the users in an effective way. Storage of data using artificial intelligence and databases for the benefit of the users of the library is the aim of this paper.

Findings: The application of computer technology in collecting and disseminating data has taken a great importance in all fields. The huge growth of unsystematic data and its process for utilising for the organisation requires techniques through which the right kind of data can be used for right purposes. So for this, the concept of data mining is very much essential which has been reflected in this article through various examples, tools and techniques to make the process good and productive. In the same way, this paper has proposed one more models to understand knowledge sharing system so that, facilities among the insiders and the data miners could be correlated.

Key Words: Data Mining, Data Ware House, Data Source, Data Analysis, Data Management, Data Export, Data Filtration

INTRODUCTION

Data Mining is a new concept in the field of Library and Information Science. It is highly required for applying, identifying and utilising the resources for the benefit of the users in a simple and effective manner. It is an analytical process to explore large amount of data with a good relationships between the variables justifying the right findings through the new sub-sub circle data processing. As per the suggestion of the statsoft.com website (<http://www.statsoft.com/>), it can be classified in three parts, namely, Initial exploration, model building and deployment. Model building is the mean for identifying data through verification and the deployment for new data in order to generate predictions.

LITERATURE REVIEW

Taking to the importance and emerging trends of the data analysis and its utilisation in the true sense of the term, it is highly required to find out the importance of the data mining process in the library and information centres. The following social scientists have given their true views on data mining and utilisation and analysis of the data in their perfect manner in the modern day information scenario. According to Dwivedi & Bajpai, (2004), the three components of data mining are artificial intelligence, statistical analysis and database management system through which extraction of data from the data warehouse is possible. In the same way, they overviewed that, the use of data mining in the field of library and information science has some impact on library administration and management in providing services to the users. In this regard, suffixing to the concept, Pal (2011) said that, data mining is a concept to facilitate innovative plans and helps in making decisions. It focuses on information retrieval, analysing unstructured texts, web-usage mining and to take good decisions for the library services. Ann (2013) added some of the important advantage points to this concept that, data mining provides research opportunities and helps in analysing data to strengthen the analysis of a researcher.

So for this, data mining is most important for the library in the field of research and helps in making good decisions for the library professionals to provide services keeping in mind that, the synthesization and the extraction of data from the data warehouses for unsolved things could be possible to the solved decisions.

CONCEPT OF DATA MINING AND THE OBJECTIVES OF THE STUDY

Data Mining is a process generally used by companies for utilising raw data through analysis into useful information (<https://www.google.co.in/>) for the organisation. Weaver in the year 2004 said that, application of data mining in the library leads generation and optimisation. It is also called knowledge discovery which helps in discovering data and analysing them for useful information for developing revenue, cost-cut or both (<http://www.anderson.ucla.edu/>). It also helps in correlating among many fields in large relational databases.

Data Mining performs following tasks like, Association, Classification and Clustering for the perfect management of the data and its maximum utilisation for the organisational

excellence. The Association, Classification and Clustering (ACC) concepts of the data mining are enumerated in the following manner.

Association: It is a defined rule for data mining for finding out correlations, associations and frequent patterns from data sets such as relational (<http://www.techopedia.com/>), transitional and other forms of data repositories. It is generally represented in the format of $X \rightarrow Y$. The association among the elements can be defined on the basis of support and confidence.

Classification: It is required for analysing the input data and to develop an accurate description for each class of features present in the data mining. It is difficult to know the set of supervised classification and the name of the class (www.eecs.wsu.edu/) could be defined after the class is finalised. The unsupervised class is commonly called as clustering. There are two categories of measures for a classification rule they are subjective measures and Objective measures.

Clustering: It is helpful and makes similar objects into one group separating from the dissimilar objects and then the label can be assigned to the group(s). It is adaptable to change and very much helpful to distinguish from the other groups. Clustering is requires for scalability, ability to deal with different kind of attributes (<http://www.tutorialspoint.com/>), discovery of clusters with attribute shape, to deal with noisy data and interpretability etc.

From the above description, it is observed that, the main objectives of the data mining is to help in the field of developing library and information centres keeping in mind the demand of the users and the way of providing right information as and when users need in the following ways.

- To extract data from the data warehouse
- To monitor on data and analyse the data
- To synthesize the data for right information
- To help in making decision making in the field of library services
- To reduce irrelevant data prior delivering to the library customer
- To save the time of the users and attract the users for better utilisation of the resources of the library

Analysing the above concept, it is important to understand the concept of data mining through the architectural framework for the operational and extraction of the data from the data source using computer gadgets for transformation of data to the library clients as a whole. The author is really interested to discover the things relating to the data mining for the purposeful utilisation of data for the customer of the library in the wide spectrum of the research and development comparing other fields of the world.

