Research on The Method of Encoding Geography Information Based on XML

JianChao Wang
China Aero Geophysical Survey And Remote Sensing Center For Land And Resources
29 College Road, Beijing, 100083, P. R. China
jianchaowang@sohu.com

XuWen Qin
China Aero Geophysical Survey And Remote Sensing Center For Land And Resources
29 College Road, Beijing, 100083, P. R. China
qin-xuwen@163.com

Abstract: This paper analyzes the advantage of the XML, and studies the simple feature object model and Geography Markup Language that proposed by the Open GIS Consortium (OGC). We discussed the means of encoding the geographical data based on XML.
Keywords: XML, GML, WebGIS.

1. Introduction

The WebGIS is the main way to publish and share the geographical data. Now there are many WebGIS productions, which have already been applied in some projects extensively. The main problem existed in the WebGIS is that there isn't enough interoperability of data and functions among different systems. The XML is a very valid means to resolve the main problem existed in WebGIS because it has obvious advantage in encoding and exchanging of the complex data. The Geography Markup Language proposed by the OGC is based on XML. Many relational databases support to store and query the XML data. It's easy to integrate those distributed heterogeneous geospatial data if the geographical data is encoded with XML.

2. Characteristics of XML

XML is an extensible markup language for documents containing structured information. The XML is a subset of the Standard Generalized Markup Language (SGML). Its goal is to enable generic SGML to be served, received, and processed on the Web in the way that is now possible with HTML. XML has been designed for ease of implementation and for interoperability with both SGML and HTML. The XML has the following characteristics:

1. Extensible

The XML only provided the basic phrasing, but not define the accurate tags. Tag can be definition according to the application purpose by anyone.

2. Separation of the manifestation with phrasing

The XML and HTML differentiation is obvious at this point. The HTML defines many markups used for the data manifestation. Almost each HTML page contains a lot of those markups. The XML manifestation is defined in an absolute style-sheet document. A XML document can be showed in different style according to the different style-sheet document.

3. Internationalization

The XML specification has already become an international standard. A XML document can use any letter and word defined in the ISO/ IEC 10646 and Unicode3.0 standard.

4. Concurrent with internet

The XML can be straightforwardly usable over the Internet. The World Wide Web is a successful global infrastructure, and XML emerges as its data interchange model. The XML provide a means of encoding spatial information for both data transport and data storage, especially in a wide-area Internet context.

3. Geospatial data model design

The purpose of geographic data model design is abstract the complicated geography things to simple models which can be identified by the computer. The OGC uses the Abstract Feature Model to descript the geographic things. Feature is abstraction of the thing of realistic world. If the thing has relation with geography position, that feature is called the geographic feature. A geographic feature is essentially a named list of properties. Some or all of these properties may be geospatial, describing the position and shape of the feature. The Simple Features model represents a simplification of the more general model described in the OpenGIS Abstract Specification. There are two major simplifications:

1.Features are assumed to have either simple properties (Booleans, integers, reals, strings) or geometric properties.

2.Geometries are assumed to be defined in two-dimensional SRS and use linear interpolation

between coordinates.

There are some consequences follow from these simplifications. For example, simple features only provide support for vector data; and simple features are not sufficiently expressive to explicitly model topology. This version of GML addresses the first of these limitations in that it allows features to have complex or aggregate non-geometric properties. Such complex properties may themselves be composed of other complex and simple properties. Common examples include dates, times, and addresses. It is expected that future versions of GML will address the second of these limitations and provide more elaborate geometry models.

Each feature has a type, which is equivalent to a class in object modeling terminology; such that the class-definition prescribes the named properties that a particular feature of that type is required to have. So a Building might be defined to have a name, a surface-construction, an owner, and a polygon (see Fig.1).

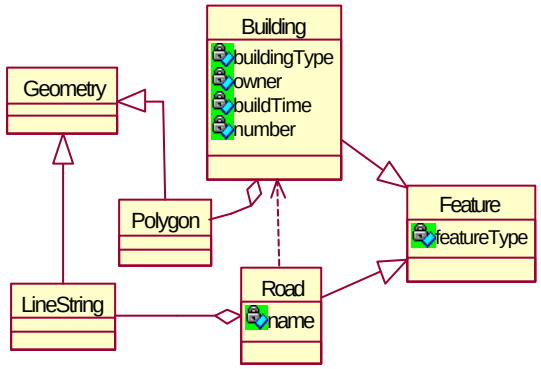

Fig. 1. An Example of simple feature model.

4.XML encoding schemas

XML schema defines the tag name, type, relationship of XML document. OGC has defined feature schema, geometry schema and Xlink schema in the Geography Markup Language specification (see fig.2). With these three simple schemas it is possible to encode a wide variety of geospatial information.

