

Study on XML-Based Heterogeneous Agriculture Database Sharing Platform

Qiulan Wu, Yongxiang Sun, Xiaoxia Yang, Yong Liang, and Xia Geng

School of Information Science and Engineering,
Shandong Agricultural University,
Taian, Shandong Province, P.R. China, 271018
wqlsdau@163.com

Abstract. At present, there are large quantities of heterogeneous agriculture database in the network. Agriculture information resources sharing are the tendency of modern agriculture development. Being aimed at the problem of “information island”, “duplication of similar projects” and urgent need for agriculture information resources sharing existing in agriculture database construction, the feature of XML’s suitable for information sharing is analyzed, some key techniques of information sharing are studied, and the implementation plan of agriculture information resources sharing system is provided based on XML in this paper. The system realizes the information sharing and integration of heterogeneous agriculture databases making full use of the advantages of the XML’s independence with platform, easy extension and may format. The system has good extensibility.

Keywords: XML; Heterogeneous database; Agriculture information; Data sharing.

1 Introduction

In recent years, with the promoting of new rural construction, many agricultural research institutes and agriculture sectors have developed website of agriculture comprehensive information strongly supported by the government, at the same time a large number of agriculture information databases have been created. This agriculture information has played a positive role in Chinese socialist new rural construction to a certain degree. But most of agriculture information scattered in agricultural research institutes or agriculture sector, and these databases are Heterogeneous because of different construction time and technology, such as different of architecture, different of operating system, different of data semanteme etc. Therefore it is difficult to realize interconnection and share information, so that makes the redundancy construction and waste of resources. [1]

With the development of the rural area informationization, many information system need to access the data of these heterogeneous agriculture databases. Therefore it is necessary to establish a new system framework to solve the problem of information sharing and integration existing in heterogeneous databases. XML has the characteristic of the platform independent, easy expansion, better interactivity, strong semantic, may format and so on. So it enables the XML became a major standards of

the Web data express and exchange. According to practical applications, the necessity of the construction of heterogeneous agriculture database sharing platform is analyzed, the implementation plan of heterogeneous databases sharing system is provided based on XML with B/S mode and integration and query of heterogeneous data is implemented with uniform XML view. The system has good extensibility because of making full use of advantage of XML's data model.

2 XML's Features for Data Sharing

XML is a meta-markup language, user can define its own symbol and describe semi-structured data even Unstructured Data. It provides a bridge of data exchange between databases, it makes it easy to barrier-free communicate with heterogeneous systems. XML's advantage in data sharing aspect are as follows:

(1) Platform independent

XML is independent of platform and applications. So it is easy to solve the problem of data sharing and exchange between different application system or data source using XML effectively. Data expressed in XML format can be distinguished by other data source without any transition.

(2) Realizing multi display of data

XML is a markup language which can store and transmit data, which keep separation between store and display of data, and can display in Web browser with the help of other languages. A XML document can have the different display results with the different style sheet.

(3) Convenient for data release of Web

XML based on text format can be transmitted through the existing protocol of network communication like HTML, don't need to change anything.

(4) Reducing the network burden

Production documents writing in XML is convenient to be saved, transmitted, modified, searched and archived. Perhaps User wants only to browse or download one part of a product file through a network, doesn't want to download the whole files. If we use traditional text editor to write production document, Users have to download the entire document editing in traditional text editor, which easily cause network blocking. Any part of the production document based on XML can be easily extracted by the style, without involving the entire document. Taking advantage of the characteristics of XML, the complicated agriculture information can be determined and described standardized, and also can be transmitted and accessed effectively in the network. [2][3][4]

3 Design of XML-Based Heterogeneous Agriculture Database Sharing Platform

Firstly, in order to realize information sharing among the heterogeneous and distributed databases, a communication platform providing a uniform interface to

access the databases need to be established. When user access heterogeneous databases, they only need to specify the desired data, don't care for Data Schema, data extraction, data synthesis and data collection method. User can access heterogeneous databases transparently through accessing sharing platform.

