

Modeling Cadastral Spatial Features Based on Geography Markup Language in GIS: A Case Study in Shanghai

X. H. Tong* and G. S. Xu

Departments of Surveying and Geo-informatics, Tongji University, 1239 Siping Road, Shanghai 200092, P.R. China

ABSTRACT. Significant progresses have been made in the establishment of cadastre databases in developing countries and the use of cadastre as a base to the development of land information systems based on Geographic Information System (GIS). How to interoperate and share these homogenous data is still a big problem in the applications of cadastral databases. This paper presents a GML (Geography Markup Language) approach to model cadastral spatial features for data interoperability and sharing. The advantages of GML in data interoperability are that (1) it provides an open and standard way to share and interoperate spatial data, and (2) it enables on-line and variety of spatial data interoperability on the Web. The cadastral spatial feature models are first defined based on geography markup language. An application schema is then defined, in which the cadastral features are described as CityModel, StreetModel, BlockModel, LandModel and BuildingModel, and the relationships between them are defined as CityMember, StreetMember, BlockMember and Landmember. The creation of a GML document is discussed in details. Furthermore, a GML-based cadastral WebGIS system is developed based on Java, Document Object Model (DOM) and Simple Application Interface for XML (SAX), and the implementation of the system is illustrated through a case study, which shows the feasibility of the proposed spatial feature models and data interoperability based on GML.

Keywords: Cadastral spatial feature, geographic information system, geography markup language, interoperability, modeling

1. Introduction

One of the significant problems in the developments and applications of Geographic Information System (GIS) is how to share and interoperate heterogeneous data by integration of geographical data from different sources and in different formats, scales and precisions. In which, the difficulties lie in (1) the differences in spatial data contents and sources. For example, vector data, image data and terrain data are all used to describe spatial position and status, while the data formats and capture methods for these data are different, thus the data transfer between them has been rather difficult, and the data quality would be decreased during the conversion; (2) the differences in spatial data models. For example, topological or non-topological model, and two, three or four dimensional spatial-temporal models are heterogeneous, the data with different models has hardly been shared; (3) the differences in software platforms and storage structures. Different GIS software system has its own software designs, data storage structures and processing methods, and each vendor is developed based on the detailed, independent and enclosed platform, therefore, the data semantic description in one vendor would have an unresolved contradiction with another, and the data communication between these different systems cannot be realized directly. Above difficulties seriously limit the rapid developments of the applications of geographic information

technologies such that the existed data resources cannot be re-used effectively, resulting in low efficiency on data sharing and high cost on data acquisitions.

With the development of the society, economy and information technology, plenty of progresses have been made in the establishment of cadastre databases in developing countries and the use of cadastre as a base to the development of land information systems based on Geographic Information System (GIS) (Dale and McLaughlin, 1988; Williamson, 1997; Silvaa and Stubkjærb, 2002; Tong et al., 2005). The principle goal of establishment of cadastral information system is for the effective and efficient land registration and management with the aid of new technologies in the acquisition, storage, update and analysis of cadastral information, which is the basis for planning and implementation of variety of real estate related operations in many areas.

Cadastral databases have now been developing in each city in China. These databases are distributed in each city, and developed based on different GIS software systems which have their own proprietary software designs, data models and database storage structures. Therefore, during the applications of cadastral databases, data sharing becomes a biggest obstacle. For example, a provincial government cannot further integrate and analyze these cadastral data stored in different places (each city) and in different formats because of the heterogeneity of these cadastral information systems in terms of data modeling concepts, data encoding techniques and storage structures, etc. (Devoegele et al., 1998; Zhang et al., 2003).

* Corresponding author: tongxhtj@yeah.net

How to interoperate and share these homogenous data is also an important issue in the applications of the cadastral databases.

Three stages have been experiencing in order to communicate and share cadastral database resources among these homogenous systems. The first stage is data conversion method, which transfers one kind of data format to another between two cadastral databases stored in different GIS systems. This method results in two major problems: one is the data precision and information loss during the conversion process, and the other is the high cost on the conversion since a conversion tool usually converts directly the data from one format to another in two specific systems, while another conversion tool is needed for data exchange between other two systems. In addition, the conversion method still cannot realize distributed data exchange and sharing (Zhang et al., 2003).

