

基于网络环境的地理信息系统整合与知识发现*

赵霏生 杨崇俊 刘冬林

(国家遥感应用工程技术研究中心, 北京 100101)

摘要 网络技术,尤其是 Internet 的迅速发展为广泛的空间信息共享、综合和知识发现提供了新的机遇。为了适应网络的分布式计算环境,地理信息系统必须走开放式道路。因此传统的地理信息系统软件必须在体系结构、数据存储模式上作出变革。本文从应用整合和数据整合入手,提出了在分布式计算平台上以空间信息显示、服务、获取、存储为基础的层次化 GIS 软件框架,以便提供数据分析、共享、知识发现等不同水平的空间信息服务。

关键词 网络 整合 分布式 知识发现

0 引言

地理信息系统(GIS)经过 30 多年的发展已逐步发展成描述和处理空间实体及其相互关系的信息系统。在 GIS 取得巨大发展的同时,其缺点也越来越明显,主要表现在:传统的 GIS 的计算管理模式是集中式的,各个 GIS 系统采用的软件、数据格式、数据存储和数据处理方法有着很大的差异,并且在数据语义上存在着不可调和的矛盾^[1],系统处于一个相对封闭、孤立的状态,不能有效地进行信息交换和共享,难以进行综合的、多维的数据分析,造成了空间分析和计算游离于主流计算之外;另外由于空间数据的特殊性,空间数据及其属性数据难于进行一体化的管理,加大了数据综合与分析的难度。从各种地学专业领域到复杂的企业地理分析,都需要更广泛的空间信息共享和综合及其多维、动态的应用分析,以便对空间实体的综合理解。因此将能够跨越时空障碍的现代信息技术和以计算机网络为依托的分布式处理策略引入到 GIS 中,走开放式的道路,打破传统 GIS 的封闭性,进行数据整合和应用整合,将多源空间信息与各种相关信息无缝连接,即空间信息的共享与综合,将空间计算融入到主流计算之中,多角度、多侧面地展示数据之间潜在的关联(知识),将成为 GIS 今后发展的主要趋势。

随着信息时代的到来,尤其是 Internet 的出现,人们的生活和工作方式发生了根本上的改变。万维

网(WWW)以其方便、简洁友好的用户界面吸引了无数的人,使计算机的应用体系结构发生了改变,成为当前网络时代人们采集、共享、分析数据的最常用的方法。Internet 应用特别适合当今空间数据类型繁多、数据量大、分布广等特点和多维、动态的应用分析需求,而且易与别的应用集成。基于 Internet 环境下的 Web-GIS 为 GIS 提供了一个新的标准的计算平台,使 GIS 在开放性方面向前迈了一大步。

本文首先讨论了 GIS 的应用整合的问题,即如何在 Internet 环境下将空间计算融入到主流计算之中;然后描述了 GIS 的数据整合,即空间数据和属性数据的一体化管理。在文章的最后,给出了一个原型系统,具体探讨了 GIS 如何为知识发现实现应用整合和数据整合。

1 应用整合

GIS 的应用整合就是要打破传统 GIS 封闭性的体系结构,充分利用 Internet 这个开放的、分布的、全球性的信息基础平台,在当前主流的开放式的计算平台上,以 Open GIS 的标准为参考,重构 GIS 软件的体系结构,使之实现以下几点:

(1) 屏蔽硬件、操作系统、网络和数据库的差异,达到不同应用和数据源之间的互操作,提高计算机系统的开放性,强化多源数据的共享与综合,展示不同层面上数据之间潜藏的知识;

(2) 支持构件化的应用开发,复用已有构件,灵

* 本研究受国家“863”计划 306 项目(No. 863-306-ZD02-03-4)资助

收稿日期:1999-11-03

活方便地定制应用,提高应用开发的效率;

(3) 协调网络资源,使之方便地与其他处理和信息技术协同工作;

(4) 可以随着技术的发展不断融入新的技术。

软件的体系结构主要指系统框架,包括软件的成分及其选择、成分间的相互作用和组合模式、系统的环境和功能等^[2]。本着开放、高效、协同、可扩展、可伸缩的整合原则,整个 GIS 体系结构在纵向上采用层次化的结构,即下一层为上一层提供服务。通过空间信息表达、服务、获取和存储 4 个层次实现空间信息的可视化表示、资源和功能的动态组合。同时在每一层次中,采用构件化的开发方法,通过明晰应用和服务的逻辑结构,详细划分功能,合理规划系统构件,提高构件的通用性。在横向上,GIS 作为一个功能服务以标准的接口规范与其他服务对等地无缝连接,协同进行工作。如图 1 所示。