3.1 System Architecture

Most of the agriculture information database is distributed in the agricultural research institutes and agriculture sectors. Databases of all departments are independent entities which have the characteristic of independence. In the sharing, users only have the privilege of query, can't modify the sharing database. In order to realize the function of query, the best way is servicing for the user through the B/S mode. Therefore, the three-level system structure is adopted based on B/S mode[5]. The architecture chart is shown in Figure 1:

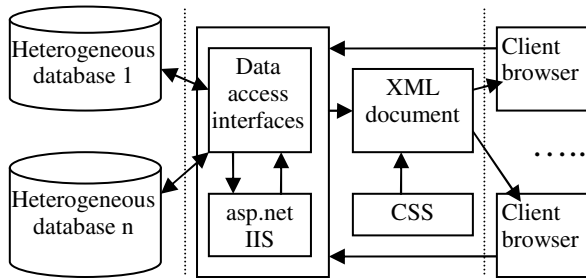


Fig. 1. Three-level architecture chart based on B/S mode

(1) User layer

On the client, users can access the agricultural information databases through submitting corresponding query on the Web, as long as users have the browser.

(2) Middle layer

A Web services middle layer established between multiple heterogeneous data sources and users acting as the communication platform for sharing information, which connect with heterogeneous data sources by using ADO data access technology and so on. The layer is the core part of the three-layer structure, which is responsible for receiving the remote or local of the query request and then converts the query into XML documents which will be uploaded to the application server. The application server deal with the logic request, deliver the corresponding SQL statement to the database server and obtain data from heterogeneous database; The query results are converted into XML documents, and translated into HTML or a variety of Script language, and then sent to the query browser. Similarly, the browser will also send request of changing, deleting and adding records to Web server, Web server complete these tasks through accessing the database.

(3) Data layer

Data layer consists of heterogeneous databases and files.

3.2 Model of Converting Heterogeneous Database Data into XML Data

In order to provide with user heterogeneous data of variety of different structures in a standardized format, the first thing we should do is to convert these data into standard XML data model format. Conversion process is shown in Figure 2:

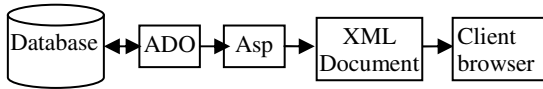


Fig. 2. Process of Conversion data into the XML document

The reference model of converting the query results for heterogeneous database into XML data through Web. The application program of ASP.NET extract data from the database through ADO, convert ADO dataset into XML document and then pass them to the Web server. The data is processed and displayed through Web browser as soon as data is stored in the server with XML standard format. The XML is shown exactly in DTD-defined format on client browser, so user can access heterogeneous database through Web.

3.3 Generating XML Pages Using ASP.NET

The relational data source convert into XML data source usually using ASP.NET, rather than creating the static XML document. The XML document generated dynamically from database can update automatically along with the update of database. ASP files can be used to create HTML pages and XML data source. The biggest difference between the them is the basic structure of a page one is the HTML page, the other is the XML data source. The output of ASP file is XML data source just like HTML page. [6]

The way of creating XML data source has two advantages:

Firstly, XML data sources can be generated using the script, which is convenient for converting a large number of relational data or general document data. Secondly, XML data sources of query can updated automatically with the update of database, because the XML data source is generated dynamically according to data in the database. So the query results of user can keep unanimous with the data of current database only through the data maintenance and management of every heterogeneous databases management system.

The key of the type of dynamic page generating in ASP.NET is to set the content type property of the response object. The dynamic HTML page is generated when content type property is text/html; the dynamic XML page is generated when content type property is text/xml. All text of XML page must be generated dynamically when XML page is generated dynamically using ASP.NET. Each statement of the page can be generated with response object of ASP.NET. For example, the following script can generate the XML text of large family of the major information:

```

<%@language=VBSCRIPT%>
<%
Response.write <?xml version='1.0' encoding='GB2312'
Response. Write <MajorFamily>
Response.write <no=2010011001>
Response.wrlte <name>WangNiu</name>
Response.write <sex>Male</sex>"
Response.write <speciality> Milk Cow Feeding
                </speciality >
response.write </ MajorFamily >
%>

```

When users query databases, the system generates dynamically SQL statement according to the sharing data structure of system database which is defined by the administrator of heterogeneous data sources. In general, the definition of table structure has represented the meaning of the data saved, when the structure of the sector system database is defined. So each table field names can be used as the tag of the data. Therefore, the generation of the tag can be simply expressed as:

The tag name= Rs (TabName) & Rs (ColName)

Rs is the result set which is returned by the local connection object when it queries the database.

In this format, two kinds of data need to be generated dynamically: one is the generation of the tag and the other is the generation of specific data.

