

基于 XML 的 WebGIS 通信协议的设计与实现*

刘昕鹏, 罗英伟⁺, 汪小林, 许卓群

(北京大学 计算机科学技术系, 北京 100871)

Design and Implementation of XML-Based Communication Protocols for WebGIS

LIU Xin-Peng, LUO Ying-Wei⁺, WANG Xiao-Lin, XU Zhuo-Qun

(Department of Computer Science and Technology, Peking University, Beijing 100871, China)

+ Corresponding author: Phn: +86-10-62763373, Fax: +86-10-62765805, E-mail: lyw@pku.edu.cn, <http://gis.pku.edu.cn>

Received 2003-03-17; Accepted 2003-09-04

Liu XP, Luo YW, Wang XL, Xu ZQ. Design and implementation of XML-based communication protocols for WebGIS. *Journal of Software*, 2004,15(6):899-907.

<http://www.jos.org.cn/1000-9825/15/899.htm>

Abstract: Concerning the features of complex objects and massive data transmission, a new XML-based method to design and implement communication protocols for WebGIS is presented. With the aid of UML, the typical requiring and responding protocols of WebGIS are analyzed through object-oriented concept. Based on object oriented analysis of the protocols, the mechanism of designing communication protocols following W3C's XML Schema specification is illustrated. Finally, the main flow of embedding the protocols into WebGIS is given by packing and parsing XML-based protocols in a WebGIS application prototype. This kind of protocols can be used in spatial information exchange among heterogeneous WebGIS platforms in distributed environment.

Key words: XML; XML schema; WebGIS; validator; protocol

摘要: 针对 WebGIS 通信中面向复杂对象及大容量传输的特点,给出了一个新的基于 XML 的 WebGIS 通信协议的设计和实现方法.使用 UML 工具以面向对象的方法细致分析了 WebGIS 典型的请求和响应协议,并据此详细说明了使用 W3C 的 XML Schema 规范设计 WebGIS 的通信协议的基本方法.最后结合一个 WebGIS 应用原型,给出了在 WebGIS 系统中通过对基于 XML 的通信协议的打包和解析,完成嵌入的主要流程.该通信协议能够适用于分布式异构平台下多种 WebGIS 系统架构的空间信息交换.

关键词: XML; XML schema; WebGIS; 验证器; 协议

* Supported by the National Natural Science Foundation of China under Grant Nos.60030005, 60073016, 60203002 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant Nos.2001AA113151, 2002AA135330, 2002AA134030 (国家高技术研究发展计划(863)); the National Grand Fundamental Research 973 Program of China under Grant No.2002CB312000 (国家重点基础研究发展规划(973)); the National Research Foundation for the Doctoral Program of Ministry of Education of China under Grant No.20020001015 (国家教育部博士点基金); the Natural Science Foundation of Beijing of China under Grant No.4012007 (北京市自然科学基金)

作者简介: 刘昕鹏(1980 -),男,湖北武汉人,硕士生,主要研究领域为 WebGIS;罗英伟(1971 -),男,博士,副教授,主要研究领域为分布式计算,软件 Agent,GIS;汪小林(1972 -),男,博士,讲师,主要研究领域为构件技术,软件 Agent,GIS;许卓群(1936 -),男,教授,博士生导师,主要研究领域为科学计算可视化,并行计算,人工智能,GIS.

中图法分类号: TP393 文献标识码: A

WebGIS 是一个将地理信息及其处理分布于 Web 计算平台的网络化 GIS 系统.目前国际上对于 WebGIS 软件技术的研究基本上集中在空间数据模型、空间数据结构、网络环境下的空间信息组织管理、通信协议、分布式策略等方面,本文所涉及到的方面是 WebGIS 的通信协议问题^[1].WebGIS 的通信协议主要分为控制命令协议和空间数据传输协议.以往的 WebGIS 通信协议大都通过参数帧描述控制命令,这需要仔细定义协议格式并实现相应的协议解析,因而不具有通用性.比如,一个数据请求协议的参数帧可能定义如下:

请求用户	请求时间	是否要整个图层	图层名	是否要索引	索引 id	其他域段...
------	------	---------	-----	-------	-------	---------

而实际上,请求交通线图层的协议可能仅仅填充如下(但这样也造成了很大的浪费):

"lxp"	2002/6/10	True	"交通线"	False	Null	其他域段均为空
-------	-----------	------	-------	-------	------	---------

