http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 9 (5): 927-934, 2010
ISSN 1812-5638

Development of an Agricultural Spatial Information Sharing Platform for Supporting User Personalization

[1]Wensheng Wang, [2]Qingtian Zeng, [2]Xiangshan Hao, [2]Xiaoyu Tang, [1]Nengfu Xie and [1]Xiaorong Yang
[1]Agricultural Information Institute, China Academy of Agricultural Sciences, Beijing, 100820, China
[2]College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong, 266510, China

Abstract: In order to facilitate the sharing of agricultural spatial information, a platform is designed and developed, which supports user personalization. The framework and the register center of the platform are introduced with brief first. Using this platform, without user's attention, the user's data to be shared can be standardized. It means that the users can publish the shared agricultural spatial information in the original schema. To support user personalization, an approach is proposed to map the standardized data to the user's data. Although, the agricultural spatial data are organized and stored with a standardized schema, the platform can provide the shared data to the users in their own schema providing data.

Key words: Personalized information service, agricultural spatial information, web service, data mapping, data sharing

INTRODUCTION

Spatial data plays more and more significant role in delivering effective government services, informed decision makings and creating business opportunities (Mohammadi *et al*., 2009). According to statistics, 80% of various kinds of information is involved with spatial locations (Teng and Zeng, 2001) and thus spatial information plays a significantly important role in people's daily life. With the development of GIS and network technology, sharing of the spatial information on the Internet has becoming hotspot and focus in the field of resource sharing research. Wang *et al*. (2009) proposed a component-based management platform for multi-source spatial data. To develop the management platform, a set of components has been developed including the integration component for multi-source spatial data, the role-based security management component for spatial data, the user-friendly mapping component and the sharing component for spatial data in the study of Wang *et al*. (2009). To facilitate the exchange of the spatial data, an ontology and GML-based exchange model and SVG-based information visualization model are proposed by Wang *et al*. (2007). There are also many spatial data in geographic information system format on agricultural domain (Carpentier *et al*., 1998; Battaglin and Goolsby, 1995).

At the same time, with the development of the information technology, there are more and more researches focusing on personalized information service (Zeng *et al*., 2009; Wu *et al*., 2009). However, previous research on agricultural spatial information management or sharing systems has failed to consider user personalization of the schema of information service.

In this study, we mainly investigate the personalized information service in the sharing platform for agricultural spatial information. A user-centered data sharing schema and a data resources presentation schema that support user personalization are proposed. Present research in this study can be used for references for the personalized information service of any spatial information.

FRAMEWORK OF AN AGRICULTURAL SPATIAL INFORMATION SHARING PLATFORM

In order to achieve the sharing of distributed agricultural spatial information, we have independently designed and developed a sharing component for multi-source and heterogeneous agricultural spatial information based on web services technology, which can provide unified services for data sharing and operation class. Using the web services technology, we have created a reasonable spatial data sharing mechanism enabling the existing agricultural spatial data distributed in different

Corresponding Author: Qingtian Zeng, College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong, 266510, China