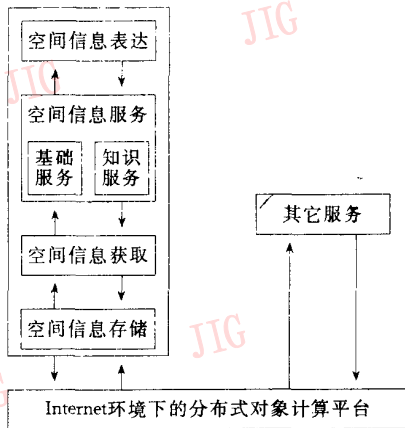


图 1 GIS 软件系统的层次化结构

分布式对象计算平台建立在构件的概念之上,为构件提供各种统一的系统服务,如 Naming、Directory、Transaction 等,屏蔽不同操作系统、数据库间的差异,实现分布式环境下的构件资源管理、通信、动态组合等功能。维护应用系统与构件之间的依赖关系,是 GIS 与其他系统交流、共享与协作的基础。可以形象地说,分布式对象计算平台类似一个软件总线,不同类型的软件构件直接插在总线上;用户可根据应用需要,动态地构成一个逻辑上的应用系统。目前国际上主要有两种关于此平台的规范,一是 OMG 组织的 CORBA (Common Object Request Broker Architecture),另一个是微软公司的 DCOM (Distributed Component Object Model)。

空间信息表达,即空间数据及其相关信息在什么样的环境下以最适宜用户操作和理解的形式展示

给用户。在 Internet 环境下,目前已经有了许多成熟的基于 Web 的应用系统,如 ERP、工作流系统等。因此,将空间信息表达在 Web 页面中,可大大扩展 GIS 的应用范围 and 与其他系统的协同性。当前最常用的客户机浏览器,如 IE、Netscape 等,只支持 GIF、JPEG 格式的图象,为了实现读取矢量图形,间接的一种方式是将矢量地图在服务器上转换成栅格形式下载到客户机上显示,但每次客户对地图的操作都要传回到服务器,等待服务器生成新的栅格图象传回来显示,这样大大增加了网络和服务器的负担。ActiveX 控件、JavaApplet 在页面中的使用扩展了浏览器的功能,不但可以实现直接在客户端读取矢量图形,而且可以以可视化的方式动态展示多维数据间的关联。每次对地图的放大、缩小、漫游等一系列简单操作可不必再求助服务器,大大减小了网络和服务器的负担。ActiveX、JavaApplet 均是以前构件的形式出现的,可以灵活配置,而且还可以回调服务器上其他构件的方法,在很大程度上增强了应用的逻辑性和兼容性。

空间信息服务由一组相互可操作的软件构件集组成,分为基础服务和知识服务。基础服务提供空间信息系统的基本功能,为空间数据提供对象管理、操作、分析、交换等服务设施^[1],为应用服务提供良好的工作基础。在这一层上要着重解决基础服务各构件的定义、结构及其实现方法,力争使各个构件具有原子性和可重用性,做到功能相对完整、接口相对开放和规范。参考 Open GIS 等规范,基础服务主要包括^[3,4]:特征表示、空间坐标转换、空间信息标注、图象处理、空间分析、图象集合模型、空间信息符号管理、图象合成、图象理解等构件。知识服务则以各种空间数据挖掘模型为基础,如泛化、聚类、互联等,探索多源数据间的潜在的关系、模式和特征,将空间信息提升为空间知识,主要包括:描述有空间对象的特征知识、区分空间数据的辨别规则、特征互联规则、描述时间变化的推导或评价规则、描述重要结构或聚类的规则等^[5]。

分布式管理解决了空间数据天生具备的空间分布的特征及决策者的管理和地理层次的不同问题。空间信息获取提供了管理分布式空间数据的能力,为在 Internet 环境下的广泛的数据共享和综合提供支持。该层主要包括:

(1) 元数据管理:元数据是关于数据的数据,可以用来表示辨认信息、数据质量信息、世系信息、空间数据表示信息、空间参考信息、应用特征目录信

息、分布信息、元数据参考信息、应用信息等^[6]。通过元数据的元素名和相应的值搜寻特征集确定数据的存在性、适宜性和位置,以获得处理所需的信息,解决数据的异构性问题。

(2) 数据搜索:用户如何在浩瀚的 Internet 信息海洋中找到所需的信息呢?源信息采集是进行网络信息挖掘的重要环节^[7]。数据搜索可通过 Agent 技术来实现。Agent 以标准的 HTTP 协议下载到所漫游的 GIS Server 上,通过查询该 server 上的元数据库从而得知局部数据的分布。随着 Agent 漫游的深入,就会得到越来越多的数据分布情况。

(3) 数据模式管理:将异构数据库中的不同的数据模式统一成用户自己的数据模式。

(4) 数据库连接管理:通过 ODBC 或 JDBC 这一访问数据库的统一界面来连接不同类型的数据库,而且还提供数据库连接池的功能,处理多用户的并发访问。

空间信息存储是指为了适应 Internet 网络环境下的多用户、安全性要求高等特点,将空间数据和属性数据进行一体化管理。这部分将在数据整合中详细介绍。

其它服务主要指其它类型的应用,如字处理、统计等,通过计算平台提供的通讯机制与 GIS 各种服务进行交互,协同完成工作。

2 数据整合

GIS 的数据整合就是要打破空间数据固有的界限,将空间数据与其它各种类型的数据融合在一起,为应用提供统一的数据存取模式,从而为空间数据共享、综合和知识发现提供更大的方便。

目前,许多 GIS 应用系统中的空间数据还是由文件系统来管理的,远远不能适应 Internet 环境下的多用户多平台所要求的多进程、多线程,内存缓冲,快速索引,数据的完整性、一致性、并发控制、安全与恢复、数据的分布存储和联机分析处理等一系列要求,而这些恰恰是目前成熟的商用分布式 DBMS 产品的优势。因此如何将空间数据高效地存储在数据库中,与其它数据进行一体化管理,更好地为知识发现提供服务,是 GIS 能不能进一步发展的关键。

当前大多数的关系型数据库不直接支持空间数据类型,因此需要定义一种数据模式将空间数据以标准的数据类型存入数据库中。该模式主要解决以

下几方面的问题:

(1) 定义良好的表结构,确定表之间的关系。最常用的表有用来存储空间特征信息及其属性信息的特征表、存储空间坐标系的空间参照系表、存储具体几何数据的空间数据表,三表之间通过主键、外键建立关联关系。几何数据可以以具体坐标的方式和二进制的方式存储在数据库中。

(2) 建立高效的索引机制。空间索引可大大加快对空间数据的检索。在建立空间特征表的同时也建立了关于此特征的空间索引。比较成熟的方法是采用固定单元大小和不固定单元大小的四叉树结构;根据空间特征的具体位置,将其划归为不同的单元格内,通过检索单元格来拾取空间特征。

(3) 根据应用的具体情况明确空间计算是以外程序还是以数据库的存储过程出现在系统中,在此基础上建立相关的具体的外部程序和存储过程。

许多 GIS 厂商,如 ESRI、MapInfo 等都推出了将空间数据无缝集成在关系型数据库中的产品,如 SDE、SpatialWare,用来统一管理空间数据和属性数据,大大提高了 GIS 应用系统的效率。目前,在世界各地已经有了很多基于上述解决方案的成功的应用例子。

随着越来越多的人认识到空间数据的重要性,几大数据库厂商已开始同 Open GIS 组织和 GIS 厂商合作改造和扩充自己的关系数据库,引入面向对象的概念,努力发展对象-关系型数据库,以便支持空间数据。在对象-关系型数据库中,通过将各种空间数据类型内置为数据库直接支持的数据类型并实现了空间索引和关于各种空间计算的 SQL 的函数,空间数据不再是个另类,而是与其它数据一样可以通过 SQL 语句进行选择、更新、删除以及计算等。用户在数据库中操作空间数据就如同操作整数、字符一样简便。Oracle 公司刚刚推出的 Oracle8I 就是以 Open GIS 的数据模型为基础,可直接支持模型中的各种空间数据,大大简化了用户的操作,为用户提供了一体化数据存储方案。