XML 能够方便地描述具有包含关系的概念模型,而且它对概念模型的表达直观易懂,形式灵活,不受类似参数帧的模式限制.使用 XML 描述协议,不仅可以定义统一的数据和控制命令格式,而且可以方便地使用已有的 XML 解析器,从而极大地方便了协议的扩展和系统对协议的集成.目前已经有了使用 XML 作为通信协议的先例.ArcInfo 在它的基于服务器网站的可发布数据平台 ArcIMS 中,已经使用了 ArcXML 作为基本的命令和数据传输协议在用户网页和后台的空间数据服务器之间进行通信^[2].更值得一提的是,W3C 在 2000 年 5 月推出了 Simple Object Access Protocol(SOAP)1.1^[3],它是基于 XML 表达的轻量级协议,用于在非集中式环境下建立一个信息交换的框架.我们基于 XML 的 WebGIS 通信协议的思想来源于 SOAP 模型,但是我们主要侧重于对 WebGIS 系统的应用.

1 WebGIS 通信协议分析

为了更好地分析 WebGIS 通信协议,我们选取 JAVA 语言开发了具有简单应用的 WebGIS 原型系统.该系统

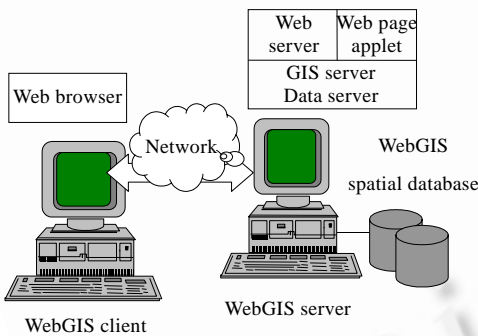


Fig.1 Architecture of applet-based WebGIS

图 1 基于 Applet 的 WebGIS 体系结构

由 3 部分组成:Applet、WebGIS 服务器和数据服务器,如图 1 所示.

数据服务器为 GIS 服务器提供了访问空间数据库的接口.WebGIS 服务器负责监听来自客户端的请求,并与合法用户建立连接.GIS 服务器有专门的数据库访问模块负责完成与数据服务器的交互,获得用户请求的数据信息.WebGIS 服务器还集成了有关 GIS 数据分析处理的各种模块,这包括拓扑分析、最短路径分析、叠加分析、数据转换等基本模块.这样,比较复杂的 GIS 数据处理将在 WebGIS 服务器端完成,而在 Applet 中保留基本的地图功能,从而当 Applet 在客户端下载运行时能够保证 WebGIS 客户端的小巧灵活,即实现“瘦客户端”.通过将 Applet 嵌

入到网页中去,并与网页一起提供各种交互界面,WebGIS 用户能够完成各种空间操作功能.

Applet 在整个 WebGIS 交互中起到了关键的作用.在 Applet 内部封装基本的 GIS 功能,形成一系列具有特定用途的工具性 Applet.例如,用于地图显示的 Applet、用于拓扑分析的 Applet 和用于属性查询的 Applet 等.通过重新定义这些 Applet 界面中的鼠标事件,并留出公有的外部接口,用户可以使用 Java Script 等语言对这些 Applet 进行类似构件的二次开发.

在 Applet 运行过程中,用户请求和 Web 服务器的数据响应形成基本的交互事务.这些交互都将通过 WebGIS 的通信协议来进行表达.为了完成 Applet 界面提供的基本 GIS 交互功能,C/S 两端必须对请求和响应的细节分层,定义良好的协议规范,以求高效、准确地完成交互,这是因为:

- 为了更最大限度地利用网络带宽,提高系统交互的并行性,必须认真分析目前 WebGIS 基础的事务交互中的协议内容,给出协议内容中涉及各类几何数据及其控制的相关性的严格表达,以定义高效的请求和响应模式.

• WebGIS 的通信协议具有鲜明的层次划分,采用以往的协议格式会定义出复杂的协议集合,从而形成最后协议编排中过多的控制字段和数据字段,导致协议开发的复杂化和最终通信效率的下降,因此,定义新的协议规范是一个需要迫切解决的问题。

• WebGIS 的通信协议必须是易于更新和修改的,新的协议规范应该考虑协议的动态扩充能力。

• 通过对 WebGIS 系统的基本交互的分析发现,WebGIS 通信协议的基本内容具有典型的树形结构,能够非常方便地使用类结构建模,并利用形式化的 UML 图规范协议细节,类继承和类包含为协议的扩展提供了非常好的结构支持,这就使得我们定义全新的协议规范成为可能。

用户与 WebGIS 系统的 Web 服务器端的交互决定了 WebGIS 系统通信协议的主要内容,下面两节给出了客户端请求和服务器响应的通信协议的详细说明,并使用 UML 图给出了形式化的描述。

