

# 地理元数据技术系统的设计与实现\*

沈体雁 程承旗

(北京大学城市与环境学系,北京市海淀区,100871)

**摘要** 介绍了地理元数据的概念、类型、标准与应用模式,以吉林省农业信息网元数据系统为例,探讨了元数据技术系统的实现方法,并提出地理元数据库设计、元数据登录系统、检查系统和查询系统等关键技术问题的解决方案。

**关键词** 地理元数据;国际互联网;万维网;分布式地理数据管理;农业信息网

**分类号** P208

元数据(Metadata),即数据的数据,是关于数据和信息资源的描述性信息。地理空间元数据是关于地理空间相关数据和信息资源的描述性信息。它通过对地理空间数据的内容、质量、条件和其他特征进行描述与说明,帮助和促进人们有效地定位、评价、比较、获取和使用地理相关数据。其中对空间数据某一特征的描述,称为一个地理元数据项。地理元数据通常是由若干复杂或简单的元数据项组成的集合。它与非空间元数据的主要区别在于其内容包含大量与空间位置有关的描述性信息。作为对事物背景信息进行描述的思想方法,元数据技术早就广为所用,如图书卡片就是关于图书馆所有书籍的简单元数据,记录了每本书的编号、题目、作者、关键字和出版日期等信息。随着 Internet 和 Web 的迅速发展,元数据技术逐渐成为分布式信息计算的核心技术之一,受到广泛重视。在基于 Web 的 GIS 解决方案中,地理元数据已经从一种数据描述与索引的方法扩展到包括数据发现、数据转换、数据管理和数据使用的整个网络信息过程中不可或缺的工具和方法之一(David,1996),是数字地球和空间信息基础设施的核心内容(Gore,1998)。

地理元数据技术系统是用以支持对地理元数据进行获取、检查、存储、处理、流转和应用的软件系统。一般包括工具系统和可运行系统两类。目前,已经有一些厂商开始提供商业化的元数据工具软件,以支持用户建立自己的元数据库和元数据管理运行系统。但由于地理元数据的自动提取、知识挖掘、内容标准的制定与选择、不同标准体系的地理元数据的相互转换、元数据与空间数

据库的动态一致性维护、版本控制以及基于地理元数据的过程描述与分布式地理数据模型等一系列关键技术问题还没有很好的解决,目前,地理元数据技术系统的发展仍不能满足区域性或行业性分布式地理信息网络建设的需要。本文以吉林省农业信息网元数据技术系统为例,探讨元数据技术系统设计与实现的技术框架和逻辑过程,地理元数据库的存储结构以及元数据获取、检查与查询系统的解决方案。

## 1 系统框架与信息模式

吉林省农业信息网是一个空间型企业级网络,它以 Internet 和 WWW 为技术平台,通过实现吉林省农业相关部门的连接,促进农业信息的采集、分享和使用,为农业生产管理提供决策支持(沈体雁,1997)。它通过通信线路与中国农业信息网、中国经济信息网及国际互联网相连,并采用防火墙与外部网络相隔离。元数据技术系统作为农业信息交换中心的主要构件,为各个部门的信息共享提供了统一的资源定位与搜索机制。其工作模式涉及到 Web 浏览器、Web 服务器、元数据服务器、GIS 服务器和关系数据库服务器等软件部件之间一系列的请求和响应过程,见图 1。一个典型的信息运作过程可以描述如下:①客户(Web 浏览器)通过元数据服务器访问元数据库,并请求查看目标地图的相关属性和略图;②元数据服务器检索该地图的属性,传输到浏览器;③元数据服务器请求 GIS 服务器生成略图,GIS 服务器得到请求后,调出相应的图形数据,抽样并转

收稿日期:1999-04-05。沈体雁,男,28岁,博士生,主要从事 GIS、技术创新与区域发展研究。

\*国家自然科学基金资助项目,编号 79730030。

换成 HTML 支持的图像格式,传输到客户端;  
 ④客户根据元数据记录中的属性描述和略图,决定  
 是否下载该地图数据,并填写订购单,发出下载  
 请求;⑤元数据服务器记录客户下载数据的日志  
 信息,GIS 服务器按客户的要求处理地图数据,并  
 将结果返回客户;⑥在每一个网络信息服务过  
 程中,元数据服务器协调处理并记录客户及其所  
 请求的 GIS 服务器之间的响应关系,不断更新地  
 理元数据记录的版本。

