SPATIAL DATA REPRESENTATION MODEL ORIENTED TO GEOGRAPHICAL PROBLEM SOLVING

Hongjun Su*, Yongning Wen, Min Chen, Hong Tao, Jingwei Shen

Key Laboratory of Virtual Geographic Environment (Ministry of Education), Nanjing Normal University, Nanjing,Jiangsu Province 210046, P.R.China - hjsu1@163.com

Technical Sessions WgS–PS: II/6

KEY WORDS: Geography, GIS, Data Representation Model, Geography Problem; Data Integration

ABSTRACT:

Geographical problem solving environments are becoming the most popular tools which used to solve sophisticated problems. In order to overcome the obstacles caused by mutli-sourced heterogeneous spatial data, a data representation model (DRM) is designed. A data representation model is a description used to provide identification of all data elements within a system, including their attributes and the logical relationships among all data elements. In object-oriented terminology, this is viewed as a class hierarchy, and described through a graphics-based design tool. The DRM provides not only a clear description of the data, but also defines the relationships between the data that are critical to ensuring correct interpretation by users. DRM also provides a common data model to define a data representation structure for traditional data and spatial data. By the experiments, it has shown that it can satisfy the requirements of geographical modelling, analysis for geographical problem solving under today's open, heterogeneous environment.

1. INTRODUCTION

The acquisition ways of geographical data are extended quickly in the past years, and they have provide mutli-typed data source for geographers and researchers to understand, observe, model and analyze the geographical phenomenon. One of the important approaches to study geographical phenomenon is solving the geographical problems, and in fact, geographical problem solving is defined as any activities that use GI technologies to process geospatial information, with the purpose of attaining certain geographical application goals (Junyan Luo, 2007). The construction of geographical problem solving environments can provide an open, flexible tool for geographical problems research and analysis for geographers and common users, a visual user program interface for decreasing the program burden, an integration platform tool which facilitates the using of spatial data for geographical problem researches.

Geographical problem solving is a useful tool for geographic modelling, simulation and analysis, while geographic data interpretation and representation are the base of geographic modelling, simulation, analysis and geographical problem solving. In the reality, geographic modelling and model integration need the support of mutli-sourced heterogeneous spatial data, which will play more and more important roles in geographic simulation, analysis based on VGEs. Unfortunately, more and more heterogeneous geographical data are created in the past years, which hindered the development of geographic modelling, model integration and analysis. So it is urgent to find a new solution especially a new data representation strategy for understanding, describing and representing geographic data perfectly, and it will facilitate the integration of mutli-sourced heterogeneous data which support geographical problem solving in VGEs. It is useful and meaningful for geographic modelling and analysis.

Now, the development of GIS has spread into the stage of virtual geographical environments (VGEs), and VGEs, which is the opposite of really environment, is the reflection of really environment in mental scene, and it is the concentrate and representation of really environments. VGEs is a new type of GIS, it not only emphasizes on the 3D visualization, but pay attention to explore the geographic nature laws by simulating and analyzing graphical phenomenon more particularly. As a platform for geographic scientific researches, it has become a knowledge sharing platform which emphasizing database and model database synchronously, that is to say it can create new knowledge by itself. It is meaningful for the study on geographic modelling, simulation and problem solving.

According to the problems raised in the process of geographic data interpretation, representation, the data representation method for mutli-sourced heterogeneous spatial data is studied under the geographical problem solving environments. For the most important, data representation model and semantic representation specifications for spatial data are discussed. Our purpose is aims at providing a credible data source for geographic modelling. At last, some experiments are designed and it has proved that the data representation model proposed in this paper can be used to interpret and represent spatial data effectively.

2. BACKGROUND

2.1 Geographical problem solving

As we all know that GIS has been recognized as a technology in the initial development stages, and it also can be described as a double faced coin. In the nowadays, people have more and more perspectives on the problem solving ability for GIS. But

* Corresponding author: Email: hjsu1@163.com, hjsurs@hotmail.com, Tel: +86-13851706937

as just one of the technology, GIS cannot solve all the problems; it is necessary to integrate the other technologies with GIS for the sophisticated problems. In fact, there are some challenges for integrate GIS with other technologies. Such as the challenges of open software integration, flexibility versus ease of use, knowledge production and understanding and so on. Thus geographical problem solving environments which can overcome those challenges are urgent needed.

