

Design and Application of Foreign Literature Information Resource System

Lin Wang*

College of Foreign Languages
Zhejiang Normal University
Jinhua, Zhejiang, China, 321004

Abstract — The rapid development of information technology is a challenge and opportunity for foreign literature information development and utilization. If convenient conditions and high technology means brought by digital environment can be made good use of, foreign literature information can make a breakthrough and play a wider role. Based on the analysis of foreign literature service requirements, foreign literature service management system based on network is researched. First of all, the system functional requirements are put forward. Secondly, service management system of foreign literature centre based on network is designed in detail, which completes comprehensive document processing system and information service system in the cross-platform, distributed, heterogeneous database environment.

Keywords - Foreign literature; Network; Heterogeneous database

I. INTRODUCTION

Foreign literatures are records of scientific and technological research activities. It contains a huge number of wisdom and research results, and can also reflect the process of studies. The degree of the exploiting and utilizing of it, which is directly related to the level of scientific research and national economic processes, and is essential to the progress of science and technology, the actualization of research achievements and the ability to innovate[1,2]. Therefore, research on how to promote, improve and protect the exploiting and utilizing of science and technology foreign literature information in digital environment is of great significance both in theory and practice considering that the e-science springs up and the achievement of the value of foreign literature is weak now[3].

In the traditional environment, foreign literature information resource is stored in each storage organization in scattered form[4-6]. The barriers of space and means results in the dual narrow channels of the development of foreign literature sources and the usage. Thus it reduces the value of exploitation and utilization of literature of science and technology, limiting its influence of science and technology, and also it causes great inconvenience for the development of information. The development of network technology and database technology breaks the boundary of time and space, and allows the worldwide science and technology document resource to be in front of developers and users [7,8]. Under digital environment, the development and utilization of science and technology foreign literature has the requirement of resource integration, allowing different parts of the patent document, conference documents, together with the standard and various types of foreign science and technology literature information included in shared scope, which implements the integration of resource development and unity [9,10].

In the traditional environment, due to the lack of technical means, the research on the development and utilization of literature of science and technology, carrying

out diversified service, the modernization of management methods, and building a service system such sloganeering strategy lacks of specific practice guiding significance. Also it makes the service work of foreign literature always stay on the surface level[11]. Under digital environment, science and technology development and utilization of literature is no longer just focus on literature resource itself, but embodies the people-oriented modern thought[12-16].

In the section 2, the requirement analysis of overall system function is investigated. In Section 3, system application architecture is proposed. In section 4, system overall design is carried out. In section 5, some conclusions are given.

II. THE REQUIREMENT ANALYSIS OF OVERALL SYSTEM FUNCTION

Database resource and a variety of document resource of information resource management platform must meet the release of the data resource, the full text retrieval and data sharing. At the same time, various management database and document resource must also get timely maintenance and update, so it is necessary to establish a set of effective data collection, transfer, exchange platform, in order to solve all kinds of system data collection, maintenance, distribution, replication and data synchronization issues. For a system to keep the lasting vitality, it must be able to guarantee the data timeliness, accuracy, and the timeliness and accuracy is guaranteed through data collection and maintenance. Data centre as centralized data management, data service centre and data exchange centre of information resource service platform, the data collection and data maintenance should have many kinds of ways and means.

Schematic diagram of hardware environment is shown in figure 1. System provides the user-defined data calculation and distribution process based on resource section. The user can formulate different editorial process according to the importance of the resource information. Data storage and management is the core of centralized and distributed storage of system data and the system data storage management

platform adopts literature resource database RMS system. RMS DB Server can directly manage digital documents such as Word, PPT, PDF, XML and HTML. Full-text retrieval is the essential function to obtain the user information accurately in vast amounts of unstructured, semi-structured and structured resource information. Core function of RMS DB Server resource database server is to achieve integrated management of structured data, semi-structured data, unstructured data and the document, and provide full text retrieval function at the same time. The system can run independently with relational database, and can be integrated with relational database seamlessly, and provide full text retrieval function for unstructured data in relational database. The system can also realize extended stored procedure through the My SQL Server to realize the seamless integration with MS SQL Server. It can realize literature data exchange standard and management standard.

