Research Article

Construction and Optimization of Distributed Electronic Archives System Merging Control Messages

Yang Zhao and Weiming Du [ID]

Wuhan Huaxia Institute of Technology, Wuhan 430223, Hubei, China

Correspondence should be addressed to Weiming Du; 12081177@wit.edu.cn

This paper analyzes the whole of the electronic archives management system, describes the basic features and functions of the system, and gives the flow of each functional module of the system. Techniques for implementing important functions in the system are discussed and compared. Using the object-oriented idea, this paper designs the electronic archives management system and gives the hybrid structure model of the system. According to the needs of the electronic office, an official document transmission module is established to realize the electronic transmission of official documents, and the specific implementation process is given. According to the characteristics of archives information management and utilization, SQL Server is used as the background database and the database design is carried out. This paper introduces the full-text retrieval technology that realizes the query and retrieval function of electronic archives and gives how to configure the full-text retrieval service in SQL Server. This paper proposes a K-means-based merging control message clustering algorithm to solve the problem of communication topology changes caused by large-scale, high-mobility merging control messages. Based on the idea of K-means clustering, the calculation method of distance in K-means is improved by combining the distribution of the initial electronic archives message, and the fitness function of the electronic archives is designed by using the information of the electronic archives in the cluster and the average distance to the neighboring electronic archives. According to the maximum fitness value, the cluster head is selected to establish clustering. The file management system is tested and analyzed in a simulated environment. Based on the functional positioning of the system and the business environment, the functional performance and performance of the analysis system are examined and analyzed, and the research and development results are inspected according to the system development goals to check whether the development results of the system have reached the advanced stage. At the functional level, it can automate the management of the core business of the distributed electronic archives management department and support the distributed storage management of massive archive data, which effectively improves the informatization level and development efficiency of archives management.

1. Introduction

With the development of society, the number of documents in both enterprises and government departments has grown exponentially [1]. Traditional paper documents can no longer meet the needs of use and management. At the same time, due to the birth of information technology, electronic documents have begun to appear [2]. Under this background, the dual-system management mode of electronic records has gradually become popular, and a series of relevant policies have been introduced one after another, which

has played a certain role in promoting the implementation of the dual-system management of electronic records. It is undeniable that, in a certain historical period, the dual-system management of electronic files has its rationality, which effectively makes up for the shortcomings of the single paper version of document preservation in the past [3]. However, with the rapid development of information technology, the electronic archives preservation system has been gradually improved, and the development of its migration and backup technology to the improvement of security performance has provided the necessary guarantee for

the preservation of electronic archives [4]. In contrast, the problem of the dual-system management of electronic records has gradually been exposed, and people have gradually realized the negative effects of resource waste and duplication of work in the dual-system management of electronic records.

Under the background of the technological revolution, document management has gradually changed from paper to electronic form [5]. Electronic archives management has become the main method of document management and has gradually become the focus of scholars' research [6]. In the aspect of system management, there are relatively few studies on the single system management of electronic records, but the single system management mode of electronic records has been advocated in China [7]. Therefore, the research on the single system management strategy of electronic records in this paper has theoretical significance. Theoretically speaking, the current research on the management of electronic files is more focused on the "dual system," and few scholars have discussed the "single system" management model [8]. The research focus is to enrich the research content of electronic archives management, and it is also an extension of archives in the Internet age. In addition, the practical significance of this research cannot be ignored. For the archives management staff in our country, it is difficult to complete the work with high quality and quantity due to complicated work content and inadequate management, so it is urgent to improve the archives management work. In order to further promote the efficiency of archives management, the reform of archives management mode is carried out, and it is combined with information technology, which is beneficial to improve work efficiency and reduce management costs and promote the development of archives management mode.

Based on the on-the-spot understanding of distributed electronic archives management work, this paper establishes the main functions of the electronic archives management system, divides the system function modules, and describes these functions. The calculation mode of the electronic file system management system is a system structure combining B/S and C/S. According to the actual situation, the design principle of the database is described, and the large-scale database system of SQL Server is established to process the data of this management system. In order to balance the information of the electronic archives network, the calculation method of the distance in K-means is improved according to the distribution of the initial electronic archives messages, and based on this, the K-means algorithm is used to cluster the initial electronic archives. After the clustering is formed, the fitness function of the electronic archives is designed according to the information of the electronic archives in the clustering and the average distance to the neighboring electronic archives. After the clustering is established, the maintenance rules of the clustering are given. This paper develops and implements the functions of the file management system, briefly explains the development environment configuration of the system, and analyzes the implementation of the cloud storage service function in the background of the system in detail. On this basis, the

software function modules of the system are implemented. This paper develops and realizes the functions of the electronic archives management system in detail and introduces the testing work of the system. In the research process, a detailed development process and technical key points are introduced for the cloud service functions and internal function modules of the targeted system, and the actual operation effect of the system is shown. At the same time, the test environment is built according to the actual deployment configuration, and the function and performance test of the system are analyzed.

2. Related Work

Relevant scholars pointed out that the main components of the shared service archive system are the existing digital information resources of archives and network-related archive information resources [9]. You can obtain dynamic archive information services such as consultation, search, download, and transmission. The classic representative is the Digital Text Archive of the National Archives of the United States. The development of information technology has provided technical support for archives management departments in all walks of life at home and abroad. Archives management has been evolving from traditional paper to digital. Various archives management systems have also been developed to meet the different needs of archives management business [10]. These file systems have different technical routes and different implementation methods [11].

From the implementation mode, the current file information system can be divided into Client/Server mode and Browser/Server mode. Relevant scholars believe that the C/S file management system is generally used in file business departments with high professionalism and fewer users [12, 13]. In this way, the server usually adopts a high-performance PC, workstation, or minicomputer. The client needs to install special client software. The maintenance cost of the file management system developed in this way is high, and the load pressure of the client is large, so it is not suitable for large-scale file business management departments [14]. The B/S file management system concentrates the main function logic of the system into the server and only needs to use a browser on the client. This design hands over the transaction processing logic to the server, and the client is only responsible for display.

