Journal of Artificial Intelligence Practice (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2024.070320
ISSN 2371-8412 Vol. 7 Num. 3

Research on Holographic Retrieval and Analysis System for Scientific Research Data Based on SSH Framework and Lucene Engine

Boyang Liu[1,a,*]

[1]*Operation Department, ONUS Global Fulfillment Solutions, 7419 Nelson Rd Unit 130, Richmond, BC V6W 1G3, British Columbia Province, Canada*
[a]*p7908686@gmail.com*
[*]*Corresponding author*

Keywords: Scientific Research Data Management, Full-Text Search, Decision Support System, SSH Framework, Lucene Engine

Abstract: With the rapid growth of scientific research data, traditional data processing methods are no longer able to meet the needs of efficient retrieval and analysis. To address this challenge, this study designed and implemented a holographic retrieval and analysis system for scientific research data based on SSH framework and Lucene engine. The system relies on Oracle data warehouse and combines OLAP technology to achieve multi-dimensional data analysis and display; By using the Lucene full-text search engine, the efficiency and accuracy of data queries have been improved; And with the help of Mahout data mining framework, multiple algorithms are integrated to support deep mining of scientific research data. This study first analyzed the shortcomings of existing decision support systems and identified the core requirements of scientific research management systems. With the support of the SSH framework, the system has achieved efficient data storage, retrieval, analysis, and visualization, forming a complete scientific research data management and analysis solution. After testing, the system has shown high accuracy and stability. The research results indicate that the system significantly improves the efficiency and decision support capability of scientific research management. The development model based on open source technology not only reduces costs, but also enhances the scalability and maintainability of the system, with broad application prospects.

1. Introduction

With the rapid development of information technology, the level of informatization in scientific research management has significantly improved, and the amount of scientific research data has shown exponential growth. How to efficiently utilize these massive amounts of data to provide support for scientific research management and decision-making processes has become a core issue that urgently needs to be addressed. Traditional manual analysis and semi-automatic tools are inefficient in dealing with complex and massive scientific research data, making it difficult to meet the needs of modern scientific research management for efficient and intelligent data processing.

Decision support systems (DSS), as an important tool for assisting decision-making, have been widely applied in fields such as enterprise management and financial analysis, and have achieved significant results. However, in the field of scientific research management, especially in multi-source data integration, holographic retrieval, and deep analysis, applied research is still relatively lagging behind. To fill this gap, this project has designed and implemented a holographic retrieval and analysis system for scientific research data based on SSH framework and Lucene full-text search engine, aiming to enhance the scientificity and decision support capabilities of scientific research management.

This study relies on multiple major scientific research projects to construct a scientific research management database and data warehouse, and combines online analytical processing (OLAP) technology to achieve in-depth analysis of multidimensional data. The system has significantly improved the efficiency and accuracy of data retrieval by introducing the Lucene engine; Based on the Mahout framework, multiple data mining algorithms are integrated to achieve intelligent mining and pattern recognition of scientific research data, providing strong support for scientific research management decisions. The study first reviewed the current development status of decision support systems at home and abroad, identified the shortcomings of existing systems in scientific research management, and proposed system design goals based on the actual needs of scientific research management. Based on the SSH framework, the system has achieved full process optimization from data storage, retrieval to analysis and visualization, and constructed a comprehensive and scalable decision support framework. The system's functionality and stability have been verified through rigorous testing.

The results of this study not only significantly improve the efficiency and decision support level of scientific research management, but also provide scientific and technological solutions for the informationization construction of scientific research management, which has important theoretical significance and broad application prospects.

2. Related Research

In the context of the big data era, the surge of technological information poses a challenge to traditional single modal data retrieval systems, prompting us to need more advanced retrieval technologies. A cross media technology information retrieval system based on deep semantic features has emerged, which can more effectively process and analyze technology information from different sources. Y Jiang's research focuses on how to optimize the retrieval process through deep semantic features, providing theoretical support for the accuracy and adaptability of the system, demonstrating its importance in the trend of technological development [1].

