Knowledge Portal Construction and Resources Integration for Hydropower Corporation

Li Wang, Lu Liu, and Wei-jia You
School of Economy and Management, Beihang University, China
37 Xueyuan Road, Haidian District, 100083 Beijing, P.R. China
seaskyinter@163.com

Abstract. Based on the scientific and technical research project "Data Collection and Decision Support System for Hydropower Dam Construction", which is set up by the Chinese National Development and Reform Commission, this paper, which aims to solve the problems in managing the construction of large-scale hydropower project, focuses on the knowledge portal using Web Services and J2EE.

After studying China Gezhouba (Group) Corporation (CGGC), which is building the largest hydropower project of the world, Three Gorges Project, a multi-layer architecture and a supporting system for the knowledge portal is put forward. The method for integrating existing resources is also proposed. Enterprise application integration (EAI), which is platform-free and improving the interoperability of heterogeneous system, is realized by interoperability in distributed environment, software combination technology, heterogeneous system integration, and application mid-ware specification. All the resources are in the open and unified environment and are easy to share. The process of designing and developing knowledge portal for CGGC Operation Management System is further provided to illustrate the architecture proposed above.

Key word: Knowledge portal, Web Services, EAI, Hydropower Corporation.

1 Introduction

Reviewing the history of applying information technology to management, the trend of development is from data management to information management and now towards knowledge management. Since Decision Support System was firstly put forward in 1970's, computer and information technology in knowledge management and knowledge processing is more and more discussed, especially over the recent decades [1].

Relative researches mainly focus on building Knowledge Management Systems, which is one of the hotspots of computer and Artificial Intelligence (AI) researches, and providing computer as tool to support knowledge management [1].

Technically speaking, the rapid development of computer technology in recent decades promotes the improvement of enterprise knowledge store and knowledge process. For example:

Development of network technology and distributed computing pattern, especially the development of Internet/Intranet, has profound impression on the knowledge management supported by computer. On one hand, network helps to improve the availability of information and inspire the demand of processing the information and retrieving useful knowledge, On the other hand, network provides the infrastructure for distributed knowledge share and collaborative innovation.

The development of Artificial Intelligence (AI), especially the development of Agent technology, provides favorable technical preparation for computer-aided knowledge processing [2].

Other development of information technology, such as groupware, data warehouse and data mining, further supports the knowledge management system technically [3, 4].

With the development of E-Business, integration of information system, e-business and knowledge management is more and more discussed [5].

Knowledge portal, which is one of the important components of decision support system, is developed to improve the knowledge share, knowledge application, and knowledge creation. The ability of analyzing and judging are enhanced and therefore decision-making is supported.

The organization location, operation data, experience and knowledge of Hydropower Corporation are all distributed, which is also the character of the industry. Especially in the large scale hydropower projects, layouts are dispersed; quality assurance nodes are distributed and related information is mass. Above characters make it necessary to apply network technology and information technology to the management and decision-making, especially combing the special experience and knowledge of experts from different departments and different industries.

China Gezhouba (Group) Corporation (CGGC), which is one of the largest hydropower corporations of China, has accomplished many large hydropower projects, including the largest one of the world, Three Gorges Project. CGGC owns rich experience of huge hydropower dam construction in special geography environment. At the same time, various information systems have been built after several years' information construction.

After information construction for several decades, more than twenty local area networks (LAN) have been built and PC terminals are more than 2,000. A management information system (MIS), which supports production, operation and management, has been developed, together with Beihang University. Three Gorge Dam process control system, shutoff statistical analysis system, project cost and quality management system and supervision system for concrete production, transportation, which are all domestic-advanced, have been successfully developed.

Regarding the universal characteristic of Hydropower Corporation and based on the actual conditions of CGGC, Beihang University, together with CGGC, applied for scientific and technical research project "Data Collection and Decision Support System (DSS) for Hydropower Dam Construction", which helps collecting and managing the information of operation of hydropower project. The operation is supervised real-timed, analyzed and guarded. The experience of dealing with exception of operation is collected and used in making decision and in supporting the quality and safety assurance.

Building model knowledge portal for Hydropower Corporation and studying the method of integrating various heterogynous resources are important contents of the research project, which will be further explained as follows.

2   Knowledge Portal System for Hydropower Corporation

2.1   Brief Introduction to the System

Knowledge portal for Hydropower Corporation, which is the entry to the corporation knowledge platform for users, is the centralized and unified provider of data, information and knowledge resource.

According to the characters of existing information system, unified access to information and knowledge resource is designed. A series of standards and specifications for knowledge share via net are defined and an integrated knowledge portal for the corporation is developed.

The construction of knowledge portal system consists of criterion system, portal, Mete data Directory Service, Unified Authentication and Resource Access Control (RAC), retrieval and navigation, assessment and supervision system, virtual organizational collaboration environment, application service integration environment and resource connection.