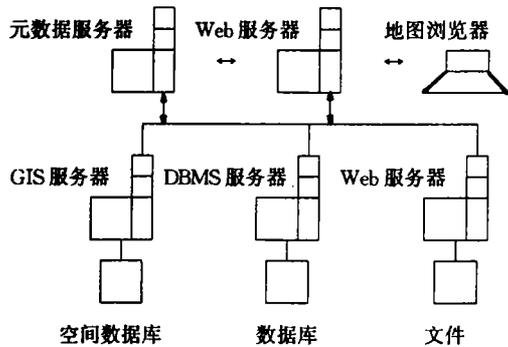


图1 地理元数据技术系统的运行框架

Fig.1 The Framework of Spatial Metadata System

吉林省农业信息网元数据技术系统以省农业  
 厅结构化布线系统为技术平台,以微软 Internet  
 Information Server 2.0 作为 Web 服务器,以微软  
 SQL Server 6.5 作为元数据的数据库管理系统,  
 采用数据库连接器 (IDC)、Java、JavaScript 和  
 Transact SQL 等开发工具,并结合农业厅业务运  
 行模式开发而成(沈体雁,1997)。整个系统分为  
 登录、查询、浏览、维护和综合应用等几个子系  
 统。登录子系统供各个部门进行元数据登录;查  
 询子系统用于网络系统的数据发现与数据定位;  
 浏览子系统提供按权限的数据浏览功能;维护子  
 系统用来对元数据进行删改、备份、索引和版本  
 控制等操作;综合应用子系统则可对数据进行格  
 式转换、下载、收费和业务模型运算等。

元数据技术系统的创建与实现包括以下几个  
 主要过程:①源数据集的标定与需求调查。对所  
 要描述和管理的地理数据依专题和空间范围进  
 行标定与细分,确定哪些数据资料将被收集和  
 描述,调查该数据集内现有元数据的情况、数  
 据集的哪些特征将被描述以及元数据的应用  
 领域、用户情况和使用要求。②选择和制定空  
 间元数据标准。分析和比较现有的空间元数据  
 标准,并根据地理信息工程的实际情况,选择  
 和制定合适的元数据标准。③设计和创建空间  
 元数据的结构。确定元数据的实现方法、创建  
 工具、记录格式和元数据项。如果采用关系数  
 据库管理系统来建立元数据库,则要确定库结  
 构。④获取(登录)和检查元数据记录。按照  
 元数据结构的要求,将空间数据源文件逐一登  
 录到元数据库中,并按照严格的数据规范进行  
 质量控制与版本控制。⑤发布并维护空间元  
 数据库。⑥对空间元数据和数据进行查询、搜  
 索、预览、显示、下载、转换及其他应用。

## 2 元数据库设计

选择地理元数据存储策略需综合考虑网络信  
 息系统的特点和需要、所涉及的地理元数据类  
 型以及所参照的元数据标准等要素。存储策略  
 的优化可减小系统开发难度,提高系统整体运  
 行效率。

根据描述对象的不同,地理元数据可以分成  
 数据库级元数据、要素/记录级元数据、属性元  
 数据、数据文件元数据、数据集元数据等几种  
 类型。对于不同类型的元数据,由于它们与相  
 应空间数据的相对关系不同,可以选择不同的  
 存储策略。一种策略是采取分散的方式,另一  
 种则采取集中管理的方式,见表 1。

地理元数据标准是为了便于不同信息系统之  
 间相互交换和共享元数据而制定的可供参考  
 和遵

表 1 元数据与数据的相对关系

Tab.1 Relation Between Metadata and Geographical Data in Storage

	分散方式	集中方式
特点	以数据集为基础,每一个数据集对应地有一个元数据文档。	以数据库为基础,即一个数据库只有一个元数据文档,文档中的每一个记录对应于一个数据集。
图形示意		
优点	元数据和数据的关联性较好	管理方便,增删和修改容易
缺点	元数据文档太多,不便于管理	必须有一种技术机制进行统一管理