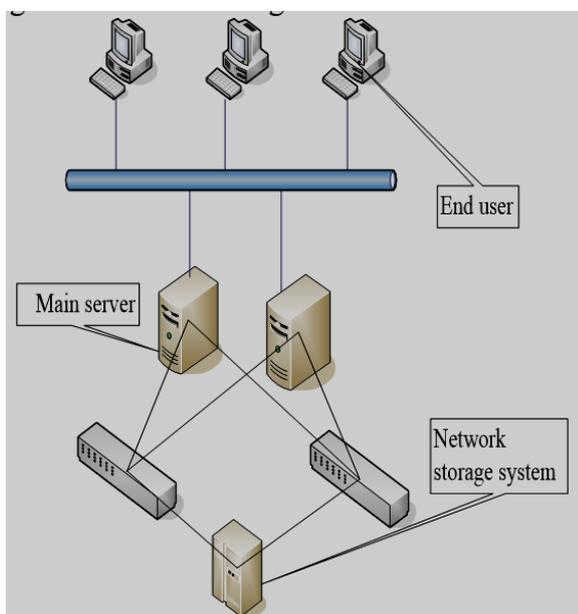


Figure 1. Schematic diagram of hardware environment

Advanced data indexing technology can support 10 kinds of indexing way and retrieval speed has nothing to do with the size of the database. The keywords retrieval has 100% recall ratio and precision. It support content management based on the full text retrieval. The structured database management technology is used to manage the document of unstructured data. There are built-in nested format commands, that users can configure output format desired by OAI, OpenUrl management. It support the physical cross-database retrieval and segmentation index management technology, which can index the incremental data real-time. It provides development platform based on component technology and supports Windows and Linux operating system. It provides independent DB Server, supports COM+ application, and Java Bean development interface, which has strong system scalability.

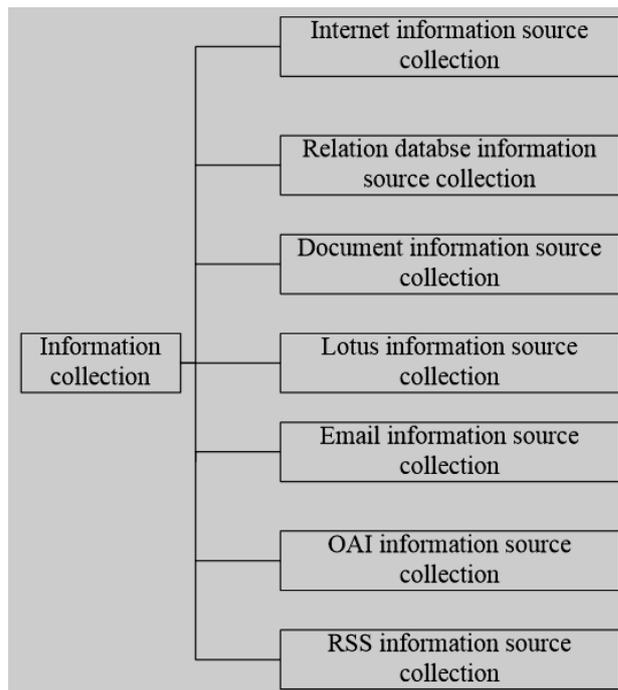


Figure 2. Schematic diagram of information collection subsystem

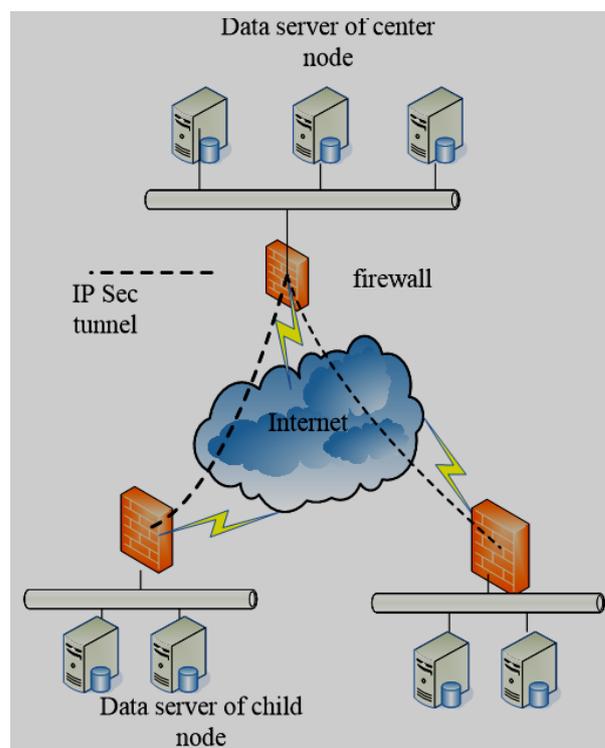


Figure 3. Document sharing network topology

Application service parts adopt WFI-Server architecture, in order to realize the integration of multiple data sources and personalized features. The system can use data source

of relational database, RMS data source, network data source and OpenUrl data source for the establishment of network information integrated service system. At the same time, we can use the system personalized service functions to complete all kinds of personalized customization service function, such as personalized interface, personalized retrieval, personalized push, etc.

