Research on Geospatial Information Web Sharing Language — GeoML

Aijun Chen[1] Qi Li[2]

[1] *CyberGIS Studio, the Institute of Remote Sensing & GIS, Peking University, Beijing, China, 100871*
Tel: 86-10-62751962(O), Fax: 86-10-62751964
E_Mail: Chaj@263.net

[2] *CyberGIS Studio, the Institute of Remote Sensing & GIS, Peking University, Beijing, China , 100871*
Tel: 86-10-62751964(O), Fax: 86-10-62751964
E_Mail: Liqi@pku.eud.cn

ABSTRACT Combining the latest Internet technology -- XML with the analysis to the characteristic of geospatial object, this paper brought out geospatial information web sharing language – GeoML, which can be easily used to publish and browse geospatial information on WWW to achieved the network sharing of geospatial information. Using the GeoML, repeatedly gathering and processing geospatial data can be avoided, and the updating and maintenance of geospatial data became more easily because of the convenience of editing and gathering geospatial information.
Key words XML, GIS, Geospatial Object, Geospatial Information,GeoML

1. Introduction

The development of computer technology and network technology have led to great changes in the field of GIS, which has changed from the initial step (PC GIS) to a mature one – Enterprise GIS, further to current Socialized GIS. The socialized GIS, that is to say, the networked GIS which can achieve the geospatial information web sharing, urgently required the sharing of geospatial information. And the networked GIS has unavoidably become the basement of building Information Highway, founding National Spatial Data Infrastructure (NSDI), and constructing “Digital Earth”(Al Gore 1998). The important reasons is that more than 80 percent of information which we are processing is geospatial information which have relation with the geospatial position, and the essential meaning of “Digital Earth” is that the Earth is digitized or is described by information. However, most of now-available GIS system, which are mostly used to process the geospatial information, are Desktop GIS rather than Networked GIS, such GIS products as MapInfo Co.’s MapInfo, WTUSM’s web publishing. In addition, network covers anywhere on the global and “Digital Earth” is being widely paid attention to now, therefor it is extremely valuable to make research on a new kind of formatted language which is compatible with the internet popular language like HTML and Java. And also this language should have the more powerfully capability of expressing geospatial information except that it can express general hypertext information, and should become the world language of “Digital Earth”, so that geospatial information web publishing and browsing is as easy as common hypertext pages. Using this language, some more GeoStar, and PKU’s CityStar etc.. All of these system have not solved the problem of geospatial information web sharing. Some GIS products, such as ESRI’s Internet Map Server (IMS) for ArcView & MapObject, Intergraph Co.’s GeoMedia Web Map, Autodesk Co.’s MapGuide, Bentley Co.’s ModelServer/ Discovery, have given some solution for geospatial information web sharing to some degree. These system are based on Client/Server model, user who want to browse geospatial information stored in a specified server by Internet must have a strengthened client such as installed plug_in or strengthened the capability of showing graph etc (Autodesk 1999, Bentley 1999). So it is impossible that all users can freely browse the geospatial information on Internet. And also the most important thing is that many professional departments, which want to publish their geospatial information on Internet, can not directly publish their information on Internet, because they need the publishing server which specified on supporting the geospatial information--

professional geospatial information can be processed and published on Internet by professionals, different professional department may interchange, supplement and repeatedly use the existing geospatial data each other. Consequently geospatial information web sharing can be actually achieved.

The appearance of eXtensible Markup Language (XML) brought the light to the information publishing of professional field on Internet. Using the XML to different fields, studying and designing a new markup language for describing professional information web publishing and implementing the

interchange and sharing of the whole specified field is the urgent requirement of various professional fields. Therefore, in order to solve the web publishing and sharing of the geospatial information, we brought the XML into the geospatial information fields to design a new markup language for describing the geospatial information. This markup language was named as Geospatial Markup Language (GeoML).

2. Introduction for XML

The appearance of XML makes it possible to implement professional various information web publishing. According to the XML standards, specified professional departments can customize their own markup languages which will be convenient for describing their own professional information. For example, Microsoft's Channel Definition Format (CDF) is used to publish specified information in Pushing technology, Netscape's Meta Content Framework which played an important role on communication of two-way pages content, Chemical Markup Language (CML) is defined by chemist and used to describe chemical formula, MathML is defined by World Wide Web Organization and used to express the mathematical formula, etc..

