

# Increasing Performance of Knowledge Discovery of Databases through Digital Curation

A.Elezabeth Dyana

Network Manager, Holy Cross College Library,  
Holy Cross College (Autonomous), Tiruchirppalli

## Abstract:

Digital data and technologies have become the integral part of the organizations in the today's world. The major part of the digitalization of organization consists of collecting the information regarding the every aspect of the entire organization. Looking up the information and socializing the communications between the information are the important activities of the digitalized organization. That is why creating the knowledge discovery of databases is essential to achieve the good optimization and needs of the user and customization capabilities in the organizations. This paper proposes and addresses the concepts related to build the knowledge discovery of databases through the digital curation policies including the traditional KDD processes for the benefit of the organizations in order to perform the tasks related to the increasing demands and adds more functionality and reliability to the digitalized databases and tries to provide a new view for doing the knowledge management decision support systems

## Keywords

digital curation, knowledge discovery of databases, data mining, data warehousing, digital repository

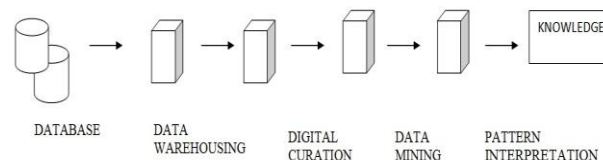
## Introduction

Today's organizations use the knowledge discovery databases to collect and store the data and get the required useful information from the resources. Knowledge discovery of database creation involves finding the knowledge in data and presents high-level of extraction of useful information from the raw data. This paper introduces the digital curation technologies in the ongoing process of the development of the knowledge discovery of databases. The management of knowledge resources comprises of the tools such as data warehousing, pattern recognition, data curation and data mining.

This paper presents the detailed study of the significant role of the digital curation in the KDD

processes and organizes the detail architecture of knowledge management systems. It also guides to build the knowledge discovery of database includes the set of processes and interpretation of patterns to get the high-level of useful information and knowledge from the given data through doing the digital curation in the correct stage of traditional KDD process.

## Proposed steps of the KDD Process



## Data Warehousing

Initially we need to build the digital repository of the organization based on the principles of data warehousing from the database. Then only we find and evaluate the useful knowledge from the repository to do the successful knowledge management systems. It is the overall compile and coordinates the raw data in order to take the knowledge based decisions.

Data warehousing means the central repository for the data in business support systems. It is purely oriented with the decision making concept. It is useful to gain new knowledge from the database in the knowledge decision support systems. It stores not only the data and results but it develops the success of the entire business development databases. The definition of the data warehousing explains “**Data Warehouse** is a subject-oriented, integrated, time-variant, and nonvolatile collection of **data** in support of **management's** decision-making process.” From the above definition of data warehousing we know that data warehouse based on the data-driven decision support system.

The tasks involved in data warehousing are as follows:

1. Specification of the user needs
  2. Implementing meta data
- Collection
  - Maintenance

- Deployment
3. Recognizing and understanding the existing data
  4. Recognizing the entities
  5. Building organization support system
  6. Outsourcing the data warehousing

## Digital Curation

The second stage of the KDD process is to propose the knowledge of doing digital curation. It is nothing but an establishing the process of selecting, preserving, maintaining, collecting and archiving the digital assets. It improves the quality of the given stored information and data. It is mainly focus on the increasing capacity and capability of the data-driven knowledge management systems. The resulted curated data in repository may be shared for the current and future references of queries of the users in the business organizations. Performing the digital curation in the digitalized data, it automatically increases the value and clarity added to the digital repository for present and future use.

The steps included in digital curation are as follows:

1. Conceptualize the objects
2. Create the needed meta data
3. Access and use the data
4. Appraise and select the digital objects
5. Dispose the digital data not selected
6. Ingest
7. Preservation action
8. Reappraise
9. Store
10. Access and reuse
11. Transform

## Data Mining and pattern evaluation:

The third stage of building knowledge repository is to perform the data mining process and pattern interpretation and evaluation. The third stage of the KDD process involves many data mining tasks which are applying to the database for getting relevant user-needed queries. The definition of data mining is to search for patterns of interest in a particular representational form or a set of such representations as

classification rules or trees, regression, clustering, and so forth. From the definition we know that the data mining truly means to analyze the data and information from the conceptualized data-driven knowledge repository. It also explains the relationship between the experimented data in the knowledge discovery databases.

The steps involved in data mining in knowledge process

1. To develop and understand the resulted database
2. Select and create a data set from the curated database
3. Pre-process and cleanse the selected data set
4. Transferring the data
5. Choose the appropriate the data mining task i.e. ( classification, regression or clustering)
6. Choose the mining algorithm for related data mining task
7. Employ the selected algorithm
8. Evaluate and interpret the resulted mined patterns
9. Discover the knowledge from the evaluated system

## Conclusion:

Building the successful knowledge management systems involves the good collaboration of the data integration, information and knowledge patterns. We present the new aspect of making knowledge discovery and decision based systems to provide better accurate results to queries from the user needs. The proposed view of the knowledge discovery database is useful to get and achieve better quality and optimized patterns from the digitalized repository for the present and future use. Thus the study enhances the process of getting valid, potential, ultimately reliable ,and understandable patterns in digital data in knowledge discovery in digital databases.

## References:

1. [http://www.dfki.uni-kl.de/~aabecker/Freiburg/Final/Erdmann/dwh\\_km.doc.html](http://www.dfki.uni-kl.de/~aabecker/Freiburg/Final/Erdmann/dwh_km.doc.html)
2. <http://www.knowledge-management-tools.net/data-warehousing.html>

3. <https://blog.udemy.com/knowledge-discovery-in-databases/>
4. [https://www.researchgate.net/profile/Larry\\_Kerschberg/publication/2414511\\_Knowledge\\_Management\\_in\\_Heterogeneous\\_Data\\_Warehouse\\_Environments/links/004635298959135856000000.pdf](https://www.researchgate.net/profile/Larry_Kerschberg/publication/2414511_Knowledge_Management_in_Heterogeneous_Data_Warehouse_Environments/links/004635298959135856000000.pdf)
5. <https://arxiv.org/pdf/1108.4041.pdf>
6. <http://onlinelibrary.wiley.com/doi/10.1002/meet.2014.14505101074/pdf>
7. [https://www.google.co.in/url?sa=t&rct=j&q=&e src=s&source=web&cd=17&cad=rja&uact=8&ved=0ahUKEwjRheuPwp7UAhUBJZQKHwvZCjwQFgh5MBA&url=http%3A%2F%2Fdownloads.alcts.ala.org%2Fce%2F03072012\\_Digital\\_Preservation\\_Workflow\\_Slides.ppt&usg=AFQjCNG-eaqGZv0YMI\\_LrA6cyWxD9p8ekg](https://www.google.co.in/url?sa=t&rct=j&q=&e src=s&source=web&cd=17&cad=rja&uact=8&ved=0ahUKEwjRheuPwp7UAhUBJZQKHwvZCjwQFgh5MBA&url=http%3A%2F%2Fdownloads.alcts.ala.org%2Fce%2F03072012_Digital_Preservation_Workflow_Slides.ppt&usg=AFQjCNG-eaqGZv0YMI_LrA6cyWxD9p8ekg)
8. [http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1\\_kdd.html](http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html)