III. SYSTEM APPLICATION ARCHITECTURE

Foreign literature centre mainly include network storage system, network server, retrieval server, configuration and deployment of database server and the construction of security and data backup system. In view of foreign literature information resource has various types of database. Its hardware construction needs to meet with large capacity storage space and provide widely distributed users with fast data access requirements. Therefore, network structure of the whole storage system adopts SAN based on the fiber channel technology.

Through the investigation and argumentation, EMC CLARiiON CX700 SAN storage system is adopted and high-speed fibre channel is taken as transmission media. Application protocol FC+SCSI is taken as storage access protocol. CLARiiON CX700 network storage system is the top model of CX series, which is specifically designed for high capacity and high application performance environment. CX700 is very suitable for high load database application, which can be expanded to 117 TB.

There are four servers, respectively named san1a, san1b, san2a, san2b. San1a and san1b constitutes cluster1. San2a and san2b constitutes cluster2. Cluster 1 is configured with IIS to realize dual computer hot standby of network site. Cluster 2 is configured with IIS and SQL Server to realize dual machine hot standby of network site and database. In CX700 storage system, configured fiber channel hard disk is 300G. There are 12 pieces of hard disk, capacity of which is 3.6 TB. There are 48 pieces of 146G fiber channel hard disks, capacity of which is 7 TB. There are 45 pieces of ATA hard disks, capacity of which is 22.5 TB. The total capacity of the whole system is 33.1 TB. The available disk space is about 27 TB.

TABLE I. NODE CONFIGURATION PARAMETER

parameter	value
Node number	1
Server name	San1a
WLAN IP	218.69.114.37
WLAN Subnet mask	255.255.255.0
WLAN gateway	218.69.114.2
WLAN DNS	218.69.114.37
heartbeat cable IP	172.28.210.1
heartbeat cable subnet mask	255.255.255.0
Node number	2
Server name	San1b
WLAN IP	218.69.114.38
WLAN Subnet mask	255.255.255.0
WLAN gateway	218.69.114.2
WLAN DNS	218.69.114.37
heartbeat cable IP	178.28.210.2
heartbeat cable subnet mask	255.255.255.0

TABLE II. CLUSTER CONFIGURATION PARAMETER

parameter	value
Cluster name	Cluster 1
The number of node	2
administrator account	fsan1/ administrator
Administrator password	*****
Virtual IP	218.69.114.39
Virtual IP subnet mask	255.255.255.0
IIS virtual IP	218.69.114.48
IIS virtual machine name	iis-cluster 1
Drive letter	H
Shared drive	F, G

In the system, RAID is regarded as a logical partition, which is made up of multiple hard disks. Storing and reading data from more than one hard disks at the same time greatly improve the data throughput of storage system, which has a comparatively complete mutual check/recovery measure, or direct mutual image backup. Specific set instances are as follows. Node configuration parameter is shown in table I. Node 1 is controller and DNS server, the domain name is fsan1.com and DNS repeater is 218.69.114.3. Node 2 is member server of fsan1.com. Cluster configuration parameter is shown in table II.

The hardware deployment of foreign literature application service system includes network server, retrieval server, database server and information gathering server, etc. Specific server functionality and configuration. Data availability can be improved by using Mirror View software of EMC company and EMC SAN Copy software supports high speed data copy between multiple suppliers storage system, which can rapidly copy data between CLARiiON and Symmetrix system, as well as in EMC, HDS, IBM, Sun and HPQ system.

The network server completes online browsing and query service of foreign literature information. The database server completes data storage, query, update, business management, index, high speed cache, query optimization, security and multiple user access control. Retrieval server completes foreign literature directory browsing, query and retrieval. Information collection server completes basic data collection, review and upload.

