On Knowledge Excavation Based on

University Library Knowledge Base

Ying Song

School of Library, Tianjin Polytechnic University

3-3-501 yang guang apartment, bin shui road, Nan kai district, Tianjin 300381, China

E-mail: tjban@163.com

Abstract

Knowledge excavation is a new information processing technology. Nowadays, knowledge excavation has been widely applied to the university library management. Knowledge excavation in university library knowledge bases has already been playing an important role in improving the management of university libraries and inevitably will play a more profound role in the future. In this paper, the meaning of knowledge excavation in university library knowledge bases and other related issues are studied on the basis of introducing university library knowledge bases and knowledge excavation.

Keywords: Digital university library, Knowledge excavation, Knowledge base, Data mining

1. Introduction

Knowledge excavation origins from an extreme growth on the data amount in a worldwide databank. People's demands require more than simple searching and maintenance. They want to make predictions on the overall character of the data and the development tendency on the basis of a higher-level processing and analysis of them. If there is no effective solution for the computer and information technology to obtain useful information and knowledge, people will be at a loss towards the extremely growing amount of information data. Accordingly, they will feel information poor and data in jail compared with data surplus and information explosion. People need a kind of technology to make refined of a so large number of data. Therefore, KDD and its nuclear technology are born. Knowledge excavation is a process to recognize effective and potentially useful and understandable model from various data. Model, a formula of language, can be used to describe a subset of a data set. The so-called knowledge is a more abstract description of information in data. Intelligent and automatic, the process of analyzing a number of data includes data preparation, model searching, knowledge evaluation and repeatedly revising and refining. Efficiency refers that the model maintains certain credibility on new data. The model discovered should be new and the knowledge discovered is about to be effectively operative in the future. The economic efficiency can be increased if the potential utility is applied in decision-support system. The ultimate understandability, which is shown by simplicity, requires users' understanding of the model discovered. The combination of utility, novelty, potential utility and ultimate understandability can be called interest.

In spite of the ten year' history of knowledge excavation, it has been applied into many fields and achieves good results. These fields include scientific research, marketing, financial investment, cheating inspection, production, telecom network management and library management, etc. Explored by Jet Propulsion Laboratory of California Institute of Technology and astronomer, SKICAT (Sky Image Cataloging and Analysis is Tool) is the first successful application of data mining, which has helped the scientists discover 16 distant star-like substances.

Nowadays, knowledge excavation has been widely applied to the university library management. Greatly meeting the university library users' requirements of individuation and tacit knowledge, knowledge excavation really can support the all-round development strategy of explicit knowledge and tacit knowledge in university library knowledge management system, because it makes use of the latest intelligence tools to excavate the tacit knowledge reserved in the massive dominant information and forms the professional knowledge bases to make tacit knowledge explicit and manages it.

2. Knowledge Excavation and Data Mining

Data mining cannot be neglected when referring to knowledge excavation. The theoreticians make no distinct difference

between these two technologies. In a certain sense, these two belongs to one concept and have their own characters.

Data mining is the extraction of information or knowledge interested by way of the utilization of a series of technology and from large scale databank or material warehouse. The information or knowledge, implied and potentially useful, performs as concepts, rules and patterns. With the increasingly expansion of the data amount in the databank, the knowledge excavation is to solve the situation of information poor within material sufficient.

The development process of technology in data mining is as follows. The first data Mining software offers one or few ways of counting the materials. However, sufficient materials and frequent changes require the processing with the technology of databank or material warehouse. The first system cannot meet the demand. On the basis of the first data Mining software and the supplement of counting ways, the second data Mining software has been explored. The character is the integration of DSMS, support of databank and material collection. They are highly functional and expansive, can acquire large material collection and more complicated ones. The support of data mining schema and searching of material mining can strengthen the flexibility of the system. Similarly, there are disadvantages in the second data mining software, mainly referring to the attention on model generation. The third data mining system is arisen with the research on the collection of predictable system.

Knowledge excavation belongs to knowledge science. Some other issues include knowledge mathematical theory, logical foundation, knowledge patterns and share in knowledge. In a wider sense, the concept of knowledge excavation is larger than data mining, and requires more on the technology itself and the acquisition findings. The knowledge excavation not only requires the application of knowledge acquired, but also is useful, new, and potential in value and can be accepted by clients.

More and more fields have taken the concept of knowledge excavation, for it is newer than data mining and has obtained good results. The fields include scientific research, marketing, financial investment, production and communication network management. However, the knowledge excavation is in its early stage and has many difficult problems for the research and challenges, such as the massive amount of data, the dynamic data, noise, and lack in values and intensity, understandability of developing pattern, value, collection of applied system, inter operation, update of knowledge, and the processing of complicated databank.

3. University Library Knowledge Base and Knowledge Excavation

Knowledge base, a variety of ultra-large-scale, scalable, interpretable collection of knowledge making use of the latest computer technology, communication technology, artificial intelligence technology and digital content, is a significant contribution in the 20th century using information technology to organize and server knowledge, and it weaves the information in reticulated patterns of various relationships.