1.1 WebGIS请求协议的UML说明

在原型系统中,Applet 内部包含了地图编辑和实体查询等处理模块,这样,WebGIS 通信协议主要负责地图数据的请求和响应,地图是分层的,基本的请求为图层获取,获取一个图层的请求途径是多样的,主要包括:

- 通过提供图层名来请求图层,这是最基本的请求方式;
- 通过重定向图层数据的位置来请求图层;
- 通过提供几何数据的空间索引子树来请求相应的几何对象索引信息;
- 通过提供空间范围来请求图层中指定范围内的几何实体。

根据上面对图层请求的主要内容的分析,图 2 给出了 WebGIS 系统关于地图数据请求协议的 UML 描述。

请求开始于类 Requests,它是一系列请求 Request 的容器,每一个具体的 Request 都要使用一个 id 标记,并用 time 记录请求的时间,从 Request 派生出的 LayerRequest,表明目前的协议只包括图层请求,每一个基本的图层请求必须包括必须的图层名 name,根据上面对图层请求的细化,从 LayerRequest 派生出 GetLayer,GetRedirection,GetSpatialIndex 和 GetEntities 这 4 种请求方式。

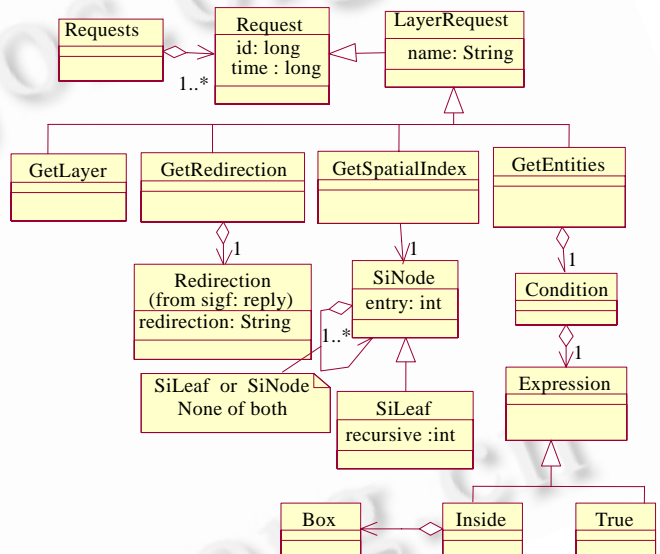


Fig.2 UML diagram of requiring protocols

图 2 请求协议的 UML 描述

• GetLayer 这种方式将得到由图层名标志的整个图层的相关信息。

• GetRedirection 需要使用成员类 Redirection 提供重定向的数据位置 redirection。

• GetSpatialIndex 需要提供空间索引子树,它用于表达用户请求的部分几何实体的索引描述,索引子树由基本的 SiNode 树节点构成,它可以直接是一个 SiLeaf 类型的叶节点,用 recursive 标明递归的层数,也可以是一棵真正意义上的子树,树中的每个非叶节点用 entry 标明入口。

• GetEntities 需要用 Condition 成员类描述用户请求的实体的条件限制,条件用 Expression 表达,它或者是 True,标明没有条件约束,或者用 Inside 下的 Box 指明请求的范围。

1.2 WebGIS响应协议的UML说明

针对上面的请求协议规范,在服务器端的响应也有相应的设计,主要包括:

- 对仅基于图层名的图层请求响应,它响应给客户端所请求图层的主要属性信息;
- 对空间索引树的请求响应,它响应给客户端相应的空间索引子树,每个节点已经携带了对应的索引信息;

- 对几何实体的请求响应,它将以集合的形式返回所有满足用户请求条件的实体信息以及重定向获取实体域段的地址信息;
 - 对重定向的请求响应,它将简单地回传重定向地址,真正的重定向数据获取将由服务器相应的处理模块完成;
 - 对几何实体域段的请求响应,它将响应重定向获取域段数据的地址信息.
- 同样,这些响应协议的分类可以对应到如图 3 所示的 UML 图.