循的地理元数据内容和范式。目前国际上著名的地理元数据标准有美国联邦地理数据委员会(FGDC)于1994年7月发布的“数字地理空间元数据内容标准(CSDGM)”；国际标准化组织第三工作组于1996年开始发布和更新的ISO/TC211元数据标准等等。其中,FGDC的标准包括标识信息、数据质量信息、空间数据组织信息、空间参考信息、实体和属性信息、数据分配信息和元数据参考信息7个部分,是美国国家信息基础设施(NII)中标定空间数据源、提供数据访问的强制性国家标准,得到了广泛的认同和遵循。显然,参照不同的标准,地理元数据的内容和库结构也会大不相同。

就记录形式而言,元数据可以采用非结构化的文本记录方式和结构化的数据库记录方式。前者所要求的计算机资源较少,适用于记载数据项不复杂、使用不频繁的元数据;后者一般需要以关系数据库管理系统为基础,查询、检索和使用元数据的效率较高。Arc/Info提供了一个简单的元数据处理系统 Document,它就是以文件的形式存储和管理地理元数据的。其中,DOC文件记录 Coverage的一般描述性信息,如形成时间、最新更新时间、主题名称等;REF文件记录数据源引用的数据;NAR文件可以存放数据用户认为有价值的描述性信息,如数据可能的用途、数据摘要等;ATT文件用来描述数据表中的对象特征数据和单个数据项;位置属性文件存放源、排序和数据分布信息。

本研究采用结构化的关系数据库管理系统 Microsoft SQL Server 6.5 管理和维护元数据库。与文本方式等非结构化的记录方式相比,结构化的存储策略有利于充分发挥分布式关系数据库系统的数据输入、查询、完整性与一致性维护、多线程与多任务机制,有利于实现元数据技术系统与GIS的整合。实现这种策略的关键在于实现关系数据库与Web的连接,并开发出面向特定数据集和特定事务处理目标地理元数据应用模式。

## 3 软件的实现

### 3.1 获取系统

根据地理元数据的形成机制,其获取方式通常有:①手工生产和录入。由数据生产者、数据管理员或数据用户用元数据技术系统支持的手工方法,直接填写一系列不能由设备或软件自动产生的元数据项;②数据生产系统或设备自动记录。如卫星

拍摄影像数据时自动记录轨道参数、时间等;③软件系统自动发掘。GIS软件或其他应用软件在使用和处理数据时,可以通过关联、量测、计算、推理、访问等方法自动获取并记录各类元数据。

本研究采用HTML的FORM元素来收集用户输入的元数据项。为了使Metadata的登录界面更为直观、简捷,根据农业数据的空间特性和元数据项的“NULL”属性把元数据的输入分成4种类型:①非空间短字节输入型。对简单文本、声音、录像等空间参考性不强的数据类型可以不输入描述空间参考的元数据项,只输入非“NULL”型的字段。②非空间长字节输入型。提供对非空间数据的完整输入。③空间短字节输入型。对图形、图像和关系数据库等空间参考性强的数据类型输入必要的描述空间参考的元数据项。④空间长字节输入型。提供对空间数据的完整输入。用户可以根据所要登录的数据类别选择使用上述输入类型,从而减少页面上不必要的FORM输入框和数据输入量。每一输入类型做成一个页面。页面之间通过JavaScript语言进行数据交换,并以HTML3.0提供的FRAME框架的形式集成在Web浏览器窗口之中。

由于地理元数据通常具有很多内容项,特别是在参照国际标准时更是如此,元数据登录系统的界面设计往往至关重要。通过精巧的程序设计,使录入工作量尽可能小及用户化,同时尽可能遵循标准,满足数据共享和信息服务的需要,从而保证系统的可操作性。