2.1. What is XML

XML is the "eXtensible Markup Language" (eXtensible because it is not a fixed format like HTML). It is designed to enable the use of SGML on the World Wide Web. XML is an idea data format for storing structured and semi-structured text intended for dissemination and ultimate publication on a variety of media (Robin Cover 1999, W3C 1998).

Actually, XML itself is not a single markup language, it's a meta-language which let you design your own markup language. A regular markup language defines a way to describe information in a certain class of documents (eg. HTML). XML lets you customize your own markup language for many classes of document. It can do this because it's written in SGML, the international standard meta-language for markup languages. It is based on this idea that the research and design of GeoML has been proposed (Neil, Bradley, 1998).

2.2. Difference between XML and HTML

XML differs from HTML in six major respects (W3C, 1998):

a) Information providers can define new tag and attribute names at will.
b) Document structures can be nested to any level of complexity.
c) Any XML document can contain an optional description of its grammar for use by applications that need to perform structural validation.
d) document.
e) The control file which contained semantic data and structured data can be produced from structured XML document, and at the same time can be downloaded by browser to be used by XML parser.
f) The more plenty hyperlink technologies have been added, the two-direct link, multi-window link, multi-resources link and various attribute links can be implemented in XML.

2.3. The Component and Structure of XML

Each XML document has both a logical and a physical structure. Physically, the document is composed of units called entities. Entity contained text and binary data, text consist of characters, some of characters buildup content of document, some buildup XML markup. An entity may refer to other entities to cause their inclusion in the docu-ment. A document begins in a "root" or document entity. Logically, the document is composed of declarations, elements, comments, character refe-rences, and processing instructions, all of which are indicated in the document by explicit markup. The logical and physical structures must nest properly.

There are six markup types that can be presented in an XML document, which are similar with the HTML markup, they are Elements, Entity References, Comments, Processing Instructions, Marked Sections, Document Type Declarations (DTD), and DTD consist of Element Declarations, Attribute Declarations, Entity Declarations, and Notation Declarations. Especially, DTD can be edited and stored in a single file, which has the file extension name 'dtd' (*.dtd), and referenced in many XML documents repeatedly (Neil, Bradley, 1998).

The other technology which are closely related with XML include XML Link Language (XLL), XML Stylesheet Language (XSL) and Document Object Model (DOM), Detailed information is omitted here.

3. Characters and Rules of Geospatial Object

Generally, geospatial object is defined as geom-etrical object which positions in geographical space, accompanied with some attributes, described by geographical data. Geographical data consist of geospatial data and attribute data. Geospatial data describes spatial position, geometrical state of geospatial object and relation with other geospatial object, for example, the

position and shape of a house can be described by coordinate data. Attribute data describe the natural and social character of geospatial object, for example the owner and the building material of a house. The attribute data often is stored in relational database, using the unique ID of geospatial object stored in relational database, the relation with the corresponding geospatial data can be established. So when user access the attribute data, user also can access the geospatial data (Open GIS Consortium,1998).

Therefore, it is very important to know about the nature of geospatial object. Only after analyzing and understanding the geospatial object, can we propose the markup language which is suited to describe the geospatial object. The following is a simple analysis to geospatial object.

3.1. Spatial Position of Geospatial Object

In the world, we can use various spatial references to position a geospatial object, for example , we can use coordinate data (X，Y，Z) to accurately position such a point as the location of a building, also, we can use a series of coordinate data (X1, Y1, Z1; X2, Y2, Z2;…) to locate such complex geospatial object as road, which is a line-type element and land using state, which is an area element (Zhu Yuan etc., 1997).

3.2. Types of Geospatial Object

According to geographic concept and attribute concept, geospatial objects have the following types: pure geometrical type, geometrical topological type, pure topological type, spatial objective type, and non-objective type In terms of the dimension used toexpress the geospatial object, detailed geospatial types are as following (Open GIS Consortium, 1998) (Liu Dong, 1996):

3.2.1. Zero-dimension geospatial object:

a) Independent point objective, it is pure geometrical type, but it is a spatial objective with corresponding attribute encoding and attribute table.
b) Pure node, it is geometrical topological element, it is not objective type, only be used to express the relationship with arc and aggregation location.
c) Node objective, it is both geometrical topological type and spatial objective, for example, the node between wires is a switcher station.
d) Annotate reference point, it is used to reference the annotate location, and can be stored in annotate data structure.
e) Polygon ID, it is the auxiliary information of polygon, can be stored in polygon's data structure.

