International Journal of Engineering &Technology, 7 (1.7) (2018) 51-53

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Data mining in personalized service of digital library

Nivedhitha G *, Rupavathy N

Assistant Professor, Department of Computer Science and Engineering, School of Computing, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai-62, TamilNadu, India
**Corresponding author E-mail: nivedhithagopal25@gmail.com*

Abstract

With the deepening of information engineering, people face increasing amounts of information resources and also the demand of information is more and more. The digital library is a wealth of information resources. Affording readers a richer layer of personalized service is a new objective for the growth of these digital libraries. The data mining techniques to elicit useful information from a bunch of clutter information, in parliamentary procedure to provide efficient technical support for personalized services of digital libraries. The feasibility of information mining technology in digital libraries is analyzed in this paper. It also discusses the information mining technology in the digital library applications and the feasibility analysis for the data mining applications in digital libraries.

Keywords: *Data Mining; Personalized Services; Digital Library; OLAP; DM.*

1. Introduction

In recent years, along with the " digital survival" approach is accepted by people gradually, The digital library has been paid more and more attention with its large information quantity, little space occupation, update speed and without the limitation of time. People enjoy the convenience brought about by the digital library at the same time, also suffered a large and diverse information resources-distressed [1], because it has a strong goal-oriented nature, the same information for different individuals exhibit different values. For individual users, it is not possible to digital library all the information resources are needed, while the same information is also not necessarily satisfy all users. Personalized service is a solution to the user “told " the key of the matter, it is the important means to adapt to the diversifying needs of users, as well as measures to address an important diversification of information resources of the library.

Digital Library personalized information service is based on information the user behavior, habits, preferences, characteristics and specific needs of users, to provide users with information to meet their individual needs and the functionality of the system services. First of all, it should be capable of meeting the needs of individual users of digital library information service, that according to the user explicitly request the provision of information services, or through the analysis of user habits and active information you may need, to provide users of its services; secondly, it should be service that developing one's personality and leading the demand for, it can help individuals to develop their personality, guide to demand found and promote social diversity and diversification of development.

2. The overview of data mining

2.1. The connotation of data mining

Now people is surrounded by ubiquitous information, and it has become more and more difficult to obtain useful knowledge from a large amount of information from the amount of increasing information. This is called by John Nethbert "information-rich and knowledge-poor" [4] dilemma. In order to solve this problem, a new technology named data mining came into being in the early 1990 of the 20th century. Data mining is a drawn from a large amount of data model of technology; knowledge can be discovered in database through data mining technology. Data mining is both a knowledge acquiring technology and a data processing. Data mining process can be divided into three stages: expression of data preparation, extraction operations, results and interpretation. During excavation is a repeated refining process as a whole, without user intervention. Data mining is related to mathematical statistics, databases, artificial intelligence, machine learning, visualization, parallel computing, artificial neural networks, such as the interdisciplinary, it is currently on the international database and decision support in the areas of research direction of the most forward position. Since then, many domestic and foreign scholars and company are devoted to research on data mining and the corresponding technologies and software development has made a lot of progress.

2.2. Present situation of study on application of data mining in library

Since in the 1990 of the 20th century, there are a lot of paper about data mining theory in abroad, especially in the United States and Europe, the application of data mining in library has very good research, development and application. The study on data mining in library is basically to specific library to research in abroad. Scott Nieholson doctor in “Gaining Strategic Advantage

Copyright © 2018 Nivedhitha G, Rupavathy N. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

through Biblio-mining for Management decisions in Corporate, Special, Digital, and Traditional Libraries" paper talked Syracuse University Library as an example about using literature mining to help traditional library management. Because in the United States and Europe data mining has been in widespread used in the library, so foreign studies are more practical and more mature than domestic.

Many foreign libraries have been successfully applying the data mining technology to the work and management. These libraries are not only to the collection of traditional institutions, but also become the information institutions and these library staff also no longer is in the traditional significance librarian, they develop into the information expert.

In domestic, the researches in these areas are mainly concentrated in the colleges and universities for scientific research personnel. In the library, data mining application in the library system is far from enough. But, the library is extremely similar by the information transmission primarily service way and the market marketing pattern in the very great degree, library reader behavior by collecting information and analysis processing, research library group or individual readers borrow some interest, habit and requirement, can predict the possible future borrowing behavior, to a large extent library can raise the information service quality. Generally speaking, the data mining technology in China's library research is only just starting and stays basically in the aspects of theoretical research, fragmentation, not mature enough and no overall strength. For example in domestic there have no book about data mining applications of library, there only have a few of copyright in digital library related to data mining technology. This shows that data mining is along with the generation of the digital library and being brought into the library, so the domestic research in this area is still very scarce.