ARCHITECTURAL FRAMEWORK OF THE DATA MINING

Architectural framework is a process of understanding into a concept and it is required to find out the correlation among the databases and the apparatus used for extracting data for synthesization.

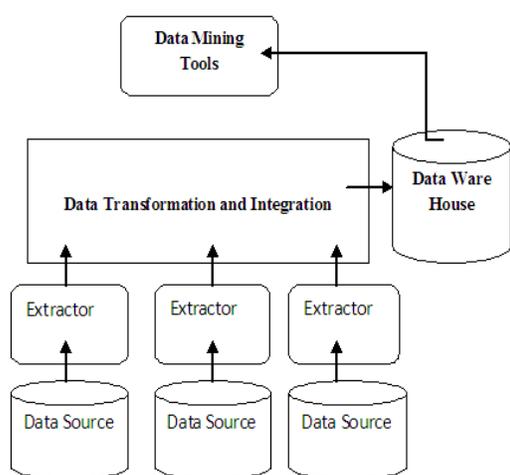


Fig.-1: Mining in a Data Warehouse Based Architecture
Source: Kargupta, Hillol & et al., PP. , 2004

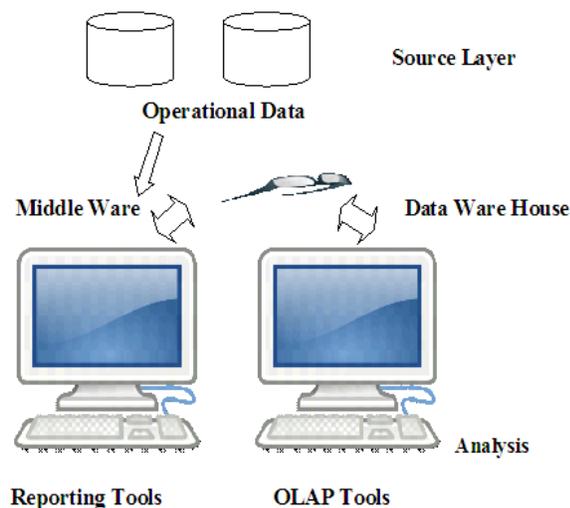


Fig.-2: Single Layer Architecture for a Data Warehouse System; Source: Holferelli, Matteo & Rizzi, Stefano, PP.7

The above Fig.-1 tells about the structure and process of the data mining. Data warehouse is the source of the data and data extraction through the data mining processes using data tools for analysis and interpretation of data into correct information is the aim of the above figure-1. But in the same way, Fig.-2 tells about that, data mining is possible through the different layers through the data filtration process. Otherwise, the conglomeration of data will never allow the library professionals to analyse the data as per the demand of the users. Another thing is that, the storage capacity of the data depends upon the capacity of the storage devices. Generally the capacity should be in the form of gigabyte or terabyte. The client server and the PC platform will support the system for the establishment of the data mining. The more powerful system will help in the query processes at a time. The application techniques sometimes differ in the installation processes among the computer hardware professionals. Some of them use extensive indexing capabilities to improve query performance (<http://www.anderson.ucla.edu/>). Also some of them use massively parallel processors (MPP) to archive order of magnitude improvements in the query time.

DATA MINING METHODOLOGIES AND ITS EFFECTIVE USEFULNESS IN THE LIBRARY

Generally, the application of theoretical methods to the field study associated with the branch of knowledge (en.wikipedia.org/wiki/Methodology) for effective and usefulness of the data

utilisation. It is the main aim of the application of the methodologies. It is mainly required to extract fine data from the complex data for the effective use of the users for divergent subjects. So here, the application of different methodologies is to be made for the benefit of the users through the data management operators. Methodology may or may not be based on one theoretical aspect to attain an objective. As the author understands about the methodology on data mining in the field of library and information centres, the following steps are to be taken into account for proper data mining processes.