There are some misunderstandings in the traditional understanding of archives management. Archives have high social value, and for a long time archives are often stored in archives management institutions as a witness to history rather than a modern push. The idea and practice of "emphasizing possessions and neglecting utilization" have continued in the development of archives management. Often more emphasis is placed on the collection, arrangement, processing, and preservation of archives, with the security of archives and collections as the priority. In addition, there is no standardized and reasonable retrieval method, so looking up archives has become an extremely troublesome thing, which hinders the formation of archives.

Therefore, it is imperative to realize the digital archives information system.

In order to further improve archives management in the United States, the National Archives and Documents Administration has specially established an electronic archives management system, which uses a comprehensive digital management system for the permanent preservation of important documents [15]. Since the start of the program in the United States, it has been implemented nationwide, and hundreds of users have endorsed the system. This system is more and more recognized by people, so in order to provide people with better services, the United States has further optimized the system, trying to further improve the storage space and security of the system [16]. Therefore, the development of confidentiality management technology is carried out.

Due to the early start of foreign electronic records management, the research on electronic records management by foreign scholars also started earlier [17–19]. Foreign scholars have also done some research on the electronic archives management mode [20]. Australia and the United States adopt a centralized management model. In Australia, the National Archives conducts comprehensive management of electronic archives. The National Archives has developed Xena software to keep and manage all electronic archives through the software. The United States implements the electronic archives management project, and the National Archives is responsible for the permanent management of archives and the maintenance and management of the National Archives management system [21, 22]. The integrated management mode of documents in western countries makes the file management feature centralized and improves the efficiency of file management.

3. Method

3.1. System Requirements Analysis. File management has formed a stricter management process, and the file inventory is relatively clear. However, the number of documents and archives is large, the classification method and the method of file grouping have changed many times, and there are many departments involved in the archives, and each department has its own catalog for receiving and transferring archives. They lack a unified interface, there are no interconnected information channels, and data is usually sealed on the computers of various departments, resulting in duplication of work and increasing the probability of errors, so management is difficult to a certain extent.

The electronic archives management system is based on the universal application of modern information technology. It is an ultra-large-scale, distributed digital information system that provides reliable information services for enterprise production, management, and decision-making.

According to the field investigation of distributed electronic archives management, the main functions of the electronic archives management system are as follows. (1) It realizes the computerization of archives description, to provide data entry, data conversion and collection, and arrangement. It can automatically convert the file entries entered by the user in the past and can also link with other systems such as office automation systems and MIS through e-mail or system modules, so as to solve the problem of transferring from file flow to the classified filing. (2) It realizes the function of computer-aided archives classification and cataloging, and the classification and cataloging conform to the provisions of the National Archives standard and the classification method of the electric power industry. (3) It provides the functions of identification, destruction, retrieval, and statistics of files and can print out various forms. (4) It realizes the network query, borrowing, return, and reminder of files and conducts statistical analysis on file query and utilization. (5) It realizes the maintenance of archive data and the management of user rights allocation and export and backup archive data to CDs, tapes, and other media for long-term preservation.

Through these functions, the system enables users to complete all aspects of file management on the computer network. The functional structure of the electronic file management system is shown in Figure 1.

3.2. System Design. According to the analysis of the business process and function of the electronic archives management system, the scheme adopted by the system is a mixture of two-layer and three-layer structures combining WWW technology and Client/Server (C/S) structure. Specifically, the business processing of archivists sends data to the main server for processing through a two-layer structure, while the retrieval query of ordinary users sends a request to the Web server through the browser, conducts the query in the main database, and returns the query result to the user.

The essence of Browser/Server (B/S) mode is a three-tier C/S mode, which is divided into three layers: browser, Web server (application server), and database server. This model centralizes all application logic on the server-side and retrieves information from the Web server on the client through an intuitive, easy-to-use browser. The Web server establishes hypertext links to internal pages and related back-end databases via HTTP, so the information on all Web servers can finally be queried with a browser. The browser constitutes the presentation layer of the system, the Web server and the application server constitute the application layer of the system, and the database server constitutes the data layer of the system. Therefore, it can also be said that the information system of the Browser/Server mode is composed of the presentation layer, the application layer, and the data layer.

When enterprises and institutions build their own information systems, they should use the C/S mode in locations with high-security requirements, strong interaction, a large amount of data processing, and flexible data queries. The B/S mode is used in a wide area where the security and interactivity requirements are not high and the location is flexible, and the two modes are combined to give full play to their respective strengths. In this way, a safe, reliable, flexible, and convenient software system with high efficiency can be developed.
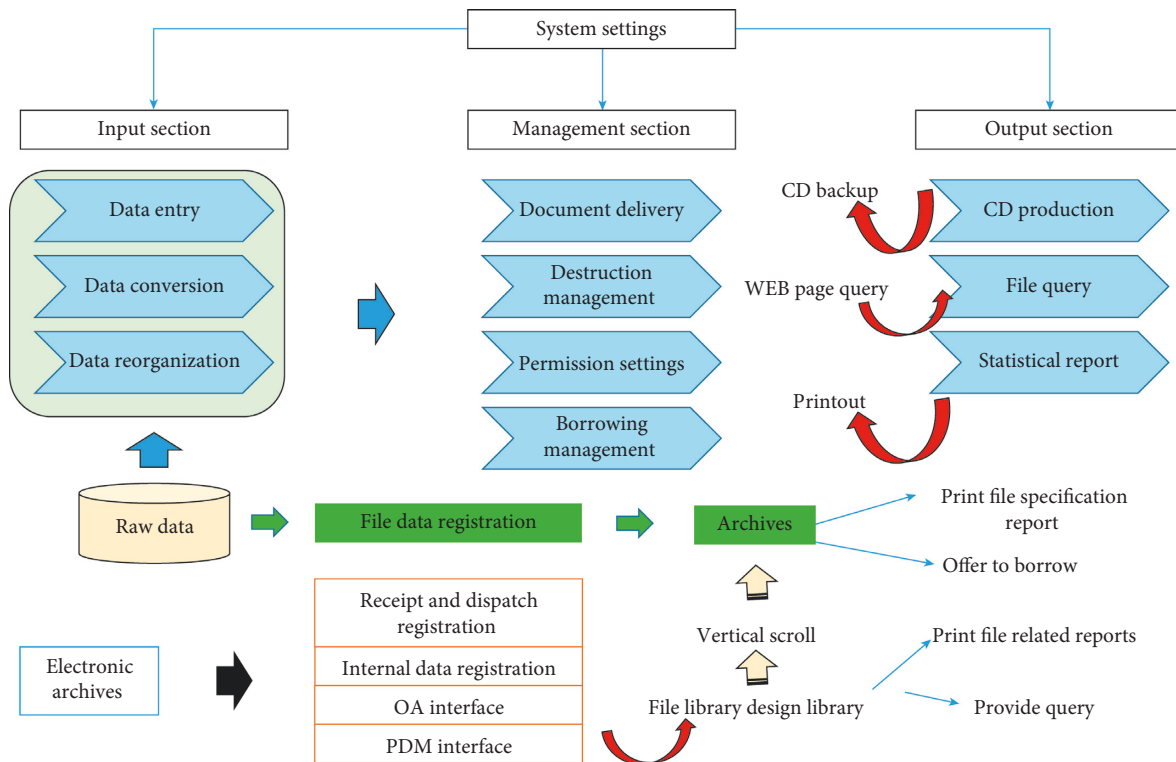
Figure 1: System function structure diagram.