A new information processing technology as well as an interdisciplinary subject exclusively concerned by researchers in multidisciplinary fields, knowledge excavation is an advanced and effective method to explore the mass data in accordance with the established goal, reveal the hidden laws among them and then model it. The purpose of knowledge excavation is to integrate a large number of unstructured multimedia information into ordered, sub-leveled, easy-to-understand information, and further into the knowledge which can be used to interfere with prediction and decision-making. In short, knowledge excavation is actually an intelligent and automated process.

With the development of information, digitization and integration, knowledge has become the most important element. How university libraries add the value of knowledge in its circulation becomes really important, so the library knowledge base systems come into being. The outcomes of knowledge base effectively promote the research of library knowledge management theory and application. It has become the consensus of the library profession that library is in a step-by-step transition from information management and information service to knowledge management and knowledge service. Objectively speaking, the current application of knowledge base in libraries also generally remains in the original form of knowledge base the scope of database. University library can be in accordance with the characteristics of its own business and subject construction to set up a reasonable classification and build a variety of knowledge bases suitable for working among different databases, with the purpose of promoting the development and exchange of knowledge. Knowledge base provides necessary conditions for libraries to plan their own organizational structure and management by the way demanded by users, and thus contributes to the transition from the traditional passive service to the service mode of network, personalization and self-service, changing into the user-centered and user-satisfying service mode.

The construction of knowledge bases in libraries is based on knowledge points. The constant updating, developing and finding of knowledge determines the dynamics of knowledge bases. Knowledge managers need to grasp the timeliness of related knowledge points, which makes the data collected in knowledge bases practical, veritable and novel, effectively serving the target users.

4. Application of Knowledge Excavation in University Library Knowledge Bases

4.1 Guiding Library Information Sources Collection, Rationalizing Its Structure

Book purchasing has been the most fundamental work in a library. With the development of book market, the sharp increase of publication, the updating cycle of knowledge, the complication of readers' needs for knowledge and reading as well as the decline of purchasing fund, book purchasing is facing great challenges. Library and information circles, on one hand, construct an integral literature resources system to replace single collection construction to adapt to the new situation and achieve co-construction and sharing of literature resources; on the other hand, emphasize the book purchasing on the basis of single library. In the past, the reference data of purchasing exist only in book catalogue and the accounts of fixed assets. Due to the information asymmetry, book purchasing cannot meet the needs of reader, behind the development of science and technology.

The implementation of library automatic systems everyday produce a large amount of data which can guide the book purchasing, such as data in the circulation of automatic system, historical purchasing data of the library, etc. However, merely data cannot be directly applied to purchasing, because it is not an easy task to analyze and then collect useful information from these data. Some data is time-and-energy-consuming to operate, while others are unable to collect and only can make some fuzzy analysis and evaluation. Knowledge excavation turns these statistics much easier. As long as the system is perfect, knowledge excavation can provide book purchasing with scientific and rational reports of analysis and prediction as well as necessary decision support from various angles, all kinds of ways to analyze different details of an interview, such as expert recommendation, CD-ROM inquiries, online inquiries and the rate of book circulation. As a same reason, knowledge excavation also can be applied to the network library automatic system, providing analysis, prediction and decision support for cooperative purchasing.

4.2 Application of Knowledge Excavation in Library Retrieval System

Information retrieval develops from traditional index retrieval, punch card retrieval and microfilm retrieval into computer information retrieval, which can be broadly divided into offline retrieval, online retrieval, international online searching, CD-ROM retrieval and hypertext network retrieval. Stepping into the 21st century, people have higher demands and expectations of a complete, accurate and fast information retrieval. It becomes an active demand to have a more in-depth, higher level and personalized as well as easy-operated automatic retrieval. The retrieval functions in the existing automatic system of libraries must advance with the times and constantly improve, hence the accessing way and retrieval of information will have more profound changes. Recently, the intelligent retrieval becomes a good case in point. Intelligent retrieval technology focused on the analysis and prediction of users' query plans, intentions and interests to provide users with effective answers, while knowledge excavation focused on extracting or identifying the unknown, interesting or useful and ultimately understandable knowledge model from the large amount of data or information with complex statistical analysis and modeling techniques. Knowledge excavation draws on database technology to make front-end processing of data and extracts useful knowledge from data after treatment through machine learning methods, namely, finds useful knowledge among the abundant data. Knowledge excavation, especially its branches, text excavation technology and web excavation technology, can be used to make a in-depth analysis of documents, extract deep-seated content in documents as well as the valuable relationship knowledge among documents, provide the searchers with the knowledge in a visible and dynamic form, and then interact with user knowledge base, with continuous feedback and adjustment. At the same time, this process embodies the intelligent and personalized search trends.

