International Journal of Advanced Research in IT and Engineering

ISSN: 2278-6244
Impact Factor: 6.690

DATA WAREHOUSING IN LIBRARIES FOR MANAGING DATABASE

Dr. Kirti Singh, Librarian, SSD Women's Institute of Technology, Bathinda

Abstract: *Major libraries have large collections and circulation. Managing libraries electronically has resulted in the creation and management of large library databases, Library to the students and teachers who are cooperating in this e-learning environment. this system in online library management system, we can use the applications of Data Mining and improve the search results on the basis of previous data searched by the users. We can also analyze the criticality of books so that library could be signaled when there is a need of adding a critical book. We propose a manner to the development of Digital Libraries, using Data Warehousing and Data Mining Techniques. I want to discuss components of the Data Warehousing architecture on which Data Mining techniques can be applied. The library is including a wireless network to a new construction project; the placement of the wireless network components needs to be considered in the libraries design.*

INTRODUCTION

Library information systems, have given libraries the ability to experience the usefulness of information system technology, and to apply this technology for the benefit of librarians and library patrons. The future, and some might even say the present, for the library professional is the digital library. Instead of waiting for the user to come to their information containers in a physical collection, librarians select high-quality materials for users to access through the Internet. It is relatively easy to put a collection of static files online; however, the library is more than just a collection of documents. A crucial part of a library is the human intermediary the librarian. This intermediary connects the users to the information needed, and can assist with advice about using the information retrieval systems and working with information.

DATA WAREHOUSING

A data warehouse is defined as a "subject-oriented, integrated, time variant, non-volatile collection of data that serves as a physical implementation of a decision support data model and stores the information on which an enterprise needs to make strategic decisions. In data warehouses historical, summarized and consolidated data is more important than detailed, individual records. Since data warehouses contain consolidated data, perhaps from

several operational databases, over potentially long periods of time, they tend to be much larger than operational databases.

DATA WAREHOUSE DESIGN

In the data warehouse design, a snowflake schema is used. There are three dimensional tables viz. book schema, issue transaction schema and return transaction schema. The return transaction schema is normalized and we derived a dimension table book dimension to provide the correspondence between book id and book name.A data warehouse is a system that stores and consolidates data periodically from the source systems into a dimensional or normalized data store. It usually keeps years of historical data and can be mined for pattern discovery for business intelligence or other analytical activities.

DATA WAREHOUSING IN DEVELOPMENT OF DIGITAL LIBRARIES

A data warehousing warehouse is defined as a "subject-oriented, integrated, time variant, non-volatile collection of data that serves as a physical implementation of a decision support data model and stores the information on which an enterprise needs to make strategic decisions. In data warehouses historical, summarized and consolidated data is more important than detailed, individual records. Since data warehouses contain consolidated data, perhaps from several operational databases, over potentially long periods of time, they tend to be much larger than operational databases. Most queries on data warehouses are ad hoc and are complex queries that can access millions of records and perform a lot of scans, joins, and aggregates. Due to the complexity query throughput and response times are more important than transaction. Data Mining Data Mining is the extraction or "Mining" of knowledge from a large amount of data or data warehouse. To do this extraction data mining combines artificial intelligence, statistical analysis and database management systems to attempt to pull knowledge form stored data. Data mining is the process of applying intelligent methods to extract data patterns. This is done using the front-end tools. The spreadsheet is still the most compiling front-end application for Online Analytical Processing (OLAP). The challenges in supporting a query environment for OLAP can be crudely summarized as that of supporting spreadsheet operation effectively over large multi-gigabytes databases.

USING A DATA WAREHOUSE IN LIBRARY

The data warehouse environment is more controlled and therefore more reliable for decision support than the previous methods. The data warehouse environment supports the

entire decision support requirements by providing high-quality information, made available by accurate and effective cleaning routines and using consistent and valid data transformation rules and documented pre-summarization of data values. It contains one single source of accurate, reliable information that can be used for analysis.

DATA MINING

Data mining "An information extraction activity whose goal is to discover hidden facts contained in databases. Using a combination of machine learning, statistical analysis, modeling techniques and database technology, data mining finds patterns and subtle relationships in data and infers rules that allow the prediction of future results. Typical applications include market segmentation, customer profiling, fraud detection, evaluation of retail promotions, and credit risk analysis."