Geographical problem solving environment was proposed based on the scientific problem solving environments (PSEs). And the idea of PSEs can be traced back to the 1960s, which did not implement due to the lacking of advanced computer hardware and advanced computer program languages. Since the mid 1980s, the once deceased idea of PSEs has been rejuvenated thanks to the advances of technologies. In 1994, PSEs as computational systems that "integrate powerful functionalities necessary to accomplish high-level tasks" and interact with users on "domain terms" was described by Gallopoulos and his colleagues (Gallopoulos, Houstis et al. 1994). Houstis et al. thought that scientific PSEs must provide a software environment that integrates different technologies for the users and guides or enforces them towards best practices (Houstis and Rice 2002). The goals of scientific PSEs well match the requirements of geographical problem solving. In order to answer the requirement of geographic problems, a geographical problem solving system, therefore, can be designed as a scientific PSE tailored for geospatial information and geographical applications, or more simply stated, as a geographical PSE (Junyan Luo, 2007).

A geographical PSE contains three key parts: an open software framework that integrates heterogeneous data and software functionalities, a visual programming interface that allows users to visually construct high-level geographical applications, and a knowledge-based system that enables users to retrieve, configure, and utilize computational resources in an automated and transparent manner. Geographical problem solving has the specific characteristics due to the factor that all of the real problems are different, so the ultimate meanings of geographical applications should always be analyzed case by case. Generally speaking, the solving method for iterative geographical problems can be categorized into three classes: incremental problem solving based on trial-and-error, solution refinement based on fast prototyping and task decomposition based on divide-and-conquer (Junyan Luo, 2007). Nevertheless, for the most important, the solving of geographical problem is based on the support of spatial data, and the heterogeneous characteristic of spatial data in the really world results in the difficult in problem solving; on the other hand, the researches on geographical problem solving are still in the beginning stages and only localize in fundamental theory, methods and framework etc.. The researches on spatial data representation which is important for geographical problem solving are rarely.

2.2 Spatial data models

Some useful spatial data models are proposed by scholars in the past years. The heterogeneous in the data model induced the difficult for data representation. For spatial data model, the typical 2D spatial data model contains entity data model which based on the point, line and face, raster data model which based on tessellation (Chen Jun, 1995); Coverage data model and object-oriented data model of Geodatabase (ESRI, 1981; ESRI, 2002); a hybrid data model based on raster and vector, also object-oriented spatial data model (Gong Jianya, 1992). For 3D data model, such as 3D vector data model , Octree data model , TEN and similar triangular prism are also studied(Chen Jun,1998; Li Deren,1997; Li Qingquan,1998; Wu Lixin,2007;Chen Penggen,2005). In addition, there are some works on the 3D spatial data models around the world, some useful models were proposed. Unfortunately, most of them are focusing on the modelling of the geo-entity itself, still pay little attention to the representation of the mutli-sourced heterogeneous spatial data and its integration model even oriented to VGEs, and it is a pity that it is not helpful for the development of the geography.

The spatial data models proposed in different fields have answered the questions in their own fields, but they are incapable in comprehensive modelling, mutli-sourced data interpretation and representation. Though the data representation method based on GML provide a choice for spatial data representation and integration, simple feature specification of OGC in it still can not satisfy the requirements of data modelling in large spatial information systems, even more GML has the intrinsic disadvantages in data organization and spatial analysis. Furthermore, data format exchange tools in FME and spatial data exchange specification such as SEDRIS which as a data exchange centre provide the support for spatial data representation and integration in technology level, but still can not be recognized as a spatial data model for interpreting, representing and integrating the spatial data seamlessly. In order to process the data which including ocean, atmosphere, topology, underground, drainage basin etc., a powerful spatial data representation method to support data modelling and integration is needed.

2.3 State of the art and problems

Different investigations of data integration have proposed different solutions and examples in respective fields. Unfortunately, most researches are localized in the follows three categories, namely data format interchange, data access directly, spatial data standard Interchange etc., and those still belong to the conventional methods in multi-sourced heterogeneous spatial data integration. All of those methods have their disadvantage in processing heterogeneous spatial data. The problems occurred should be paid more attention to. Firstly, the differentia of data model vary with the understanding of spatial data in different GIS software hampered the development of spatial data integration and interchange. Secondly, take all heterogeneous data into one format disobey the principle of distribution and independence for spatial data. Also, in order to integrate mutli-format data, the data access interface should be a common format used widely, it is not realistic in the near future. So, all the above three methods may have difficulty in representing heterogeneous spatial data.

