

Retraction

Retracted: Construction and Knowledge Mining of Traditional Chinese Medicine Ancient Books Bibliographic Abstracts Database Based on Genetic Algorithm and BP Neural Network

Mathematical Problems in Engineering

Received 18 July 2023; Accepted 18 July 2023; Published 19 July 2023

Copyright © 2023 Mathematical Problems in Engineering. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their

agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] Y. Wang, S. Ren, L. Song, and J. Zhang, "Construction and Knowledge Mining of Traditional Chinese Medicine Ancient Books Bibliographic Abstracts Database Based on Genetic Algorithm and BP Neural Network," *Mathematical Problems in Engineering*, vol. 2022, Article ID 6838714, 15 pages, 2022.

Research Article

Construction and Knowledge Mining of Traditional Chinese Medicine Ancient Books Bibliographic Abstracts Database Based on Genetic Algorithm and BP Neural Network

Yongmei Wang ¹, Shujun Ren,² Li Song,¹ and Jiang Zhang¹

¹School of Computer Science and Technology, Hefei Normal University, Hefei 230031, China

²Library of Anhui University of Traditional Chinese Medicine, Hefei 230038, China

Correspondence should be addressed to Yongmei Wang; wangymfddy@hfnu.edu.cn

Received 8 March 2022; Revised 20 March 2022; Accepted 23 March 2022; Published 6 April 2022

Academic Editor: Man Fai Leung

Copyright © 2022 Yongmei Wang et al. 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.

With the rapid development of modern science technology and Internet technology, the establishment of a unified and standardized bibliographic summary database to realize the exchange and resource sharing of ancient Chinese medicine bibliography, is the inevitable trend of ancient Chinese medicine bibliography digital service. Firstly, this paper formulates the bibliographic metadata specification of traditional Chinese medicine ancient books (TCMAB), extracts each cataloging file into an XML document in line with the bibliographic metadata specification of TCMAB. Secondly, this paper realizes the unified description of ancient book resources in the database system of TCMAB, uses the native XML database eXist to store and manage the XML documents of all traditional Chinese medicine ancient book resources, and integrates the multimedia data with XML data. Finally, genetic algorithm and BP neural network are used for knowledge mining and discovery, the overall model design of TCMAB bibliographic abstracts database system is proposed based on the construction process of knowledge map. The system platform adopts B/S mode, eXist database management system, PowerSSP streaming media and video server for audio video processing.

1. Introduction

Traditional Chinese medicine ancient books are precious resources to carry forward traditional culture. We should make full use of the resources of TCMAB to serve the development of traditional Chinese medicine culture, which will help to promote the inheritance and develop traditional Chinese medicine culture [1]. The construction of digital database is conducive to enriching online resources and promoting the transformation of ancient books potential resources into real productivity. The digital database system can provide more comprehensive information in one aspect online for the needs of readers, improve the service quality, and is conducive to the innovative development of traditional Chinese medicine [2]. The establishment of database can improve the efficiency of resource sharing. The resources of the database should be operated according to certain

standards and norms in the process of construction, which is conducive to promoting the co-construction and sharing of resources. In addition, the database plays an irreplaceable role in the process of optimizing and developing the existing literature [3]. It can systematically sort out the scattered resources and carry out data processing and preservation of traditional Chinese medicine resources according to the relevant requirements, which will help to strengthen the integration and sharing of traditional Chinese medicine resources. The needs of users are the basis for the system to carry out services, only by closely combining the needs of users can we provide efficient services for users [4]. The needs of users can generally be described from two levels, one is general and basic needs, such as users' daily access to data, on the other hand, it is to meet the characteristic needs of scientific research, such as finding relevant materials, which helps to meet the needs of multi-level social users. At

present, the traditional Chinese medicine ancient books are scattered all over the country, and there has not been a database dedicated to the bibliographic summary information of traditional Chinese medicine ancient books [5]. This paper aims to extract the bibliography and abstract information of TCMAB from the paper books by using metadata description on the basis of analyzing the paper bibliography and abstract books, and use the existing technology to transform the bibliography and abstract information of TCMAB into digital resources, and then on this basis, build a special bibliography and abstract database of ancient doctors. On the premise of network application, carry out the collection, integration and processing of online Chinese medicine literature summary information resources, provide special information retrieval services, realize the networking of Chinese medicine ancient book summary information organization and services, develop value-added products of information resources, and produce various types of special databases according to the market demand [6].

The specific contributions of this paper include the following:

- (1) Digital description and information extraction of traditional Chinese medicine ancient books bibliographic abstracts, formulating metadata specifications of traditional Chinese medicine ancient books.
- (2) In this paper, genetic algorithm and BP neural network are used for knowledge mining in TCMAB, The scientific storage and management of bibliographic data of TCMAB can effectively manage the digital content of traditional Chinese medicine ancient books in semi-structured form.
- (3) Realizing remote access to the digital resource database of TCMAB bibliographic abstracts, and form a local bibliographic abstracts system of TCMAB with local characteristics, which has practicality, network environment support and all-round retrieval function.
- (4) The research results of this paper can help establish digital standards of TCMAB, enrich the existing digital theory, make TCMAB bibliographic abstracts database more practical.

The rest of this paper is organized as follows. Section 2 discusses related work, followed by the analysis of ancient books digitization process in Section 3. The metadata specification of TCMAB bibliography is discussed in Section 4. Section 5 proposes the design of database system for TCMAB bibliographic abstracts, and Section 6 concludes the paper with a summary and future research directions.

2. Related Work

Traditional Chinese medicine ancient books are an important part of Chinese ancient book resources, and their quantity ranks first among ancient books of various disciplines. The digitization of traditional Chinese medicine ancient books has made some achievements both in theory

and in database construction. The bibliographic database of ancient books is the initial stage of ancient books digital resources construction [7]. It is a database formed by inputting the information such as the title, author, version, volume number, abstract and source of ancient books into the computer, and provides users with a large catalog database for the retrieval of relevant ancient book data resources in the form of computer network system [8]. Readers can retrieve the relevant information of an ancient book through the title and author of the book. The ancient books bibliographic database can meet the needs of modern works, and it is the basis for the further digitization of ancient books. At present, the bibliographic database of TCMAB is mainly concentrated in colleges and universities of traditional Chinese medicine, scientific research institutes and their libraries, mainly to reveal the collection, serve teaching and scientific research. At present, many traditional Chinese medicine libraries and scientific research institutions have bibliographic databases in their collections of ancient books, but due to the inconsistency of standards, the depth and breadth of documents revealed by the database are also different [9–14]. For example, the Chinese medical code has collected more than 1000 major works of traditional Chinese medicine in the past dynasties. Using the library classification method, the collected ancient books are divided into 12 categories: medical classics, diagnostic methods, materia medica, prescription books, acupuncture and massage, typhoid golden chamber, febrile disease, comprehensive medical books, clinical and syndrome disciplines, external treatment of health and diet therapy, medical theory and medical records, and others, involving all disciplines of traditional Chinese medicine [15–19].

Current research progress in studies on linked data, the challenge with linked data is that databases are constantly evolving and cached content quickly becomes outdated. To overcome this challenge, Qian et al. [20] proposed a change metric that quantifies the evolution of a linked dataset and determines when to update cached content. Zhao et al. [21] proposed a query language framework for probabilistic RDF data (an important uncertain link data), in which each triplet has a probability, called pSRARQL, which is based on SPARQL. Liu et al. [22] presented a new method of web form integration based on linked data and VDIS (View based Data Integration System) architecture. They proposed WebQuinLD, an alternative new method based on linked data principle, which can combine a single WQI into a single IWQI of a given domain.

3. Analysis of Ancient Books Digitization Process

The digitization of ancient books uses computer technology to convert the words or images in ancient books into digital forms that can be recognized by computer, make ancient books bibliographic database, ancient books full-text database and ancient books knowledge base, save and disseminate through CD-ROM, network and other forms. So as to reveal the rich information and knowledge resources contained in ancient books and documents, protecting, utilizing

and mining ancient books and documents. The digitization of ancient books is developing rapidly, but at the same time, we should also see all kinds of problems in the process of ancient books digitization.