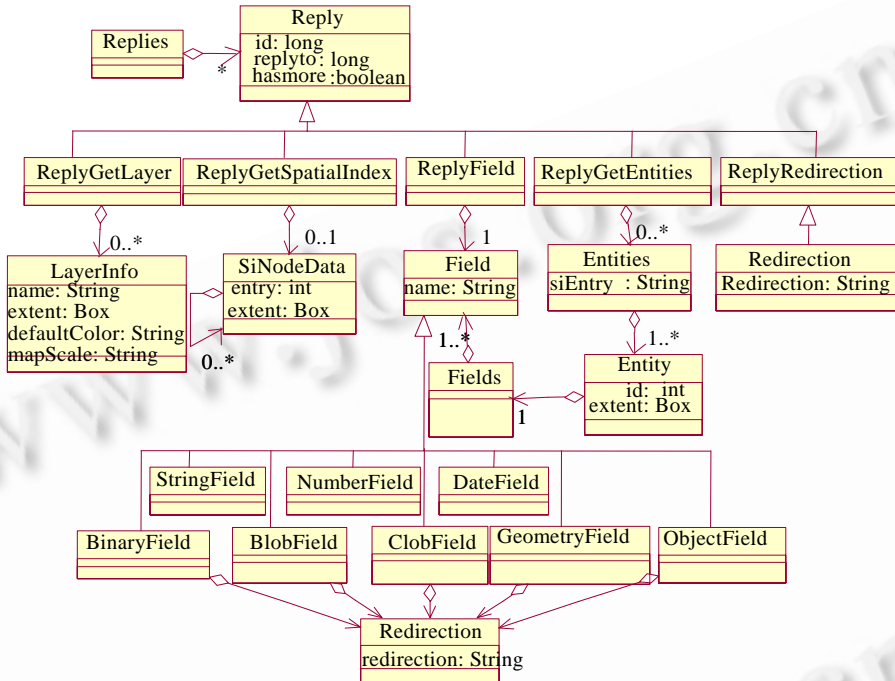


Fig.3 UML diagram of responding protocols

图 3 响应协议的 UML 描述

类似于请求协议,一组 Reply 可以被作为 Replies 的成员类,形成从服务器传来的多个响应.一个 Reply 必须用一个 id 进行标志,同时指明响应的对象 replyto,并用 hasmore 提示是否还有更多的响应或是结束.从 Reply 继承得到 5 类具体的响应:ReplyGetLayer,ReplyGetSpatialIndex,ReplyGetEntities,ReplyRedirection 和 ReplyField.

- ReplyGetLayer 是对请求协议中 GetLayer 的响应,它包含了服务器处理得到的用户请求图层的基本信息:名称(name)、范围(extent)、缺省色(defaultColor)以及比例尺(mapScale).
- ReplyGetSpatialIndex 是对请求协议中 GetSpatialIndex 的响应,它包含了以 SiNodeData 作为树节点(叶节点或者非叶节点)的空间索引子树.每个节点列出了它的入口 entry 以及该节点以下子树的范围(extent).
- ReplyGetEntities 是对请求协议中 GetEntities 的响应,它包含记录了一列满足用户请求条件的 Entity 的 Entities 成员类.每个满足条件的 Entity 由 id 标志,并注明该 Entity 的范围(extent).Entity 的数据信息被细化为用 Fields 描述.Fields 记录了一列归属于该 Entity 的 Field 重定向信息.Field 被具体化为 BinaryField,BlobField,ClobField,GeometryField,ObjectField,StringField,NumberField 和 DateField 共 8 类.
- ReplyRedirection 是对请求协议中 GetRedirection 的响应,它包含一个 Redirection,后者使用 redirection 字符串指明重定向的地点.
- 另外,也可以采用 ReplyField 直接响应某个属性域段.

需要说明的是,我们着重对基于 Applet 的 WebGIS 模型进行了通信协议的分析,这主要是为了反映在与平台无关的 Java 语言为开发背景下使用通信协议的平台无关性.由于基于 XML 的 WebGIS 通信协议自身被编排

为文本格式,而不同于以往的二进制数据流,在特定的比较成熟的平台技术下,例如微软的 ActiveX 控件、基于 C 代码的 CGI 过滤服务器模式以及 Plug in 模式等等,基于 XML 的文本解析和处理也得到很强的支持,因而该协议仍然有很广泛的应用空间。

1.3 基于XML的WebGIS通信协议和GML

OGC 在 2000 年 5 月 12 日发布了基于 XML 的空间信息表达规范 GML(geography markup language)1.0, 并很快成为业界所接受的空间信息交换格式。在许多大型的、需要集成分布式异构 GIS 平台并且构造平台之上的应用集成框架的商业应用中,GML 已经被首先采纳,作为标准的空间数据的通信格式。基于 XML 的 WebGIS 通信协议与 GML 有着紧密的联系,但是仍然存在区别:

- GML 语言(目前比较完善的版本是 GML2.0)主要面向为用于描述客观世界的几何实体定义平台无关的、标准的几何实体类框架,并提供标准的标签格式(这一点通过 GML1.0 的 DTD 和 GML2.0 的 Schema 来提供),它不包括对几何实体的面向应用的控制信息,而基于 XML 的 WebGIS 通信协议则能够描述特定的服务请求和响应;