2.2   The Architecture of Portal

2.2.1 Multi-Layer Architecture

Knowledge portal for Hydropower Corporation, which uses multi-layer architecture based on the web, is composed of five layers: client layer, web layer, application service layer, application support layer and resource layer. Multi-layer architecture is used to improve the reliability, extensibility, and retractility of the system. Various requests from users are posted to application service layer and handled in the application service layer. If the request is in need of backstage service support or information and knowledge access, application service layer will revoke the interface of service support layer and resource layer.

(1) Web Layer

Portal web layer is in the demilitarized zone (DMZ), including load balancing servers and web servers, which provides the access of request and load balance. Only portal web layer is accessed when users visit the portal. Backstage service and resource are not accessible so that the security of the whole system is assured.

(2) Application Service Layer

Application service layer, which is the kernel of portal system, includes application servers and database servers. Various services of portal system, unified authentication, Resource Access Control (RAC) and interfaces of service resource are deployed in application servers. Database system includes information database, authorization database and content management database.

(3) Application Support Layer

Application support layer includes a series of application support systems, such as information and knowledge retrieve and navigation system, knowledge map, virtual organization collaboration environment, application integration environment, evaluation and supervision system, which are integrated in the portals by portlet and Web services [6].

(4) Resource Layer

Resource layer refers to all kinds of information and knowledge resources in business systems where recourses are created and services are provided, including information in operation management system, knowledge share Portlet, human resource management system and equipment management system.

Main idea of multi-layer architecture lies in separating application service logic from data resource, and subdividing the function and roles of servers. Business logic, presentation and data store are in the different layer so that the system is robust, safe, open, extensible, capable and short-cycle. Multi-layer architecture supports multi platform, distributed calculation, load balance, centralized management of information and easy operation.

2.2.2 Design Based On Views

The knowledge portal system are be extracted to four views, presentation view, application view, data view and management view. Through the views above can knowledge resource be presented uniformly, can application resources be connected uniformly, can data be integrated uniformly and can system resource be managed uniformly.

(1) Presentation View

Presentation view refers to the portal system providing a unified net desktop system for users to access the knowledge resource platform. The distribution and heterogeneity are shielded and the various recourses of the corporation can be uniformly used.

(2) Application View

Application view refers to the integration of application service via developing an application container and application service registration center. Retrieval and navigation of information and knowledge, virtual organization collaboration and other supporting services are connected uniformly to portal system.

(3) Data View

Data view means the integration of data resource by portal system. First, metadata catalog database is constructed by the metadata catalog criterion, mass information and knowledge resources are indexed. Information and knowledge resources are accessible by resource access interface and resource integration interface.

(4) Management View

Management view refers to the unified management of internal resources, including Unified Authentication, portal security assurance, assessment and supervision system and content management, which supports unified supervision and management.

2.2.3 Logical Structure of Knowledge Portal

Via portal system, which is user-oriented, resource can be located, and construction can be guided. Requirements are acquired form users and coordinated by management department, so that the integration and application are accessible.
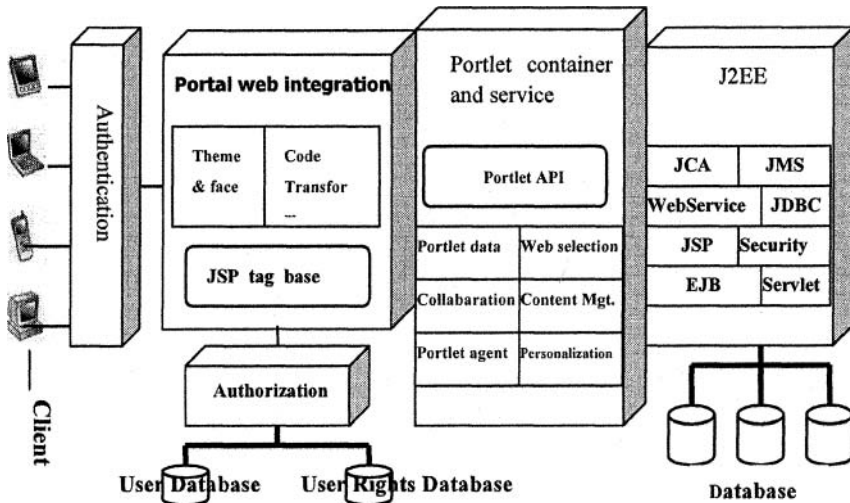
Fig. 1. Technical architecture of knowledge portal

Portal system is logically divided to four layers, resource layer, support layer, function layer and portal layer. Portal site provides unified entry to knowledge resource platform for users by using the service from service layer. Service layer provides unified presentation and access to resources available by extracting service

from support layer. Support layer provides connection to the resource of bottom layer. Resource layer includes all the resources from all the subsystems.

