

网格 GIS 中空间信息描述语言的研究

Research of the Language of Special Information's Description in Grid GIS

刘建英/LIU Jian-ying, 徐爱萍/XU Ai-ping

武汉大学计算机学院, 武汉 430079

School of Computer Science, Wuhan University, Wuhan 430079, China

[摘要] 在网格 GIS 系统中, 分布在不同位置的数据、信息、互操作、协同等需要具有一种标准的语言, 使相互之间能够进行对话和交流, GridGML (Grid Geographic Markup Language) 将是解决这些问题的关键技术。在 XML 技术、网格体系结构技术及 OGC 的 GML3 标准的基础上, 对 GridGML 进行了研究, 提出了其设计思想。

[关键词] 网格; GIS; XML; GML; GridGML (Grid Geographic Markup Language); GRSP

[中图分类号] TP393

[文献标识码] A

[文章编号] 1000-7857(2006)06-0045-03

Abstract: In the system of grid GIS, it is needed for data, information, operation each other and distributing in various places a kind standard language in order to communicate. Grid Geographic Markup Language is the key technology. Here we present researches of the GridGML on the base of XML technology, grid system structure, the standard of GML3 and OGC. A kind of idea of designing GridGML is put forward in this paper.

Key Words: Grid; GIS (Geographic Information System); XML (Extensible Markup Language) GML (Geographic Markup Language); GridGML (Grid Geographic Markup Language); GRSP (Grid Request Service Protocol)

CLC Number: TP393

Document Code: A

Article ID: 1000-7857(2006)06-0045-03

1 引言

随着 Internet 的流行和高性能计算机的利用以及低耗费、高速网络的发展, 计算机网络成为单个统一强大的计算机资源的梦想正在逐步成为可能。所谓网格(Grid), 是指将机群、超级计算机、大规模存储系统、数据库以及其他地理上分散的特殊仪器设备, 甚至个人计算机等所有的计算资源、存储资源、通信资源、软件资源、信息资源、知识资源等连接起来作为单个统一资源使用, 从而能够方便快捷地解决各种复杂的问题。继 Internet 实现了计算机硬件的连通, Web 实现了网页的连通, Grid 试图实现互联网上所有的资源全面连通, 掀起第 3 次网络技术浪潮, 因此有人也称网格为第 3 代因特网, Grid GIS 也随之应运而生。

Web GIS 是基于广域网的, 而 Grid GIS 是基于网格计算的。Web GIS 和 Grid GIS 相比, 存在以下问题。

1) 异构空间数据互操作的问题: 特定的 Web GIS 系统是特定的 GIS 数据及其应用; 各自独立、相对封闭、无法互相沟通和协作。

2) 无法实现跨平台操作问题: 基于 RMI、CORBA、DCOM 等中间件平台要求服务客户端与系统提供的服务本身之间必须紧密耦合, 无法实现跨平台的数据访问。

3) Grid GIS 是异种汇集和共享空间信息资源, 进行异体化组织与处理, 具有按需服务能力空间信息基础设施。

4) Grid GIS 能较好地实现互联网上的互操作。现有的远程访问协议 DCOM、CORBA、RMI 不能很好地用于互联网环境。Grid GIS 采用 W3C 标准, 真正的与平台无关; 不受现有的代理和防火墙的限制, 可以用 HTTP 验证模式, 支持安全套接层(SSL)。

所以, Grid GIS 将为空间信息用户对空间数据进行信息获取、共享、访问、分析、处理提供技术支持, 为空间信息提供一个强大的空间数据管理和信息处理基础设施。要实现 Grid GIS, 必须有一种标准的语言为网格 GIS 的中间件之间、智能体之间、系统之间及系统内部的空间信息传输、存储与通信等提供可能。本文将对此标准的语言进行研究和设计。