- GML 是数据定义语言,它本身不包括协议的控制信息,即它本身不能作为独立的协议,而基于 XML 的 WebGIS 通信协议则使用了 GML 作为其描述空间数据的部分。例如,图 2 中的 Box 可以使用 GML 中的 Box 标签来进行描述。可以说,基于 XML 的 WebGIS 通信协议是针对实际的 GIS 应用对 GML 的扩展。

从以上两点可以看出,基于 XML 的 WebGIS 通信协议具有良好的适应性,它既集成了面向应用的数据控制信息,也集成了标准的 GML 数据描述信息,做到了 GIS 平台无关,可以被方便地嵌入多种分布式 GIS 应用中。

2 基于 XML 的 WebGIS 通信协议的实现

通过定义协议的 XML Schema 文档,也就定义了基于 XML 表达的协议的元数据。使用元数据规范化协议,能够从数据的模式抽象层次定义基本的协议表达。对协议的元数据的修改,能够完成一类协议的重定义,这有利于协议集的更新和扩展。

2.1 WebGIS系统通信协议的XML Schema实现——定义协议的元数据

从 UML 类图结构可以很方便地转换到 XML Schema,转换一般要基于以下的规则^[4-6](限于篇幅,这里并不列出具体的请求协议和响应协议的 XML Schema):

- UML 中的一个类对应于 XML Schema 中的一个复杂类型。UML 图中的类分为抽象类和非抽象类两种。抽象类一般能够概括出由它继承得到的所有派生类的公有属性和方法,其本身不能够被实例化。在 XML Schema 中,可以使用属性 abstract 标志抽象类对应的复杂类型为抽象类型。

- UML 中的类属性对应于 XML Schema 中各种复杂类型的属性。

- UML 中的类继承对应于 XML Schema 中的类型扩展标签中的 base 属性值。该值指明了继承链中的基类型。

- UML 中的成员类通过定义 XML Schema 中复杂类型的嵌套子元素得以表达。

2.2 协议验证器

2.2.1 验证器在系统协议传输中的应用

使用 XML Schema 描述 WebGIS 系统的通信协议模式的主要目的在于利用 XML Schema 对 XML 数据文件的元数据说明能力,使用通用的基于 XML Schema 的 XML 文档验证器自动完成基于 XML 的通信协议有效性检查。验证器在系统的协议传输过程中可以被认为是在 C/S 两端的一个流程开关。为了提高系统的效率,我们基于这样的假设:

- 服务器的响应协议一般被认为已经严格遵循了上面 XML Schema 的定义,在客户端接收到响应时,可以关闭验证开关,直接对 XML 表达的协议进行解析;

- 熟悉、友好的客户端会被服务器识记。从这样的客户端提交的请求协议一般也认为是严格遵守了 XML

Schema 的定义,从而在服务器处理请求时,关闭验证开关,直接对请求协议进行解析;

- 新的客户端必须经过一个阶段获取服务器的足够信任.在这个阶段内,凡是由客户端提交的请求协议必须经过验证器的有效性检查,避免因协议本身的错误造成协议解析的不必要的资源消耗.

2.2.2 协议验证器的基本验证流程

基于 XML 的 WebGIS 通信协议的 XML Schema 所包含的类型和元素定义遵循通用的 XML 文档描述规则,但是由于其自身受应用需求限制,并没有完全涉及所有的 XML Schema 定义特征,所以采用通用的 XML 验证器会对本身就复杂的协议验证增加由于不必要的为实现通用验证能力带来的额外负荷.实现面向协议本身的协议验证器是提供系统交互性能的重要保证之一.

协议基本验证流程可以用图 4 来描述.系统采用 W3C 开发的 DOM(document object model)来解析协议的 XML Schema 和协议本身.通过 XML Schema 验证协议的有效性,尽管可以确保以后协议解析的正确性,但是这一过程本身是复杂而耗时的,所以,只要当确实有必要进行协议验证时才完成这一流程.