This system is an electronic file management system based on the network environment. Through the above analysis, according to the requirements of the system target, the system structure combining C/S and B/S is selected.

3.2.1. Data Entry. By using the data entry subsystem, archives managers can enter, edit, and modify various archive data and complete the linking of archive texts with various electronic archive originals (scanned raster images or vector graphics).

Users can arbitrarily designate a record to copy when entering projects and files. The function of refiling after modification can replace the original file with the modified electronic file so that users can directly browse to the modified new electronic file when searching and querying. At the same time, the modification information and old files are recorded for the user to track the modification of the files.

3.2.2. Collection and Reorganization. Collection and reorganization provide users with retrieval, query, and utilization of unarchived documents. For the documents generated in the circulation of official documents and the drawing documents generated in the design process, the office staff, designers, or archivists can first enter them into the temporary compilation database, and at a certain stage, the archives management personnel will sort and catalog. After cataloging, it will be transferred from the temporary library to the official archive library for retrieval and utilization.

The archives or clerical personnel can use the operations of adding, editing, deleting, adding to the current volume, withdrawing the current volume, and unwinding and cataloging the organized files in the volume and the files to be cataloged provided by the system. All the information of the above unarchived files is placed in the unarchived library of the system, and the unarchived library can be searched and queried through the retrieval query subsystem.

3.2.3. Permission Settings. The authority setting subsystem enables the file management personnel to ensure the effective use of the files by the authorized users on the Internet and to limit the user's use rights to the files, which can also ensure the safety and reliability of the files. Through this module, file managers can make a series of flexible settings for each user and their permissions on the Internet, can refine the permissions to a certain file, and can also temporarily empower a certain user.

After the permissions are set, the file management personnel do not need to perform daily cumbersome retrieval, borrowing, and return work except for temporary empowerment but only need to regularly collect statistics on the use of files. Users themselves do it on the network.

3.2.4. Identification and Destruction. The appraisal and destruction management of archives is to destroy the old archives that have been authenticated according to the archive storage period, that is, delete them from the archives, and record the destruction information in the system log to facilitate retrieval of previously destroyed operations.

Among them, the leader identification needs to print the list and submit it manually and form a paper record, and other functions can be realized through the system.

3.2.5. Borrowing Management.

Borrowing management is the process of reviewing and approving the borrowing application submitted by the user, calling out the original file, or opening the holographic information for the user to read and use until it is recovered, mainly for the utilization of paper media files. Features include borrowing, returning, recalling, and printing borrowing and recalling slips. It can also conveniently count and print the borrowing situation of users, as well as online users' inquiries and utilization of files. It mainly completes the borrowing and registration of files, so as to manage the borrowing of users. The main functions of borrowing management include borrowing management, return management, recall management, borrowing statistics, and retrieval statistics. Figure 2 shows the schematic diagram of distributed electronic file borrowing management.

3.2.6. System Data Maintenance.

The system data maintenance subsystem is to solve the problem of the backup and recovery of the electronic file data by the file management personnel. Users can perform the backup and recovery of the file data on their own clients without operating the database system on the server. This provides a good guarantee for the security of archive data.

The data software system itself can back up any file information in the system (including file type, data table structure, defined information description interface, entry information such as file files and files, and statistical reports). Later, users can restore archive information by backing up files when needed.

3.3. Database Design.

Application development mainly uses the default account of the database management system, but in the actual application of the system, one or more login identifications (accounts) of data management personnel and application personnel should be established, and necessary management rules and restraint mechanisms should be established. The interconnection technology of the Web and database is the key to the establishment of this system. The system uses the database connector that comes with Windows to realize the interconnection between the Web and the database. The database design is as follows:

(1) According to the integration of data management and the data management plan, the entire application system builds a database to meet the application system's requirements for the database; that is, the designed data serves the application system. Of course, the database will also react to the application system. Other applications of derivable databases are not described.

(2) In the definition of data tables and fields, the first pinyin letter combination of Chinese characters is used to express their meanings.

(3) The data adopts SQL Server2010 large database system to process the data of this management system.

(4) For the security application design of the system, the default rules of the data system are used for processing.

(5) In the definition of the table, it mainly highlights the name and meaning of the table, field name, Chinese meaning, field type, field length, number of digits after the decimal point in numeric fields, whether null values are allowed, the associated table and necessary comments, and other content descriptions. A cell with no value in the table indicates that the current item does not have this property.

Open Database Connectivity (ODBC) is a set of solutions for reading databases developed by Microsoft. Its purpose is to hide all the underlying operations on the database in the ODBC driver kernel. For programmers, as long as they build a connection to the database, they can use a unified Application Program Interface (API) to read and write to the database, regardless of which manufacturer the database comes from and what format it uses.

For users who use Windows operating system, the system is equipped with ODBC data source administrator, through which ADO (Access Data Objects) can be set in the system. It is specially designed for the database access method OLE DB and is used for the Internet. It includes many objects, such as Connection, Recordset, and Command. ASP achieves database access through these objects. OLE DB is a low-level interface that uses a common data access specification to handle various types of data, regardless of their format and storage method.