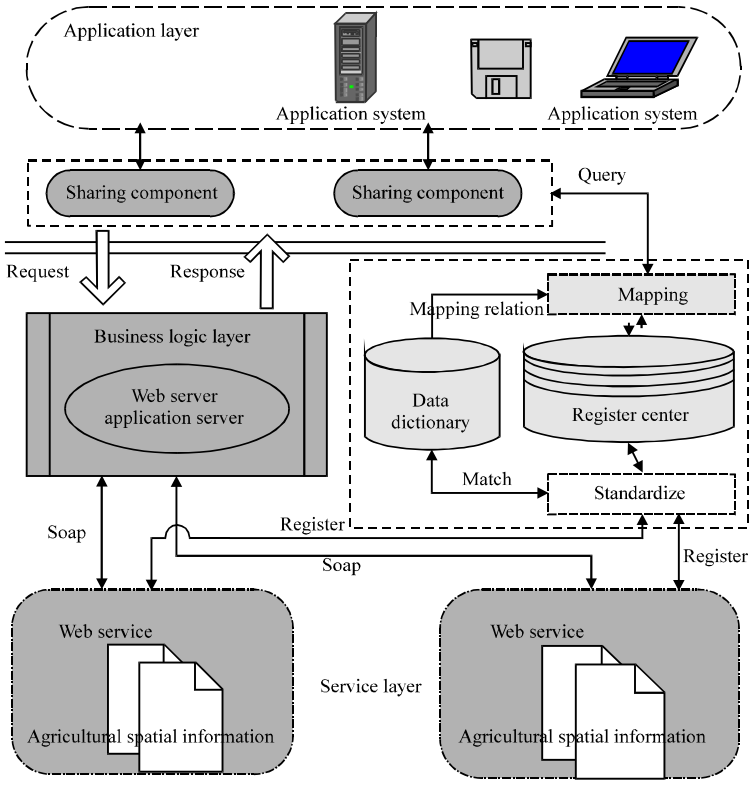


Fig. 1: Architecture of sharing platform for agricultural spatial data

departments to be integrated and assembled effectively. The sharing mechanism can also realize the establishment of interoperability platform, in which the dynamic access, query, spatial analysis and decision-making of all kinds of multi-source and heterogeneous agricultural spatial information are realized. Figure 1 shows the architecture and the main components of the platform designed to realize information sharing.

During the process of developing the sharing platform, in order to solve the problem of the data schema differences between different user nodes, it is necessary to provide the function of data transformation, i.e., the conversion of the schema of shared data to the same intermediate standardized schema. In our platform, the GML schema is adopted. Besides, a register center is designed as a part of the platform to collect data. As shown in Fig. 1, the part circumscribed in dotted line is the register center module, which plays a pivotal role in the operation of whole system.

DEVELOPMENT OF THE REGISTER CENTER FOR THE AGRICULTURAL SPATIAL INFORMATION

The services provided by register center consist of the registration and query of data. Figure 2 shows the major interfaces and the principle of the register center.

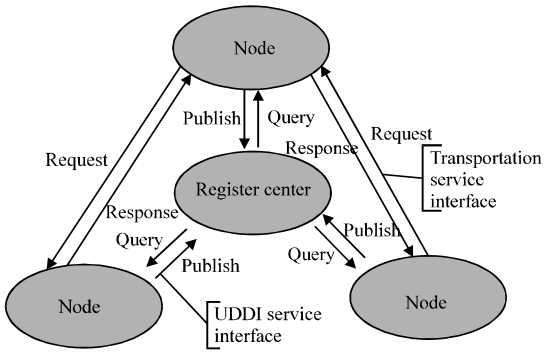


Fig. 2: Function and principle of register center

From Fig. 2, we can see that at least two interface classes are needed: interface class for publishing data and interface class for querying data. The interface class for publishing data performs the function of inputting the metadata of the spatial data files to share from subnodes into the central database, while the query interface class enables registered users to find the metadata of spatial data that they need in the central database by address.

To manage the metadata in the central database in the register center more effectively, a standard data structure specification is required. In practice, it is possible that each node has its own data structure specification defined

independently; meanwhile, data processing systems based on the independent data structure specification may already exist. In this case, the data provided when users publish metadata in register center is probably based on self-defined structure. Thus, before inputting data into central database, it is necessary for the register center to standardize the data to meet the unified specification. When a user queries metadata in the register center, if the register center can convert the data to the user-customized schema according to different users, the incompatibility between the schemas for different users' systems and the standardized schema would be avoidable and moreover, it embodies the idea of personalized information service and provides users with better services.

AUTOMATIC STANDARDIZATION FOR THE USER-CENTERED AGRICULTURAL SPATIAL INFORMATION

Heterogeneity of user-provided agricultural spatial data: In Fig. 3 and 4, agricultural spatial data from two distinct districts for sharing are presented respectively. We can see that the XML data submitted from different nodes, in the aspect of data structure, are different to each other, which means that they are defined separately and independently. In the standard specification of the register center, the fields of name, publisher, alias, abstract, key and category are obligatory; these fields are at the same node level and the sequence of them may be random and unspecified; other fields are dispensable.