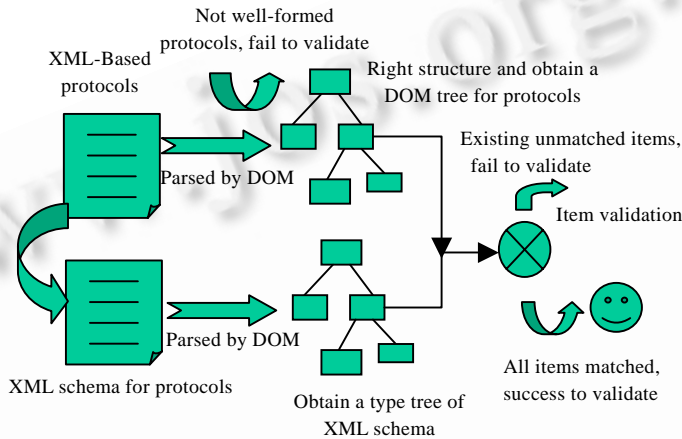


Fig.4 Work flow of XML schema-based validator for protocols

图 4 与协议相关的基于 XML Schema 的验证器工作流程

2.2.3 协议所涉及的 XML Schema 的基本验证项

实现元素类型的嵌套结构检查的主要办法是:从协议的 DOM 树和定义协议的 XML Schema 的类型树的根节点出发,采用深度优先的树周游算法,用类型树的节点信息逐一验证协议的 DOM 树节点.元素类型的嵌套结构检查是以下验证项的框架.伴随着树周游,每个节点都将对以下的验证项作依次检查.

- 元素名称检查

出现在 DOM 树的每个元素的标签名称必须在 XML Schema 类型树的相应类型中有定义,而且因为 XML 是大小写相关的,所以在进行验证时,DOM 树被周游到的当前元素必须和类型树中相应元素定义做区分大小写的字符串匹配.

- 属性继承和 use 属性检查

类型树的构造方便了在属性继承时的检查.验证器所要做的是将 DOM 树的元素结点中的属性列表和类型树中对应的元素定义的属性列表进行匹配.在匹配时主要是验证那些必需属性是否在协议的表达中出现,属性名是否匹配,属性类型是否对应.必需属性在 XML Schema 中用属性的 use 值加以说明,这一点也被记录在类型树节点信息中.

- 元素出现次数(minOccurs,maxOccurs)检查

验证器为协议的 DOM 树中的每个元素建立一个计数器,随着树周游的进行,计数器记下每个元素在协议的实际表达时出现的次数,最终和类型树中相应元素定义的阈限制进行比较,如果出现不符,即说明协议无效.

- 复杂元素的内容组合(sequence,choice,all)检查

XML Schema 允许的内容组合主要有 sequence,choice 和 all 这 3 种.在 XML Schema 的类型树中,sequence,

choice 和 all 都将作为独立的树节点出现.验证器将为这 3 种节点的子节点生成一个元素链,并考虑协议中相应元素出现的顺序和次数来完成验证.

2.3 协议的 XML 打包及类解析的实现

通过定义协议的 XML Schema 给出了协议的元数据描述.通过实现与协议相关的简单验证器保证了通过 http 传输的协议的正确性.然而要真正实现 WebGIS 系统对基于 XML 的协议的集成,必须在 C/S 两端将具体的请求或响应协议从对象集合打包为 XML 数据流,并且相应地能够将 XML 数据流解析为协议的对象集合.

2.3.1 一个简单的请求回话

图 5 给出了一个具体体现协议打包和解析的通信会话流程.客户端在执行用户显示“北京市交通线”图层请求时,首先需要用请求协议的 UML 对象类建立内存中的协议对象树.对象树经过 XML 打包,而成为根据请求协议的 XML Schema 规范书写的 XML 字符流.字符流通过 C/S 之间的 Internet 连接到达服务器,被接收下来而重新形成请求协议的对象树.服务器识别出这是 GetLayer 请求,将它分派给 GetLayer 处理模块,后者通过 GetLayer 对象的属性从数据库中查询得到“北京市交通线”图层数据.