3.4. K-Means Clustering Algorithm.

K-means clustering algorithm is the simplest unsupervised learning algorithm commonly used in data mining to solve clustering problems. The algorithm has the characteristics of fast calculation speed, uniform clustering, and easy implementation. However, since the k centers of the initial clustering of the algorithm are randomly selected, the effect of clustering is greatly affected by the selection of the initial clustering center. However, for scenarios that do not require very high clustering effects, such as the merge control message scenario in this paper, the number of clusters in the formation can be defined before the formation, which is beneficial to the formulation and planning of the merge control message task. The general steps of the K-means clustering algorithm are as follows:

(1) The algorithm first randomly selects k samples as the initial cluster center.

(2) Calculate the distance (such as Euclidean distance) from each sample to the cluster center, and assign each sample to the cluster corresponding to the nearest cluster center.

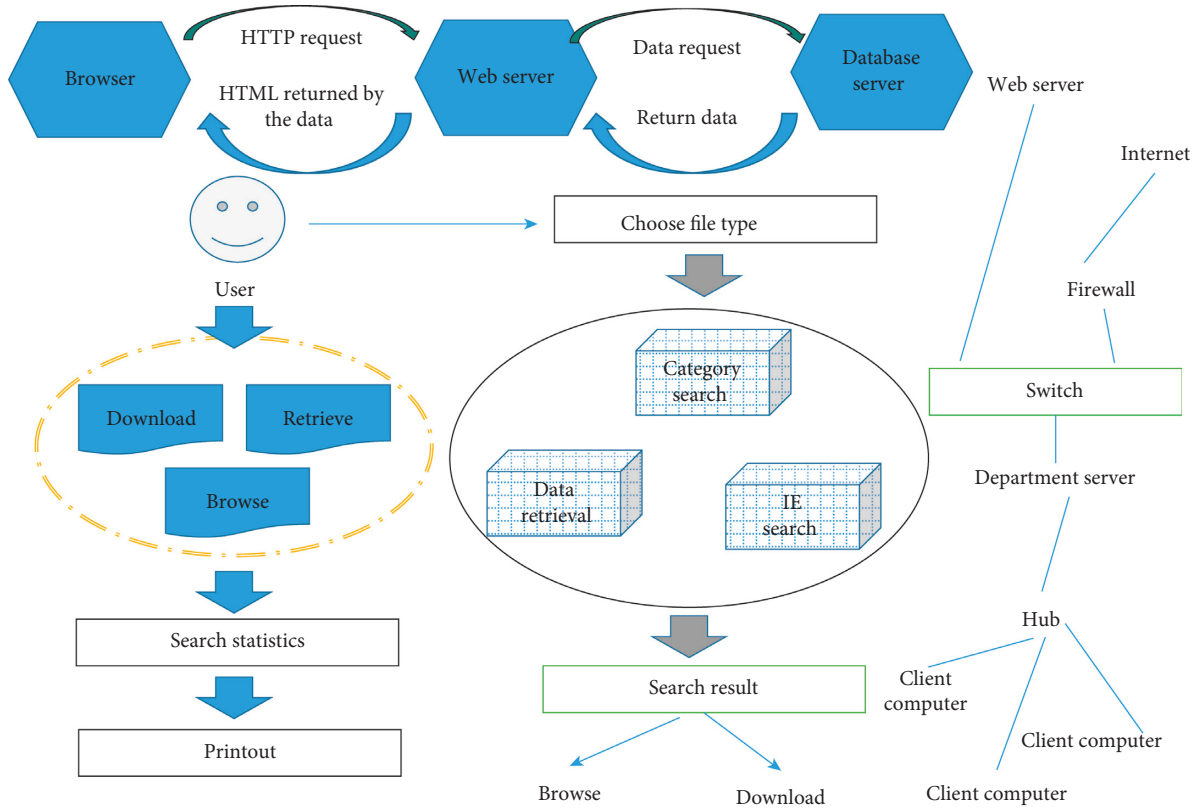(3) Use the updated samples of each cluster to recalculate the cluster center of the cluster, as shown in

FIGURE 2: Schematic diagram of distributed electronic file borrowing management.

the following formula, where m is the number of samples of the ith cluster.

$$c_i = \frac{1}{m-1} \prod_{j=0}^{m-1} (x_i + x_{i+1}). \tag{1}$$

(4) Calculate the sum of squares error of the cluster center and the original sample:

$$e = \prod_{i}^{n} \prod_{j}^{k} (x_i - c_i)^2. \tag{2}$$

(5) If the square sum error reaches the set threshold, the clustering is completed, and the cluster center and the clustering label corresponding to each sample are output, otherwise, go back to step (2), and continue to iterate until the conditions are met.

Figure 3 shows the effect of the K-means clustering algorithm for clustering. It can be seen that the distribution of the clustering centers is relatively uniform, so when it is used to merge the control message clustering, it can avoid the danger of being too dense among the clusters.

3.5. Clustering Optimization Algorithm for Merging Control Messages Based on K-Means. Different from the existing research on clustering algorithms, the clustering algorithm in this paper is not only used for network clustering

management of merged control messages but also used for clustering management of merged control message formation. The clustering of the merge control message refers to dividing the adjacent electronic files in the merge control message into several groups, and each group is called a cluster. Cluster members can select the cluster head through defined rules. The cluster head is responsible for maintaining the formation structure and communication topology of the cluster and forwarding the communication data of the electronic archives in the cluster. Among them, the communication link between cluster members and the communication link between cluster heads are relatively independent. Therefore, the information exchange of electronic archives between different clusters needs to be forwarded through the cluster head.