3.1. Inconsistent Standards. In today's digitization of ancient books, the problem of inconsistent standards is very serious and widely exists in the world. The non-uniform standards mainly focus on: version standard of digital ancient books, bibliographic classification standard, character set standard, storage format standard, digital image standard, retrieval system standard, metadata description standard, etc. These problems are perplexing the benign development of ancient books digitization.

The problem of non-uniform bibliographic classification standards, that is, what kind of bibliographic classification standards should be followed in the digitization of TCMAB, has become a problem that must be solved before digitization. Combined with the actual situation of TCMAB, the simple use of four part classification method or Chinese library classification method is not appropriate. In the digitization of TCMAB, we can refer to the classification methods of "The General Summary of Traditional Chinese Medicine Ancient Books" and "Xin'an Medical Books", which are divided into "Medical Classics" (including the basic theory of traditional Chinese medicine), "Typhoid Fever" (including febrile diseases), "Diagnostic Methods", "Materia Medica", "Acupuncture and Moxibustion", "Treatise on Prescription" (1-5), "Medical Records", "Health Preservation", "Series" (including reference books of medical history) "Textual Research on Medical Records" and "Appendix", a total of 15 parts. In this way, different classification standards can be avoided, which provides a prerequisite for the production of bibliographic summary database, and is also the basis for the construction of retrieval database in the future.

3.2. Technical Difficulties. The digital processing of ancient books mainly includes two aspects: the simple digital conversion and reproduction of the external form or content form of ancient books; after completing the text to digital conversion, related processing work, such as retrieval, content relation and so on. The main technical problem at this stage is the conversion of text from image data to text data. If such conversion cannot be realized, the related work of later text and content relation cannot be completed. OCR optical character recognition technology need to convert ancient book images into characters. OCR technology can convert the characters in the image into characters that can be seen in the computer through recognition.

There are many kinds of ancient books, but the current OCR software design only aims at an extremely limited part of them, which makes the reliability of this kind of software very limited in the text process. It is also necessary to have a deeper understanding of different ancient books and provide targeted technical processing methods for more different types of ancient books, in this way, OCR technology can better deal with a variety of ancient books in different

situations. In order to achieve a more comprehensive identification, we need more development in technology, so as to further promote the development of ancient books digitization.

Knowledge base is the combination of artificial intelligence and database. It stores knowledge in a unified form. The knowledge of knowledge base is highly structured symbolic data. Users can carry out deep knowledge mining to realize the relation retrieval of multiple knowledge points from bibliography to full text, or one author can retrieve other relevant authors. Therefore, we can find the causes and solutions of these problems in the continuous development of ancient Chinese medicine. The overall model design of TCMAB bibliographic abstracts database system is based on the construction process of knowledge map. The system platform adopts B/S mode, eXist database management system, and PowerSP streaming media and video server for audio video processing.

4. Metadata Specification of TCMAB Bibliography

Metadata's purpose is to provide an intermediate level description according to people can determine information package, which they want to browse or retrieve without retrieving a large number of irrelevant full-text. The metadata of ancient books can be simply defined as information object. Ancient book metadata is divided into three types: descriptive metadata, management metadata and application metadata.

4.1. Metadata Specification of Ancient Books. In the field of ancient books and documents, common standards related to metadata include the metadata specification of ancient books, the metadata specification of special digital object description, and the international Dublin Core metadata set (DC). The cultural industry standard of the people's Republic of China WH/T 66-2014 metadata specification for ancient books issued and implemented by the Ministry of Culture in 2014 is the metadata specification for ancient book resources. It is formulated to unify and standardize the description of the ancient books content characteristics and better manage ancient book resources. It is proposed by China's Digital Library and standard specification construction project. This specification is used to describe the content and appearance characteristics of paper books and digital ancient books. The resource objects described in the standard are ancient books, which are similar to the resource object types described in this paper. At the same time, the standard reuses some elements in DC as the core elements, and individual elements in ancient document types as the core elements of resource types. At the same time, several modifiers are added to the core elements according to the characteristics of Chinese ancient books, it provides an important reference for this paper to formulate the bibliographic metadata specification of TCMAB.

4.2. Formulating the Metadata Specification of TCMAB Bibliography

4.2.1. Standard Design Principles of TCMAB Bibliographic Metadata. The design of TCMAB bibliographic metadata specification follows the design principles of China's special metadata specification, that is, simplicity and accuracy, specificity and universality, interoperability and easy conversion, scalability and meeting the needs of users. Simplicity mainly means that the metadata standard of bibliography of TCMAB should be easy to master in the practice of description. The accuracy of TCMAB bibliographic metadata specification can improve the accuracy of cataloging of TCMAB. Specificity refers to determining the corresponding metadata specification according to the specific resource entity requirements, while universality requires that the metadata specification be universal within a certain range. The metadata specification designed in this paper is designed with reference to the cataloguing documents of ancient Chinese medicine books such as "General Outline of Ancient Chinese Medicine Books" and "Xin'an Medical Books", so as to achieve universality in the field of ancient Chinese medicine books. The interoperability of metadata refers to the feature that data can be shared among different systems. This feature ensures that metadata can be operated by application systems established by other organizations or institutions while providing services for itself. This requires that the metadata specification designed in this paper needs to meet the sharing requirements of TCMAB resource database. When designing the metadata specification of TCMAB, we need to carefully consider the definition of elements and modifiers in the metadata specification. Therefore, when designing the metadata specification of TCMAB, we refer to the "Metadata Specification of Special Digital Object Description", "Metadata Specification of Ancient Books" and the Dublin core element set widely supported abroad. The extensibility of metadata specification means that user-defined elements or modifiers can be expanded according to the specific application needs of users. Based on this feature, when formulating the metadata specification of TCMAB, corresponding elements and modifiers will be added according to the characteristics of TCMAB cataloging files collected in this paper. Considering that the purpose of formulating metadata specification is to more fully display information resources to users, user requirements are taken as an important standard to weigh the quality of metadata specification. Therefore, when designing the bibliographic metadata specification of ancient Chinese medicine books, reference is made to the bibliographic cataloguing books of ancient Chinese medicine books such as "Chinese Medical Books" and "Xin'an Medical Books", in order that the designed bibliographic metadata specification of ancient Chinese medicine books can meet the needs of users in all aspects.

4.2.2. Establishment of Metadata Standard for Bibliography of TCMAB. The metadata specification of TCMAB takes traditional Chinese medicine ancient books as the root node.

In this paper, we study the sub-node of paper TCMAB. The structure of elements and their modifiers in the bibliographic metadata specification of ancient Chinese medicine books designed in this paper is shown in Figure 1.

Dublin Core terms are reused in the node (<http://purl.org/dc/terms>), at the same time, metadata object description scheme is reused (<http://www.loc.gov/mods>), the location element definition in. The contents can be summarized into book title (including alias), dynasty, author, source of the book, volume number, preservation and loss, content summary, edition collection, notes, etc.

4.2.3. Data Modeling of Bibliography of TCMAB. The data model mainly displays the entities and attributes involved in TCMAB and the relationship between them in a more intuitive way. This paper makes a link research on the information of the responsible person, version, publishing place and so on. In the data model of TCMAB in Figure 2, it is marked with URI. The version part is the link inside the Drupal site, which is marked with node. Figure 3 shows the specific process of HTTP content negotiation mechanism for related data.

4.3. Abstract Information Extraction of Ancient Books Based on Metadata Specification of TCMAB Bibliography. This paper formulates the metadata specification of TCMAB bibliography. How to use this specification to extract the information of TCMAB bibliography and realize the digitization of the bibliographic abstracts of ancient books is the focus of this section. Abstract is a brief introduction and evaluation of the content, thought, author and version source of the literature, so as to effectively help readers use the literature correctly and play the role of reading guide. The "General Summary of Traditional Chinese Medicine Ancient Books" (referred to as "General Summary") contains well-documented medical literature before 1949, including existing literature and lost literature. "Xin'an Medical Books" collects ancient Chinese medical books from Shennong era to the middle of Qing Dynasty, including famous authors, ages and volumes, to provide help for the study of traditional Chinese medicine bibliography. "Xin'an" is the name of a region in Chinese history. It was first seen in the first year of Taikang in the Western Jin Dynasty (280 years). There are numerous medical books written by famous experts in this area in Chinese dynasties, of which the Ming and Qing Dynasties account for a large proportion, which naturally forms an important position of "Xin'an Medical Books" in the field of Chinese medical literature.