4.3 Application of Knowledge Excavation in Library Reference and Consulting Department

In a traditional library information services, the general procedure requires the librarians to answer users each request for information. Later, customized services emerged, namely, libraries provide users with a regular basis relative with their subject-related information. Both are proposed by the information needs of users first, before they can receive the corresponding service. With knowledge excavation, we can understand their interests and hobbies through analyzing and excavating the information they borrowed before and the websites they frequently visit, hence determine the content of personalized service, initiatively sending the relevant information to them.

With the growing complexity of social life and academic research, the search tasks and information needs become higher in comprehension, complexity and orderliness during the process of library information reference. People are no longer satisfied with the relevant information, clues, document bibliography data or information reference data, but pay more attention to access the information content of depth, the full text or a comprehensive and profound response to inquiries. However, it is quite difficult for general database technology to excavate, analyze and optimize the user-satisfying consulting results in the face of a large number of disorder information resources of many media and various forms and formats.

Knowledge excavation can make an in-depth excavation of ample data and information, analyze their correlation and ensure the consistency, integrity and security of data. Its highly efficient analysis and query together with interactive

graphical interface can meet users' specific needs of information in a high quality and fast way, which is also the theoretical principle for the wide application of knowledge excavation in decision support system.

4.4 Application of Knowledge Excavation in Digital Libraries

Digital library is a library by way of digit technology processing and storing of various books. It is a multimedia information system. The digital library stores different information sources together in digital forms no matter where they are. Thus, the information can be spread across different region and into internet. It refers to the processing, storing, searching, transmission and utilization of information.

Digital library is a convergence center of excellent knowledge resources. Its website is also the website of digital resources. Due to the central management and transference of data, the digital library can provide convenient resources to those who need them. The feeling of using digital library is the same with that of in a real library. Readers can fully experience the freedom of discovering knowledge and acquire large number of resources in one website. Therefore, the service system of china digital library engineering market is to combine all the constructive results in the whole society and form a super large and high quality Chinese digital rescores group. A market share system should be also constructed for the convergence and spread of resources, and highly effective and convenient service should be provided to the nation by way of public network.

Conceptually, digital library include digitalized library and digital library system. Its two kinds of work include the changing of digital books from real books and the storing, interchange and communication of digital books. Several international organizations have made their contributions to this. Some institutes in China have also participated in the construction of digital library. Formulated under the reference of international standard, China' National Digital Library Standard is a very important standard which has both the characters in and abroad. Commissioned by China National Standardized Committee, China's National Library and Beijing Municipal Library set up the Standard Association of China National Digital Library.

With the development of earth concept, technology and its application field, digital library has been a member of digital earth family. It provides an information resource essential to information expressway and is a main carrier of information resource in knowledge economy society. Digital library is a brand-new technology and also a new social career. In short, it is a highly service system which have many media content and can provide convenient and fast information. In other words, digital library is a virtual library, a knowledge network system based in the environment of network. It has a large scale, separated and convenient, without limitations of time and place. It is a knowledge center of intelligent searching. Digital library is an entire knowledge-oriented system and also a management pattern of future internet development. In a word, digital library can be applied into all kinds of organizations and institutions of society, culture, lifelong education, mass media, business consultation and e-governmental affairs.

Digital library is one of the most important issues of library automation study, and also an essential part of national information infrastructure building, which attracts world attention. By integrating and using the latest computing technology, communication technology and digital content, it builds ultra-large-scale, scalable and interoperable clusters of knowledge bases. Exposed to the Internet, digital libraries organize information resources in an orderly and efficient manner to meet the needs of users with modern high technology. At present, information on the Internet is vast, ample valuable information coexisting with junk information. The more data, the harder it is to find useful information. A large amount of disorder and random data is worthless and only the really valuable information behind data is actually needed by users. Therefore, it becomes a matter of close attention to excavate the value hidden behind abundant data and information, which is also the problem to be solved by knowledge excavation.

In short, knowledge excavation in university library knowledge bases has already been playing an important role in improving the management of university libraries and inevitably will play a more profound role in the future.

References

Feng, Xinmin and Wang, Jiandong. (2008). The Concept Dilemma of Knowledge Mining and the Broad-sense Knowledge Mining. *Journal of Information*, 7.

Gilea, M, Foody. (2003). Uncertainty, Knowledge Discovery and Data Mining in GIS. *Progress in Physical Geography*, 27(1), 113 – 121.

Qiu, Zhuangli and Li, Xuelian. (2005). Application of Knowledge Excavation in Files Management. *Lantai World*, 7.

Zhang, Jinli and Wu, Guoxing. (1998). National Digital Library-Body of Information Resources Construction in China. *Journal of the China Society for Scientific and Technical Information*, S1.

Zhao, Kaiwei and Zhang, Guiling. (2003). An Overview of Overseas Digital Libraries. *Information Science*, 5.

Zhou, Ruiping. (2000). On Digital Libraries. *Information Science*, 11.