The K-means-based clustering algorithm for merging control messages designed in this paper is mainly divided into three parts, including the formation of clusters, the selection of cluster heads, and the maintenance of clusters. The establishment of clustering means that the merge control message is used to cluster the electronic file cluster based on the initial position and message of the electronic file before the formation starts, and the K-means algorithm is used to form a cluster. The initial cluster can be handed over to the ground control center or merged. The control center of the control message is carried out. After the clustering is formed, it is necessary to select the appropriate cluster head to manage the clustering and establish connections with other cluster heads to ensure the intercluster
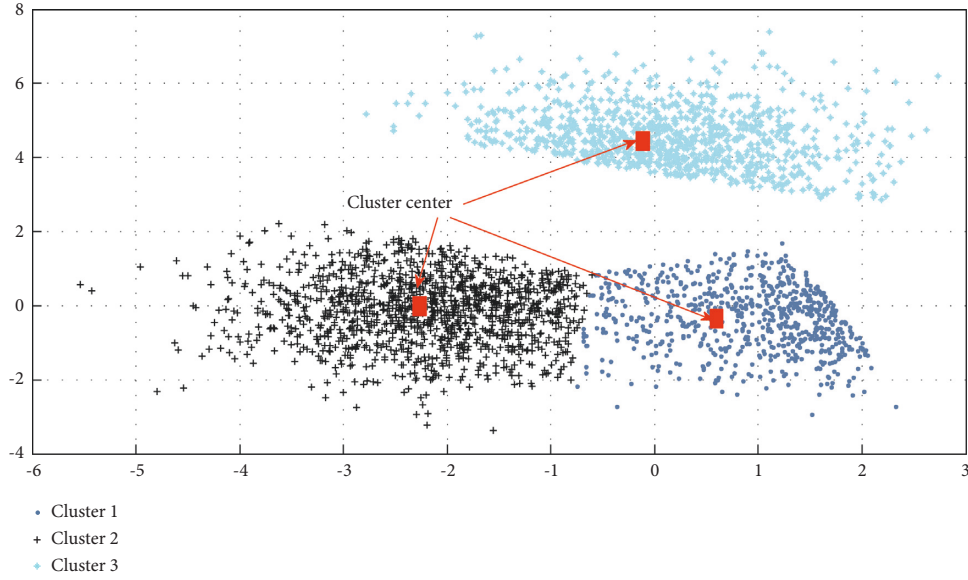
FIGURE 3: K-means algorithm clustering results.

communication of electronic archives. After the cluster head is determined, it is necessary to perform periodic maintenance on the cluster, for example, update the communication topology of the cluster and the formation of the electronic file when the electronic file is disconnected and added, and the electronic file link is disconnected. When the cluster head message is insufficient or the cluster members are too few, the clustering needs to be adjusted. Moreover, the formation of electronic archives is closely related to the clustering of the electronic archives network. Whenever there is a clustering that needs to be maintained and adjusted, the merge control message also needs to be maintained and adjusted accordingly to ensure that the electronic archives network and the merge control message are in the cluster.

3.5.1. Formation of Clusters.

For ad hoc networks that merge control messages, message balancing of the network is crucial to improve the lifetime of the merged control message network. When we perform clustering, we combine the remaining messages of the electronic archives to form appropriate clusters, so as to balance the messages of the electronic archives network. Define the remaining messages of the electronic archive:

$$E = (l - 1)E_0 - sE_c + E_m. \tag{3}$$

Among them, E_0 is the remaining messages of the previous control cycle of the electronic file, E_c represents the messages consumed by the electronic file to maintain a communication link in each control cycle, and l represents the number of links maintained by the electronic file. E_m represents the message consumed by the electronic file moving unit distance, and s represents the distance moved in the electronic file cycle.

In order to make the K-means algorithm make the information of the electronic file network more balanced when the electronic files are clustered, the following improvements

are made to the calculation formula of the distance between the electronic file and the cluster center in the K-means algorithm:

$$D = (1 - \alpha)d_{ij} + (1 - \beta)\sigma_{ij}. \tag{4}$$

Among them, d_{ij} represents the Euclidean distance between cluster center i and cluster center j, and α and β are weighting coefficients. σ_{ij} is the variance of the average message of all clusters after electronic file i is added to cluster j:

$$\sigma_j = \frac{1}{k - 1} \prod_{i=0}^{k-1} (e_i - E_c)^2, \tag{5}$$

where e_i is the average message of cluster i. Considering the actual situation of merging control messages, the number of electronic files in each cluster is often limited. Therefore, for the sake of simplicity, the maximum number of electronic files per cluster of merge control messages is defined as $\deg_{max}$, and the maximum number of electronic files that can be accommodated is $\deg_{max}$. The number of files is N_{max}; then, the maximum number of clusters of the K-means algorithm is

$$k_{max} = \frac{(\deg_{max} - N_{max})}{E_m}. \tag{6}$$

In the process of merging control messages, because the merging control messages change dynamically, electronic files are often added and removed, so the number of electronic file clusters also needs to be adjusted accordingly to the change of the formation. Next, the clustering algorithm of merge control message based on K-means is given. In the algorithm, the electronic files before the formation are divided into clusters by region and message, and the excessive number of clusters is adjusted so that the number of a single cluster is not greater than the given maximum number $\deg_{max}$.

3.5.2. Selection of Cluster Head. After the clustering is formed, it is necessary to select the appropriate cluster head in the clustering. The average distance from i in the cluster to other electronic files in the cluster is defined as d_{avg}, and the fitness function of i in the cluster is obtained through the average distance and the remaining messages:

$$F_i = \frac{\left(\lambda_1 d_{avg} E_c\right)}{\left(\lambda_2 E_i\right)}. \tag{7}$$

Among them, E_i is the remaining messages of i, and λ_1 and λ_2 are the weight coefficients. It can be seen from the design of the fitness function of the electronic file that the higher the message in the cluster, the closer the distance to other electronic files. The larger the value, the easier it is to be selected as the cluster head. The reason for choosing the electronic archives with higher messages as the cluster head is to prolong the life of the clustering of the electronic archives. The electronic archives with the smallest average distance are selected in order to establish clusters faster.

Other electronic files in the cluster automatically become members of the cluster, and after receiving the broadcast message from the cluster head, they should send a "Request" message to the cluster head to apply for joining the cluster. After all electronic archives are added to the cluster, CH will calculate the communication topology (including directed spanning tree) and formation of the current cluster and send it to the CM. The electronic archives in the cluster use the consistent formation designed in the previous section. After the cluster is established, CH will periodically maintain the cluster.