2.3 Technical Composition of Knowledge Portal for Corporation

The technical architecture of knowledge portal as follows shows.

In the technical architecture (see Fig.1), different layers are loose coupling. Each layer uses component technology and block structure, by which can the resources and services be connected seamlessly and transited smoothly [7, 8].

2.4 Unified Authentication and Resource Access Control

Knowledge portal combine all the applications, information and workflows. Knowledge resource platform of corporation is composed of integrated knowledge nodes of distributed and heterogeneous information. By Unified Authentication technology, users can access to all the separate systems by logging into the system once. Resource access control takes the policy of Role-Based Access Control (RBAC), by which can all the resources be accessed by valid user so that all the resource are free from illegally access.

2.5 Integration Environment of Application Service and Resource Connection

Enterprise application integration (EAI), which is platform-free and improving the interoperability of heterogeneous system, is realized by message processing mechanism, interoperability of distributed environment, software combination technology, heterogeneous system integration and application mid-ware specification.

A series of criterion of application service integration are constituted and realized in this environment. So that service abstract, service deployment, service management, service registration and discovery are provided.

Service abstract extract existing application systems to service according to the criterions of integration, including business logic, deploying document and other relative files. Business Logic implements encapsulate legacy functions and communicate with legacy systems. Deploying documents, based on XML, describe all the configure information, such as access control, database configure, by predefined specification of configuration documents. Application service is easily to be loaded in this integrated environment. Other relative documents include all the required source files for service.

Service deployment refers to deploying the service to application integration environment. Whether the service conforms to the integration criterion is firstly checked. Then deploying describing document is read and current service is configured and operated.

Service management provides processing of encapsulated transactions, configuration management, multi-threads, resource connection pool and other

bottom detailed service container so that application service could be quickly encapsulated and automatically deployed. Meanwhile, bottom support services are provided according to specific application service, including security, transaction, category visiting and remote access interface. Service management also provides life-cycle management and Performance Monitor.

Service discovery and registration center provides registration center which is application service oriented. When deploying service, service is firstly registered in service registration center and can be modified when service is updated. Users can discover the service by Web or standard API and integrate them by standard interfaces.

Application service integrated environment is designed based on existing architecture of application servers, Web Services container and Web servers. And a resource container is developed in accordance with engineering application. So that service abstract, service deployment, service management and service registration and discovery are provided. To share all the information and knowledge resource, application services are unified presented in the architecture.

The architecture is divided into 5 layers, which are information and knowledge resource layer, resource connection layer, service management layer, service abstract layer and service application layer.

3 Operation Subsystem

Operation management subsystem is one of the most important components of CGGC portal system, by which could the statistical data be collected in time and programming and decision be made timely. Users can access to the basic information and communicate with other related departments on line, share the knowledge and value-added service with them, by which could knowledge be innovated.

The implementation of the system will improve the integration of information and knowledge from branch companies and project departments, the study and reuse of various resources, the relationship between different projects. Detailed information of production and operation will be collected timely so that the management is improved, Various methods, such as graph, text, video and virtual reality model, are used to present the rich resources and knowledge of corporation and to improve the knowledge transmission and application.

3.1 Layered Architecture

To make the platform effective and flexible, operation management subsystem takes layered model and mid-ware technology. It is logically divided to three layers, which are data layer, application layer and connection and presentation layer.

Separating the connection layer from application handle layer makes the platform robust and powerful. By using the advanced construction technology, this platform not only performances stably and powerfully, but also well opened. Applications are loaded to the platform easily and processed by workflow engine, so that each

application can be maintained conveniently and new application can be added quickly.

3.2   Resource Connection and Integration

Operation management subsystem has to integrate a widely heterogeneous data resource. The procedure is as follows:

(1) Analysis of Heterogeneous Data

Because different companies are in the different phases of the information construction, they stored their data about production and operation in different formats, so do their data management systems. From simple file to complicated net database, different storage makes the resource of heterogeneous data. The existing data satisfy the current users of separate systems, but doesn't meet the need of corporation users, who want to be aware of all the data located in different format. Thus a distributed application system, built upon the heterogeneous data, is necessary [9].

The heterogeneity of data lies in:

1) System heterogeneity means that the application systems, database systems or operation systems, where the data are stored, are heterogeneous.

2) Heterogeneous patterns, means data are stored in different formats. Storage patterns including relation patterns, object patterns, object-relation patterns and file nest patterns, in which relation pattern is most popular. There could be differences in structure even the same storage pattern is used. For example, the data types are not totally the same in different relation database, such as DB2, Oracle, Sybase, SQLServer, and Foxpro.

3) Heterogeneous source means the heterogeneity of external data and internal data.