In terms of improving the performance of holographic data storage systems, R Chen et al. proposed a technique that combines spectrum spreading and dynamic sampling [2]. This method not only increases data storage density, but also accelerates the extraction and decoding of phase information, demonstrating how to optimize holographic data storage systems through advanced spectral processing techniques. This study provides new ideas for improving the efficiency and accuracy of holographic data processing.

L Pensia's research has made breakthroughs in system design, proposing a new method that utilizes multifunctional holographic optical elements (MHOEs) to improve system resolution and expand the field of view [3]. The experimental results show that this method increases the field of view by 423%, providing significant performance improvements for holographic retrieval and analysis systems for scientific research data. This study emphasizes the potential of integrating optical components in hardware design to achieve more efficient data acquisition and analysis.

In terms of theoretical model application, T. R. Mongan provided a theoretical basis for the

application of elementary particle models in holographic retrieval and analysis of scientific research data by analogy between the energy density of electron neutrinos and the energy density of cosmic vacuum [4]. The application of this theory predicts the mass of neutrinos and provides a scientific explanatory framework for data models in the system.

The advancement of artificial intelligence has also significantly promoted the development of holographic retrieval and analysis systems for scientific research data. Hikov's research applies generative AI and large-scale language models to the retrieval and analysis of scientific research texts, achieving efficient information extraction and automated analysis by combining the Retrieval Enhanced Generative (RAG) architecture and prompt engineering. This indicates the enormous potential of AI technology in enhancing system intelligence and processing complex data tasks.

In the field of computational urban science, T. R. Mongan explored the application of computer vision technology, particularly convolutional neural networks (CNN), in real-time traffic data management [5]. The potential of this technology in holographic retrieval of scientific research data demonstrates the effectiveness of computer vision technology in processing large-scale data and real-time monitoring.

The automated equipment data retrieval and analysis platform proposed by J Burgis obtains and analyzes data through remote processing of application instances and control messages in the environment [6], generates data feature archives, and identifies potential malicious activities. The design of this platform improves the automation and accuracy of the system, providing an effective solution for the automated processing of holographic retrieval and analysis systems for scientific research data.

J Hirt's research analyzed the overlap between multi database retrieval and citation tracking, and found that MEDLINE, CINAHL, and Web of Science Core Collection have significant overlap [7]. This discovery emphasizes the importance of utilizing these databases and citation tracking in a comprehensive search, providing practical guidance for optimizing the evidence synthesis process.

These research results provide valuable references for the holographic retrieval and analysis system of scientific research data based on SSH framework and Lucene engine, promoting the development of the system in performance optimization, intelligent processing, and automated analysis, and demonstrating its broad prospects in multi domain fusion applications.

3. System Architecture and Key Technologies

3.1. System Architecture Design Based on SSH Framework

This system is based on the SSH (Struts, Spring, Hibernate) framework and adopts a layered architecture design, as shown in Figure 1, to efficiently support holographic retrieval and deep analysis of scientific research data. The overall architecture of the system covers the data layer, model layer, application layer, and presentation layer, ensuring the efficiency, scalability, and ease of maintenance of the system.

In the data layer, the system is divided into a production database and a data warehouse. The production database directly stores actual data from scientific research management, while the data warehouse provides structured data support for analysis and decision-making through the processing and integration of this data. This processing enables subsequent analysis and decision-making to be based on comprehensive and accurate data.