The key of the generation of tag is to obtain the data and field name of the shared data structure which is defined in the various databases. Using ADO local connection object, the table and field names can be obtained from the sharing information tables, like the database query statement is generated dynamically. A tag name is composed of table name and field name. It can avoid duplicate tag name because there are the same field name in the different tables.

The specific data is obtained from the heterogeneous database. Therefore, the ADO remote connection object is needed; the query results are saved in the record set. The table of tags and the information of field are easily combined to the query results because the query statement of the heterogeneous database is generated according to the information of local database sharing information table. The data generation can be expressed as:

Data=R_Rs (TabName.ColName)

R_Rs is the record set which save data result returned by ADO remote connection object.

3.4 Display of XML Document on the Client

The correct display of XML document on the client browser not only need the XML document itself to be get the correct interpretation, but also need define XML's display style on the client. The structure content of XML document and display control is separated. There are two display methods of controlling XML documents: CSS (Cascading Style Sheets) and XSL (Extensible Stylesheet Language).

CSS provide a technology to make a Web page more fascinating on the browser, unlike the HTML that need to add the standard mark of control display frequently.

XML allows user to use any custom tag in documents, but the browser does not know how to display specific elements according to document tags in advance. So when the XML document is sent to the client, the information of how to display the document elements needs to be sent also. One of the ways to solve this problem is organizing the information of controlling document element's display into the mode of style sheet. By the style sheet, the browser knows how to format the element and display data exactly.

In the sharing systems of heterogeneous database, when the data in the heterogeneous database is converted into XML document, the output format needs to be controlled by the corresponding CSS on the client. The generation of tag is uncertain in the process of generating dynamic XML document, the definition of the display format of each tag becoming a difficult problem in CSS. The tag of root element is fixed by system in the process of generating XML document. CSS has the characteristics that the child elements inherit the parent element's attributes automatically. It is not necessary to consider the specific definition of each tag as long as the display format of the root element is defined and the CSS link declarations is included at the beginning of the XML document.

XML data is correctly shown on the client browser under the constraints of CSS. The different performance ways of the root element can be defined according to different needs. They are mainly font, color, effect and so on.

4 Conclusion

The sharing of agricultural information resources is the development tendency of modern agriculture. The construction schema of heterogeneous agriculture database sharing platform is provided using the existing network resources, and it is easy to be popularized. Some related technologies about the implementation of schema are introduced. Basing on the national and international research results, and according to the actual situations of Dingzhuang town, Dongying city and Xiangyin village, Taian town, a large number of agriculture information databases are created, such as practical agrotechniques database, agriculture policy information database, agricultural scientific talents database, produce market supply and demand database. On this basis, the XML-based agriculture database sharing platform is constructed. The platform adopts B/S three-tier architecture mode, and users on the client can access easily heterogeneous database only through Web browser.

Months of test operation at Dingzhuang town and Xiangyin village shows the system is stable and reliable, and it can meet the requirement of agriculture information resources sharing. The system effectively resolve the problem of "information island" and "duplication of similar projects" existing in the construction of agricultural informationization. It has a great significance for carrying out the national policy of "information Huimin", realizing perfect docking between the small peasant and big society and speeding up the process of the construction of socialist new countryside.

Acknowledgements

This study has been funded by State Bureau of Surveying and Mapping. The project name is the construction and demonstration of village and town comprehensive

information service Platform based on spatial data. It is supported by Shandong Agricultural University. Sincerely thanks are also due to Dingzhuang town and Xiangyin village for providing the data for this study.

References

1. Shang, M.H., Qin, L.L.: Data Share and Implementation for Heterogeneous Database. *Agriculture Network Information* (11), 70–72 (2008)
2. Chan, S., Dillon, T., Siu, A.: Applying a Mediator Architecture Employing XML to Retailing Inventory Control. *Journal of Systems and Software* 60(3), 239–248 (2002)
3. Rezayat, M.: Knowledge-based Product Development Using XML and Computer-Aided Design (32), pp. 299–309 (2000)
4. Li, S.: Land Using Data Sharing Research Based on XML Systematized Techniques, p. 18. Wu Han University (2005)
5. Xu, X.Q.: Design and Reality of Heterogeneous Database Data-sharing System Based on XML. Shengyang Ligong University (2008)
6. Dhamankar, R., Lee, Y., Doan, A.: Discovering complex mappings Between Database Schemas. In: *ACM SIGMOD (SIGMOD 2004)*, Paris, France, pp. 383–394 (2004)