Data Mining in Library and Information Services

Gopala k 1
Asst, Librarian
Dayananda Sagar College of Engineering
Bangalore-560078

Devendrappa T M 2
Professional Assistant
Jawaharlal Nehru University
New Delhi, India.

Harish C K 3
Librarian
Govt, Fist Grade College Barkur, Udupi.

Abstract:

Data mining involves significant process of identifying the extraction of hidden predictive information from vast array of databases and it is an authoritative new technology with potentiality to facilitate the Libraries and Information Centers to focus on the most important information in their data warehouses. Library and information services in schools, colleges, universities, corporations and communities obtain information about their users, circulation history, resources in the collection, and search patterns.

Keywords : Bibliomining, Data Mining, Knowledge Management, Library.

Introduction

The importance of collecting data that reflect in business or scientific activities to achieve competitive advantage is widely recognized now. Powerful systems for collecting data and managing it in large databases are in place in all large and mid-range institutions and organizations.

We are in an age often referred to as the information age. In this information age, because we believe that information leads to power and success, and thanks to sophisticated technologies such as computers, satellites, etc., we have been collecting tremendous amounts of information. Initially, with the advent of computers and means for mass digital storage, we started collecting and storing all sorts of data, counting on the power of computers to help sort through this amalgam of information. Unfortunately, these massive collections of data stored on disparate structures very rapidly became overwhelming. This initial chaos has led to the creation of

structured databases and database management systems (DBMS). The efficient database management systems have been very important assets for management of a large corpus of data and especially for effective and efficient retrieval of particular information from a large collection whenever needed. The proliferation of database management systems has also contributed to recent massive gathering of all sorts of information.

Data mining and concept

Data mining is the practice of automatically searching large stores of data to discover patterns and trends that go beyond simple analysis. Data mining uses sophisticated mathematical algorithms to segment the data and evaluate the probability of future events. Data mining is also known as Knowledge Discovery in Data (KDD).

The key properties of data mining are:

- Automatic discovery of patterns
- Prediction of likely outcomes
- Creation of actionable information
- Focus on large data sets and databases

Data, Information, and Knowledge

Data

Data are any facts, numbers, or text that can be processed by a computer. Today, organizations are accumulating vast and growing amounts of data in different formats and different databases. This includes:

- operational or transactional data such as, sales, cost, inventory, payroll, and accounting
- nonoperational data, such as industry sales, forecast data, and macro economic data
- meta data - data about the data itself, such as logical database design or data dictionary definitions

Information

The patterns, associations, or relationships among all this data can provide information. For example, analysis of retail point of sale transaction data can yield information on which products are selling and when.