The second stage is the WebGIS-based distributed geographic data sharing on the World Wide Web. Many cadastral applications are developed by the internet-based GIS systems such as the ESRI's ArcIMS, the MapInfo's MapXtreme and the AutoDesk's MapGuide. On the Website, we can download and share cadastral data for viewing, querying, analysis and manipulation at any place and at any time. However, there exist still the problems of unique software designs, data models and storage structures for these WebGIS-based applications (Choicki, 1999). Thus, the data interoperability between these internet GIS applications is still rather difficult.

The open GIS specification provides new methods to solve the geo-spatial information interoperability, which is also a new stage for cadastral database sharing. The Extensible Markup Language and Web-based data interoperability is capable to realize data sharing in the multi-hardware and software platforms, in which, the most important issue is how to create an open geo-data model through describing spatial features based on XML language according to spatial data sharing models and definition specifications.

XML is a markup language defined by W3C (World Wide Web Consortium). Currently, XML has become the universal data format, playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere. XML allows developers to easily describe and deliver rich and structured data from any application in a standard, consistent way. With its well extensibility, many XML languages are derived in different fields. In geographic and cartographic fields, SVG (Scalable Vector Graphics), X3D (Extension 3D) and GML (Geography Markup Language) are the most important languages for describing two and three-dimensional graphics in XML. SVG is an open language for describing two-dimensional vector, raster and vector/raster hybrid graphics in XML, whereas X3D is for describing three-dimensional vector graphics.

OpenGIS Consortium (OGC) released the first spatially informational representation specification Geography Markup Language (GML) version 1.0 based on XML, which was soon accepted as a common spatial information format, in 2000.

GML2.0 (GML, 2002), in which the XML Schema was a replacement for XML DTD (XML Document Type Definition) used in the previous version, was set forth in 2002. With the development of XML, GML3.0 with an ISO/TC211 standard was released in 2003 to further strengthen GML2.0 version. This version is supported by many countries and GIS software manufacturers, and will be the international standard of spatial data transferring format. China is also going to be adopted and consulted to establish a national standard.

GML is an XML encoding for the transport and storage of geographic information, including both the spatial and the non-spatial properties of geographic features. This specification defines the XML Schema syntax and mechanisms, as well as conventions to (1) provide an open and vendor-neutral framework for definition of geospatial application schemas and objects, (2) allow profiles that can support proper subsets of GML framework descriptive capabilities, (3) support the description of geospatial application schemas for specialized domains and information communities, (4) enable the creation and maintenance of linked geographic application schemas and datasets, (5) support storage and transport of application schemas and related data sets, and (6) increase the ability of organizations to share geographic application schemas and the information being described (GML2.0, 2001).

The advantages of GML in data interoperability are that (1) it provides an open and standard way to share and interoperate spatial data, and (2) it enables on-line and variety of spatial data interoperability on the Web. Recently, many efforts have been given to investigate the research of GML and applications. GeoTools (Geotools, 2001) is a free java-based toolkit to display spatial data compatibly with GML1.0 specifications on the Web. In this toolkit, GML data and corresponding DTD metadata are interpreted by the Xerces-J Parser, the parsed data are further processed and graphic data are displayed with being viewed and zoomed solely. Houlding (2001) analyzed the opportunity and superiority of XML for data standards in the geosciences. Badard and Richard (2001) presented a new method using XML for the exchange of updating information between geographical information systems. Luo et al. (2002) studied the method on development of GML-Based WebGIS. Xu et al. (2002) discussed the methods using SVG and GML to store, exchange and display spatial geographic data. Xie et al. (2003) presented the method on designing and implementing an open location based service system, in which SVG-based graphic data are used in the web server and mobile client. In the national high technology research and development program of China (863 Program) - web spatial information standards and sharing service key technologies, an application schema describing the geometric object models for spatial features in GIS based on GML was designed, and the standard - China geo-spatial data transfer format was defined based on this schema.