4. System Test Results and Analysis

4.1. System Test Environment and Method. The test environment of the file management system adopts the software and hardware environment shown in Table 1.

In the test work of the file management system, the black-box software testing method is adopted. By compiling corresponding test cases for each function of the system, the LoadRunner 11.1 tool is used to record the function test automation scripts of the system according to the test cases and record them in the system's Web server host. On the one hand, the automated test script is used to analyze and verify the functional performance of the system. On the other hand, the concurrent script running mechanism of the LoadRunner tool is used to examine the performance of the system under multiuser concurrent conditions.

For the functional test of the distributed file system in the system cloud, after completing the cloud hardware deployment and the installation of the Hadoop software service tools, we use the test tools provided in the Hadoop For.NET SDK development kit to log in to the master node in the cloud. Only the functions and performance of the system Web service function system are tested and analyzed in this section.

4.2. File Borrowing Management Test. The file borrowing management module implements management operations such as borrowing of internal files and at the same time

TABLE 1: System test software and hardware environment configuration.

Project	Environment configuration
Web services tools	Internet information service
Performance monitoring tool	Siege Web
Concurrency testing tool	LoadRunner
Web environment	Internal network
Hadoop platform	Hadoop Cloudera CDH
Client host	Development environment (4 units)
Cloud node host	Development environment (37 units)
Database system	SQL Server
Web browser	Internet Explorer/Chrome

maintains and manages the entire borrowing process based on the file borrowing list.

From an overall point of view, the functional logic of the file borrowing management module is implemented based on the file borrowing list data object maintained within the system. The system updates and maintains the file borrowing data recorded in the file list according to the specific event processing results and provides the file management personnel with functional support for file utilization statistics based on the maintenance results. Therefore, the function realization of the file borrowing management module mainly includes the application and approval of file borrowing, the maintenance of the file borrowing list, the management of file borrowing status, and the statistical management of file borrowing utilization rate.

In the process of function implementation, the borrowing application processing of electronic archives is mainly performed through the archiveApply () method in the custom class ArchiveBorrow. In this method, an archive borrowing request is created according to the borrowing request submitted by the end user and is sent to the archives manager for approval. The archive approval processing function is implemented through the audit Apply () method of the custom class ArchiveBorrow. In this method, by obtaining the file borrowing request, when the approval of the borrowing application fails, prompt information is fed back to the public host client.

In this system, all file borrowing operation information, including electronic files and physical files, is maintained and managed by means of a borrowing list. After the end user's file borrowing request is approved, the system records the borrowing list in the borrowing list. When the file is returned or expected, the system also updates the corresponding list data in the file borrowing list, so that the information in the borrowing list is consistent with the actual file borrowing request. Therefore, the file list borrowing function also includes the maintenance and management function logic of the file borrowing state.

The system counts and displays the overall filing status of all archives in the current system and displays the borrowing information of all expired archives in a list. File managers can query all expired borrowing data according to the borrower, file number, file name, borrowing time, borrowing

status, and other parameters. In addition, all the information on borrowing, returning, overdue, and renewal of all files is displayed in a list, and similar data retrieval function support is provided. The maintenance of the file borrowing list is shown in Figure 4.

The archive borrowing utilization statistics function is calculated according to the archive borrowing list recorded in the system. We use it as the data support for the archives filing management business to optimize and update the archives information within the archives management department. In the process of function implementation, the archive borrowing utilization statistics function is realized through the utilization statistics method archiveUsageRate () in the ArchiveBorrow class structure. We accumulate the times of borrowing and display it on the page in a graphical way, in which the utilization of all the borrowed files is given. The file utilization statistics are shown in Figure 5.

In the realization of the graphical function, it mainly uses the Crystal Report function component in the .NET platform to achieve, by maintaining the retrieval and statistical data of utilization in the form of DataSet data container and associating it with the CrystalReport created by the system. The crystal report component specifies the specific display format of the graphics and associates the crystal report object with the PictureBox control in the front-end Web page to realize the associated display of the utilization statistics results.

4.3. System Function Test. Taking the function of posting and adjusting the file table in the file information management module of the system as an example, the function test example is shown in Table 2.

According to the test cases of the functional modules of the system, we record the corresponding automated test scripts in the LoadRunner tool and execute the test scripts of each functional module in the four test client hosts, respectively. The obtained system functions perform as expected, so the system passed the functional test.

4.4. System Performance Test. In the performance test of the system, the automatic concurrency mechanism of the LoadRunner tool is mainly used, and the functional test script is executed concurrently in the client test host in an iterative and concurrent manner according to the concurrent user support capability of the system and is installed in the system Web server host. The performance monitoring point used in the system performance test is the management function in the file table in the file information management module of the system. The initial concurrency of iterations is 10, the interval between each iteration process is 30 seconds, and the number of test regressions is 5 times. We get performance test results from Siege on the system Web server host. The results of the first round of testing of
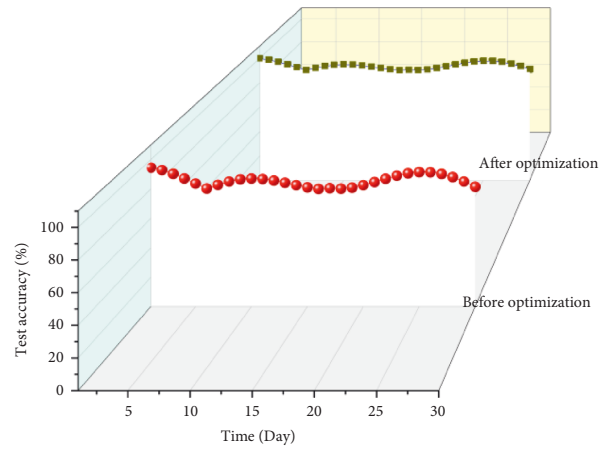

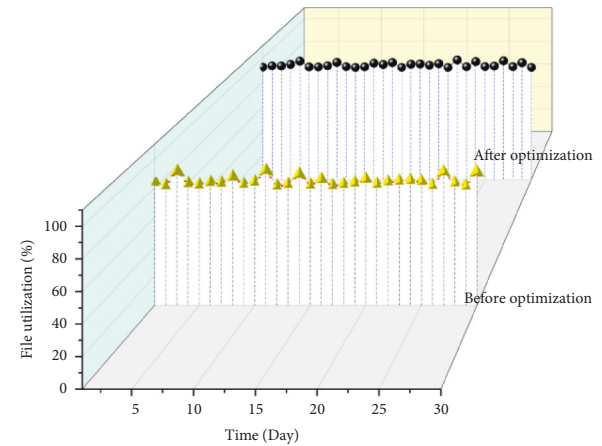FIGURE 4: Maintenance of file borrowing list.