"General Summary", "Xin'an Medical Books" and the cataloging documents of ancient Chinese medicine books all have the characteristics of ancient books cataloging. Because the ancient books bibliographic information belongs to semi-structured data, the composition is relatively simple and has obvious regularity. The digitization process of traditional Chinese medicine ancient books is mainly to scan the "General Summary" of paper books into pictures and store them as PDF files. Then, through OCR (optical character recognition), the text file and picture file are obtained, in which the text file contains the title of Chinese characters and the bibliographic content of ancient books.

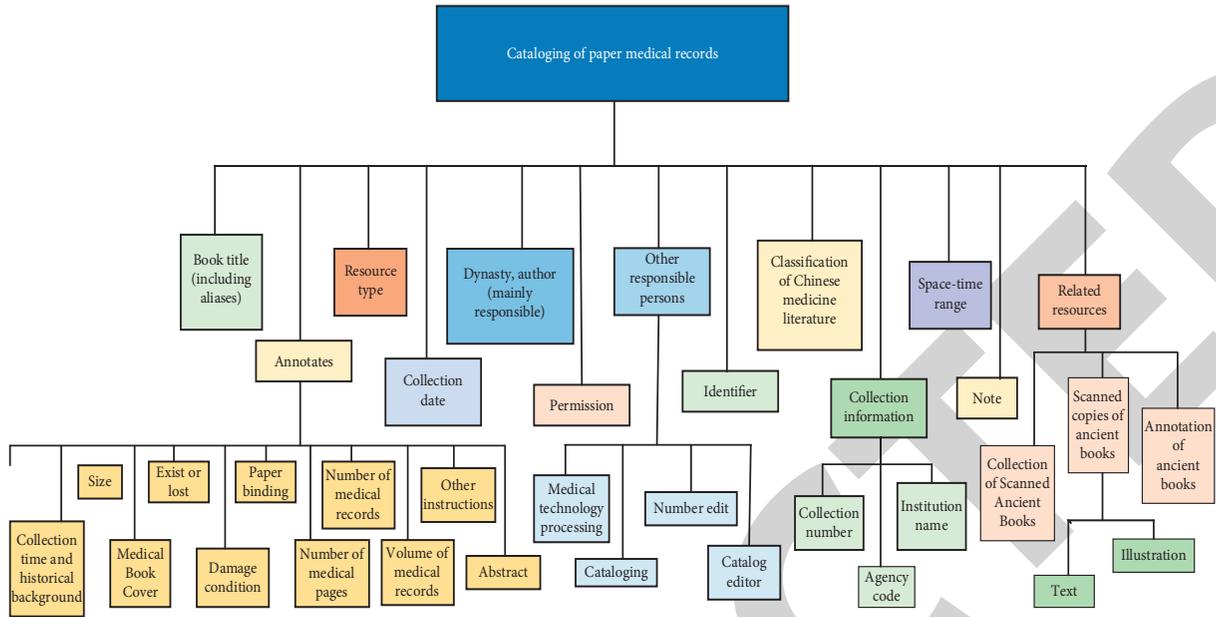


FIGURE 1: Elements and their modifiers under the nodes of TCMAB.

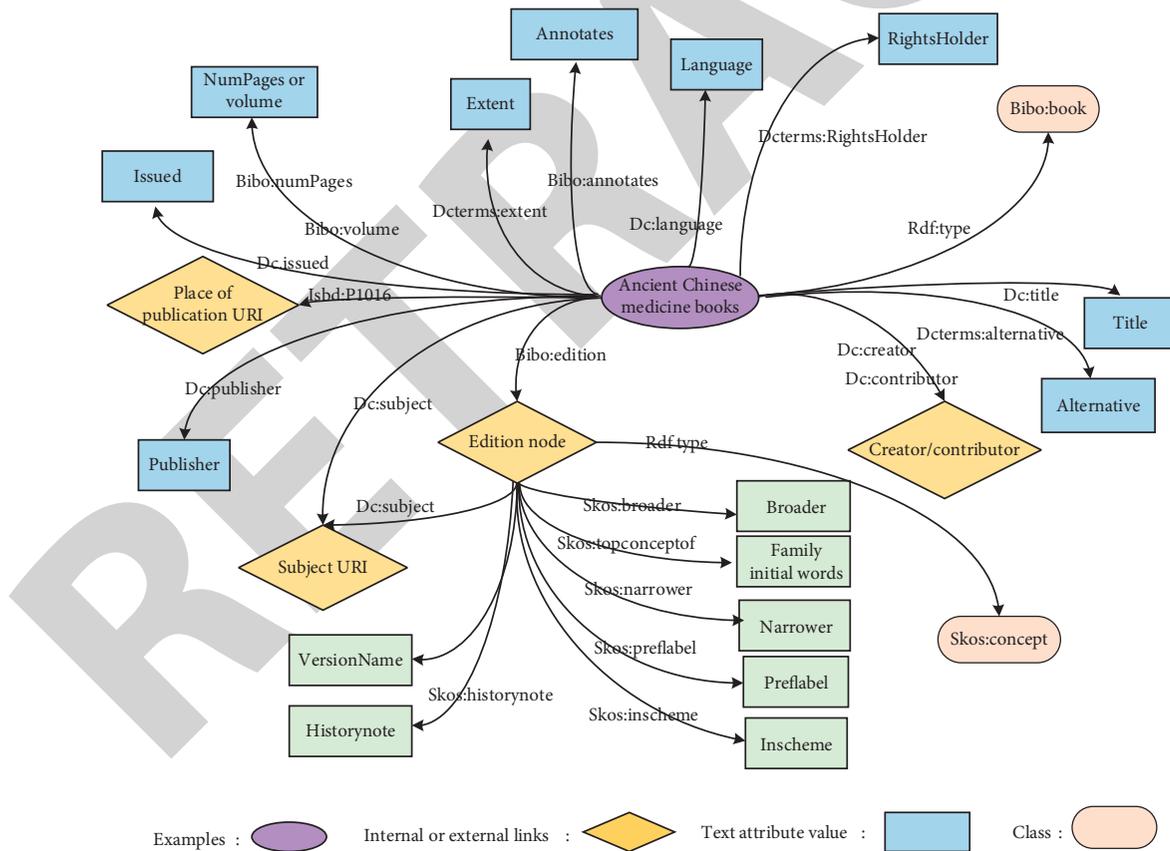


FIGURE 2: Data model of TCMAB.

Proofread the semi-structured document, and then extract the text according to the template. Name the recognized picture with the title of Chinese characters. The specific process of digitization of paper books is shown in Figure 4 [23].

The eXist is an open source native XML database management system, which has developed rapidly in recent years. Its characteristics can ensure that eXist can efficiently store and retrieve XML documents, which is also the reason why this paper chooses eXist to store and manage TCMAB.

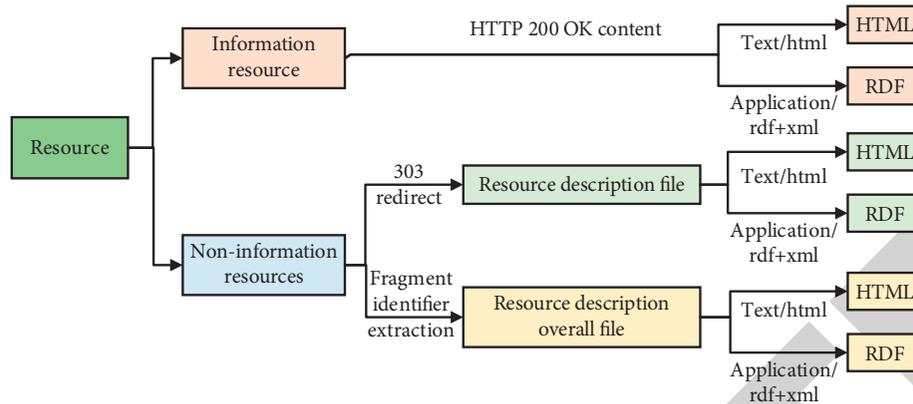


FIGURE 3: URI naming and content negotiation acquisition mechanism.

