

Object-Relational Data Structure & Organization of Massive Geo-Spatial Data

Chunbo Liu¹ Lixin Wu² Qi Li¹ Jinxing Hu² Keming Yang² Donghong Hu²

¹CyberGIS Studio, Peking University

Beijing, P.R.China, 100871

Tel: 86-10-62751962, 86-10-62751964, Fax: 86-10-62751964

E-mail: pkxcb@263.net

²Institute of Survey and Land Science

Beijing Campus, China University of Mining & Technology

D11 Xueyuang Road, Beijing, P.R.China, 100083

Tel: 86-10-62331012, Fax: 86-10-62325016

E-mail: wlx@mail.cumtb.edu.cn

Abstract Boundless, seamless, dynamic and multi-dimensional data organization of geo-spatial data is required for digital earth, which is conflicted with the traditional dual-component data structure and the digital earth data organization requirement. To fulfil this requirement, an object-relational model of geo-spatial data structure and the four-tiers-and-three-aspects data organization based on the object-relational DBMS are designed. Finally, an application prototype about the comprehensive management of massive mining map is presented.

Key Words Geo-spatial data, data structure, data organization, Mine GIS, Object-relational Model

1.Introduction

"Digital Earth" is a multi-resolution, three-dimensional representation of our planet, into which we can embed vast quantities of geo-referenced data^[1].

A new wave of technological innovation is allowing us to capture, store, process and display an unprecedented amount of information about our planet and a variety of environmental, social, economic and cultural phenomena. Today, we often find that there is far more information than we know how to use. We have insatiable hunger for knowledge, whereas a great deal of data remains unused. The hard part of taking advantage of this flood of geo-spatial information will be making sense of it -- turning raw data into understandable information. Notice that more than 80 percent of information is "geo-referenced" - that is, it refers to some specific place on the Earth's surface. We can use frame of "Digital Earth" to observe them. By using of the "Digital Earth", the geo-referenced data will be presented visually thus help people to understand them.

Above all, we must establish our digital earth, in other words, digitize our planet, on which we can then attach our other geo-referenced data. How to is based on the Poly-Vec model which is inherit from ESRI's Arc/Info system. There are several disadvantages that can't meet the demands of "Digital Earth" as following:

(1) The file system has predefined boundary of geospatial entity. The geo-spatial entity that interf-aces two or more file boundary will have paradox. Therefore the file system can't meet the

set up our "Digital Earth"? We'd first digitize the elementary data of our planet which is called SDI (Spatial Data Infrastructure) including Surface control Point, DEM (Digital elevation model), DOM(Digital Orthophoto Map), transportation network, river, administration boundary and cadaster et al. All the data above is geo-spatial data quite different from other geo-referenced data. Geo-spatial data is the frame data of digital earth while geo-referenced data is often attach on the geo-spatial entities as its attribute. The "Digital Earth" requires the geo-spatial data to be represented and organized boundlessly and seamlessly as in spatial aspect and multi thematic classification as the semantic. At the same time, the "Digital Earth" should represent the dynamic change of the spatial-temporal events and phenomena on/inside the earth.

The traditional method of geo-spatial data structure and data organization is based on the dual-component model which relies on the file system to manage the spatial data and the relational DBMS to manage the non-spatial data namely attribute data. The topology of the geo-spatial entity

boundless and seamless demands of "digital earth";

(2) The geo-spatial entity and its attribute data—geo-referenced data is stored and manipulated separately. Their consistence is ensured by one common SysID (System Identifier). However, in most file system, the SysID is base on the Geo-Spatial data file, thus it's very difficult to maintain the consistence of geo-spatial entities from

different geo-spatial data files.

(3) The Poly-Vec-model-based topological analysis is troublesome and is difficult to reconstruct and/or to extend the topology analysis function such as the management of the power network system, computer interconnect system, agent transportation et al.

(4) The file system can't compatible with the distributed computer model. The distributed computer model is core technology of DE, which will take advantage of the 3-tiers or multi-tiers C/S compute model that requires the back end be a data base system while not a file system.

(5) The safety guarantee can't meet the Digital Earth's requirement. Most DBMS support Transaction commit and roll back operation and other safety police, while the file system doesn't.

2. Object-Relational Model of Geo-Spatial Data Structure

In order to take advantage of the C/S and Internet/Intranet distributed computation mode and

the DBMS, a special data structure named Unique Object-Relational Mode Data Structure (UORMDS) is designed. UORMDS combines the graphic and non-graphic data into the unique object-relational database.