2 网格 GIS 中空间信息描述语言 GridGML 的设计思想

2.1 GridGML 的使用模式

GridGML 是在网格计算平台上描述地理信息, 它需要网格环境的支持, 图 1 是一个网格 GridGML 使用模式的设计框图^[1,5]。

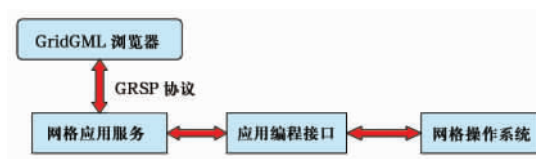


图 1 GridGML 使用模式

Fig. 1 The used mode of GridGML

2.1.1 网格应用服务器

网格应用服务器是基于网格编程接口开发的向网格最终用户提供特定服务的程序, 网格应用服务器类似于现在的 Web 服

收稿日期: 2005-12-06

作者简介: 刘建英, 女, 武汉市珞喻路 129 号武汉大学计算机学院, 高级工程师, 主要研究方向为实验教学、GIS 信息管理系统;

E-mail: ljyliu2@163.com

务器,所不同的是 Web 服务器提供的是页面访问服务 (page access service),而网格应用服务器提供的是资源访问服务(resource access service)。网格应用服务器通过网格编程接口实现对单个计算资源的访问或协同使用多个计算资源,通过网格服务请求协议 GRSP(Grid Request Service Protocol)向网格浏览器提供资源访问服务。

2.2.2 GridGML 浏览器

GridGML 浏览器是图形化的网格客户端访问设备,实现友好的资源使用环境。GridGML 浏览器使用 GridGML 服务标记语言,并通过 GRSP 协议向网格应用服务器发送资源访问请求。

2.2.3 网格请求服务协议 GRSP (Grid Request Service Protocol)

网格服务请求协议是 GridGML 浏览器与网格应用服务器之间进行通信时所遵循的规范,通过 GRSP 网格浏览器可以向网格应用服务器提交资源访问请求;在计算结束后,网格应用服务器将计算结果也通过 GRSP 返回给浏览器。GRSP 协议应该支持现有的 HTTP 协议,同时支持双向操作。

2.2.4 应用编程接口^[5]

GridGML 将是一种与平台无关的标记语言,考虑到系统的兼容性、用户界面的友好性、运行效率以及表达能力的控制与扩展等问题,GridGML 的实现中应有一套应用编程接口软件支持,它应该包括以下几个部分。

1) GridGML 编辑器:使 GridGML 的编写更加可视化,就像 FrontPage, DreamWeaver 等软件方便了 HTML 的编写一样。

2) 资源映射器:管理的是以“网格社区”为单位的资源。整个网格中的资源是没有边界的,网格是一个自组织的动态系统,而社区概念的引入使一个没有边界的系统变成了一个有界的组织,这样做为网格资源的管理提供了有效的手段,也为网格用户的界定提供了一种以社区为单位的可能,为针对网格资源的访问控制提供了主体的认证标准,社区概念对网格资源的共享模式、组织和管理方式提供了极大的便利。

3) 数据库服务中间件:它将不同的资源统一成若干种 GridGML 可以表述的对象,使 GridGML 用户可以不必关心资源的细节而直接使用对数据库的访问。每个数据库中间件除了提供数据库访问服务以外,还要定期运行网格资源注册协议,将本地资源注册到前面提到的网格社区中,就像 HTML 需要一个支持 HTTP 协议的 Apache 系统一样,GridGML 也需要一个支持 GRSP 协议的软件,传送 GridGML 文本。

4) 部署 GridGML 页面、提供 GridGML 服务的功能的接口软件。

2.2 GridGML 的技术框架

GridGML (Grid Geographic Markup Language) 是用于网格系统环境下异构、跨组织与区域系统之间的互操作、信息共享、信息传输、信息存储的网格地理信息标识语言。GridGML 以 OGC 公布的 GML 规范为基础(目前为 GML3),并对其进行扩充,增加系统互操作、网络通信等内容,应用于网格 GIS 环境。