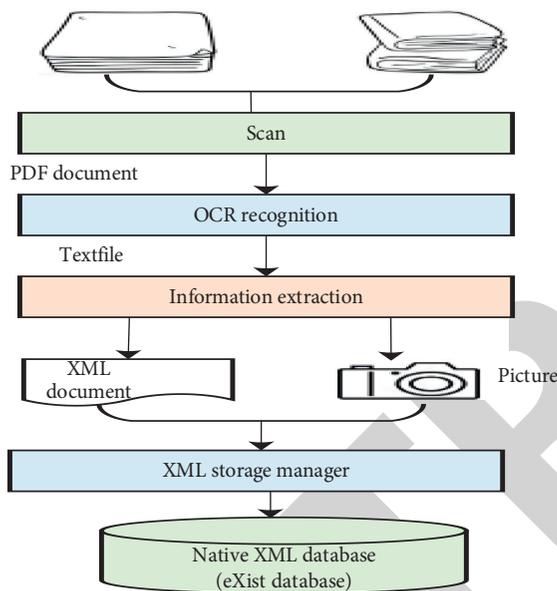


FIGURE 4: Digital process of paper books.

The goal of information extraction of TCMAB is to extract the contents of paper books and text files to form semi-structured data with standardized metadata of traditional Chinese medicine.

Before extracting information from TCMAB, we should first analyze the text characteristics of ancient books cataloging bibliography according to the different contents of metadata. The information extraction method used in this paper is the template extraction method based on regular expression. Taking “Materia Medica Preparation” in the catalogue of TCMAB in Anhui as an example, the contents can be summarized into book name (including alias), dynasty, author, source of the book, volume number, preservation and loss, content summary, edition collection, notes, etc. A large number of medical books written by doctors before the end of the Qing Dynasty have been verified, corrected, supplemented and studied from the aspects of book name, author, volume number, preservation and loss, version collection, content and notes.

Secondly, according to the writing rules of regular expression, the extraction template of ancient books in the “General Summary” is designed. According to the extraction template of books in the “General Summary”, the information of the target ancient book catalogue is extracted, and after the extraction result is compared with the content of ancient books before information extraction, the XML document after information extraction is finally obtained.

Finally, in order to test the performance of the information extraction method proposed in this paper, analyze the performance of the extraction method by using the recall rate and accuracy rate, select some ancient books in the “General Summary” as the test set for information extraction, calculate and analyze the recall rate and accuracy rate. If the results are correct, it shows that the template based information extraction method proposed in this paper can correctly extract the contents of ancient books according to the characteristics of TCMAB bibliography in the “General Summary”. Through this method, the information of the books that need to be digitized in this paper is extracted, which provides a data basis for the follow-up research work of this paper.

5. Application of Knowledge Discovery in the TCMAB Based on Genetic Algorithm and BP Neural Network

There may be a lot of noise in the data in the database, and knowledge mining tools can efficiently mine the laws and values hidden in the data, and find the qualitative relationship between data attributes, such as dependency. The trained neural network can give the quantitative and qualitative description of data attributes. In this paper, BP neural network and genetic algorithm are used for knowledge discovery and trend analysis in the database. Database technology can quantitatively analyze the management information system with technology as the core needed by knowledge discovery. In the transaction processing, data is the most processed in the database. Knowledge among database attributes is an important problem, knowledge discovery is an accurate reflection of functional relations.

5.1. BP Neural Network. BP neural network is a feedforward network where neurons are distributed in layers: input layer, output layer and hidden layer(s). The output of neurons in each layer is transmitted to the next layer. This transmission can have enhancing, weakening or inhibiting effects on the outputs through connection weight. Apart from the neurons in the input layer, the weighted sum of the output of neurons in the previous layer when the net input of neurons in the hidden layer and output layer. Each neuron is activated by its input, activation function and threshold. Its working process consists of two periods, one for learning and the other for working. The former covers two processes: forward propagation of input information and back propagation of error. In the first process, the input data are processed layer by layer, the input, hidden and output layers sequentially. The state of neurons in each layer merely affects that of the neurons in the next layer. If the output of the output layer is inconsistent with the expected output of the given sample, the output error is calculated, transferred to the error back propagation process, and the error is returned along the original connection path. By modifying the weights between neurons in each layer, the error is minimized. Through training with large quantities of learning samples, the connection right between neurons in each layer is fixed and transferred to the working period. There is only forward propagation of input information during the working period. The forward propagation is calculated according to the working process of the previous neuron model. Thus, the key to the calculation of BP neural network lies with the error back propagation process in the learning period, which is completed by minimizing an objective function. Usually, the objective function is defined as the sum of the error squares between the actual output and the expected output or the error function, and the calculation formula can be derived by using the gradient descent method.

The specific process of BP algorithm is described as follows:

The initial value of learning times is $t = 0$, and the network weights $W_{ij}(t) \in [-1, 1]$, $W_{jk}(t) \in [-1, 1]$ and threshold $\theta_j(t) \in [-1, 1]$ and $\theta_k(t) \in [-1, 1]$ are initialized with random numbers.

- (1) Enter a learning sample (X_p, T_p) , where $p \in (1, 2, \dots, N)$ and N is the number of samples, $(X_p \in R^n, T_p \in R^m)$.
- (2) Calculate the output value of each node in the hidden layer.

$$Y_j^2 = f \left(\sum_{i=1}^{n_1} W_{ij} Y_i^1 - \theta_j \right) = f \left(\sum_{i=1}^{n_1} W_{ij} X_{ip} - \theta_j \right). \quad (1)$$

it's action function is log_sigmoid type function.

$$f(x) = \frac{1}{1 + e^{-\lambda}}, j \in \{1, 2, \dots, n_2\}. \quad (2)$$

- (3) Calculate the output value of each node of the output layer.

$$Y_k^3 = f \left(\sum_{j=1}^{n_2} W_{jk} Y_j^2 - \theta_k \right). \quad (3)$$

its action function is linear Purelin function.

$$f = ax, k \in \{1, 2, \dots, m\}. \quad (4)$$

- (4) Calculate the correction value of the connection weight between the output layer node and the hidden layer node.

$$\sigma_k = (T_k - Y_k^3) Y_k^3 (1 - Y_k^3), k \in \{1, 2, \dots, m\}. \quad (5)$$

- (5) Calculate the correction of the connection weight between the hidden layer node and the input layer node.

$$\sigma_j = Y_j^2 (1 - Y_j^2) \sum_{k=1}^m \sigma_k W_{jk}, j \in \{1, 2, \dots, n_2\}. \quad (6)$$

- (6) Correct the connection weight W_{jk} of the output layer node K and the hidden layer node J , and correct the threshold of the output layer node K . The following is the error correction amount calculated in step (5).

$$W_{jk}(t+1) = W_{jk}(t) + \alpha \sigma_k Y_j^2 \quad (7)$$

$$\theta_k(t+1) = \theta_k(t) + \beta \sigma_k.$$

- (7) Modify the connection weight W_{ij} of hidden layer node J and input layer node I , and modify the threshold of hidden layer node j . The following is the error correction amount calculated in step (6).

$$W_{ij}(t+1) = W_{ij}(t) + \alpha \sigma_j Y_i^1 \quad (8)$$

$$\theta_j(t+1) = \theta_j(t) + \alpha \sigma_j.$$

- (8) If the complete learning sample is not taken, go to step (2).
- (9) Calculate the error function E and judge whether it is less than the specified upper limit of error. If it is less than the upper limit of error or the limit of learning times, the algorithm ends. Otherwise, update the learning times ($t = t + 1$) and return to step (2).

The following Figure 5 is the flow chart of BP neural network.