3 实例

基于上文所述的 Internet 环境下的 GIS 整合的思想,国家遥感工程中心开发了一套面向二次开发的 Internet GIS 软件开发平台 GeoBeans(如图 2 所示)。整个平台基于当前最先进的 Internet/Intranet 的分布式计算环境,采用了平台无关性的 Java 构件

模型 EJB(Enterprise JavaBeans),具有良好的可移植性,比较容易达到同操作平台的无关性;在数据管

理上达到了空间数据和属性数据的一体化管理。

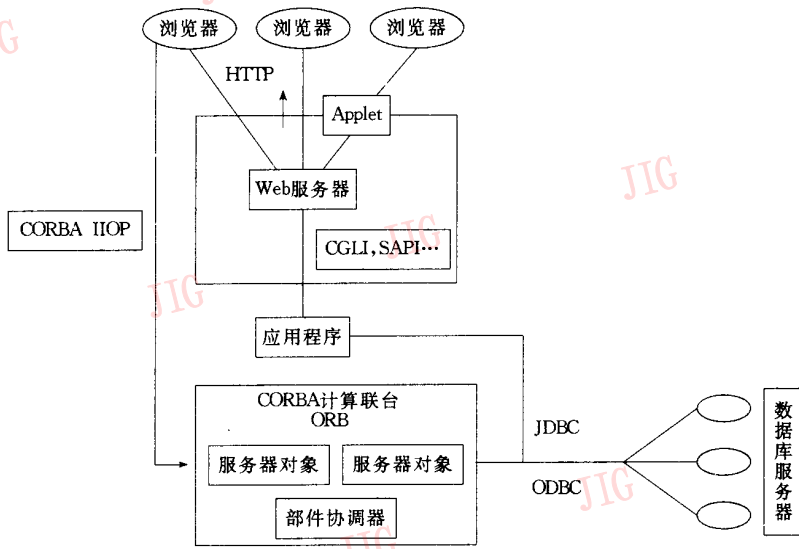


图2 GeoBeans体系结构

GeoBeans 采用了3层客户机/服务器体系结构,分为客户端、中间层、数据层。

客户端:第1层是传统的Web浏览器和以Web为中心的台式机。空间图形显示和可视化分析功能均是由Java Applet来提供。Applet嵌入在HTML页面等可移式容器中。用户通过HTTP下载页面即可使用这些对象,在浏览器中对空间矢量数据进行放大、缩小、漫游、更新、多维统计分析等一系列操作。Applet客户机到服务器对象和服务器对象到客户机的通信采用了IIOP(Internet Inter-ORB Protocol)协议。

中间层:第2层主要为HTTP和CORBA客户机提供各种应用功能服务。这种HTTP/CORBA组合几乎在所有的服务器操作系统平台上都受到支持,包括各种Unix、NT、OS/2、Netware、MacOS等。服务器对象把空间信息服务、空间信息存取中的各种应用逻辑封装起来,在中间层起着应用服务器的作用。这些对象通过IIOP与客户端的空间信息显示对象通信,彼此之间通过CORBA ORB通信。服务器对象的管理主要是通过部件协调器来实现,主要包括预先启动对象库、均衡负载、提供容错功能、协调多个组件运行等。

数据层:第3层主要是利用大型商用数据库存储空间数据,对空间数据及其属性数据进行一体化的管理。GeoBeans采用了目前应用比较广泛的微软公司的SQL SERVER6.5。在进行数据库设计时主

要参考了Open GIS for SQL的规范和ESRI公司的产品SDE的白皮书。具体设计如图3所示^[8,9];