In this paper, taking cadastral data as an example, the cadastral spatial feature models are described based on GML. An application schema is then presented, in which the cadastral features are defined as CityModel, StreetModel, Block-Model, LandModel and BuildingModel, while the relationship

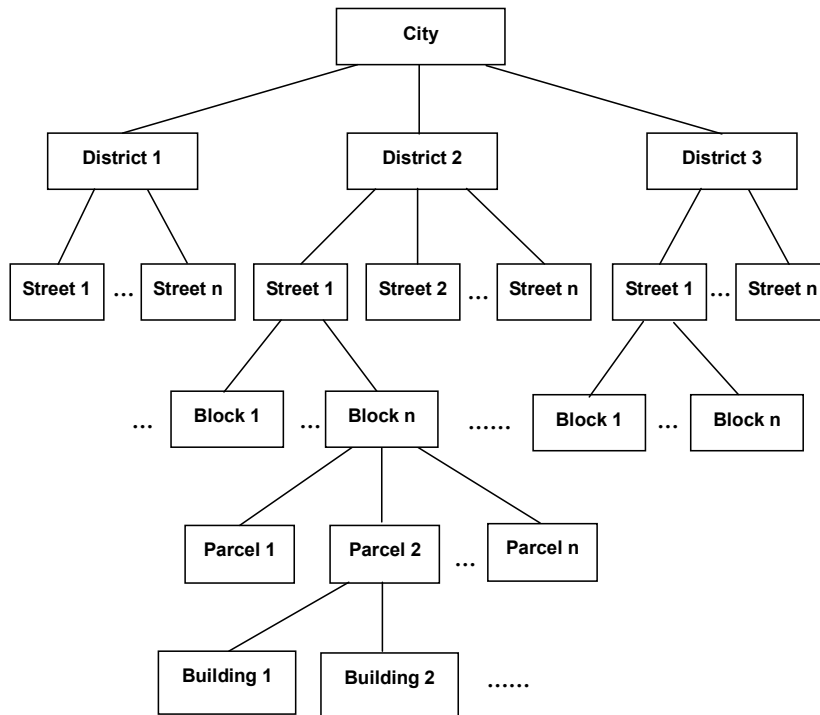


Figure 1. The hierarchy of cadastral features in a city.

between these features are defined as the CityMember, StreetMember, BlockMember and Landmember. Based on such a schema, the creation of the GML documents is therefore discussed in detail in order to store the cadastral features in Shanghai, China. A GML-based WebGIS application is, thus, developed based on Java, Document Object Model (DOM) and Simple Application Interface for developing XML (SAX). The feasibility of the proposed spatial features description based on GML is tested through a case study in Shanghai, leading to the conclusions.

2. The Application Schema Describing Cadastral Features Based on GML

For encoding spatial information, three basic schemas are defined in GML: (1) the Geometry schema, including the detailed geometric components as Point, LineString, LinearRing, Polygon, MultiPoint, MultiLineString, MultiPolygon and MultiGeometry, (2) the XLinks schema, which provides the XLink attributes, is used to implement the linking functionality, and (3) the Feature schema defining the general feature-property model. Based on the derivation from these schemas, an application schema is then defined to declare the actual feature types and property types for cadastral spatial features, including the base elements and types from the standard GML schemas, and the complex types derived from the base types in standard GML schemas.

The cadastral features are the geographic objects describ-

ing land property scopes, statuses and characters. According to the definition of feature classifications, for a city, the cadastral features include the City object, District object, Street object, Block object, Parcel object and Building object in GIS. Figure 1 shows the hierarchy of the cadastral features in a city. Based on this, a conceptual model for the cadastral features is designed based on Unified Modeling Language (UML) which is shown in Figure 2. In the Figure, the most basic and important unit in a cadastral database is a parcel which is a closed polygon constituting by a sequence of cadastral coordinate points, and the ownership, area and number representing by the parcel nominator and denominator are its main attributes. In a parcel, building features and terrain features are included. A parcel belongs to a block which is composed of many parcels without any overlaps among them. Among blocks, there are roads to separate them, thus a block is a relatively independent region, and the area of a block is always equal to the sum of the areas of all the parcels in this block. A block is one of the components of a street that is composed of many blocks and roads. All the streets constitute one district, and a city is composed of all the districts.