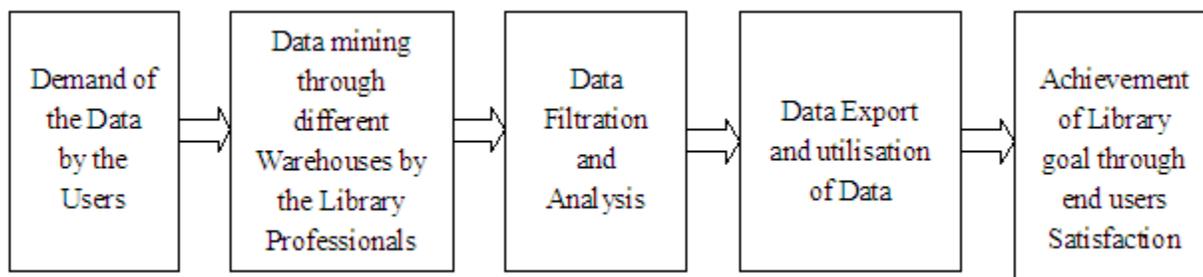


Fig.-3: Data Mining Mechanism to satisfy the end users in the Library

To understand the data mining in the library, the above figure wants to clarify that, the demand of the library users are the day to day approach for the library professionals. The approach may be done on divergent subjects in the form of top-down or bottom-up approach and the collection of data may not be in the same way to identify the problem of the users. So, it is important for a library professional to identify the problem of the users systematically and effectively. Accordingly, the collection of data could be done through the various warehouses and the duty of the library professionals are to analyse and synthesize the data for making the data useful. This is the responsibilities of each librarian or any library professional satisfying the demand of the users as the aim of establishing the library could be successful in the present day information scenario.

THE KNOWLEDGE DISCOVERY DATABASE (KDD) PROCESS AND THE VIRTUOUS CYCLE OF DATA MINING

It is a periodic development of the computational intelligence applying tools and techniques in extracting useful information from the unsystematic volume of digital data. According to [Tianrui Li](#) (2007), KDD could be used in the field of information security, medical treatment and in the information management processes. In the same way [Zhengxin](#) (1997) said that, it is important to study on knowledge discovery in human-computer symbiosis and system user adversarial partnership. The knowledge sharing and the useful models in helping to share the information and transmission of data from one field to other is a necessary tool to expand the business or providing services in each and every field. According to [Wang & Wang](#) (2008), the useful model in knowledge sharing in the data mining process is a dynamic transformation of explicit and tacit knowledge for business intelligence. The use of models in the field of transformation of knowledge and data are the essential tools for each and every organisational aspect.

OUTLINE STEPS OF THE KDD PROCESS

The main aim of the KDD is to discover useful knowledge from data or different database. This is possible mainly through the process of separation, evaluation and interpretation. The other patterns of discovering knowledge from data or different database is choice of encoding schemes (<http://www2.cs.uregina.ca/>), sampling, pre-processing and projections etc applying algorithms Prior to the data mining in the KDD process.

The following figure tells about the synthesization of the data following different steps into knowledge.

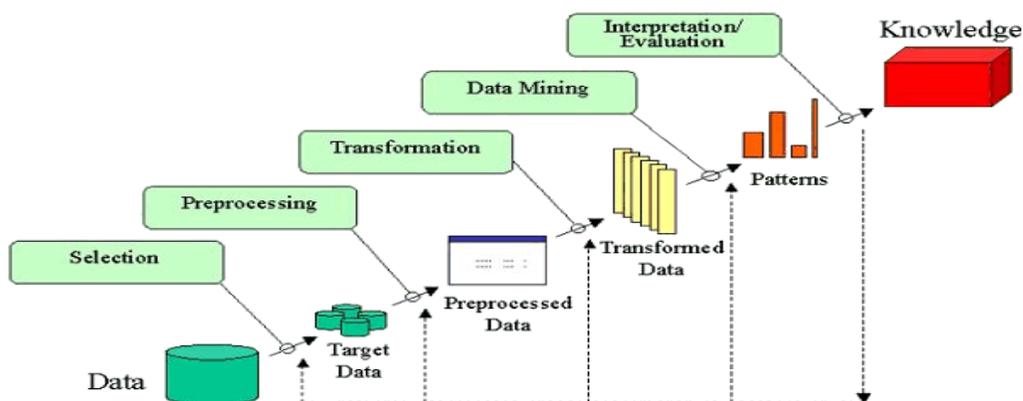


Fig.-4: Outline Steps of the KDD Process; Source: <http://www2.cs.uregina.ca/>

Data Mining and discovery of knowledge are two separate entities. The above figure tells about the different steps of data mining for discovering knowledge. First the data need to be separated from the Jung data and for that sake, different patterns of evaluation and interpretation are required which are enumerated in the above figure. The last step of the process is interpretation and evaluation of data through which knowledge discovery is possible from the whole gamut of the data mining process.