5.2. Genetic Algorithm. In genetic algorithms of each generation, individuals are chosen in accordance with the fitness of individuals in the problem domain and combined crossover and mutation are performed aided by genetic operators so as to generate a population which represents the new solution set. The process will cause the population's offspring to be more adaptive to the environment than their previous generation, just like what happens in natural

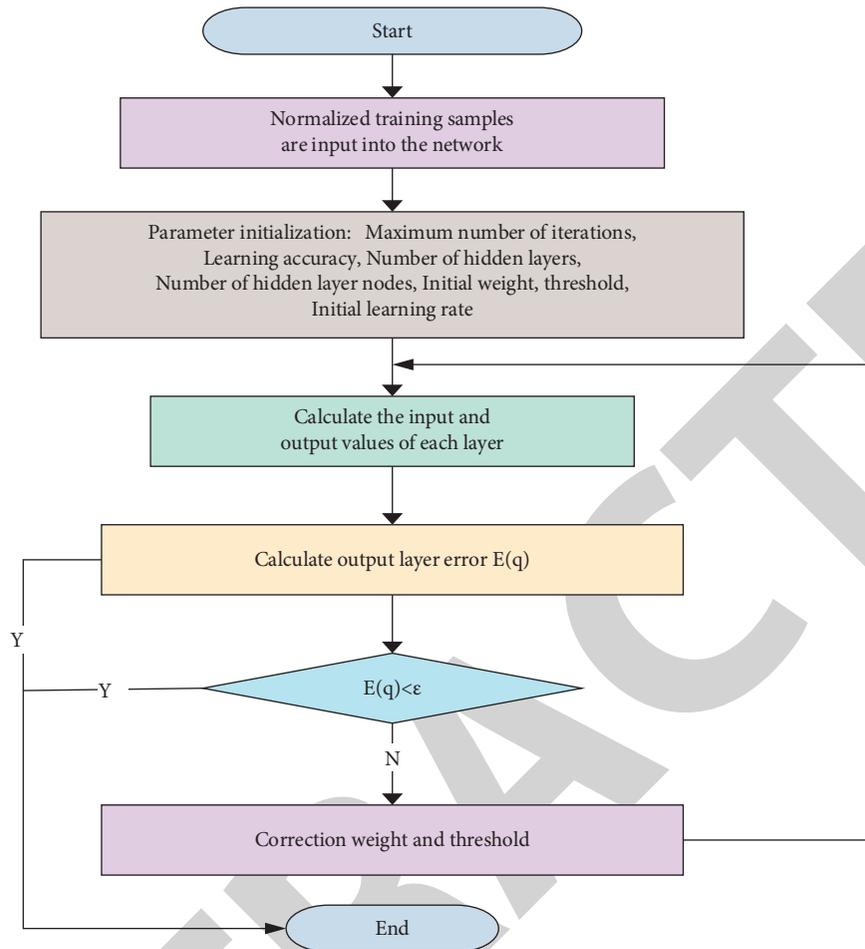


FIGURE 5: BP neural network flow chart.

evolution. When the decoding is completed, the optimal individual in the last generation of population can be regarded as the approximate optimal solution of the problem. Genetic algorithm undergoes the following basic steps:

- (a) The coding strategy is selected to transform the parameter set (feasible solution set) into chromosome structure space;
- (b) Define fitness function to calculate fitness value;
- (c) Determining genetic strategies, including selecting population size, selection, crossover and mutation methods, and determining genetic parameters such as crossover probability and mutation probability;
- (d) Randomly generate initialization population;
- (e) Calculate the fitness value of individuals or chromosomes in the population after decoding;
- (f) According to the genetic strategy, the selection, crossover and mutation operators are used to act on the population to form the next generation population;
- (g) Judge whether the population performance meets a certain index or has completed the predetermined

number of iterations. If not, return to step 5, or modify the genetic strategy and return to step 6.

The flow chart of genetic algorithm is shown in Figure 6.

5.3. BP Neural Network Optimized by Genetic Algorithm.

Because the initial weight and threshold in BP neural network are randomly generated, the model obtained through learning may not be optimal and may fall into local optimal solution. At the same time, BP neural network is able to perform self-learning and negative back propagation of training error in the internal structure of the network. Therefore, genetic algorithm can be used to optimize and analyze the weight and threshold of BP neural network, the sum of the output absolute error of the optimized BP neural network is deemed as the fitness function of the genetic algorithm. The lower the fitness value of the response value, the better the approximation effect of the neural network, and the smaller the sum of the corresponding absolute error, thus optimizing the BP network and iteratively obtaining the optimal BP network weight and threshold. Finally, the global optimal solution is obtained by fine tuning with BP algorithm. The flow chart of BP neural network optimized by GA is shown in Figure 7.

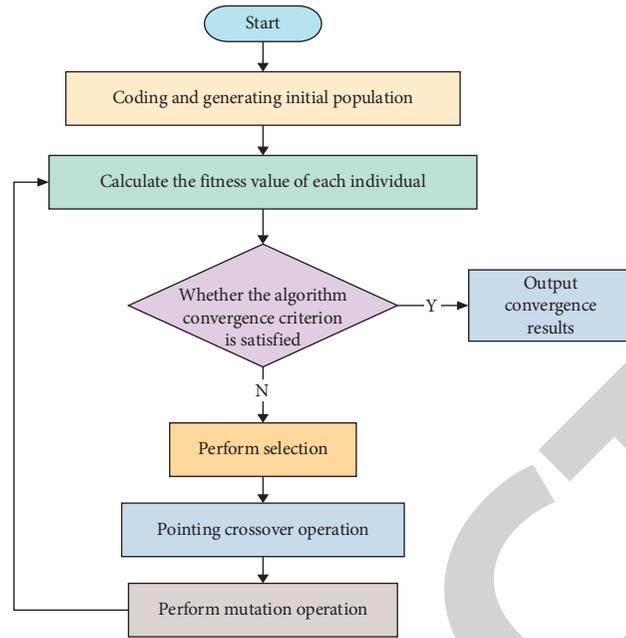


FIGURE 6: Flow chart of genetic algorithm.

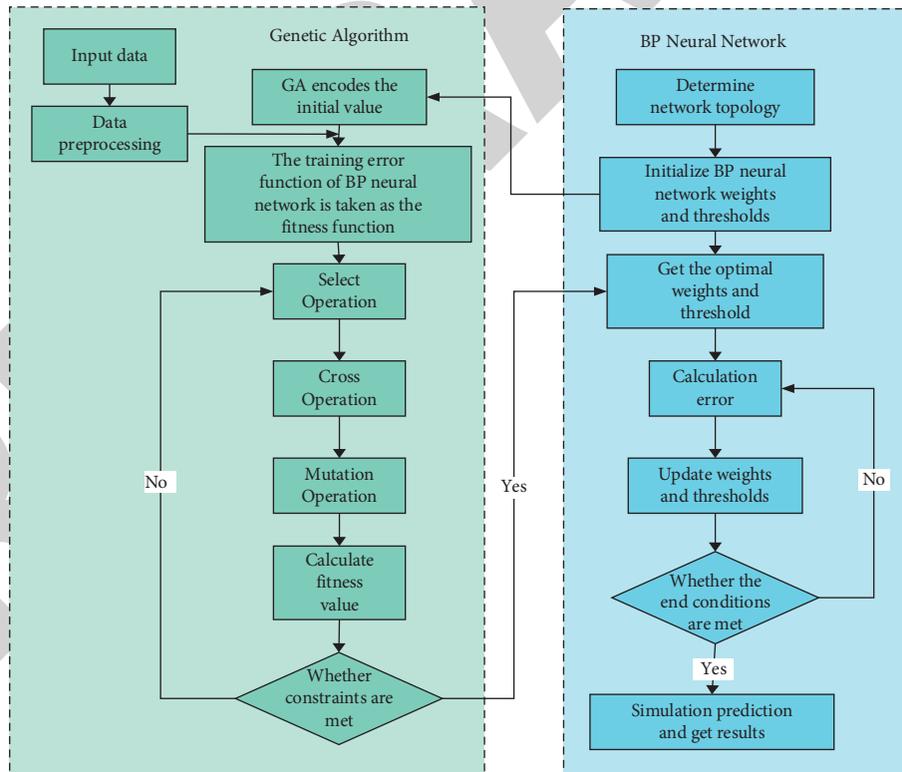


FIGURE 7: Flow chart of BP neural network optimized by GA.