In order to define each cadastral feature and describe the relationships among the features in the GML-based application schema, two different kinds of object types - ModelType and MemberType are presented in this paper. The ModelType defines the spatial position and the attribute information of the cadastral features. In addition, the MemberType describes the memberships between the cadastral features. Thus, we define

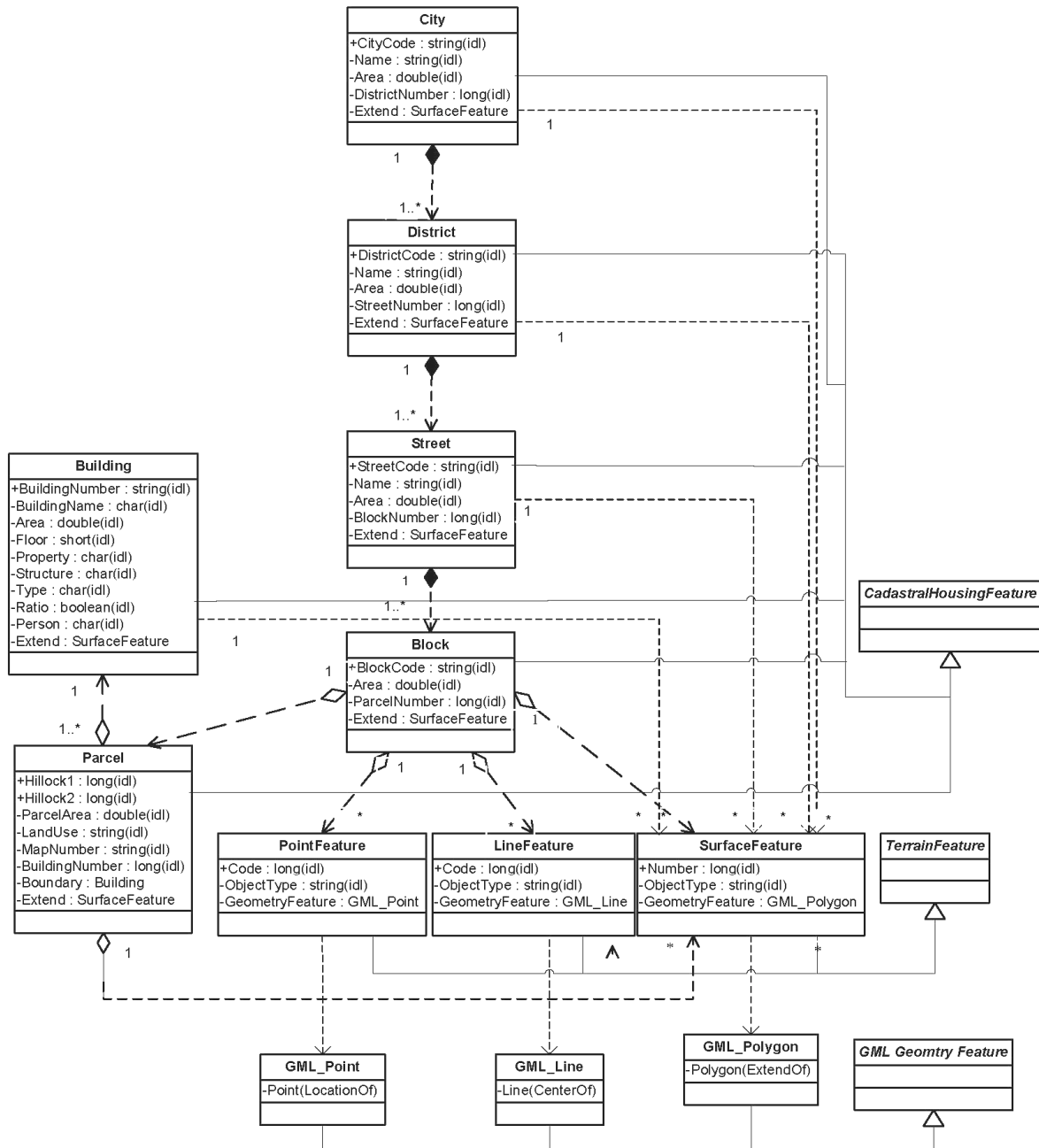


Figure 2. A conceptual model for cadastral features based on unified modeling language.