### 3.2 检查系统

检查系统的功能在于对用户登录进库的元数据记录进行规范性检查和过滤,保证地理元数据的有效性。其职能主要有以下3个方面。

1)数据完整性与一致性检查。由于元数据系统是一个分布式的多用户操作平台,所以必须保证元数据在并发处理过程中的数据完整性与一致性,这包括实体完整性、域完整性、参考完整性、用户定义的完整性和分布式数据完整性等5个方面的要求。本研究主要采用3种策略保证地理元数据的完整性:①充分利用SQL Server数据库管理系统所提供的数据完整性功能;②按照吉林省农业厅的信息业务规则,使用JavaScript语言对用户输入的元数据项进行一致性检查;③利用IDC脚本中的SQL存储过程,见表2。

2)数据安全性控制。元数据系统的安全保密控制主要通过Windows NT 4.0操作系统的安全机制、SQL Server的安全机制、数据库连接器脚

表2 元数据完整性与一致性检查策略

Tab.2 Strategies in Checking Integrity and Consistency of Metadata

数据完整性的5个方面	含 义	策 略
实体完整性	保持一个 Metadata 记录的惟一性	SQL Server 数据库的主键定义;用户标识码元数据项的定义
域完整性	保证一个 Metadata 字段的合法输入值	SQL Server 限制一个列的数据类型、格式及取值范围;FORM 输入元素的值域定义;JavaScript 条件判断
参考完整性	保证 Metadata 库的多表之间的数据一致性	SQL Server 的外键定义;以触发器和存储过程定义的限制条件
用户定义完整性	保证 Metadata 系统符合吉林省农业厅的业务规则	Java 和 JavaScript 编程
分布式数据完整性	跨数据库、跨服务器或分片数据的准确性与一致性	SQL Server 的高速锁定与并发处理机制

本文件中的用户名与保密字实现。

3)元数据版本控制。元数据记录在时间维上的一致性维护与版本控制是元数据系统检查与维护的一个难点。这里采取对“登录时间”元数据字段做标记和进行排序的方法及实现登录库与查询库分离的策略解决这个问题。当用户从浏览器提交元数据时,这些元数据记录首先被存储在登录库里,经过版本检查后,从中抽取一个现时、惟一和完整的元数据记录转存到查询库里,这个查询库就是元数据查询系统的操作对象。

### 3.3 查询系统

元数据查询系统包括对元数据记录的查询及对查询得到的元数据记录所描述数据的预览和查看。前者主要通过 IDC 脚本中的 SQL 语句来实现,后者则需要针对不同的数据类型以不同的方式完成。若是 Web 浏览器支持的文件类型,如 HTML 文本、GIF 图像、JPEG 图像等,可直接由浏览器进行解释和显示;对于 Web 浏览器不直接支持的文件类型,如各种格式的声音、动画和图形文件,则需要相应的插件、控件或 Java 小程序读取和播放。元数据查询系统提供一个文件关联器或对照表,以建立文件类型与浏览工具之间的对应关系,从而使用户以点击的方式就能查看元数据记录中所描述的数据。

查询系统包括3种查询方式:①按图形方式搜索元数据记录。包括在索引图上按矩形区域、圆形区域、任意多边形区域和以键盘方式输入一定坐标范围,检索分布在该区域范围内的地理实体所代表的 Web 信息站点的数据情况。②按元数据库字段查询元数据记录。包括按数据标题、用户标识码、关键字、摘要及所有字段查询网络数据的情况。③对文本文件的全文检索。

下面是按关键字查询元数据记录的数据库连

接器(IDC)脚本文件:

```
Datasource: meta
Username: sa
Password: xxxxxx
Template: query. htx
Required Parameters: User. ID
SQL Statement:
+ SELECT User. ID, Originator, Title,
  Abstract, Keyword, URL
+ FROM meta. dbo. meta ID
+ WHERE Keyword cn % keyword %
```