3.2.2. One-dimension geospatial object:

a)Topological arc, it is geometrical topological type, it does not has branch, but has start node and end node, maybe is part of line-type objective or border of area-type objective, even it is both the border of area-type objective and part or whole of line-type
b) Non-topological arc, it is pure geograp-hical objective, also can be named as spag-hetti, for example, contour need not think about its start node and end node, anddoesn't have left-right polygon.
c)Line-type objective, it consists of one or more arc, the branch and crossover are allowed for extending it to deal with problems which related with river valley and traffic.

3.2.3. Two-dimension geospatial object:

a） Simple polygon, including outer borderline, but without island.
b） Donut polygon, consisting of outer borderline and one or more island.
c）Compound polygon, which is composed of more than one polygon and donut polygon.
d） Generalized polygon, which only has borderline of island, but has not outer borderline.
e） Pixel element, two dimension pixel, it is the smallest, impartible image unit. objective. It may be polyline, circle, arc, smoothing curve.
f） Grid unit, it is two dimension unit.

3.2.4. Aggregation object:

a）Image, which forms a pattern, it is a two-dimension matrix which is regularly arrayed in spatial position.
b）Grid, it consists of regular or near regular area, often it is rectangle or square, for example, GIS or DEM base on grid.
c）Layer, it consists of one or more type objective, it can be vector data aggregation, image or grid.

3.3. Spatial Relation between Geospatial Object

Spatial relation is the relation among geospatial objects, it contains the following types (ESRI, 1998, Liu Dong, 1996):

a) Topological Spatial Relations: it will not be changed when making topological transition such as rotation, pan, zoom-out or zoom-in, for example, the intersection and disjoint between geospatial object.

b) Metric Spatial Relations: it is used to express metric quantity in metric space, for example, distance relationship between geospatial object.
c) Order Spatial Relations: it expresses whole example, before-back, upper-down, left-right The spatial relation is closely concerned with the dimensions, state and size of geospatial object , also relates with the dimensions of space itself, and all of topological relations concerns with and affects each other.

4. The Design of Geoml

Based on the above simple introduction, we did some research and designed the GeoML, i.e., according to the standard syntax of XML, we created elements such as ENTITY, ELEMENT etc. to express geospatial data. Because XML is simplified from SGML and extended from HTML, we had made a various reference to the standard and technology of SGML and HTML during the course of designing.

4.1. Our Definition for GeoML

GeoML is a group of descriptive aggregation of markup language which coincides with standard of XML and can be used to describe geospatial object. According to GeoML, geospatial objects consist of paths which are described by connected line or curve, GeoML markups contain the semantic information and expression information of these paths, and such objects as gsSHAPES, gsGROUPS described in GeoML can be edited further like such objects as FORM, TABLE in HTML. GeoML document can be edited on any text editor. Preliminary defined GeoML mainly describes geospatial position and geospatial object. Simple geospatial information can be described and published on Internet. So geospatial information web sharing had been achieved. The further research about this topic are under the way, and will be introduced lately.

4.2. The Design Goals of GeoML

There are some requirements must be considered during the course of design, it is the following:

a) Compatible with XML and HTML, it makes GeoML, XML and HTML existing in the same pages.
b) Using the existing mechanisms of HTML and CSS, it facilitates implementation of GeoML and ensures that implementations can reuse existing code and techniques.
c) Supporting interchange of data between applications, ii is the basis of acquiring geospatial information interchange and web sharing.
d) Retaining the information required for further editing of GeoML, this requirement has the important consequence that GeoML must be extensible.
e) Providing efficient representations of vector graphics, textual representations tend to be verbose.
f) To support scripting, including the require-ments of animation.

4.3. The Design for GeoML

4.3.1. Simple introduction of GeoML

The whole structure of GeoML can be explained by two main GeoML elements which defined according to the standard syntax of XML.

a) gsSHAPE

gsSHAPE elements describe the visible vector graphic element by defining a path, a path is an outline which consists of a series of lines and curves. Using the attributes and gsSTROKE sub-element of gsSHAPE, the outline has various attributes, for example, color, style etc., at the same time, using the gsFILL sub-element, the interior of outline can be filled with various color, style etc..