thirteen types for complex cadastral features, which are City-ModelType, CityMemberType, DistrictModelType, District-MemberType, StreetModelType, StreetMemberType, Block-ModelType, BlockMemberType, LandModelType, LandMemberType and BuildingModelType. And four kinds of terrain models are further defined for describing point, line, surface and annotation objects, which are TerrainPointModel, TerrainLineModel, TerrainSurfaceModel and TerrainTextModel.

The corresponding MemberTypes types of the cadastral features are derived from an abstract type named Feature-AssociationType in the standard GML schema. By referencing the member types, the cadastral features create the memberships with the subordinate features. The detailed data organization is listed as the follows: the CityModel type manages and describes all street features with the CityMember type, the StreetModel type manages and describes all block features

with the StreetMember type, the BlockModel type manages and describes all parcels as well as terrain features with the BlockMember type, the LandModel type manages and describes all building features with the LandMember type, and the BuildingModel type is the smallest feature unit without any member type.

3. The Application of GML Data in Cadastral Features

3.1. The Creation of GML Document for Cadastral Data

Cadastral databases and applications based on GIS have been developing in Shanghai, China. The characteristics of the cadastral data in Shanghai are that (1) the data comes from variable sources such as paper map digitization, total station field survey and aero photogrammetry, etc.; (2) the data are stored in different formats such as AutoDesk's DWG, MapInfo's Tab and ASCII text file; and (3) the data amount is huge. In Shanghai, there are total 21 governmental districts, for example, in which Xuhui district as a typical commercial and residential district and located in the southwest in Shanghai, have 14 cadastral streets, 500 blocks, 5000 parcels and 1225 map sheets with 1:500 scale whose dimensions are 50cm × 50cm (paper size). According to the defined feature models above, GML documents are created by transferring kinds of data into a universal data format. In each document, the hierarchy of the cadastral features is organized as City, District, Street, Block, Parcel and Building, and the spatial and attribute information of the cadastral features are stored. The steps of creation of GML documents are as follow: (1) Reading the spatial boundaries and attribute information of Shanghai City and creating a CityModel type in the GML document; (2) Under the framework of CityMember type, all the district models in Shanghai City are created one by one according to defined DistrictModel type, and the corresponding DistrictMember type is created; (3) For each district, under the framework of DistrictMember type, all the street models in this district are created; (4) For each street, under the framework of StreetMember type, all the blocks in each street in the district are created; (5) For each block, under the framework of BlockMember type, all the parcels and terrain features are created; (6) For each parcel, under the framework of LandMember type, all the building features are created. In fact, this step is a time consuming because of description of the most detailed features. The created GML documents can be stored either in ASCII files or in databases. Figure 3 shows the steps of creation GML documents from the City to the Building in Shanghai.

3.2. The GML-Based Cadastral Graphic Data Parsing

In the GML-based applications, the GML data needs first parsing for further use. XML Parser can check the validity of the GML data and interpret the spatial information contained in the GML document. There are two programming interface standards for XML Parser: SAX (Simple Application Interface for XML) and DOM (Document Object Model). DOM is

document-driven to convert an XML document into a tree structure in memory and to process the node in the tree. The work mode of DOM-based XML Parser tallies is composed with the logistic thinking habitat of people, so that the DOM-based XML Parser can be used easily. Because the DOM-based XML Parser will build a tree in memory, it exhausts too many computing resources and is suitable in client. SAX is time-driven to trigger events for the beginning and end of a document, elements and so on, so XML Parser has to process too many events and will be burdened with them. At the same time, SAX-based XML Parser takes XML document as not a tree but a data stream, so it can only process XML document sequentially and will only consume little computing resources. SAX-based XML Parser is suitable in server (Luo et al., 2002).