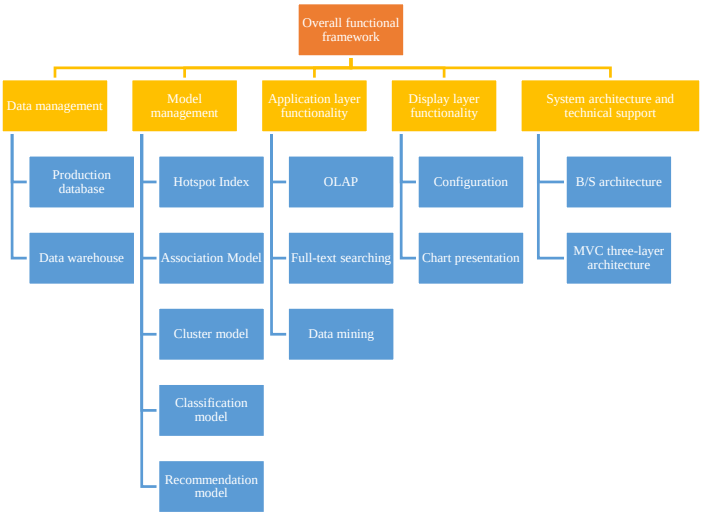

Figure 1: Overall functional framework

The model layer covers various analysis models, including hotspot indexing, association models, clustering models, classification models, and recommendation models. Hotspot indexing is used to mine hot topics in the scientific research field, association models analyze the cooperative relationships between data, clustering models evaluate the similarity of data, classification models are used to predict project completion time, and recommendation models provide suggestions for future cooperation. Although these models operate independently, the system manages them uniformly and provides services through integration, supporting comprehensive data analysis and decision-making.

The application layer integrates online analytical processing, full-text search, and data mining technologies. Online analytical processing utilizes Multidimensional Query Language (MDX) to operate data warehouses and achieve comprehensive statistical analysis; Full text retrieval utilizes the Lucene engine to search document indexes and analyze research hotspots; Data mining involves deep processing of data by calling different models. This layer interacts with the data layer and model layer through interfaces and provides analysis results to the presentation layer.

The display layer is responsible for user interaction, including parameter configuration and chart presentation. Parameter configuration allows users to specify analysis objects and personalized options, generating customized analysis reports. Chart presentation utilizes data visualization plugins such as D3 and Highcharts to present analysis results in a graphical format, making the data more intuitive and easy to understand. The system adopts B/S (browser/server) architecture, and users can directly access the system through the browser without installing additional applications. This design makes system maintenance and expansion more convenient.

3.2. Integrated Application of Lucene Engine and Data Mining Algorithms

The application of classification prediction focuses on predicting the project completion time. Through logistic regression models, the system can predict the project's completion time based on various attributes of the project, such as project type, responsible person's professional title, and the university where the project is located. This feature enables managers to plan project time and resource allocation more reasonably, optimizing the project execution process. At the same time, personalized recommendation systems are based on collaborative filtering algorithms to provide users with suggestions from potential partners. By analyzing the collaboration network of researchers in historical data, the recommendation system can provide the most scientific collaboration suggestions based on the collaboration history and research field of researchers, thereby improving the efficiency

and effectiveness of scientific research collaboration.

The Lucene engine and data mining algorithms are integrated through a unified data processing architecture, achieving efficient data analysis and decision support. At the data layer, the Lucene engine is responsible for establishing and maintaining full-text indexes to ensure efficient retrieval performance; At the data mining layer, various algorithms interact with the data layer through interfaces to extract the required data from the data warehouse for processing. The data is first processed and organized by ETL tools, stored in a data warehouse, and then analyzed and mined by data mining algorithms. The algorithm model is centrally managed in the system and interacts with the business layer through a unified interface to achieve the application and management of algorithm functions. Through this integrated architecture, the Lucene engine provides efficient retrieval capabilities, while data mining algorithms enable deep data analysis and prediction. The combination of the two enables the system to provide comprehensive and accurate solutions for scientific research data analysis and decision support, significantly improving the efficiency and scientificity of scientific research management.

4. System Development and Performance Evaluation

4.1. Implementation of Data Holographic Retrieval and Analysis Function

The system first constructs a multidimensional data warehouse model to achieve storage and management of multidimensional data. Specifically, the data warehouse adopts a star shaped model structure, which is jointly constructed by dimension tables such as personnel information, institution information, achievement information, and project fact tables. The update function of the data warehouse is triggered by the system administrator by executing a custom SQL script, which can extract and load data from the production database into the data warehouse. Through a user-friendly front-end interface, users can easily perform data warehouse updates. After receiving the trigger request, the server will automatically connect to the database, locate and execute the corresponding SQL script. The execution process is completed through batch processing, ensuring the consistency and integrity of transactions.

