

# Research on Information Resource Sharing System Based on Semantic Discover and Query Technology\*

Yan Li<sup>1</sup>, Sheng-Yang Li<sup>1</sup>, Huai-zhong KOU<sup>2</sup>

<sup>1</sup>Information Center of Yellow River Conservancy Committee of Ministry of Water Resources, Zhengzhou 450003, China

<sup>2</sup>Office of General Engineer of Yellow River Conservancy Commission, Zhengzhou 450003, China  
liyannju@163.com

**Abstract** - With the development of “Digital Yellow River” project, Yellow River Conservancy Commission accumulates a lot of information resource. There exists heterogeneous phenomenon in data. It is hard to integration and share. In order To solve this problem, the information resource sharing system is put forward to construct. The system provides a good technical solution to share and integrate information resources. The framework of semantic metadata based on ontology and formal descriptions of semantic information are introduced to the database system with heterogeneity, thus they are in favor of the users querying intelligence and improve the relevance of the query results. This paper carries through useful experiment from the point of theory and practice.

**Index Terms** - Yellow River, information sharing, ontology, semantics discover and query

## 1. Introduction

For a long time, Yellow River Conservancy Commission (YRCC) accumulates a wealth of information resources, which are of equal importance to the information infrastructure and the system of hardware and software as important factors in production and intangible assets in fathering the Yellow River. These information resources play an irreplaceable role in the overall framework of “Digital Yellow River” project. But these information resources have not been effectively integrated and shared and the rate of them is very low in utilization because of lacking uniform planning and adequate funds which is very difficult to meet the need of construction in information.

The information resource sharing system is a premise to achieve integration and sharing of information resources about the Yellow River and provides a portal used for management, query, download and other operations. The system is designed with the technology of semantic ontology metadata and the technology of Web Service, which provides users with a virtual and single information view logically through the conversion between domain (node) ontology and common ontology and masking the distribution of information resource and heterogeneous characteristics of information resource under the support of underlying database, specialized database and ontology metadata repository. So the semantic conflicts can be eliminated to achieve the operation between information resources mutually, which provides users with the safe, open and standardized service of sharing information resources.

## 2. Overall Framework

The information resource sharing system is a distributed framework with perfect control to access which integrates the distributed network database including the existing central data node and the sub-nodes fully. The system integrates the heterogeneous information easily and manages it effectively by dint of the normal information interface, the metadata described with information resource and the sharing agreement. Thus each scattered information source can turn into a part of the system expediently and can be accessed through a variety of service modules to sharing information.

With the help of global ontology, to establish shared vocabulary database and domain knowledge to be integrated areas, and all the distributed data sources can make use of shared vocabulary and shared knowledge of the global ontology to more effectively reduce the semantic heterogeneity of the various data sources. Catalog system of the Yellow River data resources is constructed with uniform standards, achieved distributed storage and logical centralization in order to truly realize the efficient allocation of resources and management. In accordance with the unified data resource classification identifies and metadata standards, the meta-database is establishment; the global ontology database of data center and the local ontology database of the sub-centers are established shared semantic ontology mapping between them. Ontology Knowledge database is key part of the framework, it provides users with intelligent search navigation, and at the same time to shield the semantic heterogeneity of distribution data resources, on the other hand, it can provide knowledge representation of data resources to support knowledge discovery and reason, and to support the establish of data resources data exchange and data sharing.

The distributed network database structure is composed of the central node and sub nodes. The central node stores the common ontology metadata repository and other databases. The sub nodes store the local ontology metadata repository and specialized database of corresponding sectors. The central node achieves access of sub-node database in virtue of the portals in the information resource sharing system. The administrator in the central node can manage and update uniformly the ontology metadata repository under the appropriate security permissions. The organization and management of sharing information in the central node is

\* The National Science & Technology Pillar Program during the 12th Five-year Plan Period (2013BAB05B01).

divided into two systems, that is the catalog system and the semantic search system, which are realized all through the metadata management. The catalog system is classified and managed according to the criterion of information classification that YRCC has constituted. The semantic search system has the function of automatic search and disposes the semantic conflicts by the global ontology database established in the central node that is supplemented by descriptions in context.