FIGURE 5: Statistics of file utilization.

the performance of the file management system are shown in Figure 6.

According to the performance test monitoring data of the system, the system can reach the system pressure of 100 concurrent users within 20 minutes. During the 5 rounds of performance testing, the operating status of the system Web service is normal. At the same time, the maximum response time of all functional operations does not exceed 3 seconds, and the performance of the system in the simulated test environment is in line with expectations. Since the test environment of the system adopts the same hardware and software configuration and network configuration as the actual deployment environment, based on the above analysis, it can be concluded that the performance test of the system meets the expected requirements and the performance test passes. Figures 7 and 8 show the results of the second round of testing and the third round of testing of the performance of the file management system, respectively.

TABLE 2: Test cases for file information management function.

| Project/software	File management system development project
Testing purposes	Check the consistency of the operation and design of the archives information management function
Preconditions	Be able to log in to the system
Features	File sending, receiving, directory management, malicious code checking, etc.
Test data	File sending, receiving, file table data
Reference information	Archival information management module outline design and detailed design
Functional module	File information management module
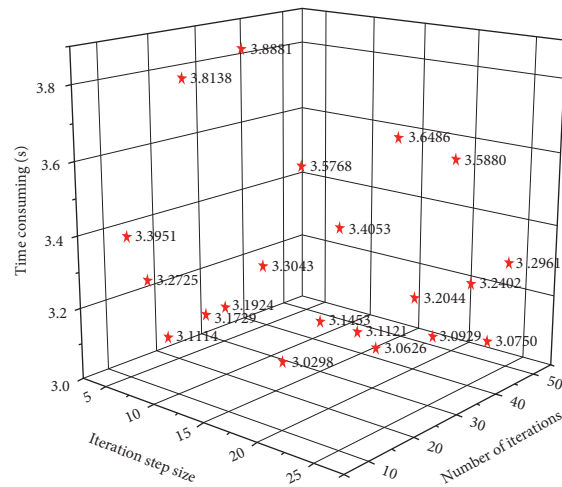
FIGURE 6: Results of the first round of testing of the performance of the archives management system.
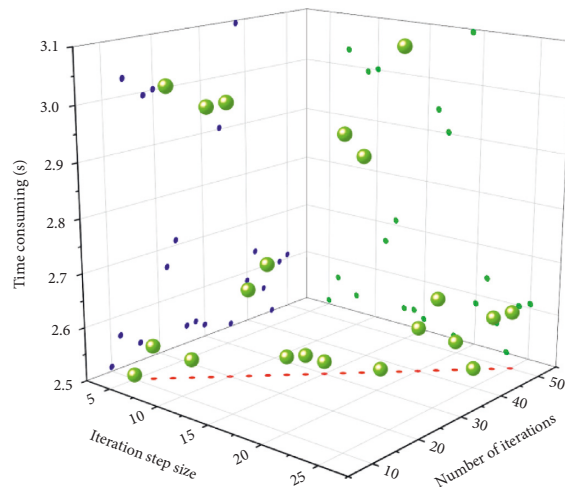

FIGURE 7: Results of the second round of testing of the performance of the archives management system.
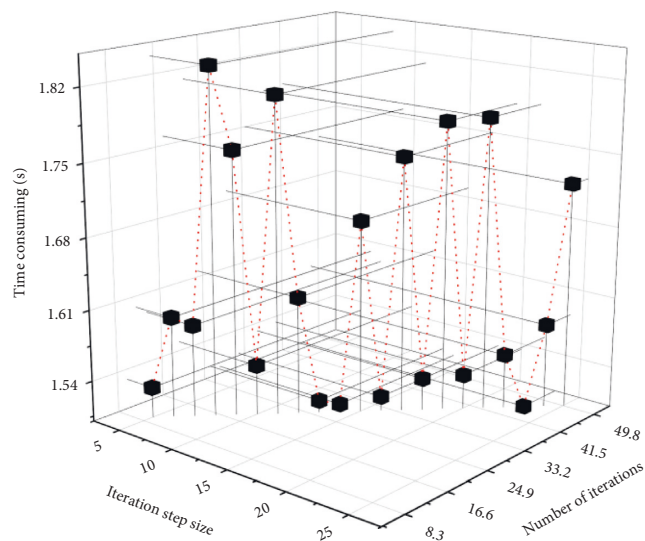
FIGURE 8: Results of the third round of testing of file management system performance.

5. Conclusion

In order to establish the electronic archives management system, the system analysis, system design, and data collection and arrangement are carried out in this paper. We completed the design and development of the database and system program and carried out system tests. The system improves work efficiency, ensures efficient, fast, safe, and convenient document processing, and greatly improves the office efficiency and management level of the entire network dispatching system. SQL Server is selected as the background relational database of the archives management information system, which can easily realize full-text retrieval of plain text documents and documents in various binary formats, which is very beneficial to archives management query and greatly improves the retrieval efficiency. According to the requirements of the new architecture, a set of common standard protocols is proposed to solve the heterogeneous and Web-oriented distributed computing Web service. We discuss the integration of archives management information systems and other management information systems through XML and Web service. The clustering algorithm in this paper makes the electronic archives with similar distances in the three-dimensional space form clusters and elects the cluster head in the clustering. Each subcluster is formed according to the formation given by the cluster head, and all the cluster heads are also formed according to a certain formation. As long as the proper positional relationship between the cluster head formation and the subcluster formation is maintained, a complete formation can be formed. In order to balance the information of the electronic archives network, the distance calculation method in the K-means algorithm is improved based on the distribution of the initial electronic archives messages, and the initial electronic archives are clustered. In this paper, the fitness function is designed by using the average distance of the neighbor electronic profiles in the cluster. According to

the maximum fitness value, the cluster head is selected to establish a cluster, and the maintenance rules such as cluster entry and exit, cluster head update, and merging are defined, respectively. The distributed electronic archives system adopts a distributed file system based on cloud technology in the storage of archives and documents, which can make full use of the idle information hardware resources of its colleges and can also effectively improve the system's ability to deal with massive archives data. At the same time, the functional system of the system can cover the daily business of the archives management department, provide convenient business operation support for archives management personnel and other users, and improve the implementation efficiency of the archives management business of the college.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Z. Chen, R. Wu, M. Liu et al., "Tunable electronic coupling of cobalt sulfide/carbon composites for optimizing oxygen evolution reaction activity," *Journal of Materials Chemistry*, vol. 6, no. 22, pp. 10304–10312, 2018.