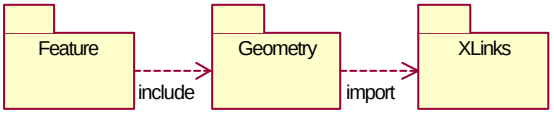

Fig. 2. Three schemas of GML.

The feature schema includes the geometry schema because that feature contains geometry properties according to the simple feature theory. The feature schema defines the general geographic feature model, and the geometry schema defines the geometry elements.

1.Geometry schema

The GML Geometry schema includes type definitions for abstract geometry elements, concrete (multi) point, line and polygon geometry elements, as well as complex type definitions for the underlying geometry types.

2.Feature schema

The Feature schema uses the <include> element to bring in the geometry schema and make them available for use in defining feature types: <include schemaLocation="geometry.xsd"/>. A geometric property is modeled as an association class that links a feature with geometry.

3.Application schema

The base characteristic of XML is extensible. Those three schemas provides only base type and structure definitions. In the idiographic GIS application actual geographic feature and new properties can be defined according to the base schemas. The schema defined by the user is called the application schema. The base schemas can be viewed as the components of an application framework for developing schemas or sets of schemas that pertain to a particular domain, jurisdiction, or information community. Furthermore, such application schemas may be developed in a more horizontal fashion to support many information communities.

A set of logically related GML schemas, which we term the GML Framework, is depicted in Fig.3.

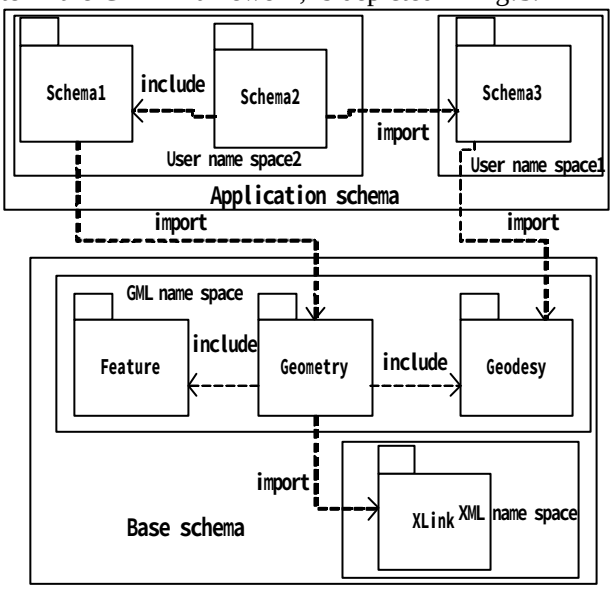

Fig. 3. Application framework of base schemas.

A conforming GML application schema must heed the following general requirements:

1. An application schema must not change the name, definition, or data type of mandatory GML elements.

2. Abstract type definitions may be freely extended or restricted.

3. The application schema must be made available to anyone receiving data structured according to that schema.

4.The relevant schemas must specify a target namespace that must not be "http://www.opengis.net/gml".

5.Display the xml document

How to display the XML document of geospatial data in the general browser such as Internet Explore or Netscape? Now there are three kind of methords to resolve the problem:

1.The server converts the geospatial data encoding in XML to the raster image,and respones to the browser through the HTML page. This methord is adopted widely at present stage. But this method obviously did not develop the function of the XML which is the next generation network language. It increases the loading of server,and some easy applications based on vector format become more complex.

2.Providing some applet or ActiveX programs runing in the client or browser which explain and display the XML document.This methord is often adopted too.But it realizes very inconvenient because of the browser's complicated variety.

3. The general browser has the component or plug-in which can explain the particular XML coding. This methord is the most ideal.Only a general browser is needed to explore the XML document like the HTML document. The Adobe company provided the plug-in to surport the SVG for the general browser. some scholar bring up the way to covert the XML to SVG which can be displayed in the browser with the plug-in.

6.conclusion

This paper gives a way to encode the spatial properties and non-spatial properties of geographic feature. It's easy to integrate those distributed heterogeneous geospatial data if the geographical data is encoded with XML. There are some problems need further research such as how to express the topology of geographic feature? The author believes that the application of the XML in the WebGIS will be more deep and mature.

References

[1] http://www.opengis.net/gml/01-029/GML2.html.

[2] ESRI ArcGIS, ArcXML Programmer's Reference Guide 4.0, 2000.

[3] Http://www.w3.org

[4] XML Schema Part0:Primer,W3C Recommendation,2 May 2001

[5] Elliotte Rusty Harold, XML Bible, IDG Books Worldwide, Inc.

[6] Michela Bertolotto, A Conceptual Model for Web-based Heterogeneous GIS.