3. System structure

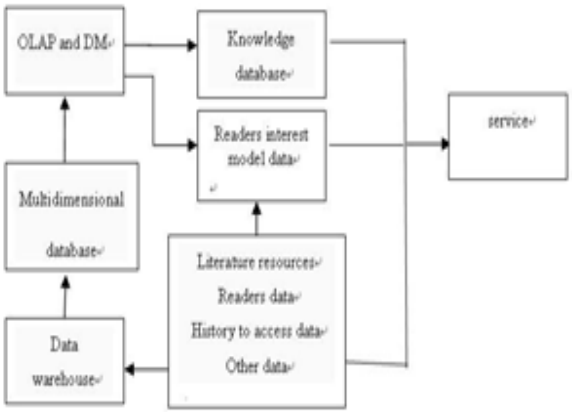


Fig. 3.1: System Structure.

Figure 3.1 shows the design of the digital library personalized service system structure.

In this system, we consider using the data warehouse, data mining and on-line analytical processing technology to carry on the analysis and mining, to readers, resources and readers access to resources such as data, and find the readers' interest and resources association, developing the individualized service for readers. The design of the digital library personalized service system structure as shown.

In the design of personalized service application system structure [5] is the application of data warehouse, OLAP and data mining technology based on the history data and current data analysis and mining provide personalized service for readers. It mainly includes the following aspects:

1) Integration of heterogeneous documents resources. As time goes on , the digital library digital information resources and the number of elements in a continuous increase. But it is not convenient to reader acquire information , because all kinds of information resources in a heterogeneous hardware and software platform of the environment, to the detriment of the. Therefore, the need for readers with unified retrieval platform, and convenience of the reader retrieval.
2) Data warehouse establishment. In the on-line analytical processing（OLAP）and data mining（DM） before, the data organization in data warehouse, according to the analysis of the different theme, the data are then organizations, such as the reader data, resource data, to access data, etc.
3) Data cube model establishment. The data in the data warehouse according to the star type structure is organized into data cube form, OLAP and DM to prepare.
4) OLAP and DM. OLAP analysis is mainly based on multidimensional organization of data section, cut into parts, ground up polymerization, drilled, and turning movement analysis, in order to analyze data, so that we can from a variety of dimensions and multiple side, a variety of data integrated degree view readers the visit to resources. On the basis of multidimensional data cube, through the data mining found the reader's interest, access mode and resource association, etc., formation rule library and readers interested in library.
5) Web personalized service. The module is dealing directly with readers; it is to point to heterogeneous literature resources one-stop personalized network service. It is according to the reader's interest, the different relationship between literature and the correlation between disciplines, as a model for carrying out the information resource service for readers.

4. Application of data mining in digital library feasibility analysis

With the development of network and database technology the rapid development and wide application, our library is to change the direction of informationization and digitalization. In recent years, our country has increased the investment strength to the library, the library database and storage device has been expanded and expansion, network and server also get further popularization and promotion. In under this kind of automated network environment, the use of data mining technology can be better for the digital library resources for better reorganization, collection, extraction and prediction; the use of data mining technology can be more convenient and more quickly to get data and transformation from the Internet, so the data mining technology can provide digital library development to provide better decision support and to provide readers with better service.

4.1. Information demand played traction

In the informationization society, digital library better survival and development and the application of advanced technologies are inseparable. Looked from the last period of Library Theory and practice, digital library has been spare no effort to chase the information and the development of computer technology. Information technology not only has greatly promoted the modernization of libraries, but also brings the information explosion. In the era of knowledge economy, solved the massive information storage development and utilization, is related to the digital library in the future survival and development of the major issues. Using database technology to realize the mass data storage and utilization in digital library, supporting various levels of scientific decision-making services, achieving efficient information industry cooperation mode is the common demand from information of the external pressure and internal mechanism of the development of digital library.

4.2. Material base formation

On one hand, our country's library system passes through many year automated construction, has had the suitable physical condi-

tions [4] and the reserve of capable people, and accumulated the mass data, has laid certain material base for the database application. On the other hand, the digital library is the development of China's informationization [2] construction important constituent; countries gave high attention and provided a number of policies and financial support, the country has provided the good policy environment and the economical safeguard for the professional database implementation and the data mining development application.