The Geo-graphic data is divided into two components, the location/shape data and topology data. The Geo-spatial data and the abstract of geo-spatial entity will be discussed in subsections.

Firstly, the presentation of geo spatial data and the topology of the spatial data will be described in the following two succeed subsections.

2.1. Geo-Spatial Data and Geo-Spatial Entity

Geo-referenced data is usually attached to the geo-spatial entity thus is often called attribute data. It may be text, common relational database record and/or multimedia data such as sound or video. The geo-spatial data maybe divided into vector data and raster data. Fig. 1 illustrated the classification of GIS data. The raster data is abstracted as a common image LBOB (Large Binary Object) despite whether it's DEM, DOM or RS Image. This paper mainly discusses the vector data structure and data organization based on the object-relational model.

As to the vector data, there are five elementary entity types: point, linear, area, text annotate and complicated entity, which is comprised of point and/or linear and/or area. These five elementary entity types are the objects that are used to represent, in other words, to reconstruct our planet or Digital Earth. Our task is to define, manipulate and management these objects efficiently base on

the object-relational model. Fig. 2 illustrated the basic entities and its relationships.

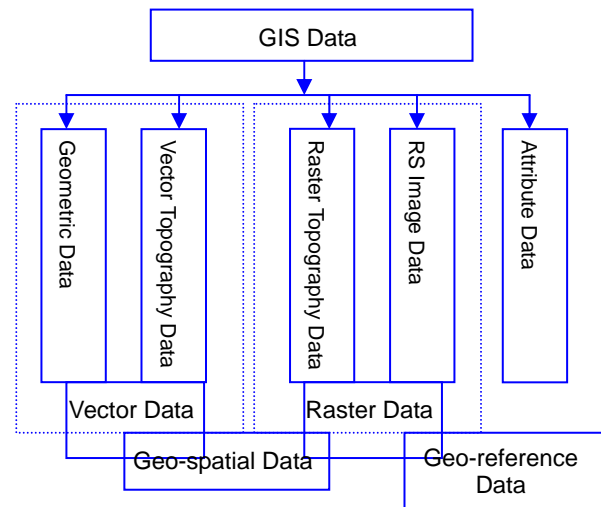


Fig. 1 The basic GIS data types

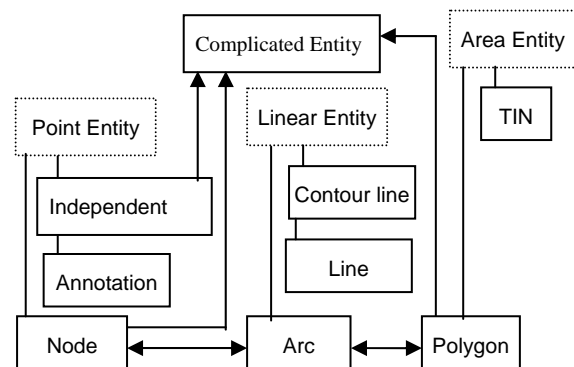


Fig. 2 the basic geo-spatial entities

Representation of Geo-Spatial Entity

Five LBOB (Large Binary Object) abstract data types (abbreviated as ADT), namely Point-Coordinate class (CCoordinate), PointID-list class (CPointIDList), ArcID-list class

(CArcIDList), EntityID class (CEntityID), and EntityID-list class (CentityIDList), are designed to abstract and characterize the location/shape data of Geo-spatial entities. Therefore, the Geo-spatial entities' location/shape data is characterized by using of the object-oriented database mode. Fig 3 illustrated the detailed physical storage structure of the five ADTs.

Class CCoordinate is constructed to record the 3D coordinate. CPointIDList Class is to manipulate the points composed of the arc, while the CArcIDList is to manipulate the arcs composed of

the polygon. CEntityID and CEntityIDList are to present the complicated entity.