In terms of online analytical processing (OLAP), the system is based on Mondrian and MDX technology frameworks, constructs data cubes, and defines corresponding schema configuration files. The construction of a data cube generates a data model for multidimensional data analysis by mapping the fact table and dimension table in the data warehouse. Users can perform complex data queries and analysis operations through MDX query language. The system will automatically parse and execute these query requests, and return the results to the user to support multidimensional analysis and decision support of business data.

For the implementation of data mining functions, the system combines various classic algorithms, including association algorithms, clustering algorithms, classification algorithms, and recommendation algorithms. The association algorithm is based on FP Growth implementation, which effectively discovers the correlation between data by constructing a frequent pattern tree (FP Tree). The clustering algorithm adopts the K-Means method to automatically classify and cluster data based on the principle of minimizing intra cluster distance. The classification algorithm is based on a logistic regression model, which is trained on historical data to establish an accurate classification model for predicting the classification results of new data.

The system has specially designed a dynamic update mechanism for model maintenance and updates. When the model needs to be updated, the system will automatically delete outdated models and retrain to generate new models to ensure the accuracy and timeliness of the analysis results. The entire update process is managed through automated scripts to ensure that the new model can be quickly applied to practical business scenarios, thereby improving the system's response speed and

stability.

In the system design, the data holographic retrieval and analysis function integrates multiple sources of data to achieve comprehensive analysis of network logs. This feature is based on big data analysis technology, using parallel computing and distributed storage to quickly process and correlate data from multiple dimensions, thereby generating accurate analysis results. By optimizing query algorithms and data indexing techniques, the system effectively improves retrieval efficiency and ensures response speed when processing massive amounts of data. Ultimately, the system can provide users with multi-dimensional visual analysis results, support complex log data mining and deep analysis, and meet the needs of network security and operations management.

4.2. System Performance Testing and Stability Verification

As a web application system based on B/S architecture, this test is divided into two parts: client and server. The client testing mainly uses Chrome browser and Firebug plugin, while the server is based on Tomcat application server and Oracle database.

Functional testing includes verifying functions such as parameter configuration analysis, model management page redirection, algorithm model updates, analysis execution, analysis result viewing, and document export. By simulating ordinary user operations, the system can correctly respond to various operational needs of users in the university analysis and decision-making module. The testing content covers the entire process from selecting analysis content to executing analysis, to viewing results and exporting documents, ensuring the normal operation of various functions. During the testing process, the system was able to provide timely abnormal prompts when users made input errors, ensuring the robustness of the system.

In the response time testing of the system, the response time varies depending on the amount of data and model type. Especially in model update operations, due to the large amount of data and high algorithm complexity, the system incurs high computational overhead, resulting in longer response times. However, as model updates can be processed offline, there is no significant impact on users' online usage. For document export and analysis result viewing functions, the system response time is controlled within a reasonable range. In order to enhance user experience, the system has added running prompts in the front-end, allowing users to perceive the system's operating status and present visual analysis results clearly after analysis is completed.

To ensure the reliability of the system and the integrity of the functions of each module, I conducted in-depth system performance testing and stability verification on the discipline analysis and decision-making module and the researcher analysis and decision-making module. The main purpose of these tests is to evaluate the response efficiency of each module when processing large amounts of data and verify the stability of the system under high load operation. In addition, by comparing the performance of different modules in similar functions, a basis is provided for further optimizing system performance.

