

Research on Data Sharing and Integration of Multi-source Information Systems

G.F. XU, P.W. XU, Z. WANG

Equipment Academy, Beijing, China

ABSTRACT: A data sharing and integration architecture of multi-source information systems is designed and this architecture is composed of 4 layers. A metadata format, DUBLIN CORE, is introduced in detail and the design idea is introduced as well. Besides, the process of data sharing and integration is introduced in detail.

KEYWORD: Sharing; Integration; Multi-source; Information System

1 INTRODUCTION

Nowadays, with the quick developments of technology about network and information system, information systems are widely used in our daily life and playing more and more important roles[1]. Due to the weak points of management system, lack of unified data standard and imperfect technology method, a lot of 'Data Island' or 'Information Island' got out among kinds of information systems in some particular areas[2]. The comprehensiveness, uniformity, correctness, sharing degree and integration degree of data seriously restrict the deep analysis, excavation and value added utilization of information resource, which cannot meet the strongly increasing data requirements.

In this paper, a method combined with XML protocol standard[3] was researched to solve sharing and integration problems among information systems. Metadata[4] format was studied as well. And detailed steps of data sharing and integration process would be introduced.

The rest of this paper is organized in the following sequence. In section 2, data sharing and integration architecture of multi-source information systems will be introduced in detail. In section 3, metadata format and the design idea will get elaborated in detail. In section 4, the detailed data sharing and integration process will get described. Concluding remarks are summarized in the final section.

2 DATA SHARING AND INTEGRATION ARCHITECTURE OF MULTI-SOURCE INFORMATION SYSTEMS

At present, data is stored and represented in a variety

of ways. The types of database are SQL SERVER, ORACLE, MYSQL, KINGBASE etc[5]. There are structured or unstructured text information and data in other formats.

In order to realize the sharing and integration of multi-source information systems, the heterogeneous data sources should be unified and standardized at first. And data access method should get automatic adapted as well. Then the user's application requirements would get meet. The data sharing and integration architecture of the multi-source information system is designed and shown in Fig. 1.

There are 4 layers in this architecture. They are application layer, data access layer, meta-database layer and heterogeneous data resources layer.

1. The heterogeneous data source layer is the basics. It is a logic set of all kinds of heterogeneous data source of all information systems. It contains all the related data source.

2. The meta-database layer is a unified representation form of heterogeneous data source based on XML protocol standard. The meta-database is consisting of many metadata records, which is the simplest form to express the information. The design ideas of meta-database will get described in section 3.

3. The data access layer provides a dynamic adaption method for different application systems to choose the optimal and appropriate data access method[6]. Because each information system may has different data access method, such as JDBC, ODBC, and SOCKET etc.

4. The application layer refers to the various application systems access the metadata and meet the users' application demands.

The detailed data sharing and integration process will be described in section 4.

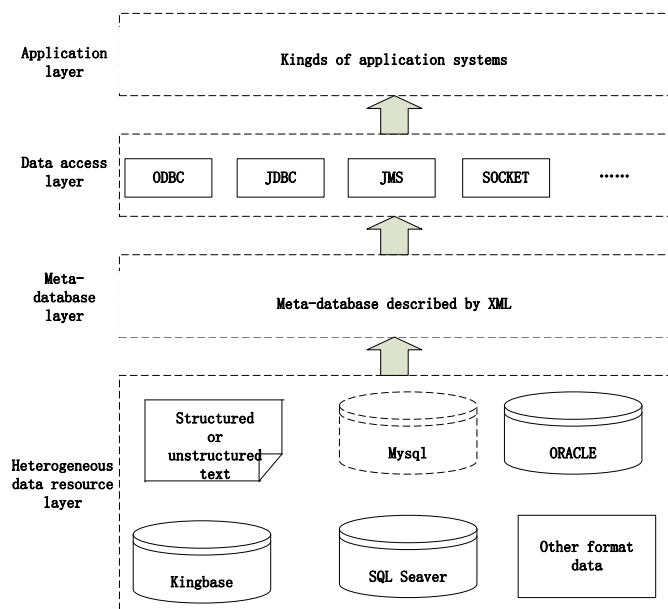


Fig.1 Data sharing and integration architecture of multi-source information systems

3 METADATA FORMAT AND THE DESIGN IDEA

Metadata can solve the problem of data inconsistency among heterogeneous data sources. In the current mainstream, there are two main metadata standards. They are DUBLIN CORE (DC)[7], GILS-information resource location service standards[8].

DC and GILS are all descriptive metadata standards. They can not only describe the virtual information resources in network, but also can describe the physical resources. They both emphasize the discover function of virtual information in network. Due to the advantages in scalability, interoperability and resource discovery function of DC, XML protocol standard will be used to describe and design the meta-database based on the standard specification of DC.