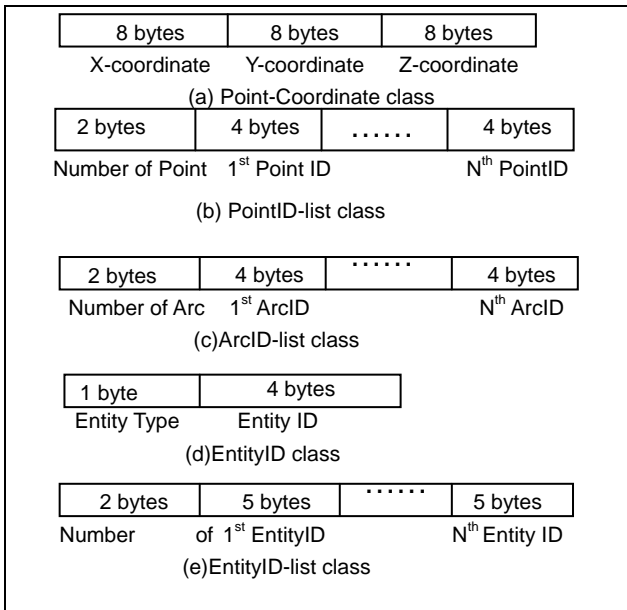


Fig. 3 The binary-stream structure of the five Geo-ADT

Class CCoordinate is constructed to record the 3D coordinate. CPointIDList Class is to manipulate the points composed of the arc, while the CArcIDList is to manipulate the arcs composed of the polygon. CEntityID and CEntityIDList are to present the complicated entity.

G-SQL (abbreviated from Geo-SQL), the extension of SQL to manipulate the spatial data, is designed to access the geo-spatial data in the object-relational database.

3. The Topology of Geo-Spatial Entity

There are three types of topology: point-related topology, arc-related topology and polygon-related topology. Fig. 4 illustrated the relationship between them and the entity type.

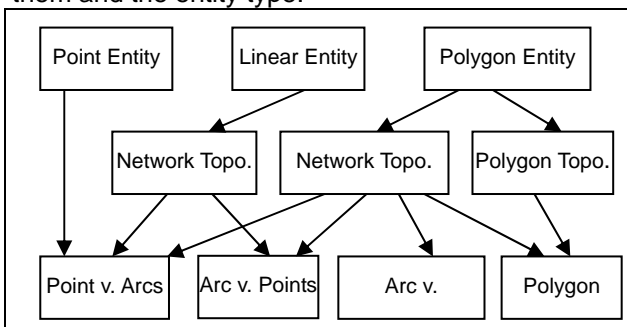


Fig.4 The topology between the Geo-spatial entities

Unlike the location/shape data of the geo-spatial entity, the topology data is characterized by using of the conventional relational database mode. Some of the topology data is unstructured as well as the

location/shape data, such as arcs related to specified nod point and so on. The components of the unstructured data are not equal. It may have its own extra different attributes. These extra different attributes may play very important role in the spatial analysis especially in the network topology analysis. So the traditional relational database is used to present the topology of the geo-special entities. It means that in the topology table one geo-spatial entity may has several records to represent the topology-relationship respectively, not that the location/shape data being abstracted as one LBOB as a whole.

The arc vs. points and arc vs. polygons topology relationship can be merged into one table to be represented. There are three types of polygon related topology: polygon vs. arcs, polygon vs. inner-polygons, and polygon vs. outer-polygons. The polygon vs. arcs topology is included in the representation of the polygon's location/shape data. Table 1~4 illustrated the presentation of the topology data of point vs. arcs, arc vs. point/polygons, polygon vs. inner-polygons, and polygon vs. outer-polygons respectively.

Table 1 The definition of point vs. arc topology table.

Field name	Description	Field type
Point SysID	SysID of the specified node-point	unsigned long
ArcID	The ArcID topo-related to Node-point SysID	unsigned long
.....	Extra attribute	

Table 2 The definition of arc vs. points/polygons topology table

Field name	Description	Field type
ArcID	SysID of the specified arc	unsigned long
StartNodeID	The start node pointID of arc ArcID	unsigned long
StopNodeID	The stop node pointID of arc ArcID	unsigned long
LeftPolygonI	The left polygonID of arc ArcID	unsigned long
RightPolygo	The right polygonID of arc ArcID	unsigned long
.....	Extra attribute	

As presented above, the location/shape data and the topology data of geo-spatial entity are all manipulated in the unique object-relational database system. Fig.5 illustrated the data

presentation of geo-spatial entity including location/shape data, topology and their relationship. Table 3 The definition of polygon vs. inner-polygons topology table

Field name	Description	Field type
SysID	SysID of the specified polygon	unsigned long
Inner-PolygonID	The polygon's SysID located in the polygon SysID	unsigned long
.....	Extra attribute	

Table 4 The definition of polygon vs. outer-polygons topologies table

Field name	Description	Field type
SysID	SysID of the specified polygon	unsigned long
Outer-PolygonID	The polygon's SysID the polygon SysID	unsigned long
.....	Extra attribute	

3.Object-Relational Based Data Organization Of Massive Geo-Spatial Data

The organization of massive geo-spatial data based on the object-relational model includes four tiers feature, feature class, theme, and map layout. On the other hand, the organization includes three aspects according to the character of geo-spatial data, namely spatial data organization, thematic classification organization and temporal organization. All the three Fig.5 The abstract and representation of geo-spatial data based on the object-relational model organization aspects are based on the four-tier organization.