In the testing of the discipline analysis and decision-making module, the response time test results showed that due to the significantly increased amount of data processed compared to the university analysis and decision-making module, the execution time of various functions was also correspondingly extended. The testing covers updates to hotspot indexing, classification models, recommendation models, association models, as well as analysis result viewing and document export functions. The test results are shown in Table 1, where the average response time for hotspot index updates is 8.44 seconds, classification model updates are 37.6 seconds, recommendation model updates are 12.6 seconds, association model updates are 2.94 seconds, analysis result viewing is 4.37 seconds, and the average response time for document exports is 8.61 seconds. Although the execution time of functions has been extended with the increase of data volume, these results are within an

acceptable range. In addition, model updates support offline processing, allowing users to update without affecting system operations. At the same time, prompt information is provided during document export and analysis result viewing, ensuring a good user experience.

Table 1: Results of Response Time Test for Discipline Analysis and Decision making Module

Number	Function	Processing data volume (pieces)	Average response time (seconds)
01	Update hotspot index	Hundreds to tens of thousands	8.44
02	Update classification model	11106	37.6
03	Update recommendation model	6494	12.6
04	Update the associated model	3551	2.94
05	View analysis results	Without	4.37
06	Document export	Without	8.61

For the testing of the analysis and decision-making module by researchers, the data processing volume is larger and the response time is significantly prolonged, especially in the updating of clustering and association models. The test results are shown in Table 2, with an average update time of 126000 seconds for the clustering model, 37.6 seconds for the classification model, 134 seconds for the recommendation model, and 25200 seconds for the association model. Despite the huge amount of data and the system running in a standalone environment, model updates can still be processed offline, and users will receive corresponding prompts when viewing analysis results and exporting documents. This design ensures that the system can still provide a good user experience under high loads.

Table 2: Results of Response Time Test for Analysis and Decision Module by Researchers

Number	Function	Processing data volume (pieces)	Average response time (seconds)
01	Update clustering model	246149	126000
02	Update classification model	11106	37.6
03	Update recommendation model	24656	134
04	Update the associated model	168881	25200
05	View analysis results	Without	1.94
06	Document export	Without	3.67

5. Conclusion and Prospect

This article provides effective reference for scientific research management by studying the decision support system for scientific research management and using modern computer technology to comprehensively analyze scientific research data. The research content covers multiple aspects such as requirement analysis, technology selection, system design, implementation, and testing. The results indicate that the developed system can effectively support research management decisions of universities, disciplines, and researchers, and has high reference value. Although the system has

achieved its expected goals, there is still room for improvement in data processing capabilities and user interaction. In the future, the system's functionality can be further improved by introducing distributed computing frameworks, expanding algorithm types, and optimizing application scenarios. With the continuous advancement of computer technology, decision support systems will play a more important role in scientific research management and other fields.

References

[1] Jiang Y, Xue Z, Li A. Research on Cross-media Science and Technology Information Data Retrieval. 2022. DOI: 10. 48550/arXiv. 2204. 04887.

[2] Chen R, Hao J, Wang J, et al. Phase retrieval in holographic data storage by expanded spectrum combined with dynamic sampling method. Scientific Reports, 2023, 13 (1). DOI: 10. 1038/s41598-023-46357-9.

[3] Pensia L, Kumar R, Kumar M. A compact digital holographic system based on a multifunctional holographic optical element with improved resolution and field of view. Optics and Lasers in Engineering, 2023, 169 (Oct.): 107744. 1-107744. 9.

[4] Mongan T R. Standard Model Fermion Masses and Charges from Holographic Analysis. Journal of Modern Physics, 2024, 15 (6): 8. DOI: 10. 4236/jmp. 2024. 156035.

[5] Pi Y, Duffield N, Behzadan A H, et al. Visual recognition for urban traffic data retrieval and analysis in major events using convolutional neural networks. Computational Urban Science, 2022, 2 (1): 1-16. DOI: 10. 1007/s43762-021-00031-w.

[6] Burgis J, Johnson R, Nunes E, et al. Automated Device Data Retrieval and Analysis Platform: WO2021US55107. WO2022081930A1 [2024-08-27].

[7] Hirt J, Bergmann J, Karrer M. Overlaps of multiple database retrieval and citation tracking in dementia care research: a methodological study. University Library System, University of Pittsburgh, 2021 (2). DOI: 10. 5195/JMLA. 2021. 1129.