On condition that the six obligatory fields and any number of other optional fields are provided, two aspects of the heterogeneity between the standardized and user-defined data structure can be permissible:

The tags of the fields can be named by users separately: For example, in Fig. 3, the metadata of the node is rainfall distribution, of which the tag name is defined as Topic here, while the standardized name of the corresponding fields should be name. In Fig. 4, the name of the same field is defined according to the standard specification as name. In this way, the method of naming the field in Fig. 3 is an example of the heterogeneity of naming a particular field.

The sequence of fields can be specified by users: For instance, the order of the metadata items in Fig. 3 is:

- <Topic> <VersionDate> <Alias> <Abstract> <Provider> <Version> <Progress> <RepresentationType> <SpatialResolution> <Key> <SubjectType> <DataVolume> <Category>

```
<?xml version="1.0" encoding="utf-8" ?>
- <Document>
- <MetaDataofNode>
  <Topic>Rainfall Distribution</Topic>
  <VersionDate>2009-4-28</VersionDate>
  <Alias>Rainfall Distribution Chart</Alias>
  <Abstract>rainfall</Abstract>
  <Provider>Jinan Agriculture Bureau</Provider>
  <Version />
  <Progress />
  <RepresentationType />
  <SpatialResolution />
  <Key>rainfall</Key>
  <SubjectType />
  <DataVolume />
  <Catagory />
  </MetaDataofNode>
- <MetaDataofNode>
  <Topic>Crop Distribution</Topic>
  <VersionDate>2009-4-28</VersionDate>
  <Alias>Crop Distribution Chart</Alias>
  <Abstract>cotton, soybean</Abstract>
  <Provider>Jinan Agriculture Bureau</Provider>
  <Version />
  <Progress />
  <RepresentationType />
  <SpatialResolution />
  <Key>cotton, soybean</Key>
  <SubjectType />
  <DataVolume />
  <Catagory />
  </MetaDataofNode>
  </Document>
```

Fig. 3: Data of Jinan Agriculture Bureau

```
<?xml version="1.0" encoding="utf-8" ?>
- <Document>
- <MetaDataofNode>
  <Name>Agricultural Pest</Name>
  <Summary>Pest</Summary>
  <VersionDate>2009-4-28</VersionDate>
  <AliasName>Pest Distribution Chart</AliasName>
  <Owner>Qingdao Agriculture Bureau</Owner>
  <Version />
  <Progress />
  <RepresentationType />
  <SpatialResolution />
  <Keyword>pest</Keyword>
  <SubjectType />
  <DataVolume />
  <Classification />
  </MetaDataofNode>
- <MetaDataofNode>
  <Name>Crop Distribution</Name>
  <Summary>Wheat, Maize</Summary>
  <VersionDate>2009-4-28</VersionDate>
  <AliasName>Crop Distribution Chart</AliasName>
  <Owner>Qingdao Agriculture Bureau</Owner>
  <Version />
  <Progress />
  <RepresentationType />
  <SpatialResolution />
  <Keyword>Wheat</Keyword>
  <SubjectType />
  <DataVolume />
  <Classification />
  </MetaDataofNode>
  </Document>
```

Fig. 4: Data of Qingdao Agriculture Bureau

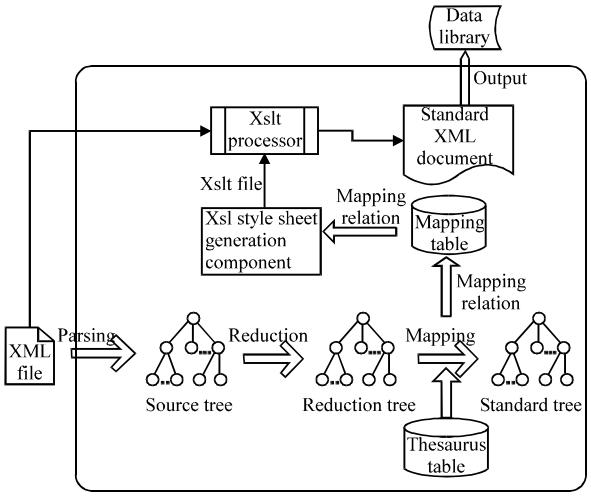


Fig. 5: Flow graph of standardization process

But, in Fig. 4, the sequence of the metadata items is:

- <Name> <Summary> <VersionDate> <AliasName> <Owner> <Version> <Progress> <RepresentationType> <SpatialResolution> <Keyword> <SubjectType> <DataVolume> <Classification>

From this aspect, the discord among different sequences of data fields is an example of the heterogeneity of different orders of sibling nodes.