In this paper, genetic algorithm and BP neural network are used for knowledge discovery and trend analysis in TCMAB. It has a fast response to the analysis between linear and nonlinear uncertain attributes. It can effectively manage

the digital content of semi-structured ancient books of traditional Chinese medicine, realize the remote access of TCMAB, and form a practical, leading and omni-directional retrieval function of TCMAB.

6. Design of Database System for TCMAB Bibliographic Abstracts

The overall model design of the database system of TCMAB bibliographic abstracts is based on the construction process of knowledge map. The system platform adopts B/S mode, eXist database management system, PowerSSP streaming media and video server for audio video processing. This section first analyzes the design objectives and functional requirements, then expounds the structural design based on XML database and the construction scheme of TCMAB bibliographic database system.

6.1. Design Goal. The database system of TCMAB bibliographic abstracts is the premise and foundation of TCMAB resource sharing. How to effectively manage the resources of TCMAB is the key to construct the resource management system of TCMAB. The design principle of TCMAB bibliographic abstracts database system should be able to provide efficient storage and management for all kinds of ancient Chinese medicine books cataloging documents, and provide users with convenient fast storage and retrieval functions at the same time. The design shall meet the following requirements.

- (1) According to the “Metadata Specification for Ancient Books”, “Metadata Specification for Special Digital Object Description” and international DC metadata, formulate the metadata specification for the bibliography of TCMAB, extract each cataloging file into an XML document conforming to the metadata specification for the bibliography of TCMAB, and realize the unified description of ancient book resources in the bibliographic summary database system of TCMAB, for the cross platform sharing of ancient book resources.
- (2) The native XML database exist is used to store and manage the XML documents of all TCMAB resources to ensure that the structural information in the XML document can be completely saved.
- (3) In the process of collecting TCMAB, there are relevant multimedia resources such as ancient book interpretation recording and introduction animation or video of medicinal materials, prescriptions and characters. Using the metadata specification of TCMAB, the multimedia data and XML data are integrated to display TCMAB to users in an all-round way.
- (4) The TCMAB are classified according to their different sources, and different access rights are set for different users.
- (5) According to the metadata specification of TCMAB, it can support multi-condition retrieval and keyword retrieval. At the same time, due to the particularity of XML database, users can directly use XQuery statements to retrieve, so that users can retrieve the required content accurately and flexibly.

6.2. System Functional Requirements. System users are divided into ordinary users and administrators. The following Figure 8 is use case diagram and Figure 9 is UML class diagram. Ordinary users do not need login authentication to directly enter the main page of the bibliographic summary database system of TCMAB. Ordinary users only have the retrieval function and can query the resources in the system vaguely or accurately, after the search conditions are submitted, the system will return the search result list. Ordinary users have the permission to view the details and can view the details of the ancient book bibliography, the database system of TCMAB bibliographic abstracts provides the function of resource download for ordinary users.

Compared with ordinary users, the administrator needs user login verification. After successful login, enter the home page of the administrator’s bibliographic summary database system of ancient Chinese medicine. In addition to all the permissions of ordinary users, the administrator can manage the ancient book resource data uploaded by the system, including adding, deleting and modifying. You can not only add bibliographic records of ancient books one by one, but also upload a catalogue file of ancient books.

6.3. Design of Database Model Based on Knowledge Map. By collecting the bibliographic data of TCMAB and studying the related technology of knowledge map, the knowledge map model of ancient books designed in this paper is shown in Figure 10. The model includes resource acquisition layer, knowledge unit processing layer, relation representation layer and application layer.

In Figure 10, the resource acquisition layer is at the bottom of the model, its main function is to obtain resource information through existing literature, paper and electronic data. The resource acquisition layer is the main source of database information. The knowledge unit processing layer is mainly responsible for the correlation analysis of the data obtained by the resource acquisition layer, to form an independent knowledge unit and establish an index for it. The main function of the relation presentation layer is to reveal the relationship between knowledge nodes links and knowledge relations. By establishing the correlation between each knowledge node, the knowledge units with different structures are connected in series to form an intertwined knowledge network, so as to provide basic data for cross database retrieval and correlation analysis of traditional Chinese medicine book knowledge in the application layer. The application layer is mainly responsible for providing friendly human-computer interaction interface, and different users provide different services. For ordinary Internet users, the system mainly provides users with basic traditional Chinese medicine knowledge browsing and query. For traditional Chinese medicine experts and administrators, the system mainly provides deeper knowledge maintenance, knowledge index and knowledge relation services.

As can be seen from Figure 10, this paper constructs the model from two perspectives. The first perspective is from the system administrator. The model improves the knowledge service of traditional Chinese medicine books layer by

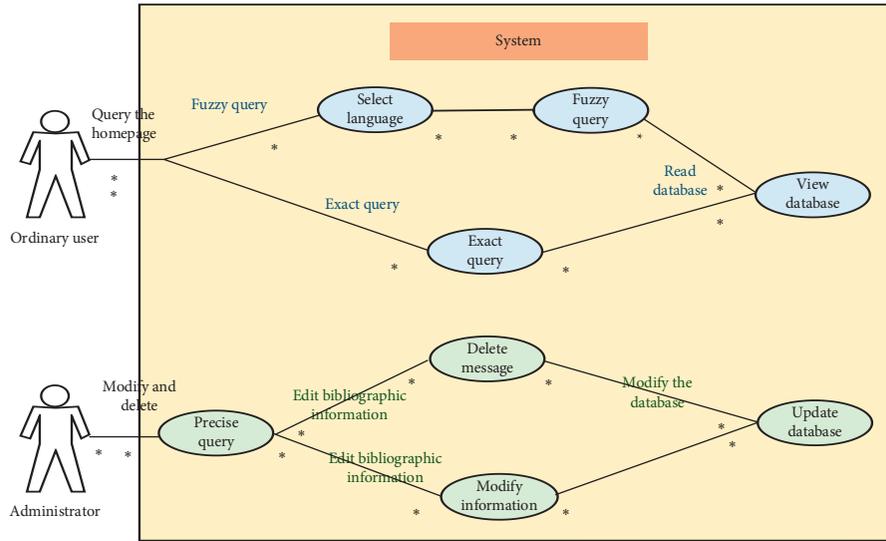


FIGURE 8: System user use case diagram.

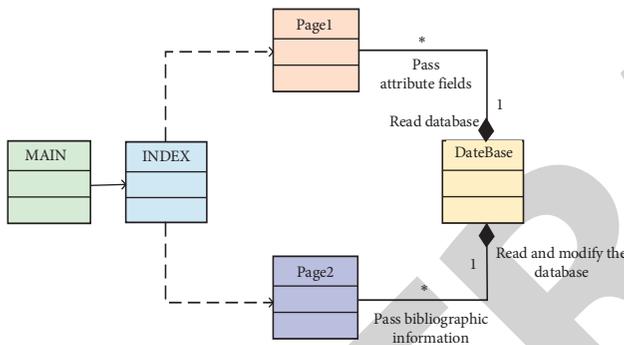


FIGURE 9: UML class diagram.

layer from resources to services, and from bottom to top by continuously deepening the humanistic knowledge service. From the perspective of user-friendly query, we can provide users with the information from the perspective of user-friendly query. The second problem is to provide users with a convenient query interface.

6.4. Retrieval Method.

- (1) Forward consistent retrieval. When the user enters any phrase in the specified search input port, the system will display all records starting with the phrase in the corresponding fields of the database. For example, if you enter "Jilin Province" in the "title" input box, the system will display all records whose "title" begins with "Jilin Province".
- (2) Arbitrary word retrieval. When the user enters any phrase in the specified search input port, the system will display all records containing the phrase in the corresponding fields of the database. Any content can be entered in the search box for search, including book name, author, keyword and other information. Advanced search can be used between multiple search items. The search result page can realize

cluster search, and can see the definition, word network and item information contained in the keyword. For example, enter "Jiutai" in the "title" input box, and the system will display all records containing "Jiutai" in the "title".