According to the characteristics of the cadastral features, the important features as city, districts, streets, blocks, parcels and buildings are interpreted using DOM, which are stored and managed in the memory to improve the query and operation speed. The terrain features which include the road, river, pipeline and wall features etc., due to the huge amount and the setup as the background, are suitable for using SAX. These data are loaded and interpreted in the real time, while not needing to be stored in the memory in advance; it will economize the memory resource.

3.3. Java and GML Based Cadastral WebGIS and Its Application

According to the GML models discussed in above section, a java-based WebGIS prototype system is developed to explore, query and store cadastral features. In the system, the data amount is rather huge, for example, it has achieved up to 2 GB in Xuhui district. The system uses DOM to interpret the cadastral features while SAX to the terrain features in GML documents, and uses Java Database Connectivity (JDBC) to connect Oracle databases to integrate the spatial and attribute data management. Figure 4 shows a user-friendly interface in cadastral WebGIS for the GML-based interoperable cadastral database. The whole Xuhui District is displayed on the left of the frame. The right of the frame will display the attribute table for is tested through the system, leading to the conclusions clicked feature. The basic functions of the GML-based WebGIS system include: (1) Map operation. User can zoom in, zoom out, pan and redraw the map; (2) Map query. User can query the features by the ways of point click, rectangle selection, circle selection and arbitrary polygon selection, and the interface will highlight the selected feature and display the attribute for the feature; (3) Attribute query. User can query the feature by searching the attribute such as the number, location, and the interface will locate the found feature; (4) Layer control. User can turn on/off a feature layer, and set the visible scope of the features on a layer; and (5) Statistic analysis. For example, user can do land area classification statistics according to the land usages etc. Figure 5 shows the results of parcel query. The attributes of the found parcels are listed in the lower right of the interface, which include parcel number,