DATA MINING LIBRARY DATA

Libraries arrange their items based on pre-defined subject areas. Although books related to the same subject are kept close to each other, related items that would be identified as in the supermarket example may not be together. For instance, the related videos and CD or even magazines or journals would be available elsewhere. Thus the chance of a reader getting to know availability of such items and making use of them is low. Data mining techniques must be applied to data to reveal such information.

Data of a digital library should be organized in a manner in which it could be analyzed later. Classification of data according to identified characteristics is one way of achieving it. Classification will allow grouping of discrete values such as by subject of a book. However, estimation would be required to deal with continuous values such as age groups of borrowers. Classified and estimated data could be used to predict future behaviors and decision making could be done accordingly characteristics could help to determine affinity groups. Services offered could target such groups. For example, reduced membership fees or special facilities may be offered to them and ensure that they are not neglected.

DATA MINING IN LIBRARIES

The applicability of above data mining techniques is highly depended on the size of the database. If the database is small, it is possible to find interesting patterns and relationships by simple inspection of results from familiar tool such as spreadsheets and multidimensional query tools. However, in order to generate classification rules, association rules, clusters, or predictions one require a large amount of data. Since the data of the library continuously

growing with an exponential rate, the main problem of referencing the required information from the large amount of redundant information of the library can be reduced using mining techniques. Searching through classification of content of the library and acquisition of books through data mining knowledge are among the main benefits that a library would gain through the management of large databases and data warehousing.

DATA MINING CONSISTS OF FIVE MAJOR ELEMENTS

1. Extract, transform, and load transaction data onto the data warehouse system.
2. Store and manage the data in a multidimensional database system.
3. Provide data access to business analysts and information technology professionals.
4. Analyze the data by application software.
5. Present the data in a useful format, such as a graph or table.

DATA MINING AND DATA WARE HOUSING

The goal of data warehousing is to support decision making with data. Data mining can be used in conjunction with a data warehousing to help with certain types of decisions. Data mining can be applied to operational database with individual transactions. To make data mining more efficient, the data warehouse should have an aggregated or summarized collection data. Data mining helps in meaningful new pattern that cannot be found necessary processing data or metadata in the data warehouse.

The goal of the Data system is just a data base that provides a single point to access to multiple software systems used by libraries. Ware housing is much larger than other kinds of data bases ranging from several gigabytes to terabytes are common. Typically a data warehouse will centralize a given kind of business data in one storage point to make it easier for management / other systems to process and access this data. The data analysis software is what supports data mining.

CONCLUSION

In this paper, we implemented such data warehouse from the library automation system data.This type of Data Ware Housing and Data Mining tools can support library management system with efficient techniques for decision support in connection with procurement of books and journals etc.A data warehouse is a system that stores and consolidates data periodically from the source systems into a dimensional or normalized data store. we successfully developed the data warehouse which is much applicable to libraries to help produce necessary, required but non readily-available reports. We analyzed the existing

needs of reports and then design the corresponding fact tables and dimension tables of data warehouse. It usually keeps years of historical data and can be mined for pattern discovery for business intelligence or other analytical activities. Libraries must compete against online booksellers, downloadable audio books, and the vast supply of "free" information of varying quality from the Internet, librarians must begin to take the initiative in using their systems and data for competitive advantage and to justify continued support and funding of libraries.

REFERENCES

1. Guenther, K. (2000). Applying data mining principles to library data collection. Computers in Libraries, 20(4), 60-63.
2. Chaudhry, A. S. (1993). Automation systems as tools of use studies and management information. IFLA Journal, vol. 19(4), 397-409.
3. Data Mining http://www.ibm.com/sfasp/locations/milan/index.html. Visited on September 2004.
4. Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals. PaulrajPonniah.Copyright© 2001 John Wiley & Sons, Inc
5. http://www.anderson.ucla.edu/faculty/jason.frand/teacher/ technologies/palace/datamining.htm. Visited on September 2004
6. J. Han and M. Kamber, “Data Mining: Concepts and Techniques,” Morgan Kaufmann, San Fransisco.
7. K Prakash, Prem Chand and UmeshGohel, 2nd Convention PLANNER - 2004, Manipur Uni., Imphal, 4-5 November, 2004 © INFLIBNET Centre, Ahmedabad page. 166-178.
8. Pujari, A. K. (2002), “Data Mining Techniques”, Universities Press, First Edition.
9. Ramez, E.; Shamkant, B.; Navathe, D.; Somayajulu, V.L.N.; Gupta, S.K. (2006). “Fundamentals of Database Systems”, Pearson Education.
10. W. H. Inmon, “Building the Data Warehouse,” Fourth Edition, J. Wiley, New York.