Automatic standardization for the user-centered agricultural spatial information: The standardization process of user's data to be shared can be mainly divided into five steps: parsing XML document, structural reduction, mapping based on synonym, generation of the style sheet for conversion and the conversion by XSLT processor. The whole procedure is shown in Fig. 5. In the following, the functions of each module and the technologies used are discussed briefly.

Parsing XML document: Based on the XML DOM, this technology starts with reading XML document into RAM and organizing it into a tree. This module can be realized based on the DOM parser interface provided in the .NET platform. The input is a set of data in the form of XML document provided by users and the output is a DOM tree, the nodes of which consisting of document elements, document declarations, comments and text. In this step, we call the output DOM tree the source tree. The XmlDocument class in the class library is used as the DOM parser interface.

The reduction from source tree to reduction tree: We are aware that the text nodes, comment nodes and declaration nodes in the source tree are useless in the following steps, so such nodes should be reduced. A reduction module to remove the source tree is constructed to reduce the affecting the operations in the following steps disadvantageously. The production of the reduction process is a reduction tree. In this step, the DOM parser interface used is also the XmlDocument class in the NET framework class library.

The mapping from reduction tree to standard tree: This is the crucial step in the standardization process, which matches each node in the reduction tree to a node in a standard tree. The mapping relations are produced and recorded in the mapping table in dictionary database. In addition, the search for synonym in thesaurus in the mapping algorithm is also needed. The programming interface involved in the depth-first traversal of reduction trees is the DOM parser interface, while accessing the mapping table and the thesaurus table are mainly database operations, which can be realized with the ADO.NET technology.

The module for the automatic generation of XSLT file: According to the mapping relations in the mapping table, we generate the templates for the conversion of each node in the reduction tree one by one, with particular way of generating template for every one out of the three types of nodes. The data needed in generating template are, the path information and the standard name of the nodes in the reduction tree that stored in the mapping table; besides, the sequence of the nodes at the same level in the standard tree is also required. This step finally produces XSLT style sheet file as the output, which is provided to the XSLT processor module as conversion rule.

The automatic conversion of XML documents: The style sheet file is loaded by the processor module as the conversion rule to transform the XML file submitted by users to the standardized schema. The interfaces used in the processor module are provided in the NET Framework already, among which the XmlTransform class is the most important one in this study.

Figure 6 shows two expanded subtrees. The first subtree represents data from Jinan Agriculture Bureau. Compared to Fig. 3, 6 shows not only different tag names (Topic_Name, Alias_AliasName), but also a different sequence of tags as <Name> <Publisher> <AliasName> <Version> <VersionDate> <Synopsis> <Progress> <RepresentationType> <SpatialResolution> <Key> <SubjectType> <DataVolume> <Group>. The naming method and sequence for tags in Fig. 6 comply with the standard specification defined in the register center. It is

```
  <?xml version="1.0" encoding="utf-8" ?>
- <Document>
+ <MetaDataofDS>
- <MetaDataofDS>
  <Name>Crop Distribution</Name>
  <Publisher>Jinan Agriculture Bureau</Publisher>
  <AliasName>Crop Distribution Chart</AliasName>
  <Version />
  <VersionDate>2009-4-28</VersionDate>
  <Synopsis>cotton, soybean</Synopsis>
  <Progress />
  <RepresentationType />
  <SpatialResolution />
  <Key>cotton, soybean</Key>
  <SubjectType />
  <DataVolume />
  <Group />
  </MetaDataofDS>
- <MetaDataofDS>
  <Name>Agricultural Pest</Name>
  <Publisher>Qingdao Agriculture Bureau</Publisher>
  <AliasName>Pest Distribution Chart</AliasName>
  <Version />
  <VersionDate>2009-4-28</VersionDate>
  <Synopsis>Pest</Synopsis>
  <Progress />
  <RepresentationType />
  <SpatialResolution />
  <Key>Pest</Key>
  <SubjectType />
  <DataVolume />
  <Group />
  </MetaDataofDS>
+ <MetaDataofDS>
  </Document>
```


Fig. 6: Standardized data in register center

easy to conclude that, the second subtree in Fig. 6 which comes from the user of Fig. 4, is as same as the first subtree structurally. The standardization process regulates them to conform to the standardized schema described in the register center.