3.1. Fur-Tier Organization Strategy

The four tiers of massive geo-spatial organization are feature, feature class, theme and map-layout. The feature that corresponds to one geo-spatial entity is the instance of feature classes. Every feature has a record in the geo-spatial object-relational database to record the geo-spatial data and the geo-referenced data. The feature class is the aggregation of features which have some similar characteristics. One feature class has one base database table. The features of one feature class have similar representation on the map, in other words, they use the same symbol system. The theme is aggregation of feature classes or a complicated feature class composed of several related feature classes. One theme may has a query of the tables of several feature class tables or a base table created for the complicated feature class. All spatial analysis is based on the theme tier.

The map-layout is the logical collection of several

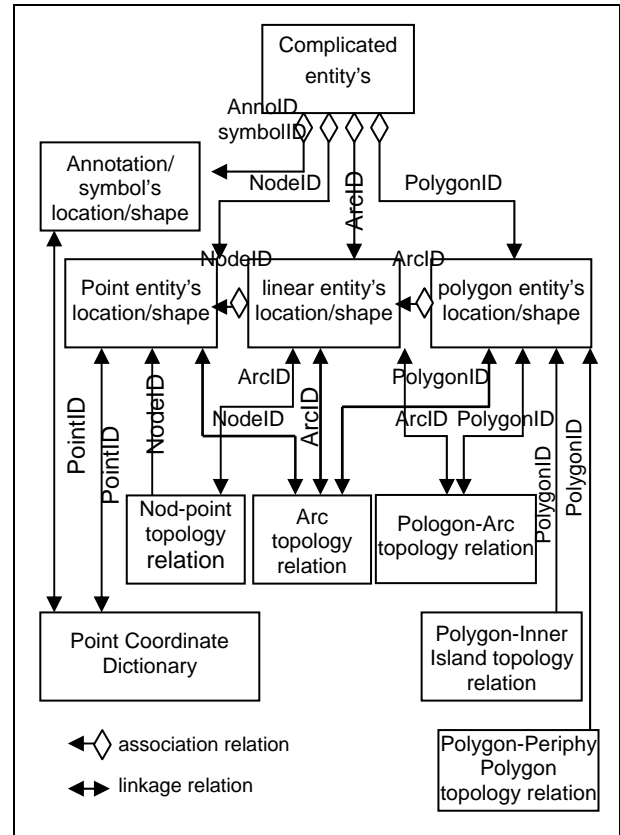


Fig.5 The abstract and representation of geo-spatial data based on the object-relational model

themes and/or feature classes. The map-layout tier is designed for mapping application and the application that based on the traditional map-layout. The relationship between themes and feature classes is defined in the theme defining table and the map-layout defining table defines the relationship between the map-layout and themes/feature classes. Fig.6 illustrated the relationship and control flow between feature, feature class, theme, and map-layout

3.2. Three Aspect of Geo-Spatial Organization

In order to effectively organize geo-spatial data of multi thematic classification, multi scales and multi temporalities, the data organization is designed including special character organization, temporal character organization and thematic classification character organization.

3.2.1. Spatial Organization

The special character organization responds to the multi scale character of geo-spatial data to meet the need of boundless and seamless organization.

The spatial organization is implemented according to the international standard of atlas (1:1,000,000). The spatial index and unique object identifier (UUID) technologies are used to support the rapid data access and unique identification of object especially for the objects that belong to two or more map-layout. A scale series is used to improve the efficient of multi scale browse.