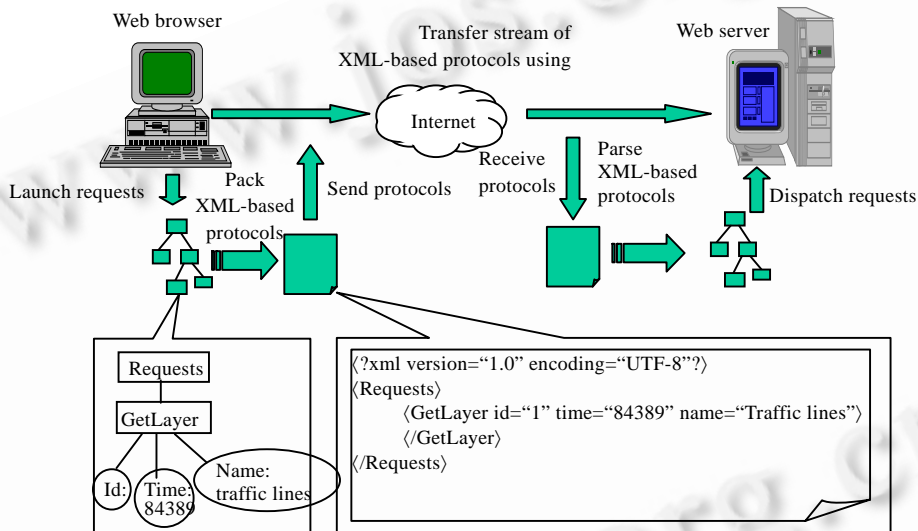


Fig.5 Packing and parsing of XML-based protocols for WebGIS

图 5 基于 XML 的 WebGIS 系统协议打包和解析

2.3.2 协议的 XML 打包

协议的 XML 打包是同时存在于 C/S 两端的协议处理模块.

在客户端,用户在 Applet 上触发的请求事件被客户端所捕获,接着被装入请求的对象树中.请求的对象树是我们根据请求协议的 UML 图设计的 JAVA 类的实例集合.它作为用户请求和 XML 协议生成的中间界面,是方便 WebGIS 系统 XML 协议集成的重要组成部分.在服务器端,服务器同样需要将各个处理模块得到的数据和响应信息装入响应协议的对象树中,再打包成 XML 协议流传给客户端.

协议的 XML 打包过程本身是简单的,这和我们从协议的 UML 图生成协议的 XML Schema 是类似的.只不过后者是在元数据的层次上完成对象类到 XML 语法的映射,而前者是在数据本身完成这一映射.

2.3.3 协议的类解析

一次从客户端到服务器再到客户端的完整会话必须包括协议 XML 打包的可逆过程——协议的类解析.它也是同时存在于 C/S 端的协议处理模块.协议的类解析从实现技术上讲和协议验证有相似处.它需要将接收到的 XML 协议流通过 XML parser 形成 DOM 树,而后从通用的 DOM 树中取出协议的参数,装入协议的对象树中.

综上所述,协议的 XML 打包和解析体现了 WebGIS 系统对基于 XML 的通信协议的集成,它们从实现层面

上保证了协议的可扩展性以及 C/S 两端会话的完整性和有效性。

3 基于 XML 的 WebGIS 通信协议的可扩展性和局限性

3.1 WebGIS通信协议的结构特点

从图 2 和图 3 可以看出,WebGIS 的通信协议的结构因为采用了类继承和类包含而具有良好的可扩展性,这体现在:

- 面向对象本身对属性和方法的封装易于协议模块的稳定.事实上,对协议中一个类内部的属性修改不牵涉其他任何类,而以往基于线性域段的协议帧则不具有这个优点.一个域段的添加或删除都会造成相关的一系列域段的偏移量的改变和处理流程的重新设计.

- 类继承对一类相似的请求或者响应作了高层次的概括,提取出公有的属性,利于协议表达的简洁.协议集在实现的初期随着应用的扩展和功能的深入必然会进行很多扩充.继承类的添加过程独立于其他兄弟继承类,最大限度地保留了原有协议的有效性.例如,在 Requests 抽象类下面添加 mapRequest 继承类只需定义 mapRequest 类本身.

- 类聚合表达了一个请求或者响应的具体内容,而这正是协议中频繁变动的地方.使用底层的类结构描述具有很好的灵活性.当我们需要引用或者修改内容时,只要置换请求或响应的成员类就可以了.一个有说服力的例子是,比较复杂的 WebGIS 通信协议会包含 MultiLayerRequest 请求.一个 MultiLayerRequest 请求是对一幅地图中多个图层的一次性请求,它可能在客户端对一幅地图中的相关图层作叠加分析时被引用.MultiLayerRequest 从 Requests 的继承而来,它应该包含 LayerRequest 作为成员类,尽管它自己可能包含新的属性来标志这些请求图层之间的关系.这时只要直接将 LayerRequest 类作为它的成员类就可以完成对整个 LayerRequest 结构的引用了.

基于 XML 的通信协议的结构特点使得它自身就带有良好的可扩展性,这也是除了使用 XML 作为基本的语法表达之外,协议设计的又一大特点.

3.2 基于XML的WebGIS通信协议的应用局限

基于 XML 的 WebGIS 通信协议的应用局限来自于对 XML 字符流的结构化处理.

比较强大的 XML 字符流解析器是 DOM,它能够把 XML 文档处理为具有典型树结构的对象模型.我们可以方便地使用 DOM 对象模型中的对象方法提取 XML 文档中的元素名、元素属性、文本内容等等.