The examples given in Fig. 3 and 4 show four items of data provided by Jinan and Qingdao Agriculture Bureau, respectively, that present the standardized schema after being normalized in the register center. Two of the four items are shown in Fig. 6: the one named crop distribution provided by Jinan Agriculture Bureau and the other one named agricultural pest provided by Qingdao Agriculture Bureau. In addition, another two items of data which are submitted to the register center in Fig. 3 and 4, are not expanded in Fig. 6: the data named rainfall distribution from Jinan Agriculture Bureau and the data named crop distribution from Qingdao Agriculture Bureau.

After being standardized, the data will comply with the standard specification defined in the register center precisely. With this module for standardization, the other modules in the register center will no longer need to deal with heterogeneity in any way, as long as their interfaces are designed following the standard data description specification. The standardization process not only facilitates the design of the register center, but it also reduces the workloads of clients, improves the friendliness to data providers andachieves full compatibility of our platform with data from any individual user. In this way, the heterogeneity between user data and the data in the form of standardized schema can be ignored by every data provider, whose custom and experience are thus fully respected.

MAPPING-BASED PERSONALIZED PRESENTATION OF THE AGRICULTURAL SPATIAL INFORMATION

Here, we emphasize the dominance of users on the aspect of data input process. From the point of user-friendly interface, we will discuss the working principle of personalized data presentation in our system and the supporting environment for the personalized data presentation.

In Fig. 7, a fundamental framework for the personalized presentation of user data is proposed, which includes three main steps: (1) request processing engine processes the query request from a user and then transfers the processed request to the query module after initiating it. Meanwhile, the query module starts the XSLT style sheet generation module as well. (2) the query module queries data in the central database. (3) Using the Mapping table and login information of current user, the XSLT Style Sheet Generation Module generates XSLT Style Sheet file for transformation from the standard data structure into the current user's customized data structure.

Because the data that meet the request is found, the Query Module converts them into standard XML data andthen loads them into the XSLT processor. The purpose for the XSLT processor is to load the XSLT Style Sheet file, to execute the relevant interfaces and fulfill the conversion from standard XML data to customized XML data. The file after being converted matches the data structure of the user exactly. When the XSLT processor finishes the process, it will send the personalized XML data to the Request Processing Engine. After that, the Request Processing Engine presents the data to the query user in a certain way. Thus, a process of personalized data presentation based on mapping table is accomplished.

In Fig. 6, the data in the register center is organized according to the standard data specification. Then, if a user from Jinan Agriculture Bureau wants to view data in the data center, what kind of result it would be? When the user checks out the data in the data center, the system will provide data on the working principle illustrated in Fig. 7. First, the system will receive query command from Jinan

Agriculture Bureau and then find all data that meet the query. Secondly, according to the user ID of Jinan Agriculture Bureau and the Dictionary Database, the mapping relations between the user's customized data schema and the standardized data schema are discovered, on the basis of which the XSLT file used for transforming the data from standard schema to the customized schema of the current user is created. As the data in standardized schema retrieved and the XSLT file generated are input into the XSLT processor, the result data that qualifies the user's customized schema can be obtained. As illustrated in Fig. 8, for the sake of simplicity, no query user interface is provided here and by default, all data in the Central Database is retrieved and showed when any user logs in. Furthermore, the formatted display of the retrieved XML file is actualized here for convenience. In Fig. 8, the login user selects Jinan Agriculture Bureau and the four data items from Jinan Agriculture Bureau and Qingdao Agriculture Bureau are all displayed. Here, a GridView control is applied for the formatted display of data result. The tags of XML file are displayed as the content of table head of the GridView control. We can see from Fig. 8 that, the sequence of the fields in the table head is:

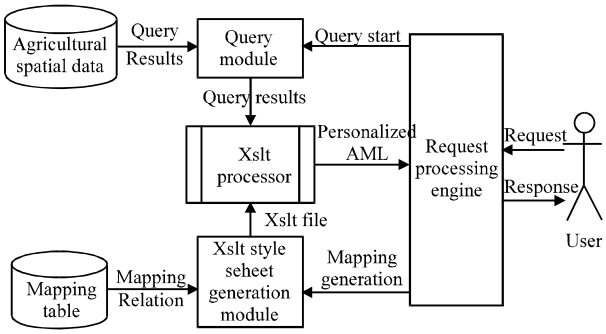


Fig. 7: Working principle of personalized data presentation

- <Topic><VersionDate><Alias><Abstract><Provider><Version><Progress><RepresentationType><SpatialResolution><Key><SubjectType><DataVolume><Category>

The order and the field names here are as same as those in Fig. 3, from which we can see, our method for the personalized data presentation based on mapping is able to recover data in the form of user's customized data schemata effectively and precisely.

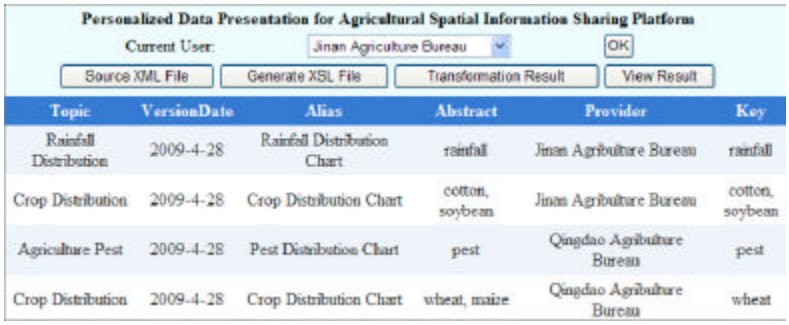
Moreover, In Fig. 8, it shows that in the column of Provider, there are two data items for Jinan Agriculture Bureau and the other two for Qingdao Agriculture Bureau. As we know, the two users provide self-structured data, but, no matter from which user, the data displayed here are organized in accordance with the user's customized data schema and vice versa.

The result for the user from Qingdao Agriculture Bureau is presented in Fig. 9. The four data items that this user receives are also equal to the data schema defined by Qingdao Agriculture Bureau structurally whether the data comes from the user or not. In Fig. 9, the order and the field names are:

- <Name><Summary><VersionDate><AliasName><Owner><Version><Progress><RepresentationType><SpatialResolution><Keyword><SubjectType><DataVolume><Classification>

As we can see, the data contents are all the same in Fig. 8 and 9. Exactly, they are the two data items from Jinan Agriculture Bureau and the two from Qingdao Agriculture Bureau, which are all stored in the database of register center. The sequence and field names in Fig. 6 are:

- <Name><Publisher><AliasName><Version><VersionDate><Synopsis><Progress><RepresentationType><SpatialResolution><Key><SubjectType><DataVolume><Group>



Topic	VersionDate	Alias	Abstract	Provider	Key
Rainfall Distribution	2009-4-28	Rainfall Distribution Chart	rainfall	Jinan Agribulture Bureau	rainfall
Crop Distribution	2009-4-28	Crop Distribution Chart	cotton, soybean	Jinan Agribulture Bureau	cotton, soybean
Agriculture Pest	2009-4-28	Pest Distribution Chart	pest	Qingdao Agribulture Bureau	pest
Crop Distribution	2009-4-28	Crop Distribution Chart	wheat, maize	Qingdao Agribulture Bureau	wheat

Fig. 8: Data view for Jinan agriculture Bureau

Personalized Data Presentation for Agricultural Spatial Information Sharing Platform

Current User: Qingdao Agriculture Bureau OK

Source XML File | Generate XSL File | Transformation Result | View Result

Name	Summary	VersionDate	AliasName	Owner	Keyword
Rainfall Distribution	rainfall	2009-4-28	Rainfall Distribution Chart	Jinan Agriculture Bureau	rainfall
Crop Distribution	cotton, soybean	2009-4-28	Crop Distribution Chart	Jinan Agriculture Bureau	cotton, soybean
Agricultural Pest	Pest	2009-4-28	Pest Distribution	Qingdao Agriculture Bureau	pest
Crop Distribution	Wheat, Maize	2009-4-28	Crop Distribution Chart	Qingdao Agriculture Bureau	Wheat