(2) The Requirement of Data Integration

The aim of integrating heterogeneous data is to provide unified, safe, timely information for operation system application to meet the requirements of query, data mining and decision making. Therefore, integrated data has to be of integrity and security in data accessing.

1) Integrated

After the integration of separate information in different isolated operation systems, query can be uniformly executed in data warehouse which integrated various data organically and achieve associated store, without accessing each isolated system. Data are no longer stored simply and separately in each database.

2) Integrity

Integrity refers to the data integrity and restriction integrity. Data integrity refers to extract data totally, while restriction integrity means the relationship between different data, which is the only character of presenting the logic of data. Restriction integrity is the premise of data publication and exchange, and can improve the processing and the efficiency.

3) Consistency

Different information differs semantically, and causes all kinds of incomplete and wrong information. From name to structure, semantic conflicts make integration result redundant and the processing, publication and exchange of data interfered. Integrated data should be transformed following the data structure rules and codes [10].

4) Data Access Security

Different data resources belong to different companies and each database owns rights control. Therefore access and security management are not centralized. In order to keep the security of accessing the data source without intrusion and shield the original data access control, designing a unified user security management is necessary.

(3) Heterogeneous Data Integration and Integrating Pattern Realization

Aiming to meet the requirements of heterogeneous data integration, data warehouse technology and data extraction tools are used to integrate heterogeneous database and heterogeneous types of files, such as text and spreadsheet [10]. Based on the existing systems, independent data transforming code secondary development is executed to collect raw data, clean wrong data, integrate heterogeneous data, transform data structure and refresh data periodically.

Based on the theory of data warehouse, heterogeneous data is integrated to integrated database by exchanging access technology.

3.3 Resource Retrieval and Navigation

System retrieval is designed differently according to different data source.

1) Mixture of metadata and distributed data interfaces are usually used for local resources.

2) Data collection method, i.e. Web virtual interface, is used when dealing with those data with high requirements.

3) Special technology for system application.

Inter-operability of heterogeneous resources focuses on platform heterogeneity, operation system heterogeneity and database heterogeneity, among which the crucial problem is the retrieval interface. Therefore, new technology, such as Web Services and grid technology are applied, based on XML/Web services/UDDI, according to current resources [6]. The structure designed above helps to improve the inter-operability of heterogeneity resources.

4 Conclusions

Knowledge portal system, which is an important component of decision support system, is the platform for knowledge share, knowledge reuse, and knowledge creation. Developing knowledge portal for large scale Hydropower Corporation should take characters of legacy systems and future development into account.

It is necessary to study the characters of legacy information systems and existing knowledge resources when building a multi-layer architecture and a supporting system for the knowledge portal based on WEB SERVICE and J2EE. Enterprise application integration environment, which is platform-free, improves the interoperability of heterogeneous system and provides an open and unified environment, is realized by message processing mechanism, interoperability in distributed environment, software combination technology, heterogeneous system integration and application mid-ware specification. By enterprise application integration environment, data, information and knowledge from all the branch departments are integrated effectively and a platform of high performance for knowledge share and innovation is also provided.

References

1. C. Despres and D. Chauvel, Knowledge Horizons: The Present and Promise of Knowledge Management (Butterworth-Heinemann, 2000).
2. A. Preece, et. al,, Better Knowledge Management through Knowledge Engineering, IEEE Intelligent Systems Jan/Feb, 36-44 (2001).
3. U. Fayyad, G.G. Grinstein, and A. Wierse, Information Visualization in Data Mining and Knowledge Discovery (San Francisco, USA: Morgan Kaufmann Publishers, 2001).
4. H. Havenstein, Web Services Drive BI Evolution, Inforworld February 24, (2003).
5. B. Junnarkar and C.V. Brown, Re-Assessing The Enabling Role Of Information Technology in KM, Journal of Knowledge Management. 1(2), 142-148 (1997).
6. N. Dhyanesh., G.C. Vineel and S.V. Raghavan, DEVISE: A Methodology for Building Web Services Based Infrastructure for Collaborative Enterprises, Proceedings of the Twelfth IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE'03), 1080-1383/03 2003 IEEE, IEEE COMPUTER SOCIETY.
7. F. Curbera, M. Duftler and R. Khalaf, Unraveling the Web Services Web: An Introduction to SOAP, WSDL, and UDDI, IEEE Internet Computing, 2002-03/04, pp. 86-93.
8. T.W. Malone and K. Crowston, What Is Coordination Theory And How Can It Help Design Cooperative System? In the Proceedings of the CSCW'90 (New York: ACM, 1990).
9. K. Sandoe, Enterprise Integration (New York: Wiley, 2001).
10. P. Windley, What Web Services Can Do For Business Intelligence, Computerworld April 14, (2003).