The Implementation of GIS Multi-source Heterogeneous Data Integration Considering of Java

Lingli Zhao, Shuai Liu, Junsheng Li, Zhongpin Lou, Xindong Mao, Enkun Xie

School of Engineering, Honghe University

Mengzi, P.R.China,661100

e-mail:zhaolingli_csu@126.com

Abstract—**With the development of information technology, GIS information has become an important part of the obtained data. The GIS data has characteristics of multi-source heterogeneous, multi-scale, large differences in format and huge data etc. It is difficult to achieve relatively joint query. So in this paper, we constructed a GIS multi-source heterogeneous data query platform to implement heterogeneous data integrity centralized management.**

Keywords- Data Integration ; Data Query; Data Index Tree; R-trees; Classified Index

I. INTRODUCTION

With the rapid development of urbanization and the rapid development of spatial and information technologies, the amount of valuable data captured in different systems and different structures grew and technical progress made it possible to link these different systems and different structures, the wish to exchange and share these data arose and became more and more important. But the combination of data done on different levels that we subsume under the term data integration (Bömelburg, 1996) requires resolving heterogeneities, which still poses research questions (Koch,2001). Database people have dealt extensively with architecture alternatives (Dadam, 1996).

Much research has been and is still done in order to develop and improve them, for example in the areas of feature matching (Gabay & Doytsher, 1995; Walter & Fritsch, 1997; Sester, Anders & Walter, 1998), eneralization (Guercke, R & Brenner, C, 2007; Weibel & Jones, 1998; Lamy, Ruas, Demazeau, Jackson, Mackaness & Weibel, 1999; Cecconi & Weibel, 2001) and semantic data integration (Birgit Kieler & Monika Sester 2007). For GIS (geographic information system) professionals and users, the urbanization has brought forward new geospatial challenges, and it forces the construction of urbanization in geospatial establishment.

These data have the characteristic of multiple types, multiple scales, large data, different data formats etc. There are a large number of studies from many scholars on different technical aspects of the data integration technology, these scholars have proposed the multi-source spatial data integration technology , such as database integration, model integration, the integration based on ontology, semantic integration etc.

In order to construct data integration system of urbanization, we need to construct a unified platform for huge multi-source data and according to the demands to realize data fusion and seamless connection between multi-source data in high precision, after integrating data resource, thus providing rules and foundation of multi-scales expression, visualization, spatial query and managements for constructing urban dynamic monitoring system.

In this paper, we adopted the eclipse platform using Java programming language and SQL Server 2005 as database as to implement data integration. The functions include data storage (spatial and non-spatial), file query, classification index query and joint inquiries based on hierarchical index tree.

II. SYSTEM ANALYSIS AND DESIGN

GIS multi-source heterogeneous data integration should have the management of map and data, and data joint query function. Therefore, the multi-source heterogeneous data integration system can be mainly divided into the software foreground and background. The main function of the foreground are providing interface to allow users to fast query to the relevant data needed, while the background is the management of a variety of data management to meet the functional requirements of the system.

We adopted the eclipse platform using Java programming language and SQL Server 2005 as database as to implement data integration. The main function of the system is divided into three parts: foreground map display module, background data management module and data query module.
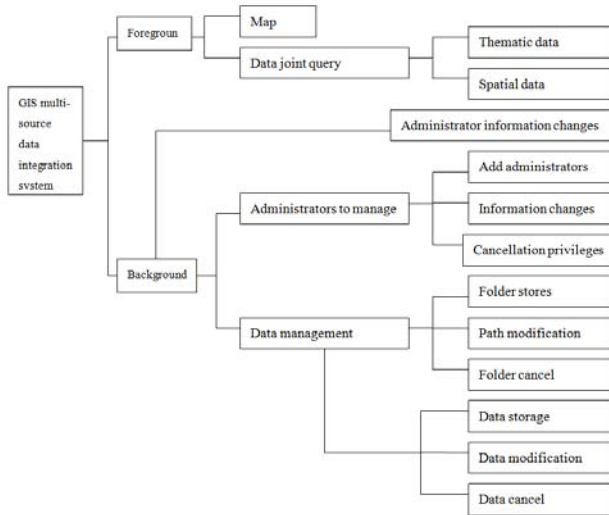
Figure 1. The functional diagram of the system

III. DATA INDEXING

Data index can be called to organize the data records, which are stored to a file in a data structure. It is a kind of description of the location information; it is stored on the data storage medium. Indexes can quickly find the data collection from the selected data records, which is an optimization procedure of data query. It takes different ways to access the same set of records. It is more convenient and efficient for us to access to data. In essence, access to data and data index is a kind of data structure for quick retrieval.

The spatial relationship of spatial data is more complex compared with the traditional data. It has characteristics of space, abstract, multi-scale and so on. Since spatial data has these characteristics, and thus the spatial data processing is a time-consuming and costly work. In order to carry out the fast real-time query for the object spatial position and effectively improve the spatial efficiency of data processing, we adopt an appropriate spatial indexing [9] in this paper.