The configuration of network server is as follows. Hp dl 580 G95000 Intel dual-core Xeon 7120 3.0 GHz*4, 4 MB three-level cache, the processor of front-end access, 800 MHZ double independent front bus, integrated iLO2 remote management, 4GB(4x1GB) 4 PC2-3200R 400MHZ DDR-II memory, 146*4 hot-plug SAS hard disk, 16 times speed DVD-ROM, two gigabit Ethernet card, integrated the iLo remote management Smart and 2 hot plug the power supply, Hp L1906 (PX850AA) 19 inch LCD monitor.

The configuration of database server is as follows. 2U Rack/WOODCREST 5140 2.33G 1333MHZ*2 (4M L2) /FB DDR2-533 1G*4/SCSI 146G 10K*4/ SCSI RAID 0/1/5 /double INTEL gigabit network card support I/O the AT/ redundant power supply/560W 1+1/DVD Campbell/USB floppy drive.

The configuration of retrieval server is as follows. 2U Rack/clovertown 5320 1.86G*2 1066MHZ/FB DDR2-533 1G*8/SCSI 146G 10K*4/ SCSI RAID 0/1/5/double INTEL gigabit network card support I/O the AT/ redundant power supply/560W 1+1/DVD Campbell/USB floppy drive.

The configuration of information collection server is as follows. Dual-core Intel Xeon 5130 processor 2.0GHZ*2, integrated 4 MB two-level cache, 2GB(2X1GB) PC2-5300 full buffer DIMMs memory (DDR2-667), integrated Smart Array P400 array controller, 256 MB high speed cache, support RAID 0/1/5, 512 MB cache with battery protection, 146G*2 hot-plug SAS hard drives, DVD drive of 16 times speed, 2 hot plug power supply, two gigabit network cards.

Document sharing network topology is shown in figure 3. In the process of construction, Intranet VPN is set up between shared units. On how to realize the VPN, fire wall VPN devices are adopted to set up network. The IP SeC technology is used to ensure internal network security of foreign literature resource.

IV. SYSTEM OVERALL DESIGN

Life cycle and main task based on information resource is divided into the following process. That is information planning, resource collection and integration, resource processing, resource management, resource release and information collaboration, resource transmission exchange and information service.

Information planning realizes information classification system planning of foreign literature information resource organization and management. Information collected through various channels is classified into relevant information type and is automatically indexed. It realizes organization and management of all kinds of heterogeneous information resource.

The information collection realizes collection and integration of the network news information source, relational database information source and document information source. It also realizes scheduling and management of collection task.

Information processing realizes foreign literature information editing, auditing, information issue, information statistical analysis. The focus of it is to complete a set of editorial business logic process, and it also can support information collection, indexing and release.

Information management realizes unified management of provided foreign literature information resource (including local source, relational database resource and virtual resources, etc.), information issue, information presentation and key topics. Information release realizes information issue and service of multiple front sites based on a set of background database. It realizes static and dynamic information portal and WAP information portal. The key of it is to solve the dynamic release information portal. Information service realizes the system public information service, personalized information service, the full text information retrieval service, unified information resource retrieval service, information email service, RSS information service, CD publishing service and statistical analysis of information service.

User permission authentication billing realizes information browsing based on group policy, authorization management and the corresponding permission check. It realizes user billing setting, cost collection and the user account management. It also realizes interface support of payment platforms including bank card and mobile phone micro payment. System management realizes parameter management, log management, system backup and recovery. The background service realizes background full-text retrieval service, realizes data synchronization between the material library and full-text index library, and realizes unified authentication service system development and registration of the user purchase function module. System realizes integration of various heterogeneous information resource based on the planning of information, information collection, information processing, information transmission exchange, information management, information release, information service, user management authority certification and billing payment. It realizes the individualized information service, value-added information service, information push service, rapid information retrieval of uniform resource, high-level information service function of the key topics and information sharing and exchange.

This system classifies all kinds of data information, and according to the needs of different users, it carries out configuration and management for all kinds of database resource. Based on unified definition and management of local resource, relationship database resource, virtual resource and linked resource, it provides function support for literature navigation subsystem and uniform resource search subsystem. Through the management of presentation and it provides support for special topic of the portal platform network site.

According to the needs of system control, it includes information management, system management and other business modules and subsystems. According to the need of component reuse, the user management, privilege authentication and billing is taken as an independent business module and subsystem. Based on classification principle of back and front platform, various business module and the management control system is taken as network front end system, background services such as the background information acquisition, full text search service and data synchronization index update is taken as independent modules and subsystems. Considering the development target and application scope of foreign literature platform, the system according to the needs of content management and information service, from the perspective of information resource management and value-added information service business, business modules and subsystems are partitioned, such as information planning, information collection, information processing, information release, information service and user privilege authentication billing system.