The following demonstration is a simple gsSHAPE (as Fig. 1) and corresponding GeoML:

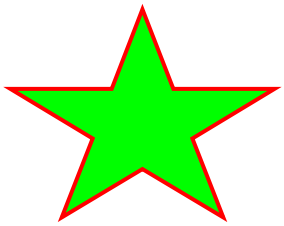

Fig. 1 Demonstration of a simple of

<g: gsshape gsstyle = 'top: 0; left: 0; width: 250; height : 250' gsstroke = "true" gsstrokecolor = "red" gsstrokeweight = "2" gsfill = "true" gsfillcolor = "green" gscoordorigin="0 0" gscoordsize="175 175">
<g: gspath g = "m 8,65 l,72,65,92,11,112,65,174,65,122,100,142,155,92,121,42,155,60,100 x e"/>
</g: gsshape>

b) gsGROUP

gsGROUP consists of one more than gsSHAPEs, used as a whole to describe complicated geospatial object.

In addition, several predefined shapes may be used as convenient alternatives to explicitly declare a shape element with a path. These predefined shapes are line, polyline, curve, rect, roundrect, oval, arc, and image.

4.3.2. Attribute Types of GeoML Element

The attributes of GeoML element use the following types:

a) Boolean: An attribute which can take values true and false.
b) String: Character data of any length.
c) Number: Numeric data, used for values that are integer or fractional numbers and for values which specify lengths.
d) Vector2D: Numeric data in the form X,Y. Usually used to list a coordinate in 2D space. May be in form "x y" or "x, y"
e) Vector3D: Numeric data in the form X, Y, Z. Usually used to list a coordinate in 3D space. May be in the form "x y z" or "x, y, z".

4.3.3. DTD Entity Definitions

Some frequently used attributes in element and sub-element have been declared as Entity in DTD, so we can directly make a reference to them in such GeoML shape elements as shape, group and the predefined shapes. The detailed information is omitted.

4.3.4. Preliminary GeoML Element

There are two preliminary GeoML elements, as following:

a) gsSHAPE element

This is a top-level element used to describe a shape. This element may appear by itself or within a <gsgroup> element. If a <gsshapetype> is referenced using the 'type = attribute', any attributes specified in the shape will override those found in the gsshapetype, and some of attributes reference to DTD entity.

b) gsGROUP element

It is a top-level element, and is used to group shapes (including other groups) so that they can be positioned and transformed as a single unit to express complex geospatial object.

The attributes of gsGROUP element have been described in gsSHAPE element completely.

4.3.5. Prototype Shape Elements and Their Advanced Attributes

The attribute element of shape element play an important role on display of vector graphic, image and text signal related with graphic, they include gsPATH, gsFORMULAS, gsSTROKE gsSHADOW, gsFILL, gsTEXTBOX, gsTEXTPATH, gsIMAG-EDATA etc., and prototype graphic elements include gsLINE, gsPOLYLINE, gsCURVE, gsRECT, gsR-OUNDRECT, gsOVAL, gsARC, gsIMAGE etc..

During the course of preliminary designing, only the geospatial position and geospatial object have been considered. The research of geospatial relation and other complicated geospatial problem are in the progress.

5. Implementation Of Geoml

The implementation of GeoML consists of two parts, one is editing the GeoML document, and another is interpreting the GeoML data document, i.e., GeoML editor and GeoML interpreter. Now the impleme-ntation is based on Java and Microsoft IE 5.0

5.1. Design and Implementation of GeoML Editor

At present, because of simple GeoML markup element, we can replace the GeoML editor with the common text editor which is used to write the GeoML document (*.gml)and DTD document (*.dtd). We can read the data from GeoML document according to the standard text format, and interpret various markups and corresponding attributes.

5.2. Design and Implementation of GeoML Interpreter

GeoML interpreter was implemented by programming with Java, and made into AcitevX control, it can automatically be downloaded and installed when IE browses the GeoML document for the first time. Lately IE may browse GeoML document.

The following figure shows the function module of GeoML Processor and the logic relation between GeoML document as Fig. 2.

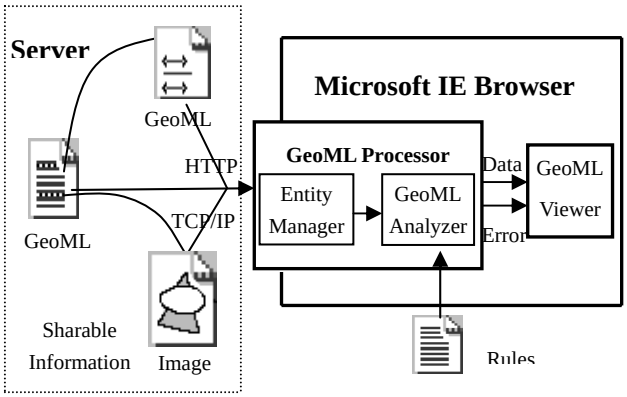

Fig. 2 Function module of GeoML Processor and the logic relation between GeoML document