Spatial index is based on the distribution of spatial objects and location, which contains basic information about spatial objects, such as the object's identity, the minimum bounding rectangle (MBR) and a pointer pointing to spatial object entity. Spatial index is arranged in order according to certain data structures, describing the storage space on the media object's physical location information. Through the optimization of the data retrieval operation, you can quickly find spatial data objects, thus can Improve the system efficiency of the data access records.

The multi-source data will be divided into spatial data and non-spatial (thematic data) data in this article, and sets up their indexes respectively and the relationship between them.

A. Thematic data index

Thematic data in various forms may be spreadsheets, document files may also be a database. For an example the urban data of One city, the project data can be classified from different angles, from different thematic perspective, the data can be divided into six major topics: basic geographic data class, population, economics, social class, tech, environment Ecology class; from the data by year category is divided into from 2000 to 2010 data from 11 different years; from the data by region, there are various different regions of the data, such as A, B , C zones; from data format type to be divided, with word, excel, .dwg, .bmp file formats and other data; data collection, the research team is divided into nine research group; from the collector to be divided, with different data collected by collectors. It can be seen, urban non-spatial data in the form of different thematic data, information scattered, to be quickly retrieved, must be based on characteristics of the data to establish an effective index, which is why this index to build classification for some reason.

Since the project data can be divided from different angles, in order to query useful information from the large amount of data required and from different angles to the original data classification, therefore, according to different standard for data classification. On the basis of the classification index, category index for different combinations of queries, so users can quickly retrieve the original data and queries. Urbanization index feature non-spatial data classification structure shown in Figure 2.
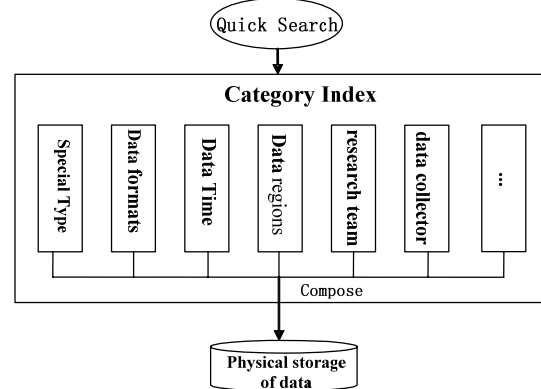

Figure 2. Category Index Chart for Urban Thematic data

B. Spatial Data Index

The adopted data index structure model [9] is a regional division based on hiberarchy index tree, by recording the minimum range for each area rectangle MBR, to create a region name and the spatial MBR based index direct link between (Figure 1 upper left); urban property data is to establish the classified index, data list range is one of the regional classification index, so it established a classification index and the links between MBR based index (upper right in Figure 1). The establishment of spatial data indexing and thematic data offers direct link for the data integration, and provides the necessary conditions for the joint query for spatial data and thematic data Figure 1 is the hiberarchy index tree and the MBR based index and the thematic data associated diagram.
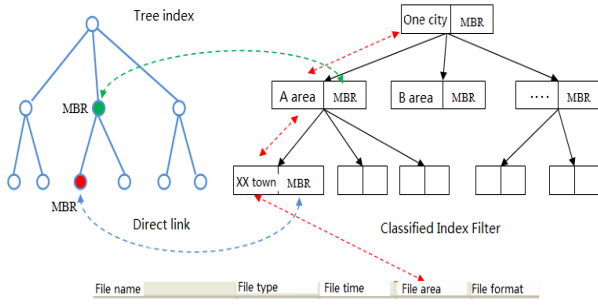
Figure 3. The Flow Chart of Building Relationships among Hiberarchy Index Tree, MBR based index and thematic data

The adopted hiberarchy index is a kind of tree structure. In order to define the tree structure, we must first define a number of the related basic structures. Define the leaf nodes of the hierarchical tree LevelTree is {name, id, MBR, LevelTList , size}. Where name is the zone name, id is the area number, MBR is the minimum bounding rectangle, LevelTreeList is a child node pointer, size represents the number of the child nodes.

Figure 4 illustrates the data type of leaf node LevelTree of the hiberarchy index tree, the structure diagram of child node table is shown in Figure 5.

The data type description of hierarchical index tree nodes

```
Class HITreeNode{
Data menber
name, id, MBR, LevelTreeList,TreeList, num, size
Oprations:
LevelTree(string name, string id, double[] mbr);
GetName();
SetName();
GetID();
SetID();
GetMBR();
SetMBR();
GetTreeList();
SetTreeList(LevelTreeList);
GetListSize();
SetListSize();
}
```

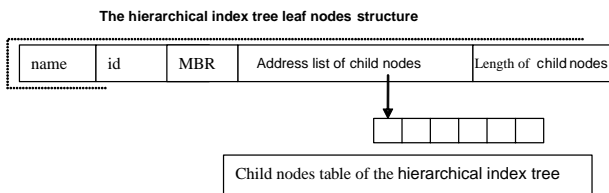Figure 4.   The data type description of hierarchical index tree