在网格 GIS 系统中,当用户提交一个任务时,网格 GIS 系统首先分析完成任务所需要的计算资源。然后,找到这些资源并分配给任务。同样地,运行任务所需要的数据也被检索出来并传送给计算资源。在这个过程中,网格 GIS 需要具备:分析任务的能力,随时掌握网格中资源的能力,执行任务程序的能力,任意传输数据的能力,判定和保障服务质量的能力,从错误中恢复的能力,记录出错情况的能力等。而在这个过程中,网格 GIS 环境中的各智能体、系统之间的数据流、任务流等如何理解和执行,即是 GridGML 的主要功能。

GridGML 标记语言是一种描述网格 GIS 的语言,GridGML 浏览器将 GridGML 描述的服务请求发送给网格应用服务器,网格

应用服务器在翻译后调用网格编程接口完成计算任务。

GridGML 既是一个网格应用编程语言,又是一个网格应用使用语言,用户使用 GridGML 编写一个 GridGML 页面并部署到网格上,然后就可以通过一个 GridGML 网格浏览器来使用这个页面,从而获得网格上的服务。在运行时,GridGML 页面的网格请求通过一个网格定义的协议 GRSP 从客户端网格浏览器传到网格端的一个网格服务器上,后者完成所需的服务后将结果传回网格浏览器。

网格 GIS 要解决的根本问题还是对地理信息的处理。因此,GridGML 首先要扮演的是一种描述空间地理信息的角色,它需要做到如下几点。

1) 管理来自不同数据源的地理数据。GridGML 是 GML 在地理领域的扩展,因此它具有 GML 所具有的对数据来源的多样性和多种应用的灵活性和适应性。GridGML 能对来自不同数据源的结构化的 GIS 数据进行合并、集成,然后统一进行处理。

2) 方便空间数据的共享和互操作。空间数据要被 GridGML 唯一地进行标志,要方便地用于统一查询和搜索,以大大提高信息资源共享性和地理信息服务的互操作性。

3) 尽量做到内容和显示形式的分离。GridGML 本身不是一种显示语言,它只描述 GIS 数据本身,而数据的具体表现形式可利用其他的技术进行处理。

GridGML 的设计是面向网络的,将地理特征采用 GML 的文本方式进行描述,目前的浏览器都可解析,用户不需要采用其他的软件来显示和交互基于 GridGML 编码的地理数据。在 GridGML 中 GridGMLSchemas 的制定非常重要,对于空间信息网络交互 GridGML 技术的解决途径,应该从研究 GIS 的概念模型和元数据开始,制定相应的规范,使各个网格中间件都相互协作,达到相互理解的目的。GridGML 技术的研究架构如图 2 所示^[2-3]。

在图 2 中,根据面向对象的思想,GridGML 把地理空间信息分成好几个大类进行描述。根据抽象类型 Object 和 GIS 概念模型得到的 GML OBJECT 对象是各类地理空间信息的父节点,由它可以衍生出特征(Feature)、几何(Geometry)实体、拓扑(Topology)关系、数值(Value)、覆盖(Coverage)、坐标参照系统(Coordinate Reference System)、时间参照系统(Time Reference System)、元数据(Metadata)等对象。基于每一对象又可以进行多重衍生,以描述相应的地理空间信息。