The components of GeoML Interpreter are detailed here.

a) GeoML Entity Manager

Entity Manager is responsible for locating GeoML fragments which have been declared in entity declaration or other files, and replace the entity reference with the actual entity declaration.

b) GeoML Parser

Parser validates the GeoML document, it compares the content of GeoML document and the prototype document structure rules to check the validation and well-form of GeoML document and send the error message to the GeoML viewer for displaying them to user.

c) GeoML Processor

GeoML Processor consists of GeoML entity manager, parer and other functional module, and is responsible for providing formatted, validated and legal GeoML documents data to GeoML viewer.

There are two different approaches to read the content of an XML document, one is named as Event-driven, the other is Tree-manipulation. Now we used the first method for our GeoML Processor.

d) GeoML Viewer

Viewer accepts the GeoML documents which come from Processor, and interprets various GeoML markups, shows the geospatial information expressed in GeoML document on the browser window.

6. CONCLUSION

The preliminary research and design of geospatial information web sharing language – GeoML has been covered in this paper. But the further research is urgently needed, the author think that the following problems have to be done in the future:

a) Deeply researching the characters and rules of geospatial objects and the relation among them, so the GeoML is designed more effectively, and the expressing of geospatial vector data is expected more perfectly.

b) Making more effort to create GeoML element to express the relation among graphic elements or image data.

c) Devoting oneself to the GeoML displaying technology -- Geospatial Stylesheet Language (GeoSL) on the GeoML Viewer.

d) Doing more research on the GeoML hyperlink technology – Geospatial Link Language (GeoLL) used in vector graphic, image, attribute text and relations among them.

e) Researching and designing the Tree-manipulation by which GeoML Processor send data to GeoML Viewer more effectively, and perfect the web processing of GeoML document. send data to GeoML Viewer more effectively, and perfect the web processing of GeoML document.

f) Continuously keeping watch on XML development, apply the newest maturely network technology to guarantee that the research and design is compatible to the new information web sharing technology. information sharing is more practical.

g) Taking reference and considering the international standards about the sharing of geospatial information, so the geospatial Taking reference the other markup language used in other various professional fields.

h) Thinking about and designing the transformer, which can be used to transform the popular geospatial data format to GeoML document, mainly for vector data.

i) Designing the managing scheme of mass GeoML document.

When the above problems have been solved, we can apply the GeoML technology in geographical field to make geospatial information web sharing come true. And geospatial information expressed by the GeoML can be freely interchanged and shared on Internet. Furthermore, GeoML will provide Networking GIS and Digital Earth with technology, perfect Networking GIS and speed up Digital Earth developmen

References

Al，Gore，1998, Digital Earth: Understanding Our Planet in the 21st Century, Given at the California Science Center, Los Angeles, California, on January 31, 1998, http://www.digtial-earth.gov/speech.html

Autodesk, 1999, http://www.mapguide.com/

Bentley, 1999, http://www.bentley.com/

Bian Fulin etc., 1996, the Principle and Methods of GIS, Survey and Mapping Press, Beijing, China.

ESRI, 1997, ESRI White Paper, The Future of GIS on the Internet, June, 1997

ESRI, 1998, ESRI Shapefile Technical Description, An ESRI White Paper -- July, 1998

ESRI, 1998, http://www.esri.com/

Joel Sklar, 1998, Introduction to XML, http://www.joelsklar.com/-xml-agnd.htm.

Laura Mengel, 1998, What is XML, http://www.fnal.gov/cd/-seminars/www6/xml_what.html .

Liu Dong, 1996, Principle and Application of Spatial Query Language, Ph.D. Thesis of Peking University.

Neil, Bradley, 1998, The XML Companion, Addison Wesley Longman Limited, Essex CM20 2JE, United Kingdom.

Open GIS Consortium, OpenGIS Abstract Specification. Version 3, 1998.

Robin Cover, 1998, The SGML/XML Web Page, Extensible Markup Language (XML) , http://www.oasis-open.org/cover/-xml.html.

W3C, 1998, W3C Recommendation, REC-xml-19980212. Seybold Publications and O'Reilly & Associates Inc., http://www.xml.com.

Xml Application Examples, 1998, http://www.inrialpes.fr/pub/-seminaires/

XML1-10.12.98/sld05000.htm.

Zhu Yue etc., 1997, the Preliminary Rules and Application of GIS, Survey and Mapping Press, Beijing, China.