FUTURE PROSPECTIVE OF THE LIBRARY THROUGH DATA MINING

According to Mishra & Mishra (2013), data, model, analysis, improvement and control are the five major aspects of the data mining to process, retrieve, implement and execute the data for the benefit of the organisational excellence. As library is the store house of knowledge and the place for disseminating data to the users, it is most important to make data mining in the proper way and utilise it for the benefit of the different kinds of users in different aspects. The compilation and synthesization of data using appropriate software is most important as analytical software are easily available in the market for data analysing and statistical interference. In the same way, it is most important to give emphasis on time factor through which the users can save their precious time through machine learning and neural networks. Data mining in the field of circulation of data semester wise as most of the colleges run in the semester system, the variance tendency of the monthly circulation and the knowledge about

the issue of books month-wise could be known. In this regard, Uppal & Chindwani (2013) said in his article that, the ratio of issuing books in the library at regular time series would be pasteurised through which the act of performance of the librarian for purchasing books and other materials and enhancement of resources would be done systematically. The other functionalities of the library like, students' ratio, intake statistics year-wise, reference and circulation of books ratio could be defined through data mining in a systematic and better management way.

LIMITATIONS IN ADOPTING DATA MINING

In the excellence service point of view, the advancement of technology has given a good benefit and security to the data mining because it enables cooperation to minimise risk and increase profit both in the private and public sectors. However, it increases the privacy of the consumer and misuses the personal information. It is a question of legality to check the corruption of personal information and still a problem to stop in making rules because, the social networking sites like, blogs and twitters etc are globally accessible. As data mining techniques are not cent-percent accurate, it is difficult to check data smuggling and people's interest. Universal acceptance of data format and the government interference in controlling data warehouses with certain standard will give the system security and self actualisation. If the safe system of the data mining would not be implemented, it would be a difficult part to protect the library bibliomining (Pal, 2011) in to a complete form. Thus, it is essential to take steps predicting the library situation otherwise the data mining of the library would go in a non-profit way.

CONCLUSION

In the concluding observation, it can be said that, the day to day importance of data in each and every field is growing in a galloping rate. The mechanism to synthesize and find out for utilising data in the field of management, business, medical science treatment and other fields in any organisation is a most important thing. It is further requires that, data mining must be integrated with the knowledge management for the improvement of the knowledge in the organisation. The involvement of knowledge workers will make the data mining more relevant and value added. It is also understood that, the important aspects of data mining is to share the knowledge and connect through planning to improve the data organisation for the organisational excellence and actualisation.

REFERENCES

1. Chen, Z. 1997. "Knowledge discovery and system-user adversary". *Kybernetes* 26 (4): 425-441.
2. Dwivedi, R. K., R. P. Bajpai 2004. "Use of Data Mining in the field of Library and Information Science: An Overview". 512-518. Retrieved on March 12, 2015. http://ir.inflibnet.ac.in/bitstream/1944/375/1/04cali_67.pdf
3. Hai Wang, S. W. 2008. "A knowledge management approach to data mining process for business intelligence". *Industrial Management & Data Systems* 108 (5): 622-634.

4. Mishra, R. N., A. Mishra 2013. "Relevance of data mining in digital library", *International Journal of Future Computer and Communication*. 2 (1): 10-14.
5. Okerson, A. 2013. "Text & data mining-A librarian overview" *IFLA WLIC*. Singapore. Accessed on April 15, 2015. <http://library.ifla.org/252/1/165-okerson-en.pdf>.
6. Pal, J. K. 2011. "Usefulness and applications of data mining in extracting information from different perspectives". *Annals of Library and Information Studies*. 58: 7-16.
7. Tianrui Li, D. R. 2007. "An extended process model of knowledge discovery in database". *Journal of Enterprise Information Management*. 20 (2):169-177.
8. Uppal, V., G. Chindwani 2013. "An empirical study of application of data mining techniques in library system". *International Journal of Computer Applications*. 74 (11):42-46.
9. Weaver, D. C. 2004. "Applying data mining techniques to library design, lead generation and lead optimization". *Current Opinion in Chemical Biology*. 8 (3):264-270.

NOTES

- en.wikipedia.org/wiki/Methodology
- http://ir.inflibnet.ac.in/bitstream/1944/375/1/04cali_67.pdf
- <http://library.ifla.org/252/1/165-okerson-en.pdf>
- <http://nopr.niscair.res.in/bitstream/123456789/11552/1/ALIS%2058%281%29%207-16.pdf>
- <http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm>
- <http://www.statsoft.com/Textbook/Data-Mining-Techniques>
- <http://www.techopedia.com/definition/30306/association-rule-mining>
- http://www.tutorialspoint.com/data_mining/dm_cluster_analysis.htm
- http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html
- http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html
- https://www.google.co.in/?gws_rd=ssl#q=what+do+you+mean+by+data+mining
- www.eecs.wsu.edu/~cook/dm/presentations/chandra.ppt