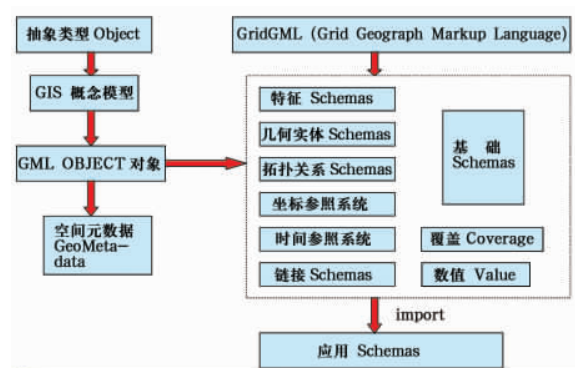


图 2 GridGML 技术研究框架

Fig. 2 The research framework of GridGML technique

2.3 GridGML 页面结构

GridGML 可由以下几个部分组成^[4-6]:GridGML 语言定义,GridGML 客户端,GridGML 服务器,通讯协议,社区系统。

GridGML 语言定义遵循 XML 规范,GridGML 自身定义了自己的文档结构和特定标记,与用户自己定义和其他系统定义的标

记共同使用。GridGML 自定义标记分数据引用、显示风格、界面输入、转移、内部函数几种功能,它支持对 Web Services 和 Grid Services 的描述和访问。

GridGML 客户端具有对 GridGML 文档解析、处理、显示的功能。它支持界面直接修改和缺省回写接口定义的功能。GridGML 客户端也支持现有 Web 资源的获取与显示。

GridGML 服务器提供 GridGML 文档的管理与访问,也提供 Web Services 和 Grid Services 访问代理功能,GridGML 服务器也提供多个客户端之间交互的通路。

GridGML 通讯协议(GRSP)提供客户端和服务器双向的高效连接与传输机制。

GridGML 社区系统支持资源的组织,提供资源虚拟视图,方便用户使用,社区具备资源地址描述的虚实转换及上下文推理功能,允许社区内部运行专有的协议,实现更紧密的联合。

一个 GridGML 页面由 HEAD、BODY 和 TAIL 三部分组成。

GridGML 的头部包含该 GridGML 文件的版权、制作日期、社区等信息;还包含认证信息和访问控制列表(ACL)等安全控制信息,以实现网络安全中的访问控制。GridGML 的 HEAD 部分与 HTML 的 HEAD 部分很相似,它们都不是需要在页面中显示的信息,一个 HEAD 的例子如下:

```
<HEAD>
<TITLE>GridGMLExample </TITLE>
<AUTHOR>zhang </AUTHOR>
<COPYRIGHT>ICY </COPYRIGHT>
<ENCODE>GB2312</ENCODE>
<COMMUNITY>ICT</COMMUNITY>
<DATE>2004-11-18</DATE>
</HEAD>
```

BODY 部分用来描述由结构化数据、非结构化数据、转移和服务等网格操作和内容组成的网格应用;TAIL 部分主要用来定义一些服务结束动作,比如清除临时文件,记录 LOG 等;网格浏览器仅仅显示 BODY 部分,BODY 部分是 GridGML 的核心,是 GridGML 浏览器执行网格操作的实体,它由以下 4 个主要的部分组成:非结构化数据,如 TEXT(文字)和 IMAGE(图片);STRUCTURE(结构化数据);TRANSITION(转移);SERVICE(服务)。

<TEXT>标签与<IMAGE>标签与 GML 中的使用方法相同。

<STRUCTURE> 标签用来显示和处理结构化数据,STRUCTURE 既可以显示静态的数据,比如点、线、面,也可以显示动态的数据,比如属性数据。数据库数据是结构化数据,它是动态的。在 GridGML 里有关数据库数据的标签不用来记录数据的内容,而是记录数据的来源。结构化数据可以以列为单位,column 的属性信息详细描述了一列数据是如何得到的,比如下面的一段 column 描述:

```
<COLUMNHOST="127 0 0 1"DB="gridDB"TABLE="grid"
COLUMN="col1"COND="">TITLE</COLUMN>
```

它表示该列来自 127 0 0 1 主机上的“grid”数据库中的“grid”表中的“col1”列,条件为空,显示的时候该列的标题是“TITLE”,GridGML 浏览器负责将如下请求通过 GRSP 协议以 GET 传送给数据库中间件:DBserver:grsp://127 0 0 1/DBserver?db=grid&table=grid&column=col1&cond=""

数据库中间件会通过网格数据库的查询返回以 XML 编码的数据结果,如下例:

```
<? GridGMLVERSION="3.0"?>
<DATASET>
<CELLKEY="1">10</CELL>
<CELLKEY="2">100</CELL>
<CELKEY="3">1000</CELL>
```