- (3) Completely consistent retrieval. When the user enters any phrase in the specified search input port, the system will display all records containing only the phrase in the corresponding fields of the database. For example, if "Xin'an Medical Books" is entered in the "title" input box, the system will display all records whose "title" is "Xin'an Medical Books".
- (4) Full text search. The system can retrieve any phrase or phrase combination in a field of the database and display their records. For example, enter "Changchun City" + "Jiutai" in the "content introduction" search entry, and the system will display the records containing "Changchun City" and "Jiutai" in the content introduction, in which "Changchun City" and "Jiutai" are highlighted in other colors.
- (5) Classified retrieval. It contains clinical knowledge base, classic prescriptions, materia medica, health preservation library and famous doctor library. It contains medical figures from ancient legends to the late Qing Dynasty, which can be classified and searched.
- (6) Knowledge map analysis and presentation. Use knowledge map technology to realize visualization, support the analysis, presentation of word network, knowledge discovery and book academic development trend map.
- (7) Retrieval approach. The system can search a record through multiple search points, such as title, author, collection place, content introduction, or combine multiple search points through the logical relationship of "and", "or". Support the multi-channel

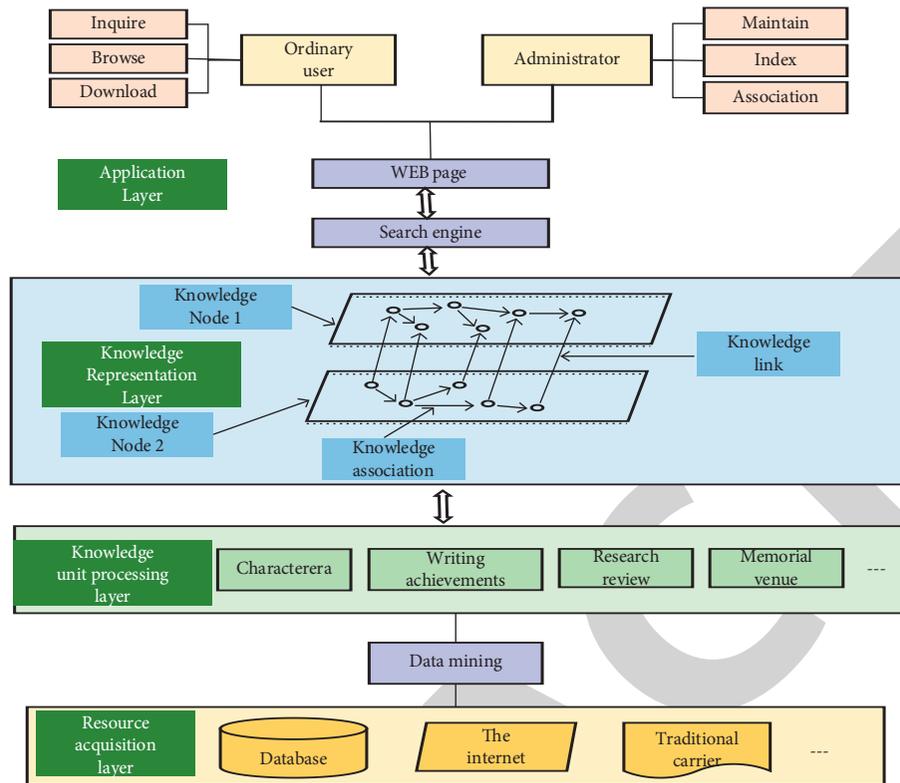


FIGURE 10: Database construction model based on knowledge map.

knowledge classification acquisition of medical books, can enter the conditions of any search item for combined search, and quickly find the required books or materials.

6.5. Streaming Media Technology. In the process of collecting TCMAB, there are relevant multimedia resources such as interpretation recording of ancient books, introduction videos of medicinal materials, prescriptions and characters. Streaming media technology integrates multimedia data with XML data to display TCMAB to users in an all-round way.

Streaming media video server is the center of media data storage and publishing. The storage capacity, number of concurrent users, stability and image quality of VOD are directly determined by the performance of streaming video server. PowerSSP streaming media video service system is based on a distributed architecture, which unifies PC streaming media, ipTV and mobile streaming media on one platform, adopts the platform structure of centralized governance and distributed services, realizes CDN content distribution, supports hierarchical program distribution and storage, adopts more optimized active and intelligent global content management and system load dynamic balance.

PowerSSP streaming media video service platform includes three layers: streaming media service layer (PowerMedia), CDN content distribution layer (PowerCDN) and customer layer. The architecture of PowerSSP streaming media video service platform is shown in Figure 11.

PowerSSP adopts loose coupling structure to realize interconnection among streaming media service layer, CDN content distribution layer and customer service layer through application program interface. CDN Content distribution layer can not only provide CDN content distribution, but also realize CDN content distribution to third-party streaming media system through API.

6.6. Function Design of Background Data Processing System. The background data processing system platform of TCMAB bibliography database is divided into four functional modules: document database maintenance, picture database maintenance, character database maintenance and audio and video database maintenance. The functions of adding, deleting and modifying document information, picture information, character information and audio video information are realized respectively. The data entry personnel log in to different function modules according to different permissions. When adding data, the system first checks the duplicate, that is, judge whether there is duplication according to the entered book name, and then judge whether to enter the next entry operation according to the query results. The specific data processing flow is shown in Figure 12.

The background data sorting of the application system can automatically realize the background data sorting without manual intervention. After setting the specific content and time interval of background data sorting, the system will process it automatically and regularly without

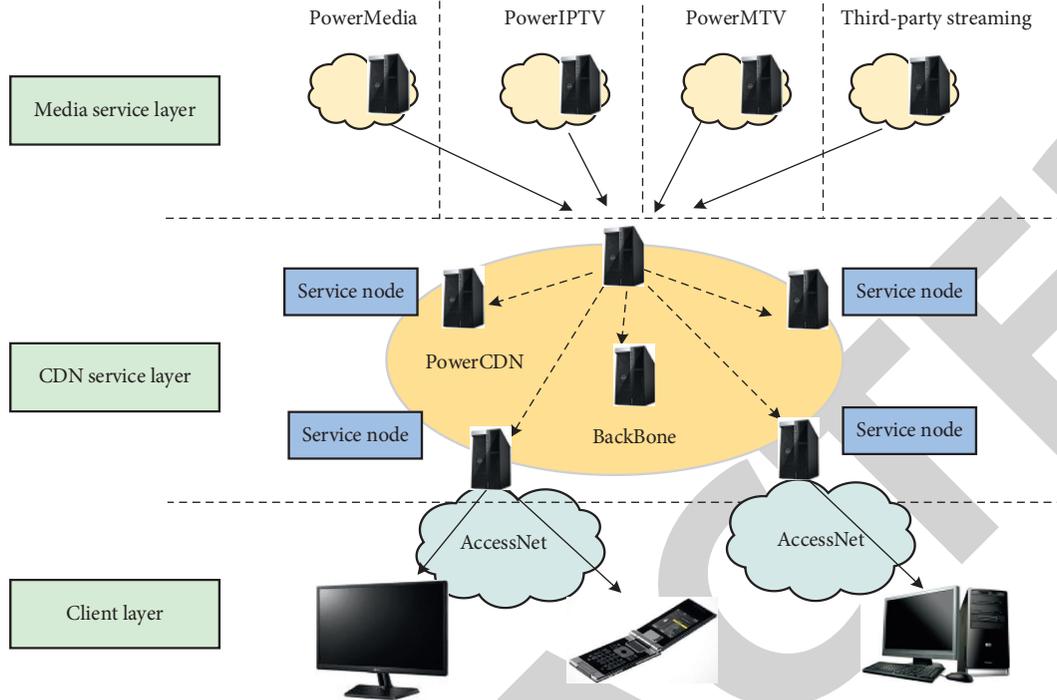


FIGURE 11: Streaming media video service platform architecture.