[2] A. Scemama, Y. Garniron, M. Caffarel, and P.-F. Loos, "Deterministic construction of nodal surfaces within quantum Monte Carlo: the case of FeS," *Journal of Chemical Theory and Computation*, vol. 14, no. 3, pp. 1395–1402, 2018.

[3] J.-S. Chou and T.-K. Nguyen, "Forward forecast of stock price using sliding-window m-optimized machine-learning regression," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3132–3142, 2018.

[4] Z. You and C. Wu, "A framework for data-driven informatization of the construction company," *Advanced Engineering Informatics*, vol. 39, pp. 269–277, 2019.

[5] C. Jing, X. Song, K. Li et al., "Optimizing the rate capability of nickel cobalt phosphide nanowires on graphene oxide by the outer/inter-component synergistic effects," *Journal of Materials Chemistry*, vol. 8, no. 4, pp. 1697–1708, 2020.

[6] J. Zhao, H. Dong, H. Yang, and Y. Zheng, "Exploring and elaborating the novel excited state dynamical behavior of a bisflavonol system," *Organic Chemistry Frontiers*, vol. 5, no. 18, pp. 2710–2718, 2018.

[7] K. Yuan, D. Lützenkirchen-Hecht, L. Li et al., "Boosting oxygen reduction of single iron active sites via geometric and electronic engineering: nitrogen and phosphorus dual coordination," *Journal of the American Chemical Society*, vol. 142, no. 5, pp. 2404–2412, 2020.

[8] L. Chen, W.-K. Lee, C.-C. Chang, K.-K. R. Choo, and N. Zhang, "Blockchain based searchable encryption for electronic health record sharing," *Future Generation Computer Systems*, vol. 95, pp. 420–429, 2019.

[9] T. Meng, B. Li, Q. Wang et al., "Large-scale electric-field confined silicon with optimized charge-transfer kinetics and structural stability for high-rate lithium-ion batteries," *ACS Nano*, vol. 14, no. 6, pp. 7066–7076, 2020.

[10] N. Padhy, R. P. Singh, and S. C. Satapathy, "Software reusability metrics estimation: a," *Computers & Electrical Engineering*, vol. 69, pp. 653–668, 2018.

[11] D. S. Levine, D. Hait, N. M. Tubman, S. Lehtola, K. B. Whaley, and M. Head-Gordon, "CASSCF with extremely large active spaces using the adaptive sampling configuration interaction method," *Journal of Chemical Theory and Computation*, vol. 16, no. 4, pp. 2340–2354, 2020.

[12] M. Dash, S. Moroni, A. Scemama, and C. Filippi, "Perturbatively selected configuration-interaction wave functions for efficient geometry optimization in quantum Monte Carlo," *Journal of Chemical Theory and Computation*, vol. 14, no. 8, pp. 4176–4182, 2018.

[13] P. R. Nagy, G. Samu, and M. Kállay, "Optimization of the linear-scaling local natural orbital CCSD(T) method: improved algorithm and benchmark applications," *Journal of Chemical Theory and Computation*, vol. 14, no. 8, pp. 4193–4215, 2018.

[14] M. Xiao, Y. Chen, J. Zhu et al., "Climbing the apex of the ORR volcano plot via binuclear site construction: electronic and geometric engineering," *Journal of the American Chemical Society*, vol. 141, no. 44, pp. 17763–17770, 2019.

[15] S. Y. Cho, "A program optimization method for embedded software developed using open sources," *International Journal of Advanced Science, Engineering and Information Technology*, vol. 8, no. 4, pp. 1692–1697, 2018.

[16] P. Hou, J. Zhu, K. Ma, G. Yang, W. Hu, and Z. Chen, "A review of offshore wind farm layout optimization and electrical system design methods," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 5, pp. 975–986, 2019.

[17] X. Chang, C. Zhang, C. Lv et al., "Construction of a ma-based DNA logic device on live cell membranes via associative toehold activation for accurate cancer cell identification," *Journal of the American Chemical Society*, vol. 141, no. 32, pp. 12738–12743, 2019.

[18] C. V. Dinh, X. Chen, and K. L. J. Prather, "Development of a quorum-sensing based circuit for control of cpcn production system," *ACS Synthetic Biology*, vol. 9, no. 3, pp. 590–597, 2020.

[19] X. Li, X. Li, C. Liu et al., "Atomic-level construction of tensile-strained PdFe alloy surface toward highly efficient oxygen reduction e," *Nano Letters*, vol. 20, no. 2, pp. 1403–1409, 2020.

[20] H. Yang, L. Sun, Z. Wang, W. Li, G. Liu, and Y. Tang, "ADMETopt: a web server for admet optimization in drug design via scaffold hopping," *Journal of Chemical Information and Modeling*, vol. 58, no. 10, pp. 2051–2056, 2018.

[21] Y. Li, Y. Xia, K. Liu et al., "Constructing Fe-MOF-Derived Z-scheme pectn and carbon sheath synergistic effect," *ACS Applied Materials & Interfaces*, vol. 12, no. 22, pp. 25494–25502, 2020.

[22] F. Caschera, A. S. Karim, G. Gazzola, A. E. d'Aquino, N. H. Packard, and M. C. Jewett, "High-throughput optimization cycle of a cell-free ribosome assembly and protein synthesis system," *ACS Synthetic Biology*, vol. 7, no. 12, pp. 2841–2853, 2018.