In order to solve the representation problems of mutli-sourced heterogeneous data, three problems should take into account (Lu Guonian, 2005): data format exchange, geographic symbols sharing and topology reconstruction. Of them, topology reconstruction is the most significant issues because it is the base of simulation, reproduction of the geographic entities and phenomenon, also is the key which can distinguish geo-data integration from others.

From the analysis above, it is no doubt that the traditional spatial data representation has the barrier in data parse, data modelling, semantic library construction and topology rebuilding and can not be used for geographic modelling and analysis. In the paper, we emphasized that the purpose of the data representation model are provide a common data representation specification for geographic modelling and analysis under the geographical problem solving environments, and it is not only limited in the data integration. So, some other useful technologies can be used and combined with the GIS to provide a data tool for sophisticated problem solving under the geographical problem solving environments.

3. DATA REPRESENTATION MODEL

Geographic environment have the characteristics of hierarchy, inherent relationships among geo-entities, and it is a whole entity. The reality world is composed by a series of small entities, and those small entities can build up more complex entities. The entity which can not be divided up is named atom-entity, and it is the minimum cell that composes the complex entity and systems and each of them has the specific data type. Based on the analysis above, Data Representation Model (DRM) oriented to geographical problem solving is designed; it is an object oriented data representation tool.

DRM, an object oriented data representation model, based on XML technology, is the key part of the framework of multi-source heterogeneous data integration, and it is the foundation of extensible geography environment modelling and integration platform. A data representation model is a description used to provide identification of all data elements within a system, including their attributes and the logical relationships among data elements. In object oriented terminology, this is viewed as a class hierarchy, and described through a graphics-based design tool. The DRM provides not only a clear description of the data, but also defines the relationships between the data that are critical to ensuring correct interpretation by users. DRM provides a common data model to define a data representation structure for traditional data and spatial data. Using it, user can design their geographic data structure in order to correctly interpret other data.

Some nameable projects such as SEDRIS, Ptolemy and ArcGIS were referenced when design DRM. The priority of it to other data models is that it can provide an effective data describing and representation method namely geographic data representation model, and also provide a data exchange mechanism for distributed geographic data. There are also two key principles when design DRM, one is separating the semantics of what something represents from the "data primitives" used to represent it, the other is factoring out the common syntax and semantics of data models used to represent similar objects. Figure 1 shows the interface of DRM.

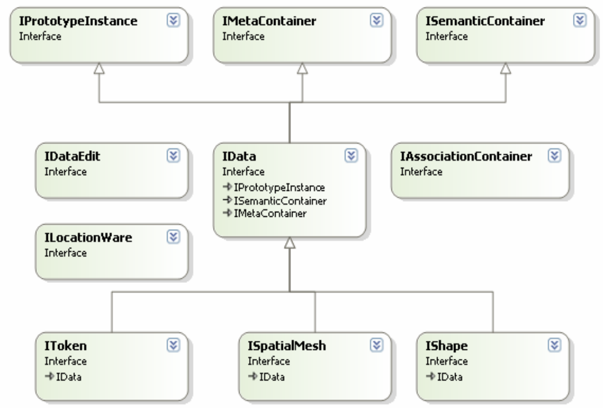


Figure 1. DRM data interfaces

DRM contains several parts such as Prototypes, Tokens, Shapes, Spatial Mesh and so on, and they can describe the really environment after be combined together. In the Tokens parts, there are some data types in common use as follows: Int8,Uint8,Int16,…,Double,Float,DateTime,String,Point,Line,Polygon,Zware etc. The geograhpic entities can be described by those data types, and in particularly, those data types are enclosed in Prototypes container as prototypes can be modified when designing.

- Prototypes: Deal with single data type, and providing root prototype for all kinds of data type.
- Tokens: Deal with complex data type such as value, list, array, table and so on, which is composed by single data type in Prototypes.
- Shapes: Represent geometric data type, mainly includes Point, Mutli-Point, Line, Mutli-Line, Polygon and complex polygon etc.
- Spatial Mesh: Describing the structure of spatial partition such as Tin, Grid, and Ten and so on.
- Locations: Tracking the original location of data.
- Associations: Storing the map or relation between different data, or topology.
- Semantic: Specification and interpretation of the data, and providing a categorization mechanism for geographic object. This work is partly referenced to ECDS of SEDRIS.
- Unit & SRS: Describing the unit and spatial reference of data, in which Unit was referenced the Unit system of Ptolemy, and SRS was referenced SRS of ArcGIS & SEDRIS.
- Meta: Data about data.