Fig. 9: Data view for Qingdao agriculture Bureau

For example, the field and the data for the tag Name in the standardized data schema are listed in the first column, which correspond to the first field Topic in Fig. 8 and the first field name in Fig. 9. The field and the data for the tag publisher in the standardized data schema are listed in the second column, which correspond to the fifth field Provider in Fig. 8 and the fifth field Owner in Fig. 9. The field and the data for the tag synopsis in the standardized data schema are listed in the sixth column, which correspond to the fourth field Abstract in Fig. 8 and the second field summary in Fig. 9. From these examples, we can see that the sequence and the names of fields that users see are different to those in the standard data description specification and in the data views for Jinan and for Qingdao Agriculture Bureau the sequences and names of fields are also different, which fully embodies the personalized data presentation.

CONCLUSION

There are many requirements about the effective sharing of agricultural spatial information resources. In this study, a platform supporting user personalization is designed and developed to facilitate the sharing of agricultural spatial information. We introduced the design and implementation of the framework and the register center in the platform. Using this platform, without user's attention, the user's data to be shared can be standardized. It means that the users can publish the shared agricultural spatial information with the original schema. To support user personalization, an approach is proposed to map the standardized data to the user's data. Although, the agricultural spatial data are organized and stored with a standardized schema, the platform can provide the shared data to the users in their own schema providing data. The approach proposed in this study can be used for any spatial information sharing system to facilitate user personalized information service.

ACKNOWLEDGMENTS

This study is supported partly by the NSFC under Grant No. 60603090, 90718011 and 50875158; the Excellent Young Scientist Foundation of Shandong Province of China under Grant No. BS2009DX004; the Research Foundation of the Technology Paper Sharing of Ministry of Education; the Open Project of Computer Architecture Lab of ICT, CAS, (No. ICT-ARCH200807); the Research Foundation of Shandong Educational Committee under Grant No. J08LJ77; Shandong Natural Science Foundation for Distinguished Young scholars of China No. JQ200816; and the Taishan Scholar Program of Shandong Province.

REFERENCES

Battaglin, W.A. and D.A. Goolsby, 1995. Spatial data in geographic information system format on agricultural chemical use, land use and cropping practices in the United States. US. Geological Survey Water-Resources Investigations Report 94-4176. http://pubs.usgs.gov/wri/wri944176/.

Carpentier, C.L., D.J. Bosch and S.S. Batie, 1998. Using spatial information to reduce costs of controlling agricultural nonpoint source pollution. Agric. Resource Econ. Rev., 27: 72-84.

Mohammadi, H., A. Rajabifard and I.P. Williamson, 2009. Enabling spatial data sharing through multi-source spatial data integration. Proceedings of the GSDI 11 World Conference, June 15-19, Rotterdam, The Netherlands, pp: 1-19.

Teng, S.W. and Z.M. Zeng, 2001. The study on GIS data sharing. The 2001 Annual Thesis Collection of China GIS Association, Chengdu, China.

Wang, Y.Q., H. Gao, C.Y. Sun and W.X. Shen, 2007. Research on multi-source heterogeneous spatial data exchange model based on ontology and GML. Proceedings of the 8th International Conference on Electronic Measurement and Instruments, Aug. 16-18, Xi'an, China, pp: 931-934.

Wang, W., C. Li, Z. Wu, Y. Luo and Q. Zeng *et al*., 2009. A component-based management platform for multi-source patial data. Inform. Technol. J., 8: 529-536.

Wu, Z.Z., Q.T. Zeng and X.W. Hu, 2009. Mining personalized user profile based on interesting points and interesting vectors. Inform. Technol. J., 8: 830-838.

Zeng, Q.T., Z.Y. Zhao and Y.Q. Liang, 2009. Course ontology-based user's knowledge requirement acquisition from behaviors within E-learning systems. Comput. Educ., 53: 809-818.