</DATASET>

元素属性“KEY”在这里是指能够准确定位一个元素的表达式,表达方式因数据库不同可以不同,在上例中,KEY 的值是表的主键,数据库可以根据 KEY 的不同来决定对数据的修改权,如果有 KEY 值,那么浏览数据的用户就可以在 GridGML 浏览器中直接修改数据库中的数据。当用户在 GridGML 浏览器中编辑一个元素的内容时,该操作是与数据库同步的,即用户修改完毕以后,修改结果立即生效。与获取数据不同的是,这次 GRSP 协议采用的是 PUT,而不是 GET 方式。

在 GridGML 里边,数据库数据的获得可以通过一种类似“索引”的方式,这种方式给 GridGML 的编写提供了一种可能,即用户不必关心数据库技术的细节,可以自由组织数据格式。因此对数据的采集再不是一种数据对应一个 CGI 了,而是用户可以根据自己的需要动态组合,数据的传递是通过 GRSP 协议实现的。

在 GridGML 中除了有 HTML 中超链的功能外,还有其他的网格特征。在 HTML 里边,连接是确定的 URL,即 HTTP 协议知道将要连接的具体网络位置;在网格系统中,同质服务可能有很多,比如提出一个计算服务的请求,查询结果有若干个结点都可以提供服务,而且网格服务具有“转瞬即逝”的特征,第 2 次访问同一个服务,服务的提供方可能不在前次的网络位置。

“服务”在网格系统中是一个相对复杂的问题,实际上网格提供的各种功能都可以用服务来描述,比如计算服务、数据库服务,还有 GridGML 服务等。GridGML 中的服务被看成是一种原子的计算,也可以看作是一个对若干个参数进行事务处理的函数。因此,GridGML 通过 GRSP 协议的 EXEC 格式发送服务请求,服务的参数可以是用户的输入,也可以是 GridGML 页面的内容,比如结构化数据。与 HTML 不同的是,HTML 通过 FORM 表单和其他静态数据构成 CGI 的参数,而 GridGML 的参数没有类似 FORM 这样的嵌套的标签限制,GridGML 的数据可以是该 GridGML 文档的任意部分,甚至可以是服务能够识别的远程资源。

3 结语

基于网格技术的地理信息系统的研究与开发,将改变传统 GIS 系统的应用模式与计算能力,使 GIS 转变为以服务为中心,提供一体化的空间信息应用服务的智能化信息平台。GridGML 描述了现实世界中实体对象的结构及其关系,与任何图形规范无关,将是网格 GIS 的交流与对话语言,为网格 GIS 的中间件之间、智能体之间、系统之间及系统内部的空间信息传输、存储与通信等提供了可能。

从目前的 Web GIS 发展到 GridGIS 还有一段相当长的路程,因此,本文所进行的研究和所提出的设计思想对 GridGML 语言标准的制定有一定的参考价值。

参考文献(References)

- [1] 骆剑承,周成虎,蔡少华. 基于中间件技术的网格 GIS 体系结构[J]. 地球信息科学, 2002 (3): 20-21.
- [2] 王继周,付俊娥. 基于 GML 的网格 GIS 空间数据交互研究[J]. 计算机应用研究, 2004 (1): 44-46.
- [3] 蔡少华,骆剑承等. 网格 GIS 中的 GML 语言技术与设计框架[J]. 地球信息科学, 2003, 9(3): 47-49.
- [4] 李丙辰,徐志伟. GSML 网格编程语言的一种实现方法[J]. 计算机研究与发展, 2003, 40(12): 1717-1718.
- [5] 徐志伟,李伟. 织女星网格的体系结构研究[J]. 计算机研究与发展, 2002,39(8): 927-928.
- [6] 丘广华,张文敏. XML 编程实例教程[M]. 北京:科学出版社, 2004.

(责任编辑 黄永明)