我们使用的是 DOM 的 JAVA 包进行开发.但是,由于这个包较大,而微软 IE 浏览器内嵌的 JAVA 核心版本较低,所以 WebGIS 的客户端必须下载这个包,才能正确完成协议在客户端的解析和 XML 打包.这不利于实现 WebGIS 的瘦客户端.DOM 解析器在客户端的应用因此被限制.目前可行的解决方案是:

- 在客户端采用另外一个 XML 解析器 SAX(simple API for XML)完成 XML 协议流的解析.SAX 比 DOM 包小得多,它提供了识别 XML 文档中各个元素的开始标签、结束标签、文本内容等能力,并通过重定义事件实现对 XML 文档的元素的解析.SAX 的功能较 DOM 而言弱许多,它不能识别整个 XML 文档的元素嵌套结构,而只是对所有的元素提供统一的事件接口,这对开发者来讲,需要自行设计事件的处理,使 SAX 解析器能够达到类似 DOM 的解析功能.

- 在客户端实现对协议的 XML 字符流的直接转换,即以每个协议的对象为单元,向连接的输出流中按照 XML 的规范和协议 XML Schema 的格式输出该对象所对应的 XML 元素字符串.因为目前协议所涉及的对象类基本对应到 XML 表达中的元素,而且对象数目不多,这个方法是可行的.但是随着协议本身的扩充,如果不能找到客户端替代 DOM 的较合适的工具,将会极大提高实现的复杂性.

4 总 结

本文结合 WebGIS 的应用背景,提出了以 XML 作为 WebGIS 系统通信协议的基本思想,并给出了详细的协

议设计和实现.本文主要做了以下工作:

- 给出了 WebGIS 系统中请求协议和响应协议的基本框架和 UML 描述;
 - 结合 XML 以及定义 XML 的元数据 XML Schema 技术,实现了协议的 UML 描述到 XML Schema 定义
- 转换;
- 给出了实现和协议相关的验证器的设计思路;
 - 结合应用实例标志了在一个完整的 WebGIS 系统协议处理流程中 XML 打包和 XML 解析的作用;
 - 分析了以 XML 表达 WebGIS 系统协议的优势和局限.

所开发的原型系统已经在客户端实现了基本的地图浏览功能.鉴于原型系统的示例特性,它只是实现并演示了对基于 XML 的 WebGIS 通信协议的应用效用,对于在真实的客户环境下需要处理大规模终端交互的实用系统,系统除了需要集成成熟的基于 XML Schema 的验证器,还需要在分布式应用环境中,考虑以 Web Service 的形式发布 WebGIS 服务,并在 SOAP 协议中嵌入 WebGIS 通信协议.图 6 是一个典型的示例界面,在地图的显示过程中,基于 XML 的 WebGIS 通信协议在下方的文本框中也被打印出来.



Fig.6 A demonstrative screenshot of prototype system at client side

图 6 原型系统的一个典型客户端界面

基于 XML 的通信协议作为新的设计思路被提出并初步加以实现,但目前只能完成简单的 C/S 交互.为了完善协议,下一步需要进行的工作包括:

- 根据协议分析 WebGIS 通信回话的细节,实现一次事务的多次交互功能;
- 实现 C/S 端 XML 协议流的并发发送以及另一端 XML 协议流的正确截取和分派,提高系统的通信效率;
- 结合 WebGIS 客户端 Applet 的二次开发需求,扩充和定义具有更高灵活性的协议集并完成协议的 XML Schema 元数据定义;
- 使用通用的基于 XML Schema 的 XML 验证器,完善协议验证机制.

References:

- [1] Li MH. Study and implementation of WebGIS as a component system [MS. Thesis]. Beijing: Peking University, 2000 (in Chinese with English abstract).
- [2] ESRI. ArcXML Programmer's Reference Guide(ArcIMS 3), PDF on CD only. 2001.
- [3] W3C Note, Simple Object Access Protocol (SOAP) 1.1. 2000. <http://www.w3.org/TR/SOAP>
- [4] W3C Proposed Recommendation, XML Schema Part 0: Primer. 2001. <http://www.w3.org/TR/2001/PR-xmlschema-0-20010330>
- [5] W3C Proposed Recommendation, XML Schema Part 1: Structures. 2001. <http://www.w3.org/TR/2001/PR-xmlschema-1-20010330>
- [6] W3C Proposed Recommendation, XML Schema Part 2: Datatypes. 2001. <http://www.w3.org/TR/2001/PR-xmlschema-2-20010330>

附中文参考文献:

- [1] 李慕华.分布式 WebGIS 的构件化研究与实现[硕士学位论文].北京:北京大学,2000.