### 3. Construction Contents

The information resource sharing system for the Yellow River is a distributed data sharing network relying on the LAN by YRCC, and is a unified portal of information sharing service facing each sector, department and user. It can exchange data between data centers and data sub-centers and provides distributed information services. The system has the function of collection and exchange, query and browse, data download, data directory, user management, accessing statistics, etc.

#### A. Data Collection and Exchange

The information described by ontology metadata in the central node and sub node is stored in the ontology metadata repository of each node through the web interface of data collection and exchange. This module has the function of registering and updating ontology metadata automatically, auditing data, managing the log of data collection and exchange, and has the function of collecting and exchanging text, picture, multimedia, spatial data and other types of data. The function of registration is responsible for monitoring semantic registration request coming from each node about resources and services, and calls the updating function to register the ontology metadata repository based on directory service. The information registered is stored in the ontology metadata repository as the form of conceptual ontology metadata for users to search or inquiry. The function of automatic update refreshes periodically through semantic ontology metadata repository and achieves dynamic and timely update.

#### B. Query and Browse

The function of query and browse can provide users with the service of searching and querying semantic information transparently in order to help them quickly find the information needed. To meet the need of different demand, this module has the function of intelligent query based on semantic reasoning and the function of fuzzy search according to multi-keywords. The emphasis of semantic search is to match in knowledge and semantics and to accomplish the task of reasoning according to the reasoning rules using the ontology information in the repository. Therefore, the precision and integrity in query can be better guaranteed.

#### C. Data Directory

The data directory has the function of searching by directory, navigation and customizing personalized active directory, and also has the function of managing the directory

of information resource categories for the developer and manager. The directory management has the operation of adding, deleting, changing directory and others. At the same time, some shared directory can be licensed to the specific users by the administrator in terms of the resource sharing management system.

#### D. Standard Catalog Database and Ontology Database Construction

According to standard for data resources catalog classification and its metadata of the Yellow River Basin, we build data resources catalog database at all levels of the catalog service node, and realize the maintenance, update and manage of the catalog database and metadata database. Local domain ontology is the professional ontology, provides the concepts and relationships between the concept and the concept of a professional disciplines. The task of local domain ontology is to establish conceptualization, explicitly formalized, shared local domain ontology. Local domain ontology as the data source metadata, it includes domain concepts and concept relationships, that is, it is a view of data resources sharing.

For the construction of multi-disciplinary local domain ontology, we can build the professional local domain ontologies, then do ontology integration according to the relationship between the professional and local field. According to the method of construction local domain ontology, we build basin ontology database, it can solve the problem of the semantic sharing conflict of the data resource as metadata of data resources. Combination with the characteristics of basin, basin ontology database which is applied to shared and operated is constructed, according to the relevant technical standards promulgated by country and the Ministry of Water Resources. Basin concept is obtained mainly including the terms selection, extract ontology concepts, basin concept is obtained from the basin dictionary, database table structure and data, the existing planning and systems. The organization of the basin concept in the processing of ontology construction, is the main step, mainly to build the ontology class level for each category and to fill instance, the basin ontology construction is completed on the basis of the basin concept collection and organization, Finally, the basin ontology is assessed and modified in order to better share and exchange.

### 4. Pivotal Technology

The information resources about the Yellow River have the characteristics of complexity and multiplicity in content and form, heterogeneous configuration, storage in mass and distribution. And there is a semantic difference between different information resources. It is the key issue of how to manage, query, and share these information resources uniformly. Therefore, the management of information resource and the mechanism of information service are introduced into the process of developing the system, and the semantic reasoning on ontology is combined with the metadata information efficiently, that is to introduce ontology to

describe the exact meaning of various terms and the relationship of them in the metadata on the basis of using metadata to describe local database information, thereby the global mode in semantic unification is established to mask the heterogeneous phenomenon of bottom database. The semantic information resource sharing system has been constructed eventually on the basis of combining the ontology metadata with the directory service organically to solve the problem of exchanging, sharing, querying the shared information between different sectors in the distributed network. Directory service is a pattern of organizing information in term of tree form and is a method of connecting the information management with the information service.