The Dublin core element set is a solution to make a standard for international standard Web resources architecture. It defines a common core standard for all Web resources and its content is less and general. It only contains 15 basic elements, which are enough to structurally describe the resource content, intellectual property and external attributes. And the field query is supported. It provides an effective method to solve the problem of organization and consolidation of the network information resources fundamentally. At the same time, the metadata standards for other types of resources basically are compatible with the DC standard, and there are some expansion based on DC. DC has become a formal standard RFC2413 of Internet and American national information standards Z39.85. DC is widely used. When DC is used, it can be extended combined with domain knowledge and the actual demand.

15 basic items of DC is shown in Table 1.

Table 1. 15 basic items of DC

	Item	Description
1	Title	Name of DLO
2	Author or creator	The original creator or organization of DLO
3	Subject or keywords	Words or sentences that reveals the theme or subject content of DLO
4	Description	Abstract or content description of DLO
5	Publisher	Organization that is responsible for the publish of DLO
6	Other contributors	Someone do not do the main work of DLO, such as the illustration author
7	Date	Publish or issue date of
8	Resource type	Type or category of DLO
9	Format	The data format of DLO, such as text, ASII, picture etc.
10	Resource identifier	String or number which can unique identify DLO
11	Source	Where the DLO derivates from
12	Language	The language used by DLO
13	Relation	The relations with other DLOS
14	Coverage	The regional scope of time and geography of DLO
15	Rights managements	Copyright statements(null, no restriction on tense, URL or other pointer)

DLO is short of document-like object, which is an electronic document that can be recorded like a traditional printing media.

4 PROCESS OF DATA SHARING AND INTEGRATION

The process of data sharing and integration is shown in Fig. 2.

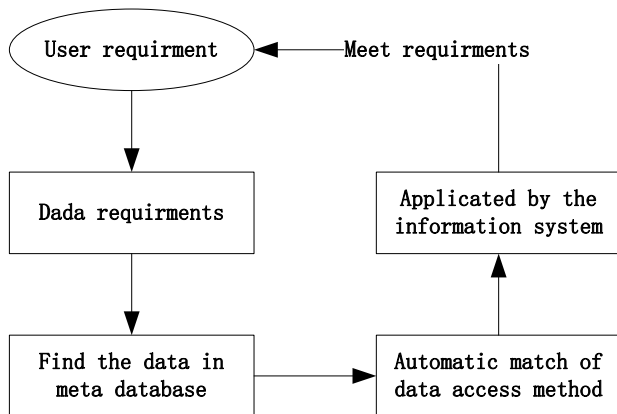


Fig.2 The process of data sharing and integration

First of all, the meta-database should be established.

In this section, the process of data sharing and integration is shown as follows:

Users put forward some specific application requirements.

The application requirements should be translated into data requirements based on the standard specification of meta-database.

According to the data requirements, data records should be found out in meta-database. Records IDs should get obtained.

Due to the data access methods of application systems are different from each others, an automatic match methods should be designed to solve the problem. Based on this method, application can directly without considering the data access methods at all.

The metadata would be used by the application systems and users' application requirements would get met.

5 CONCLUSION

In this paper, a data sharing and integration architecture of multi-source information systems is designed and this architecture is composed of 4 layers. A metadata format, DUBLIN CORE, is introduced in detail and the design idea is introduced as well. Besides, the process of data sharing and integration is introduced in detail. This architecture is a formal representation and should get further study and realization.

REFERENCES

- [1] Chang I, Chang C, Wu J, et al. Assessing the performance of long-term care information systems and the continued use intention of users. *Telematics and Informatics*, 2015, 32:273-281.
- [2] Wenxue Z. *Universal Method of XML Data Island Dynamic Generation and Data Binding*. Modern Electronics Technique, 2009.
- [3] Arikawa M, Kubota K. A standard XML based protocol for spatial data exchange - its capabilities and real applications. *International Workshop on Emerging Technologies for Geo-based Applications*, May 21-26, 2000, in, 2000.
- [4] Jones C, Blanchette C, Brooke M, et al. A metadata-driven framework for generating field data entry interfaces in ecology. *Ecological Informatics*, 2007, (3):270-278.
- [5] Zhen-lian Z. Several Types of Database Access to Technique in VC. *Journal of Nanyang Teachers College*, 2003.
- [6] Xiong-wen P, De-yu Q, Yong-jun L. Data Access Method in Globus 3. *Application Research of Computers*, 2006.
- [7] Levan R R. Dublin Core and Z39.50. *Journal of Library Administration*, 2001, 34(3):229-236.
- [8] Chun-nian L. Cognitive research on GILS. *Journal of Academic Library and Information Science*, 2004.