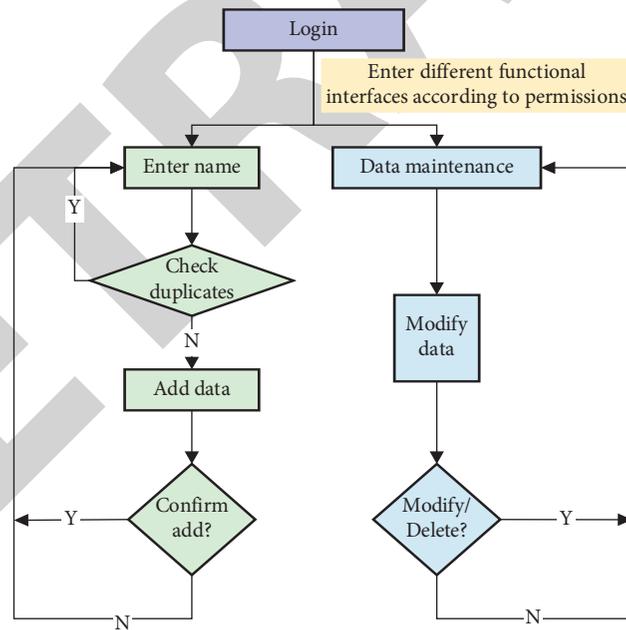


FIGURE 12: Background data processing flow of database.

affecting the foreground application and manual confirmation, it is convenient for users to manage. The content of data processing can change with the change of users' needs. The time of background data storage can be easily controlled from the foreground interface, which is convenient for users to manage. It can also effectively sort out the data. The outdated garbage data can be removed in time, which can adapt to the cleaning of various data quantities. No matter

how the content of data sorting changes, it can ensure the effectiveness of data sorting, automatic dump and storage of historical data. In order to ensure the query and calculation speed of the application system, data storage is divided into two parts: daily database and historical database. Background data sorting should automatically realize the dump of daily data from daily database to historical database, and ensure the integrity and consistency of data logical

relationship. The background data sorting of the database job mechanism realizes the automatic scheduling and execution of the job. The user modifies the data saving time in the client foreground interface, which is actually the content of the database job. Each time the user modifies the data saving time, the system automatically resubmits the database job, and the content of the data sorting can be changed according to the new needs of the user, so as to realize the purpose of controlling data sorting from the client.

7. Conclusion and Future Work

The construction of TCMAB knowledge base has become one of the future development directions of ancient books digitization. The means of sorting TCMAB by using modern information technology is gradually improving, the knowledge organization and processing of TCMAB have become more detailed, and the computer expression of TCMAB knowledge has also made a creative breakthrough. The establishment of a unified and standardized bibliographic database to realize the exchange and resource sharing of TCMAB bibliography is also the inevitable trend of TCMAB bibliography digital service. Relying on knowledge mining and other technologies, this paper establishes a bibliographic summary database of TCMAB with unified standards, rules and formats, and realizes the exchange of TCMAB through online resource sharing. The bibliographic database of TCMAB is constantly adjusted, modified and improved in the process of building the database. The field description information suitable for non-professional cataloguers is set, which can be input at one time and searched in multiple directions. This is a centralized sorting and bibliographic information mining of the existing TCMAB, which facilitates network retrieval, improves the utilization rate of bibliographic resources of TCMAB. The research results of this paper is practical and valuable for comprehensively understanding, obtaining version information, discovering and mining knowledge of traditional Chinese medicine.

Data Availability

The simulation experiment data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Scientific Research Project of the Education Department of Anhui Province (Grant no. KJ2021A0903), And The authors wish to thank the anonymous reviewers who helped to improve the quality of the paper.

References

- [1] F. Xiong and G. Song, "Investigation and research on the construction of digital resources of ancient Chinese medicine books in colleges and universities," *Inner Mongolia Science Technology & Economy*, vol. 7, no. 1 5, pp. 97–107, 2020.
- [2] X. Cao and L. Pei, "Practice analysis and countermeasures of digitization construction of ancient books of traditional Chinese medicine," *Library Science Research*, vol. 16, no. 13, pp. 42–44, 2016.
- [3] W. Shu, K. Cai, and N. N. Xiong, "Research on strong agile response task scheduling optimization enhancement with optimal resource usage in green cloud computing," *Future Generation Computer Systems*, vol. 124, pp. 12–20, 2021.
- [4] J. Yang, J. Liu, R. Han, and J. Wu, "Transferable face image privacy protection based on federated learning and ensemble models," *Complex & Intelligent Systems*, vol. 7, no. 5, pp. 2299–2315, 2021.
- [5] W. W. X. Xia, M. Wozniak, X. Fan, R. Damaševičius, and Y. Li, "Multi-sink distributed power control algorithm for Cyber-physical-systems in coal mine tunnels," *Computer Networks*, vol. 161, pp. 210–219, 2019.
- [6] Z. J. Liu, "Problems and countermeasures of digitization of ancient books," *Library Work and Study*, vol. 10, pp. 50–52, 2019.
- [7] Y. Sun, C. Xu, G. Li et al., "Intelligent human computer interaction based on non redundant EMG signal," *Alexandria Engineering Journal*, vol. 59, no. 3, pp. 1149–1157, 2020.
- [8] W. Shu and Y. Li, "A novel demand-responsive customized bus based on improved ant colony optimization and clustering algorithms," *IEEE Transactions on Intelligent Transportation Systems*, vol. 2022, 2022.
- [9] L. Wang, X. Qi, and S. Yu, "Digital research of traditional Chinese medicine literature," *Journal of Traditional Chinese Medicine Information*, vol. 22, no. 10, 2015.
- [10] U. Akhtar, A. Sant'Anna, and S. Lee, "A dynamic, cost-aware, optimized maintenance policy for interactive exploration of linked data," *Applied Sciences-Basel*, vol. 9, no. 22, 2019.
- [11] S. Chun, J. Jung, and K.-H. Lee, "Proactive policy for efficiently updating join views on continuous queries over data streams and linked data," *IEEE Access*, vol. 7, pp. 86226–86241, 2019.
- [12] J. Hernandez, H. M. Marin-Castro, and M. Morales-Sandoval, "WebQuln-LD: a method of integrating web query interfaces based on linked data," *IEEE Access*, vol. 9, pp. 115664–115675, 2021.
- [13] S. Jeter, C. Rock, B. Benyo et al., "Semantic links across distributed heterogeneous data," in *Proceedings of the Distributed Computing and Artificial Intelligence, 16th International Conference*, vol. 1003, pp. 107–115, Ávila, Spain, June 2020.
- [14] P. Chhaya, C.-H. Choi, K.-H. Lee, W.-S. Cho, and Y.-S. Lee, "KMLoD: linked open data service for Korean medical database," *The Journal of Supercomputing*, vol. 76, no. 10, pp. 7758–7776, 2020.
- [15] J. L. Sánchez-Cervantes, L. O. Colombo-Mendoza, and G. Alor-Hernández, "LINDASearch: a faceted search system for linked open datasets," *Wireless Networks*, vol. 26, no. 8, pp. 5645–5663, 2020.
- [16] A. D. Goncalves and M. D. D. Jacyntho, "A method for linked data semantic publishing of conventional databases and a real case study of academic papers," *Transinformacao*, vol. 32, 2020.

- [17] C. Xi and S. Qin, "The design and construction based on the ASEAN piano music library and display platform," *Journal of Intelligent and Fuzzy Systems*, vol. 35, no. 3, pp. 2861–2866, 2018.
- [18] H. Döring and S. Regel, "Party facts: a database of political parties worldwide," *Party Politics*, vol. 25, no. 2, pp. 97–109, 2019.
- [19] S. E. Bosch and M. Griesel, "Exploring the documentation and preservation of African indigenous knowledge in a digital lexical database," *Lexikos*, vol. 30, pp. 1–28, 2020.
- [20] Y. Qian, Z. Xing, and X. Shi, "From collection resources to intelligent data: construction of intelligent digital humanities platform for local historical documents of Shanghai Jiao Tong university," *Digital Scholarship in The Humanities*, vol. 36, no. 2, pp. 439–448, 2021.
- [21] S. Zhao, M. Tang, and Y. Sun, "Digital projects of Chinese historical local private documents: database development and exploring of text mining," *Library Trends*, vol. 69, no. 1, pp. 164–176, 2020.
- [22] Y. Liu, D. Jiang, B. Tao et al., "Grasping posture of humanoid manipulator based on target shape analysis and force closure," *Alexandria Engineering Journal*, vol. 61, no. 5, pp. 3959–3969, 2022.
- [23] Y. Wang and S. Geng, "Li Ning Digitization and data management of Dongba ancient book resources," *Journal of Electronic Measurement and Instrument*, vol. 31, no. 4, pp. 636–643, 2017.