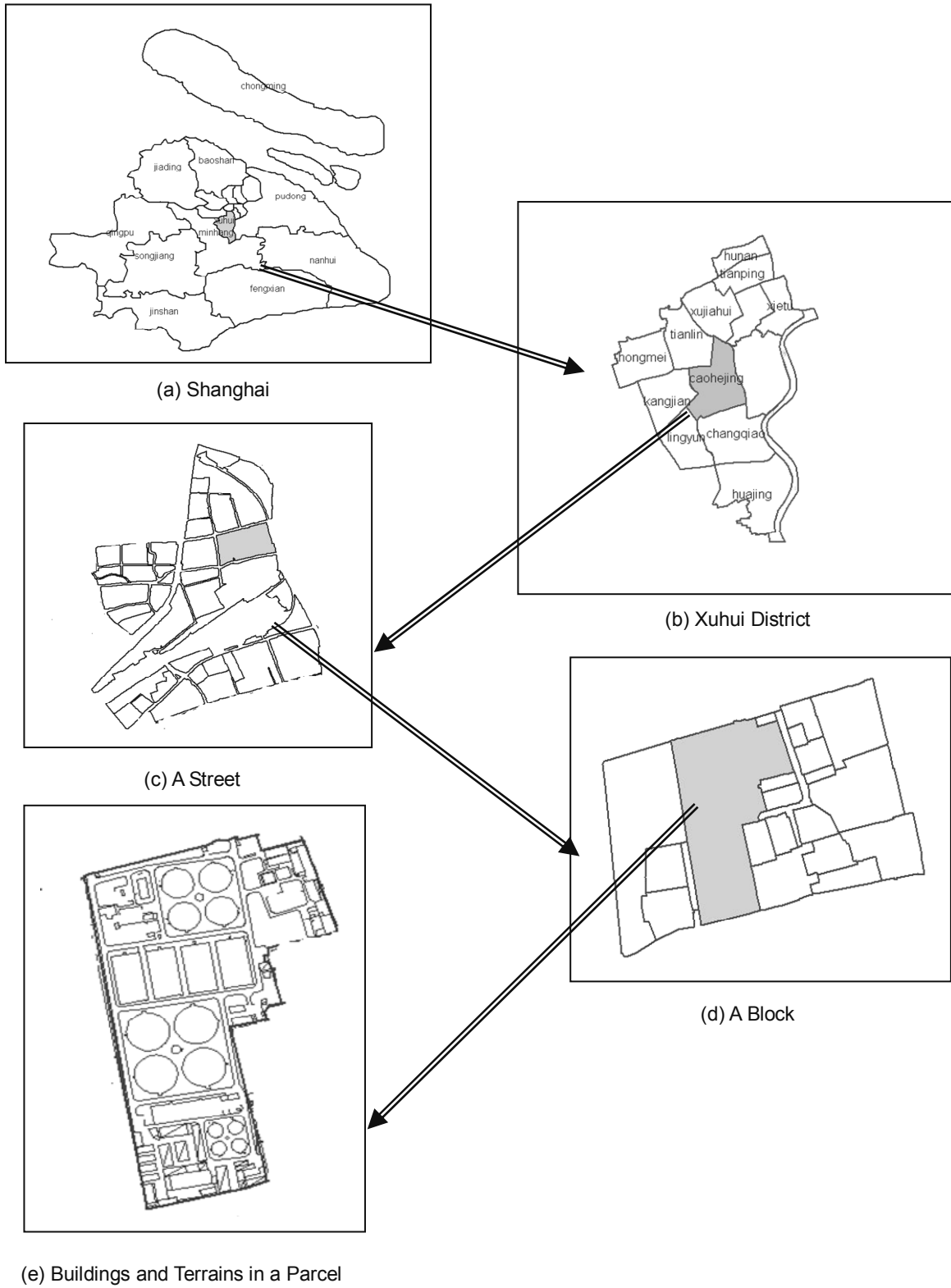


Figure 3. The Created GML documents from the city to the building in Shanghai.

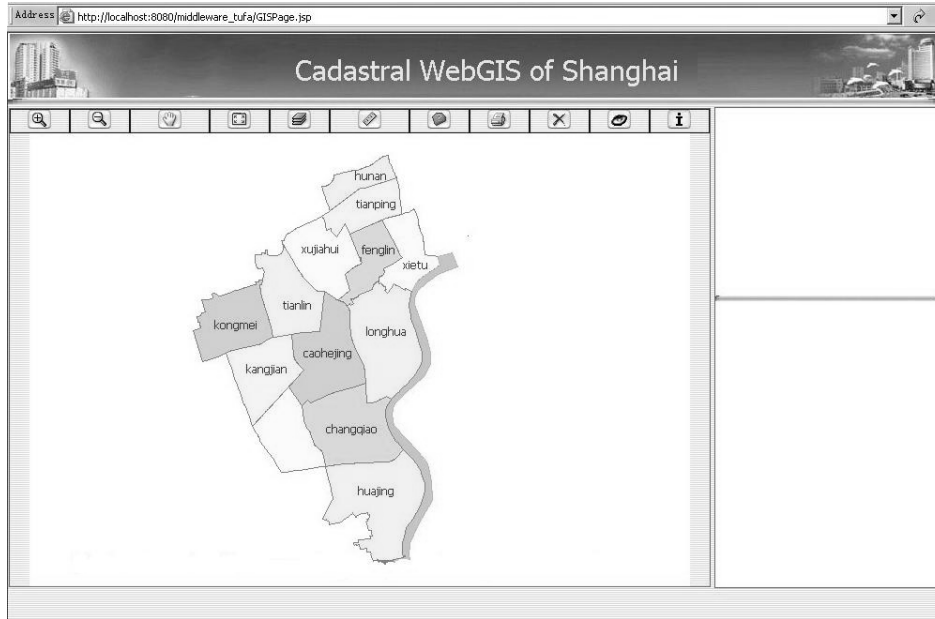


Figure 4. Interfaces of the GML-based Cadastral WebGIS.



Figure 5. Results of Parcel Query.

area, address, usage and status. The selected parcel is displayed highlight in the right center of the frame. These parcel and building data are all stored in the GML document and are then parsed and displayed through the developed system with a rapid processing speed.

4. Conclusions

Many cadastre databases based on Geographic Informa-

tion System (GIS) have been established in developing countries such as Shanghai, China, while data sharing is still a biggest obstacle in the application of cadastral databases because of the heterogeneity of these cadastral systems in terms of data modeling concepts, data encoding techniques and storage structures and so on. This paper presented a GML (Geography Markup Language) approach to model cadastral spatial features for data interoperability and sharing.