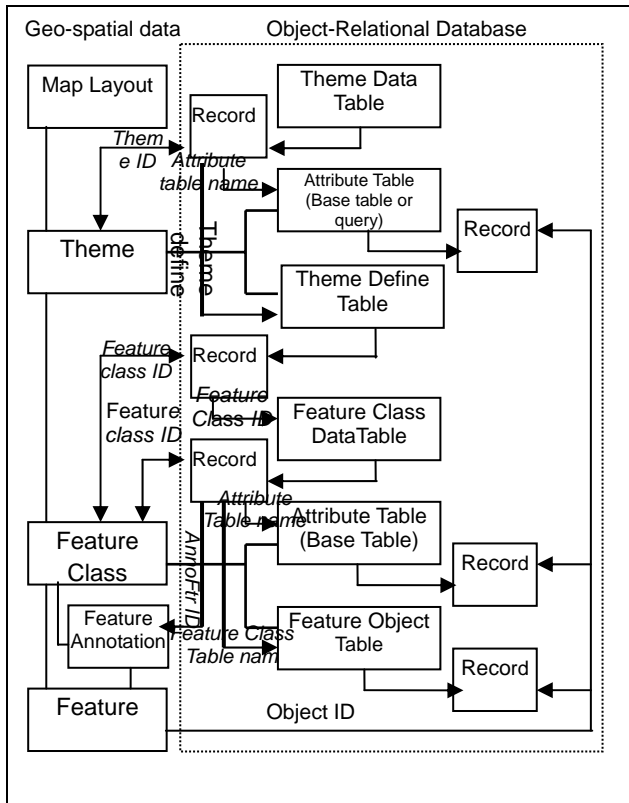


Fig.6 The control flow of four-tier organization of geo-spatial data based on the object-relational model

3.2.2. Temporal Organization

Temporal character organization is to represent the multi temporal characteristics or the dynamic of the geo-spatial data. The temporal organization is divided into three levels as illustrated in fig. 7. The temporal data is recorded in the geo-spatial entity table at the feature level. This is the most often used stratagem. Each entity can has one event time and a valid period. Thus the entities of different times are distinctive. The temporal data is recorded in the feature class defining table and the theme's defining table at the feature class and theme level.

3.2.3. Thematic Classification Organization

The thematic classification character organization responds to the massive data organization and management of multi thematic. Each thematic classification is corresponded to one feature class

in the object-relational database. At the same time, some of the thematic classification is corresponded to one theme according to the requirement of spatial analysis, or the thematic classification can't be represented as one feature class sometime.

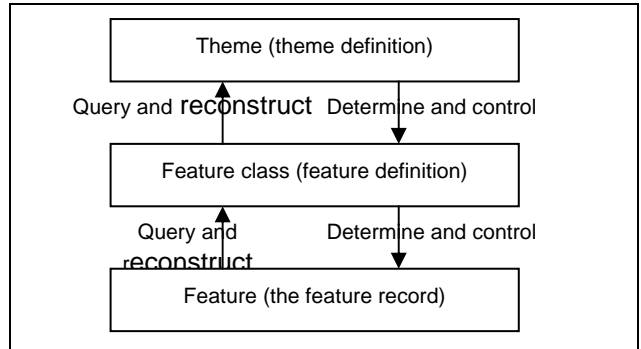


Fig. 7 The three tiers of temporal organization

4. Application Prototype

In the MGIS (Mine GIS) demonstration project in Kailuan Mining Bureau, we firstly classified the mining data into five levels, which contain 14 classes at the top lever, and encoding them uniquely. Secondly, we normalize the content of the frequently used mining maps according the unique code. Finally, the meta-data are abstracted according to the three layers: feature class, theme and drawing layout. Meantime, the three aspects: special character, temporal character and thematic classification character are suggested, by which the integrated management of massive mining maps was carried out.

4.1. Abstract of Meta-Data For Mine Map

The meta-data for mine-map includes mainly three parts besides the common meta-data. All the data are extracted from the geo-spatial object-relational database as illustrated in fig. 8.

4.2. Development Remarks

The Microsoft Visual C++, Visual J++ and Delphi developing platform are selected as developing tools, and the integrated GIS developing kit—Titan is selected to carry out the access and operation of spatial data in the object-relational database. Titan GIS is a powerful spatial data management system which contains spatial data manager (SDM), spatial data access engine (SAE) and spatial data structure query language (SSQL) 3 components. The SSQL is the extension of SQL to support the spatial data operation and manipulation, which enables the windows programmer to SAE. The SSQL command sets support the common operations of most GIS software.