4. IMPLEMENTMENT STRATEGY

In this paper, the process of mutli-sourced heterogeneous data parse and integration based on DRM is designed and it mainly includes: normalize prototype template, design data template, construct data container, create instance etc.

Firstly, the basic data types consist of prototype template. The prototype template is stored in PrototypeLib as XML format, and the prototype in PrototypeLib is abstractly and can not be used directly until assigned name and ID information and stored in XML format. Single prototype template could be converted to complex prototype template by methods of combination and nested. The prototype stored in PrototypeLib can be edited, add semantic information, and spatial reference etc according to the use's requirement. The prototypes can be seen in Figure 2.

Secondly, data template is provided to describe the structural information and semantics attached for the heterogeneous data. The structural information refers to the organization and management for the spatial data, and the semantics attached to structural information describing the meanings of the structure and data. The process of design data template includes format interchange, semantic matching, data fusion, pattern extraction, metadata interpretation etc. It is should be paid attention to is that only structural information of data are stored, and the data template can not store the data itself.

Thirdly, data container is designed to store data template and provide the interface for input data. The structural information of data can be extracted from heterogeneous data sources and put into data container, and it is the structural relationship that becomes the bridge between data template and data container.
Lastly, the spatial data and its semantics put into data container based on prototype template and data template and when it finished the instance are created. The VGEs is composed of a series of instances. Figure 3 has shown data parsing and integration process.

```
<?xml version="1.0" encoding="gb2312" ?>
- <PrototypeLibrary name="PrototypeLibrary">
  - <items name="Value" type="ValueType">
      <item name="UInt8" uid="0" type="ValueType" enumType="UInt8" Description="" />
      <item name="Shape" uid="3" type="ValueType" enumType="Shape" Description="" />
    </items>
  - <items name="Map" type="MapType">
      <item name="Map" uid="6" type="MapType" Description="" />
    </items>
  - <items name="Table" type="TableType">
      <item name="Table" uid="9" type="TableType" Description="" />
    </items>
  - <items name="SpatailMesh" type="SpatailMeshType">
      <item name="TIN" uid="12" type="SpatailMeshType" enumType="smTIN"
        Description="" />
      <item name="GRID2D" uid="15" type="SpatailMeshType" enumType="smGRID2D"
        Description="" />
    </items>
  </PrototypeLibrary>
```

Figure 2. Prototypes in prototype library

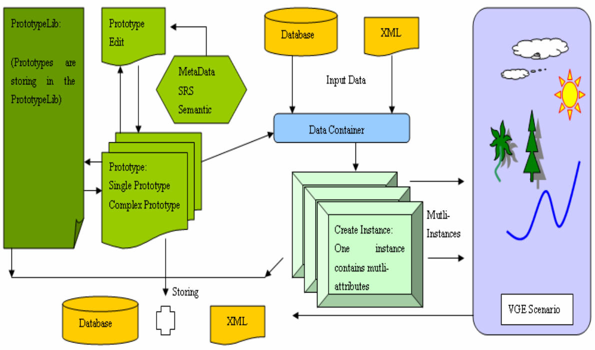
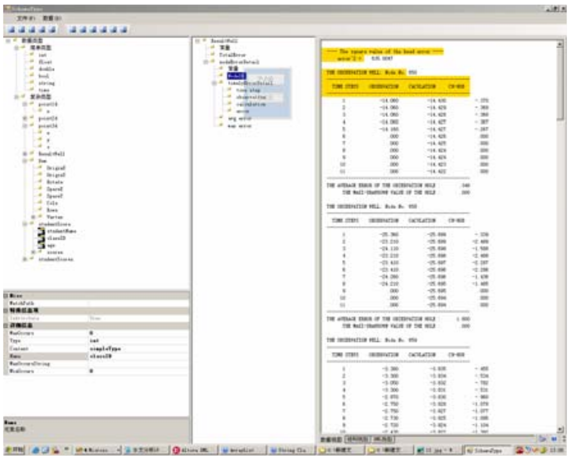