The advantages of GML in data interoperability are that

(1) it provides an open and vendor-neutral framework for the definition of geo-spatial application schemas and objects, and data sets; (2) it increases the ability of organizations to share geographic application schemas and the information being described; and (3) it enables on-line and variety of spatial data interoperability on the Web.

In this paper, an application schema based on GML is defined, in which the hierarchical cadastral feature models and the relationships between them are presented. Taking the cadastral data in Shanghai as a case study, the steps of creation of GML documents are introduced in details. A GML-based cadastral WebGIS system is further developed based on proposed models, Java, DOM and SAX. The basic functions of the system, such as map operation, map query, attribute query and statistic analysis, can be performed using the cadastral database on the Web. The case study shows the proposed spatial feature models and data interoperability based on GML is feasible and since the models are defined based on the standard GML, the models can be applicable to other kinds of cadastral data different from Shanghai area.

The limitations of the GML in data operability are (1) the GML document is a text-based file, so the file size is rather big, how to compress and decompress the GML data to promote the transmission speed on the Web needs further studying; and (2) geo-spatial information interoperability is categorized as data interoperability, software interoperability and semantic interoperability. But, presently, the GML still cannot fully solve the problem of the semantic interoperability. With the developments and applications of GML, further research needs investigating this problem for wide using GML for data interoperability and sharing.

Acknowledgments. The work described in this paper was substantially supported by the National Natural Science Foundation of China (Project No. 40301043 and 40171078) and Foundation of Shanghai Rising-Star Program (Project No. 05QMX1456). The authors would

thank the anonymous reviewers for their valuable comments on the paper.

References

- Badard, T. and Richard, D. (2001). Using XML for the exchange of updating information between geographical information systems. *Comput., Environ. Urban Syst.*, 25, 17-31.
- Dale, P. and McLaughlin, J. (1989). *Land Information Management: An Introduction with Special Reference to Cadastral Problems in Third World Countries*, Oxford University Press, Oxford.
- Devogele, T., Parent, C. and Spaccapietra, S. (1998). On spatial database integration. *Int. J. Geogr. Inf. Sci.*, 12(4), 335-352.
- Choicki, J. (1999). Constraint-based interoperability of spatialtemporal databases. *Geoinf.*, 30(3), 211-243.
- OGC (OpenGIS Consortium) (2002). Geography Markup Language (GML) 2.0. <http://www.opengis.net/gml/01-029/GML2.html> (accessed August 1, 2002).
- The University of Leeds (2001). GeoTools, The Java GIS toolkit. <http://www.ceg.leeds.ac.uk/geotools/demos/gml> (accessed December 1, 2001).
- Houlding, S.W. (2001). XML-An opportunity for data standards in the geosciences. *Comput. Geosci.*, 27, 39-849.
- Silvaa, M.A. and Stubkjær, E. (2002). A review of methodologies used in research on cadastral development. *Comput., Environ. Urban Syst.*, 26, 403-423.
- Tong, X.H., Shi, W.Z. and Liu, D.J. (2005). A least squares-based method for adjusting the boundaries of area objects. *Photogrammetric Eng. Remote Sens. (PE&RS)*, 71(2), 189-195.
- Williamson, I. (1997). The justification of cadastral systems for developing countries. *Geomatica*, 51(1), 21-36.
- Luo, Y.W., Wang, X.L. and Ma, J. (2002). A study on GML-based WebGIS application. *Comput. Eng.*, 28, 15-16.
- Xie, Z.Y., Li, Q.Q. and Zuo, X.Q. (2003). Design and Implementation of an Open Location Service System Based on SVG. *Geomatics Inf. Sci. Wuhan Univ.*, 28(1), 74-79.
- Xu, C.J., Zhou, L.J. and Sheng, X.H. (2002). Application of SVG and GML in WebGIS. *Remote Sens. Inf.*, 3, 38-40.
- Zhang, C., Li, W., Day, M.J. and Peng, Z.R. (2003). GML-based Internet geographical databases. *Cartography*, 32(2), 1-15.