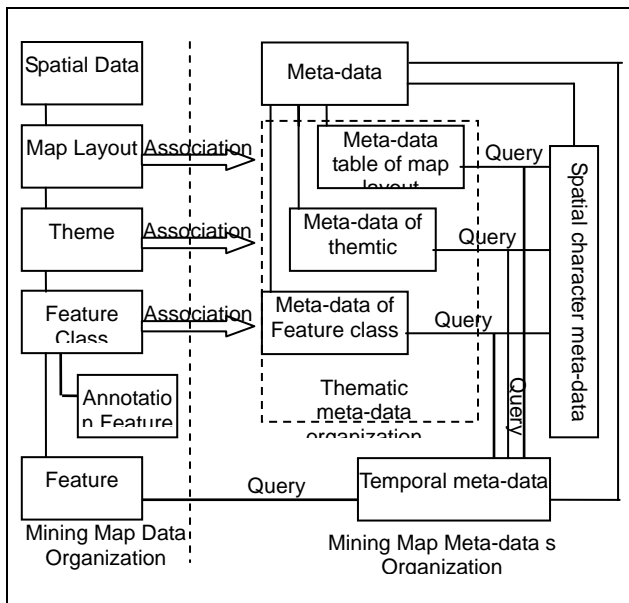


Fig. 8 The extraction of mine map meta-data from the Geo-spatial Object-relational database I

But we need to say that Titan can't support the access of object-relational database directly and all algorithm is based on the it's own *.tsf file system. So we designed a middleware to solve the problem. Fig.9 illustrated our strategy for object-relational access in order to use the titan developing kit.

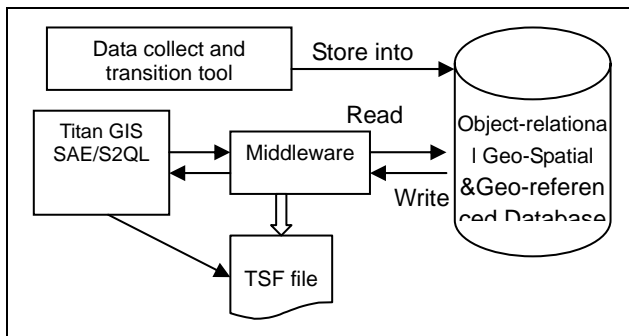


Fig. 9 The access strategy of object-relational spatial database compatible with the file system core of Titan-kit.

5. Conclusions and Further Researches

5.1. Conclusions

(1)The abstract of Geo-spatial data and the object-relational-model-based data structure is designed to represent the geo-spatial entities in the object-relational database.

(2)The four-tiers and three aspect of massive geo-spatial data organization strategy is designed

to meet the boundless, seamless and dynamic data organization demands of digital earth and digital mine.

5.2. Further Research

(1)Spatial data warehouse: Maybe we'd use the theory and technology of data warehouse to construct spatial data warehouse. Thus to move the object-relational model into the warehouse to organize and to manage the geo-spatial/geo-referenced data more efficiently in the future. In other words, we should extend the traditional data warehouse so as to support geo-spatial operation.

(2)Spatial-temporal data structure: In our current object-relational-model-based data structure, the temporal data are recorded as geo-referenced data of geo-spatial entities. This method is not efficient enough to the represented spatial-temporal event. So the spatial-temporal data structure is another direction.

References

- Al Gore, Jan. 31st, 1998The Digital Earth: Understanding our planet in the 21st Century, Given at the California Science Center, Los Angeles, California
- Chen G.R., Wu L.X. et.al., 1999, Study on design and development of a map symbol base in MGIS, Mine Surveying, 2: 8-10
- Cui Z.L., Wu L.X., 1999, Application of OODB technique in Mine GIS, China Mining Magazine, 4: 62-66
- Ren F.H., Fu K.E. et. al., 1998, Information Infrastructure management system and networking GIS, Proc. Spatial Information Technology Towards 2000 and Beyond--Geoinformatics '98 Conference: 164-172
- Sun Y.Y., Hou Y.G., 1998, Developing ActiveX-Based Web GIS Client, Proc. Spatial Information Technology Towards 2000 and Beyond--Geoinformatics '98 Conference, 173-177
- Wu L.X., Hu J.X. et.al, 1999, A ID-Coding Method Containing Quadrant Label for Non-Boundary 3D-GIS Proc. of the 2nd international Joint ISPRS Commission Workshop on Dynamic and Multi-Dimensional GIS, Beijing: 105-108
- Wu L.X., Liu C.B. et.al., 1998, A preliminary Investigation into problems relating to MGIS in China, Mine Surveying, 4: 48-51
- Wu L.X., Liu C.B. et al., 1999, The object technology and access interface of cartography symbol library, Mine Surveying, 1: 32-35
- Wu L.X., Liu C.B. et.al., 1999, The Research and Development of Networked Mining Core-GIS, Proc. of the 2nd international Joint ISPRS Commission Workshop on Dynamic and Multi-Dimensional GIS, Beijing: 359-364