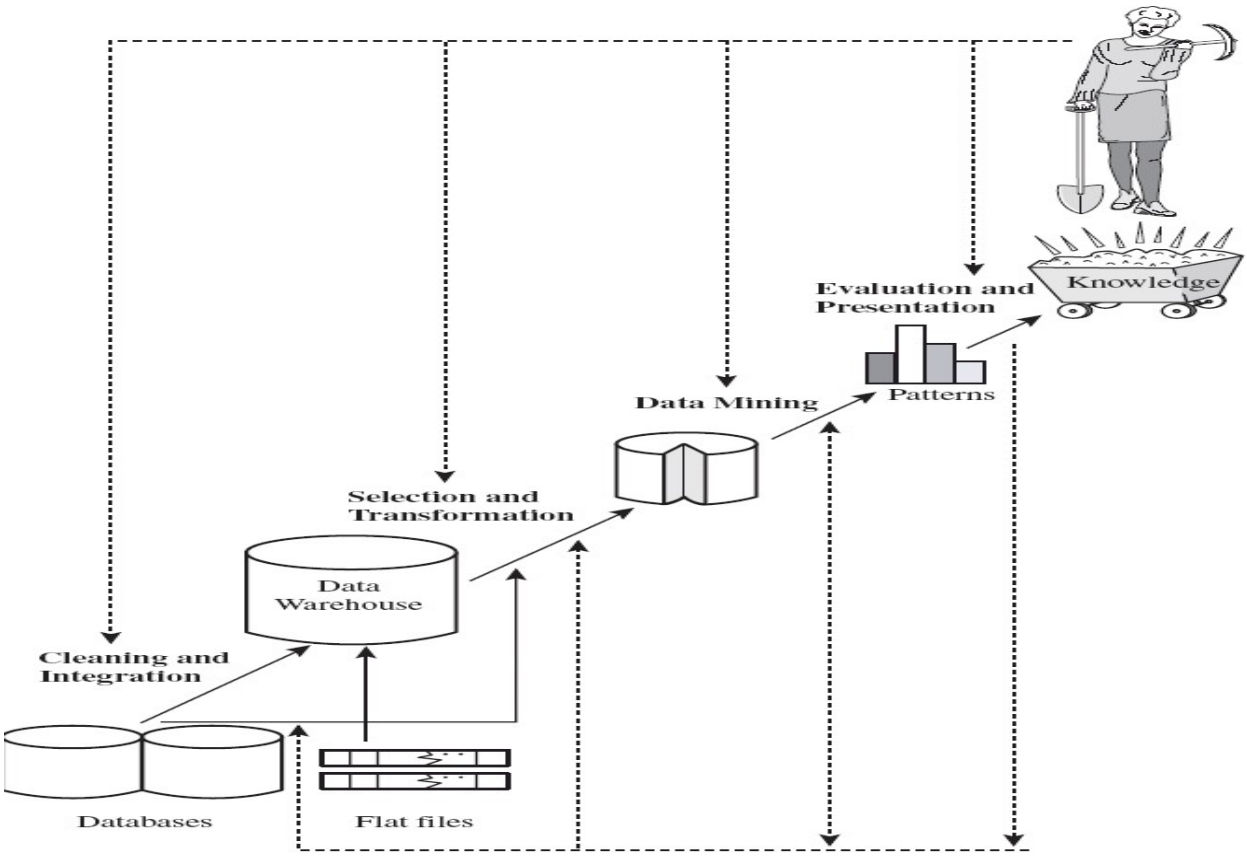
Knowledge

Information can be converted into knowledge about historical patterns and future trends. For example, summary information on retail supermarket sales can be analyzed in light of promotional efforts to provide knowledge of consumer buying behavior. Thus, a manufacturer or retailer could determine which items are most susceptible to promotional efforts.

Data mining and Data Warehouses

Dramatic advances in data capture, processing power, data transmission, and storage capabilities are enabling organizations to integrate their various databases into data warehouses. Data warehousing is defined as a process of centralized data management and retrieval. Data warehousing, like data mining, is a relatively new term although the concept itself has been around for years. Data warehousing represents an ideal vision of maintaining a central repository of all organizational data. Centralization of data is needed to maximize user access and analysis. Dramatic technological advances are making this vision a reality for many companies. And, equally dramatic advances in data analysis software are allowing users to access this data freely. The data analysis software is what supports data mining.

Data mining process



Overview of Library Workflow

Workflow in a traditional "bricks and mortar" library creates a number of data sources appropriate for bibliomining. Before a library obtains new information resources (e.g., books, databases, reference tools, electronic access, etc.), a librarian assesses the needs of the existing collection in light of available and upcoming publications. Next, acquisitions personnel obtain the information resources specified from this needs assessment. Once the library obtains requested new resources, cataloging personnel either create or purchase a catalog record for the new resource. The circulation department then makes the resource available to end-users. Depending upon the size of the library and the scope of its operations, these activities fall within the purview of one, a dozen, or possibly hundreds of different employees organized into specialized departments. The vast data stored in the databases of traditional and digital libraries represent the behavioral patterns of two important constituencies: library staff and library users. In the case of library staff, mining available acquisitions and bibliographic data could provide

important clues to understanding and enhancing the effectiveness of the library's own internal functions. Mining user data for knowledge about what information library users are seeking, whether they find what they need, and whether their questions are answered, could provide critical insights useful in customer relations and knowledge management. These kinds of information can have strategic utility within the larger organization in which the library is situated.

Bibliomining to Improve Library Services

The users of library services are one of the most important constituencies in most library organizations. Most libraries exist to serve the information needs of users, and therefore, understanding those needs is crucial to a library's success. Examining individual users' behaviors may aid in understanding that individual, but it tells librarians very little about the larger audience of users. Examining the behaviors of a large group of users for regular patterns can allow the library to have a better idea of the information needs of their user base, and therefore better customize the library services to meet those needs.

Conclusion

Data Mining is one of the most important parameters in the age of digital era. The need and application of data mining has become essential to manage, organize, and disseminate information to the right users at right time. Though it is primarily intended for the business class, still then it has got practical implications in Libraries and Information Centers due to overwhelming growth of literature especially in digital formats. Now-a-days, and more digital data are being collected, processed, managed and archived in Libraries and Information Centers to suit to the varied need of the user communities every day. Algorithms, software tools, and systems to mine it are critical to a wide variety of problems in all business, science, national engineering, and health care including Libraries and Information Centers.

Reference

Data Mining. http://www.ibm.com/sfasp/locations/milan/index.html.

http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm

Oracle® Data Mining Application Developer's Guide
11*g* Release 1 (11.1)

http://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm

M. S. Chen, J. Han, and P. S. Yu. Data mining: An overview from a database perspective. IEEE Trans. Knowledge and Data Engineering, 8:866-883, 1996.

http://webdocs.cs.ualberta.ca/~zaiane/courses/cmput690/notes/Chapter1/

http://en.wikipedia.org/wiki/Data_mining