Figure 5.   The structure schematic of hierarchical index tree leaf nodes

IV.   DATA JOINT QUERY

District name, the zone number ID and MBR are the main content of hierarchical index tree, through these elements to build spatial data and non-spatial thematic link between the data in order to achieve a joint inquiry these two types of data.
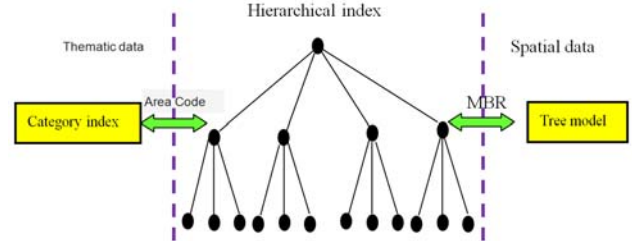

Figure 6.   The relationship between the hierarchical index tree and the data

There are three different ways for queries between non-spatial and spatial data in the hierarchical tree structure. The first is through spatial data query thematic data. The second way is through the thematic data query spatial data and relating information. The third way is through some portion of spatial data and non-spatial data to achieve comprehensive joint query. The joint query between spatial data and thematic data are showed in Figure 7.
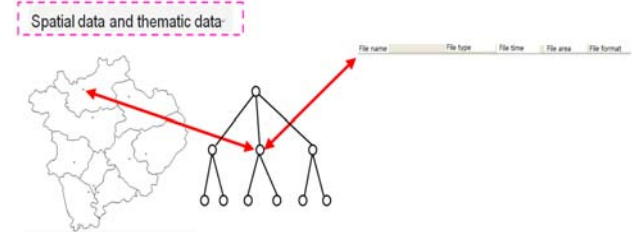

Figure 7.   The joint query between spatial data and thematic data

V.   EXPERIMENT

This design of GIS multi-source heterogeneous data integration system interface is shown in Figure 8. The top left of the interface is the original data directory, the specified file below the interface after the file information table may display the file's basic information. There are two main property pages on the upper right portion of the interface, they are respectively classification query display and map display .There is a basic information sheet of thematic data on the bottom of the interface showed in Figure 8.
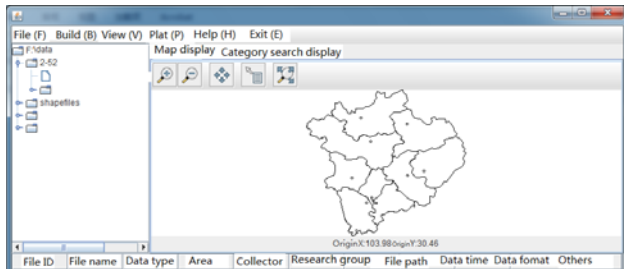
Figure 8. GIS multi-source heterogeneous data integration system interface diagram

REFERENCES

[1] Held, G., Rahman, A. A. and Zlatanova, S.. Web 3D GIS for urban environments[C]. Proceedings of the International Symposium and Exhibition on Geoinformation 2004 (ISG2004), 2004 Kuala Lumpur, Malaysia

[2] Teller J. , Keita A. K. , Roussey C. , Laurini R., 2005. Urban Ontologies for an improved communication in urban civil engineering projects. In: Proc. of the Int. Conference on Spatial Analysis and GEOmatics, Research & Devel-opments, SAGEO 2005 Avignon, France, June, 20th-23rd.

[3] Hamilton, A., Wang, H., Tanyer, A. M., Arayici, Y., Zhang, X., Song, Y., 2005. Urban information model for city planning, ITcon Vol. 10, Special Is-sue From 3D to nD modelling

[4] Gröger, G., Kolbe, T.H., Czerwinski, A., Nagel, C., 2008. OpenGIS City Ge-ography Markup Language (CityGML) Encoding Standard, Version 1.0.0, OGC Doc. No. 08-007r1, Open Geospatial Consortium

[5] CityGML Homepage: http://www.citygml.org Featherstone WE, Sproule DM (2006). Fitting AUSGeoid98 to the Australian Height Datum using GPS data and least squares collocation: application of a cross-validation technique,Survey Review 38(301):573-582.

[6] Lapierre, A., Cote, P., 2007. Using Open Web Services for urban data man-agement: A testbed resulting from an OGC initiative for offering standard CAD/GIS/BIM services. In: Coors, V., Rumor, M., Fendel, E. M., Zlatanova S. (eds): Urban and Regional Data Management. Proceedings of the 26th UDMS, October 10-12, 2007, Stuttgart, Taylor & Francis

[7] Kolbe, T. H., Gröger, G., Plümer, L., 2008. CityGML – 3D City Models and their Potential for Emergency Response. In: Zlatanova, Li (eds.), Geospatial Information Technology for Emergency Response, Taylor & Francis

[8] Lee, J., Zlatanova, S., 2008. A 3D data model and topological analyses for emergency response in urban areas. In: Zlatanova, Li (eds.), Geospatial In-formation Technology for Emergency Response, Taylor & Francis

[9] Mäs, S., Reinhardt, W., Wang, F., 2006. Conception of a 3D Geodata Web Service for the Support of Indoor Navigation with GNSS. In: Proc. of 3D GeoInfo 2006 in Kuala Lumpur, LNG&C, Springer