其中,Datasource 项定义 ODBC 数据源为“Meta”,它必须设定为 ODBC 的有效登录,IDC 才能正常工作;Username 定义用户帐号名为“sa”;Password 设定用户保密字;Template 项表明查询结果将填充到一个叫 query. htx 的模板文件,并生成格式化的 HTML 文档;Required Parameters 表示 IDC 将确保必须接收到从客户传过来的参数“User. ID”才能开始查询;SQL Statement 段是执行查询功能的 SQL 语句。

## 4 结 论

本文从概念、类型、标准、应用及技术系统的实现方法等方面,系统地介绍了地理元数据技术体系及其在吉林省农业信息网建设工程中的应用。实践表明,地理元数据技术系统以较低的开销,较好地解决了万维网环境下异型地理数据的综合管理问题,提供了一种企业级地理信息网络运行控制的解决方案。它的优点在于:①与建立数据目录的数据索引方法相比,具有良好的数据一致性、完整性和可扩展性;②支持良好的数据质量控制、版本控制和安全控制策略;③更为重要的

是,这种结构化方法支持标准的 SQL 查询和有效的网上数据搜索与定位机制,使万维网上杂乱无章的海量空间数据成为可理解、可应用的知识。

### 参 考 文 献

- 1 沈体雁. 基于 Metadata 的 WebGIS 数据管理模型及其在吉林省农业信息网的应用:[学位论文]. 北京: 北京大学, 1997
- 2 沈体雁, 程承旗, 袁文. 基于空间元数据的分布式地

- 理数据管理模型及应用研究. 测绘通报, 1999(7): 34~37
- 3 夏福祥, 曹长勇. 万维网地理信息系统. Geographic Information Sciences, 1995(9)
- 4 Medyckyj-scott D. Discovering Environmental Data: Metadatabase, Network Information Resource Tool and GENIE System. I. J. GIS, 1996, 10(1): 65~84
- 5 Gore A. Understanding Our Planet in 21st Century. The Australia Surveyor, 1998, 43(2): 89~91

## Design and Realization of Spatial Metadata System

Shen Tiyan Cheng Chengqi

(Department of Urban & Environmental Sciences, Peking University, Beijing, China, 100871)

**Abstract** Geographical metadata technologies are key to the integrated management and application of distributed heterogeneous geo-spatial data in Internet/Intranet. This paper discusses conception, classification, standards and applications of geographical metadata, then puts forward some solutions in developing a geographical metadata technical system(GMTS), including designing of geographical metadata base and realization of tool software. As a case, a GMTS of Jilin province agriculture information network(JPAIN) is introduced. Geographical metadata is descriptive information about geographical referenced data. It usually describes content, quality, conditions and other characteristics of spatial data set. In Internet/Intranet based GIS applications, geographical metadata is not only an approach to describe and index data, but also a methodology to support building the operational model of data flow in network system, including data mining, data transforming, data management and data using. So it is necessary to develop powerful tool software in order to capture, check, store, process and use geographical data more effectively. JPAIN is a spatial Intranet, based on which agriculture data is captured, shared and applied in Jilin, a province in northeast China. As a main component of clearinghouse, GMTS provides a mechanism of uniformed resources locating and searching. It works by a serial of request-response processes between WWW Browser, WWW Server, metadata Server, GIS Server and database management system Server. Here is a classical scenario in which a consumer orders some maps by Internet. Step one, he or she (by WWW Browser) accesses metadata base and sends a request of viewing attributes and delineation of the maps. Step two, metadata Server queries the metadata base and sends the result to the consumer. Step three, metadata Server requests GIS Server to create delineation from spatial database and transfer them to Browser with image formats directly supported by HTTP. Step four, the consumer decides whether to buy these maps according to attribute information or not, then fills order forms and puts forward download request. Step five, GIS Server deals with data as the consumer requests, at the same time metadata Server records this process into log file, and all the results are transferred to the consumer. Last step, after an information service process is finished, metadata version is updated by metadata Server. GMTS consists of such components as input subsystem, checking subsystem, searching subsystem, maintaining subsystem and comprehensive application subsystem. Here we used Microsoft Internet Information Server 2.0 as WWW Server, SQL Server 6.5 as metadata base management system, and Internet database connector

(下转第 350 页)