The technology describing semantic information is introduced to give users an information resource sharing system that provides the service of semantic information, utilizing the existing information infrastructure and the criterion of information network protocol, shown in Figure 1. It is semantic and distributed for information processing. The users can visit and manipulate all the shared information resources and services placidly through the information resource system.

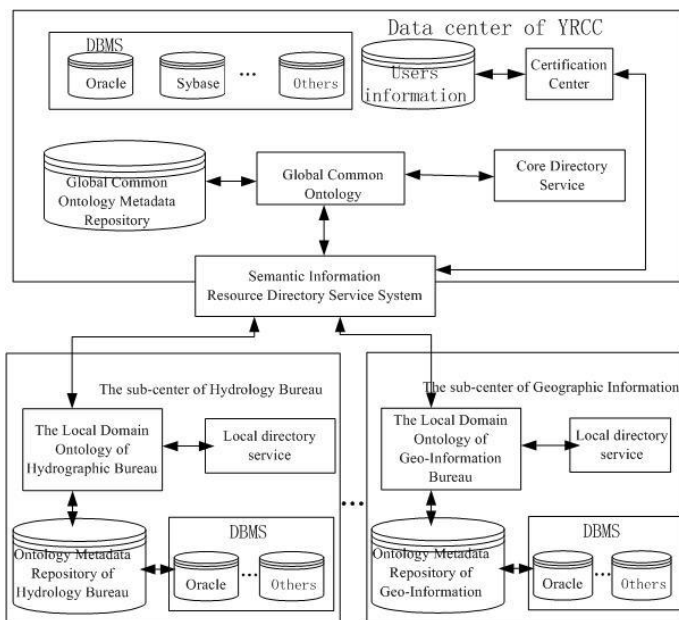


Fig. 1 The logical structure of the semantic information resource sharing system.

The core of resolving the problem of semantic heterogeneity is how to construct the ontology metadata repository including the metadata repository in global pattern (global ontology) and the metadata repository in local pattern (local ontology) and the relationship of mapping between them. The method of developing ontology knowledge from the bottom to the top is adopted to establish the ontology metadata repository following the principle of driving by application. Firstly the semantic ontology repository in nodes is constructed based on local knowledge. And then a common understanding of knowledge in this field is afforded through the analysis of local data sources and the capture of shared knowledge in local field of each node, the common terms are collected and the clear definition of these terms and the interrelationship between them is given from different levels of formal models that is described by the formal ontology modeling language to form a global ontology metadata repository. Finally a distributed network system has been established on the semantic ontology database. XML/RDF is utilized to describe metadata. OWL is used to describe ontology. And the domain ontology is established by the tool of protégé The tool of Jena developed by HP is used to parse the type of ontology document and its properties. Jena is a frame of Java that gives a portal of OWL to parse, establish and query the model of OWL.

### References

- [1] LIU Dao-fang, XU Zhao-fang, GAO Bing-sheng. Design and Research of the Metadata Management System of the Yellow River. Yellow River, 2009, 31(7):9-13.
- [2] Hubner S, Spittel R, Visse U, et al. Ontology-based Search for Interactive Digital Maps. IEEE Intelligent Systems. 2004, 19(3):80-86.
- [3] Ogden Fred L, et al. GIS and distributed watershed models, II -Modules, interfaces, and models. Journal of Hydrologic Engineering, 2001, 515-523.
- [4] Keen PGW. Decision support systems: The next decade .Decision Support Systems, 1987(3):253-265.
- [5] Michael S, Scott Morton. Management Decision Systems: Computer—Based Support for Decision Making. Cambridge, Mass: Division of Research, Graduate School of Business Administration, Harvard University, 1971:168-172.
- [6] Ogden F, Debarry P, Maidment D. GIS and Distributed Watershed Models I: Data Coverages and Sources. Journal of Hydrologic Engineering, 2001, 6 (6): 506-514.
- [7] LI Yan, LI Man-chun. A Study on the Construction of Water Resources Regulation Decision Support System of the Yellow River. Area Research and Development, 2009(5):140-144.