4.3. Realization of technology maturity

Database after years of development has already formed a relatively mature technology system, especially in the design of database; data extraction and on-line analytical processing technology have made satisfactory progress, for database application technical foundation. In addition, database technology developed in the telecommunication, manufacturing, retail, finance and other fields has deep level application and achieve great reward. These examples of successful applications provide us with precious referential experiences.

5. Based on data mining of the digital library personalized service system construction

Digital library as an information resource collection and processing service center has accumulated a wealth of information resources. Use data mining can extract and find hidden in the information from digital library large database, data warehouse and the vast network information space, the purpose is to help information workers looking for potential correlation [3] between data, found that the neglected factors .The premise of system construction must have a large number of real data accumulations, so we have to do the construction of basic database.

1) Extracting the original information and collecting user identity [6]. User's access to digital library, the system recorded user information and user behavior characteristics. User information including the user's name, sex, age, occupation, hobbies, education degree and user access IP address, etc.; Using Web log can obtain the user's behavior characteristics, such as their knowledge of the points of click rate, residence time, visits, download times, search keywords and model of information, and the user's subjective information, such as network investigation [7], BBS message, etc. An accurate grasp of the user's behavior characteristics and preference is to provide more accurate and more in line with the needs of the user information service's first condition.
2) Data preprocessing and data conversion. The collected data were processed and issued for remodeling, such as checking data integrity and data consistency, removal of noise or remove the invalid data, fill the lost domain, remove the blank data domain, the time sequence and the data changes; to find data characteristics, with the dimension transform[8] or conversion method to reduce effectively the number of variables data found, invariant, constructing topics related to data warehouse, as the next step in the process of data mining provides a basic platform, be prepared.
3) Determine the target data mining. Data mining goals should be combined with the actual situation, to refine and clear. For example, according to the constant change of the user's interest, make use of the data mining technology found that the user's latest need, or according to the user's interest degree, recommend related project information, and provide personalized interface, etc.
4) Data mining. According to the mining characteristics of target and data choose corresponding algorithm in purification and convert the data set on the data mining. Choosing a particular data mining algorithm is used to search the data model of data mining and search, or develop a particular interested mode or a particular data set; we should be on the basis of the analysis and evaluation and inspect the knowledge model of data mining.
5) Analysis and application of knowledge. The results of data mining for interpretation and evaluation, with the advance, reliable knowledge check and solve data mode possible contradictions, to improve data warehouse, into will eventually be users understand knowledge, will analyze the knowledge loading to the actual operation system. Data mining process is a repeated process, it is difficult to one pace reaches the designated position, need the mining results practice application, test, modification, comparison, until the last to achieve customer satisfaction.

6. Conclusion

In this paper, Data warehouse for different theme data integration, through the OLAP and DM to readers are classified, found associated resources, and to realize the readers based on the reader's interest model, this model through the interaction with readers, constantly receiving the reader's interest and speculated that the reader's interest, the accumulation of readers' information demand preference, realize adaptive personalized service.

References

[1] Natalia Kucirkova & Teresa ,"Personalised reading for pleasure with digital libraries: towards a pedagogy of practice and design" , CreminPages 1-19 | 2017, Published online: 25 Sep 2017

[2] Zhang, F., Gao, Z. & Ye, "Services in Digital Library Based on Cloud Computing Data Processing Technology", Aut. Control Comp. Sci. (2015) 49: 373. https://doi.org/10.3103/S0146411615060127.

[3] Xiong Yongjun, Chen Chunying, "Personalized push based on association mining technology of digital library services," library and information work, vol.54, pp.125-129, January 2010.

[4] Kang Xiao,"Data warehouse technology and middle and small scale library construction ponder," management observation,15, pp.23-26,May 2008.

[5] Zhao Yan,"Application of data mining in the digital library," Shanxi library journal,1, pp.32-33,39 January 2009.

[6] Wang Yuzhen, "Web applications of data mining in personalized services of digital library,"3, pp.54-56, May, 2010.

[7] Wu Xudong, Liu Bingxiang, "Application of clustering analysis in the management of University Library," vol.24, pp.15-16, September, 2011.

[8] Cai Wanyan, "Application of data mining in the evaluation system of University library service,"6, pp.99-101, June, 2011.