The information planning sub-system mainly completes definition of information classification and information resource configuration management, which is stored in configuration management library to serve for information collection and information processing.

The information collection system collects various kinds of information according to the defined configuration parameter of information source using the corresponding information gathering tool and the collected information is stored in the material library. Schematic diagram of information collection subsystem is shown in figure 2.

Information processing subsystem edits and classifies material information from the material library. Material information after processing is released and the editorial business process needs to follow the workflow definition mechanism, which also needs authority certification of the user privilege authentication system. Information management subsystem finishes all kinds of resource management.

V. CONCLUSIONS

Foreign literature resource system is designed which supports literature processing and information service of cross-platform, distributed and heterogeneous database environment. It uses techniques such as advanced network database, distributed database processing and data pushing.

REFERENCES

- [1] Amy Trappey, Charles Trappey. "An R&D knowledge management method for patent document summarization". *Industrial Management & Data Systems*; vol. 108, No. 02, pp. 32-45, 2014.
- [2] Betty Smith, Shirley V King, Ian Stewart. "SURVEY OF DEMAND FOR JAPANESE SCIENTIFIC AND TECHNICAL SERIAL LITERATURE". *Interlending & Document Supply*; vol. 15, No. 04, pp. 112-123, 2007.
- [3] Cambazoglu B, Karaca E, Kucukyilmaz T. "Architecture of a Grid-Enabled Web Search Engine". *Information Processing and Management*. vol. 15, No. 04, pp. 609-623, 2007.
- [4] Klink S, DENGELA, KIENINGERT. "Rule Based document structure understanding with a fuzzy combination of layout and textual features". *Int'l Journal on Document Analysis and Recognition*; vol. 4, No. 01, pp. 18-26, 2001.
- [5] Janice Bradshaw. "Evaluation of Biomed and ISTPB, two new literature databases from the Institute of Scientific Information". *Online Information Review*; vol. 7, No. 03, pp. 15-22, 2003.
- [6] William Glisson, Gobinda Chowdhury. "Design of a digital dissertation information management system". *Program: electronic library and information systems*; vol. 36, No. 03, pp. 16-34, 2002.
- [7] Chen Liming, Shadbolt N R, Goble C A. "A Semantic Web-Based Approach to Knowledge Management with Additional Information". *IEEE Transactions on Knowledge and Data Engineering*. vol. 14, No. 02, pp. 283-296, 2007.
- [8] Fujino A, Ueda N, Saito K. A. Hybrid. "Generative Discriminative Approach to Text Classification with Additional Information". *Information Processing and Management*. vol. 12, No. 02, pp. 379-392, 2007.
- [9] Huang Xianrong, Li Xiao. "Exploring copyright solutions to online-sharing of scientific literature". vol. 25, No. 03, pp. 35-42, 2010.
- [10] Kim D S, Hwang S H, Gee H. "Concept Analysis of OWL Ontology Based on the Context Family Model". *Proceedings of Convergence Information Technology*. Washington: IEEE Computer Society, vol. 5, No. 03, pp. 896-901, 2007.
- [11] Melih Kirlidog, Didar Bayir. "The effects of electronic access to scientific literature in the consortium of Turkish university libraries. *Electronic Library*"; vol. 25, No. 01, pp. 164-175, 2007.
- [12] Nico Dauphine, Mary Anderson Ochs, Nicole K. Joos. "Bringing scientific literature to the developing world: The Essential Electronic Agricultural Library (TEEAL)". *Online Information Review*; vol. 27, No. 01, pp. 34-48, 2009.
- [13] Ian Foster, Carl Kesselman. "The physiology of the Grid: an open Grid services architecture for distributed systems integration". *Chinese petroleum exploration*, vol. 2, No. 06, pp. 54-65, 2014.
- [14] Amy Trappey, Charles Trappey. "An R&D knowledge management method for patent document summarization". *Industrial Management & Data Systems*; vol. 108, No. 02, pp. 32-45, 2008.
- [15] Katsuyo Thornton, Dallas R. Trinkle. "Applying for computational time on NSF's TeraGrid-the world's largest cyberinfrastructure supporting open research". *JOM*; vol. 2, No. 06, pp. 54-65, 2014.
- [16] Marica Staresinic, Bojana Boh. "Patent informatics: The issue of relevance in full-text patent document searches". *Online Information Review*; vol. 33, No. 01, pp. 145-160, 2009.