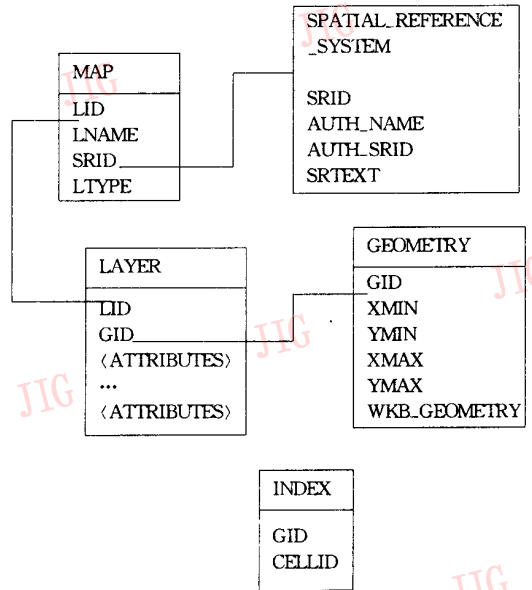


图3 GeoBeans数据库表结构示意图

SPATIAL_REFERENCE_SYSTEM表主要用来存储关于空间坐标系的信息,包括数据库中的坐标系的编号(SRID)、标准的坐标系的名称(AUTH_NAME)和编号(AUTH_SRID)、详细的文字描述(SRTEXT)。该表通过SRID与MAP表建立连接关系。

MAP表用来存放地图构成信息,包括图层编

号(LID)、图层名称(LNAME)、坐标系编号(SRID)、空间数据类型(LTYPE)。该表通过LID与LAYER表建立连接关系。

LAYER表是一个空间特征表,用来存放图层信息,包括图层编号(LID)、几何特征编号(GID)、各类属性信息(<ATTRIBUTES>)。该表通过GID与GEOMETRY表建立连接关系。

GEOMETRY表用来存放空间特征的具体几何信息,包括几何特征编号(GID)、空间范围(XMIN、YMIN、XMAX、YMAX)、空间数据(WKB_GEOMETRY)。其中空间数据是以WKB(Well-known Binary)的二进制形式存放的。

INDEX表存放索引信息,指明空间特征落在哪一个单元格里,包括几何特征编号(GID)、单元格编号(CELLID)。

在GeoBeans的设计和实现过程中,参考了Open GIS中的数据模型、服务模型、元数据、for CORBA、for SQL等一系列标准。整个系统具有良好的开放性、可伸缩性、可扩展性、可协同性、可互操作性、易于使用等特性。

4 结论与展望

Internet的迅速普及正在逐步改变人们传统的

获取信息的方式,从而为全球范围的空间信息共享和综合研究及多维、动态的空间分析提供了机遇。GIS不应再只是一个空间信息系统,而是应该通过Internet这个全球性的信息基础平台,从应用整合和数据整合入手,加强多源数据和各种应用的融合,将其提升为空间知识系统。

参 考 文 献

- 1 黄裕霞,陈常松,何建邦. GIS的互操作. 地理信息世界,1998,4:10~12.
- 2 朱 军,华庆一,郝克刚. 一个基于CORBA的图形用户界面体系结构及实例. 计算机学报,1999,22(1):79~85.
- 3 Esri Whitepaper. The Future GIS on the Internet. ESRI Company, Redlands, California, 1996.
- 4 Open GIS Consortium. OpenGIS Service Architecture. Open GIS Consortium, Inc, Wayland (MA),1998.
- 5 刘同明,刘 伟. 空间数据挖掘技术的研究和发展趋势. 遥感信息,1999,3:2~6.
- 6 Open GIS Consortium. Metadata. Open GIS Consortium, Inc, Wayland (MA),1998.
- 7 邹 涛等. WWW上的信息挖掘技术及实现. 计算机研究与发展,1998,8:1019~1024.
- 8 Open GIS Consortium. OpenGIS for SQL. Open GIS Consortium, Inc, Wayland (MA),1998.
- 9 Esri Whitepaper. Spatial Data Engine. ESRI Company, Redlands, California, 1998.

The Technologies of Conformity for Internet GIS and Knowledge Discovery

Zhao Peisheng, Yang Chongjun and Liu Donglin

(Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing 100101)

Abstract With the fast development of Internet, it is important for GIS to make it more open. The meaning of conformity for GIS is to change software of GIS in system architecture and data access in order to accord with the requirement of spatial data sharing, integrating and knowledge discovering. This paper puts forward a hierarchy for GIS based on distribute computing platform, which includes four layers those are spatial information display, spatial information service, spatial information access, spatial information storage from top to bottom. Every layer is make up of several functional components. The down layer provides service for up layer. To manage spatial and other information universally, this paper also discusses how to put spatial information into large commercial database which supports geometry type or not directly. Based on the idea of conformity for GIS and knowledge discovery, this paper also presents an example called GeoBeans.

Keywords Network, Conformity, Distribute, Knowledge discovery