Figure 3. Mutli-sourced heterogeneous data parsing and integration process

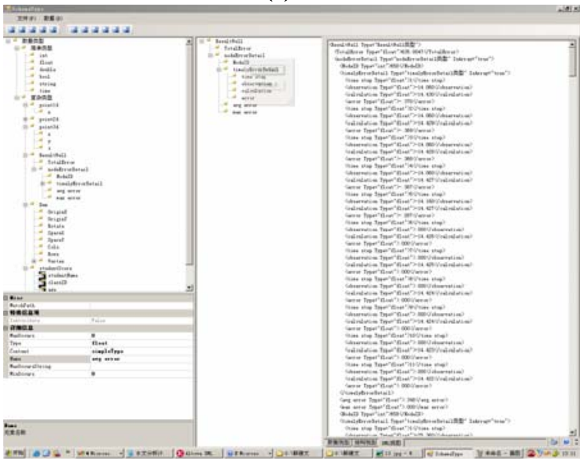
5. EXPERIMENTS AND DISCUSSIONS

The platform system of multi-sourced heterogeneous spatial data parse and integration is designed and developed under the VS.NET2005 development environment combined with XML technology and spatial data integration technology, and it can solve the bottleneck problems arise in geographic information resources which under the heterogeneous environment. Some experiments are designed and discussed, by the experiment, it has shown that the platform we developed can represent and integrate multi-source heterogeneous spatial data effectively and works better than other similar platform systems. Figure 4 has shown mutli-sourced heterogeneous data parsing and representation results.

In addition, the work discussed above is a part of the ongoing project; more works will be done in the next days. A largely virtual geographic scenario will be designed and constructed based on the DRM in the future for mutli-sourced heterogeneous distributed data integration. It is helpful for advanced geography researches. Our purpose is to provide a data representation tool to parse and integrate the spatial data which under the heterogeneous environments for geographical modelling and analysis.



(a)



(b)

Figure 4. Data parsing and representation results

6. CONCLUSIONS

Data representation plays important roles in geographical modelling, analysis. A data representation model oriented to geographical problem solving is designed and its main subparts are discussed in the paper. And the geographic data can be described in full-scale point of view. By the experiment, the data representation model can interpret, represent the mutli-sourced heterogeneous data effectively. In addition, the work discussed above is a part of the ongoing project; more works will be done in the next days. A largely virtual geographic scenario would be designed and constructed based on the DRM in the future for mutli-sourced heterogeneous distributed data.

REFERENCES

Chen jun, 1995. Key issues and research directions in GIS's spatial data models(in Chinese). *Acta Geographica Sinica*, 50(suppl.):24-33.

Chen Jun, GuoWei, 1998. A 3D topological ER model based on space partitioning (in Chinese). *Acta geodatetical et cartographica sinica*, 27(4): 308-317.

Cheng Penggen, 2005. Research on Three-Dimensional Spatial Data Models and Algorithms in Geology and Mine (in Chinese). *Ph.D. Dissertation*, Wuhan University, Wuhan, China.

ERSI,http://www.esri.com/software/arcgis/geodatabase/index.html

Gallopoulos, S.; E. Houstis, J. R. Rice,1994. Problem-solving environments for computational Science and Engineering. *IEEE Computational Science and Engineering* 1: 11-23.

Gong JianYa, 1992. An unified data structure based on linear quadtrees (in Chinese). *Acta Geodaetica Et Cartographic Sinica*, 21(4): 259-266.

Houstis,E. N.; J. R. Rice, 2000. On the future of problem solving environments. West Lafayette, IN, Computer Sciences Deparatment, Purdue Univeristy: 78.

Houstis, E. N.; J. R. Rice, 2002. Future problem solving environments for computational science. *Computational science, mathematics and software*. West Lafayette, IN, USA, Purdue University Press: 93-104.

Junyan Luo, 2007. The semantic geospatial problem solving environment: an enabling technology for geographical problem solving under open, heterogeneous environments. *Ph.D. Thesis, Doctor of Philosophy*, College of Earth and Mineral Sciences, The Pennsylvania State University, USA.

Li Deren, Li Qingquan, 1997. Study on a hybrid data structure in 3D GIS (in Chinese). *Acta geodatetical et cartographica sinica*, 27(5): 128-133.

Ptolemy. http://ptolemy.eecs.berkeley.edu/

SEDRIS. http://www.sedris.org/drm.htm

Wu Lixin, Chen Xuexi, Che Defu, 2007. A GTP-based Entity Model for Underground Real 3D Integral Representation (in Chinese). *Geomatics and Information Science of Wuhan University*, 27(5): 128-133.

ACKNOWLEDGEMENTS

The authors thank professor Lu Guonian for his instructive advices; also thank the grant from Key Program of Natural Science Foundation of China (No. 40730527).