

空间元数据管理与发布技术研究^{*}

高 睿¹, 刘修国²

(1. 武汉大学 土木建筑工程学院, 湖北 武汉 430072; 2. 中国地质大学 信息工程学院, 湖北 武汉 430074)

摘 要: 随着信息技术的飞速发展,元数据正成为地理空间信息资源共享的关键。按照 ISO/TC211 的 19115 空间元数据标准,简要介绍了空间元数据的元素组成和分类应用,提出了空间元数据建库的一些基本原则,给出了一个基于 XML 和 Z39.50 协议的元数据管理与共享系统设计方案,并对其中的关键技术提出了一些实现策略与建议。

关键词: 元数据; 标准; Z39.50; XML

中图法分类号: TP335 文献标识码: A 文章编号: 1001-3695(2005)09-0022-03

System Design for Managing and Sharing of Geo-spatial Metadata

GAO Rui¹, LIU Xiu-guo²

(1. Faculty of Civil Engineering, Wuhan University, Wuhan Hubei 430072, China; 2. Faculty of Information Engineering, CUG, Wuhan Hubei 430074, China)

Abstract: Metadata is becoming the absolutely essential solution for the sharing of geo-spatial data with the rapid development of information technologies. This paper introduce the contents and applications of metadata according to the ISO's geo-spatial metadata standard, and proposes some principles on how to create a meta-database, then emphasizes on a framework for managing and sharing of geo-spatial metadata based on XML and Z39.50 protocol with detailed description of two key components and their implementation strategies.

Key words: Metadata; Standard; Z39.50; XML

空间信息在人类生活中无时不有,无处不在。随着地理信息系统的出现,人们逐步具备了通过计算机对空间信息进行存储管理和分析使用的能力。但面对网络中分布广泛的海量空间信息,如何查找所需要的空间数据;如何使用这些空间数据等问题也接踵而至,空间元数据是解决这一问题的最好途径。空间元数据是对空间数据进行描述的数据,它以结构化的形式描述了地理数据集的内容、质量、表示方式、空间参考、管理方式以及数据集的其他特征。它不仅可以提供对空间信息数据的搜索、导航,而且便于数据的转换、维护、理解和使用。它是实现地理空间信息共享的基础,是数字地球的重要技术支撑条件之一。建立一套空间元数据管理系统,提供对元数据的采集、管理、维护与网络发布,是元数据项目实施的当务之急。目前已经开发的一些系统,往往以文本方式提供空间元数据的管理,在数据的组织上也缺乏人性化的设计,需要用户确定搜索策略。本文结合项目实践,通过对空间元数据标准的研究与分析,总结了空间元数据的内容组成与组织表达形式,探讨了空间元数据库的建库原则和方法。

1 空间元数据标准

元数据由描述数据特征的元素组成。空间元数据标准一般都包括这么几个部分:表示方法的说明、构成元数据的元素

分类及定义(元数据字典)、按层次关系的元素组织(元数据模式)、元数据扩展方法与专用标准定义。

《数字地理空间元数据内容标准, CSDGM》是美国联邦地理数据委员会(FGDC)制定的空间元数据标准,由 220 多个元素组成。在内容组织上分为部分、复合元素和数据元素三个层次。国际标准化组织(ISO)的地理信息/地球信息技术委员会(即 ISO/TC211)在 CSDGM 的基础上制定了 ISO 空间元数据标准(ISO DIS19115.3-标准草案 2001.8)。该标准定义了元数据模式,并用 UML 包(Package)的概念来表示地理空间元数据。每个包可以有一个或多个实体(Entity),实体又包含离散的元素(Element),可以是要素(Feature)和属性(Attribute)。实体与实体之间可以互相关联,自身还可以多重聚合和重复(图 1)。该标准定义了 14 个元数据包,分别是元数据实体集信息、标志信息、资源限制信息、数据质量信息、维护信息、空间表示信息、参照系统信息、内容信息、描绘目录信息、发行信息、元数据扩展信息、应用模式信息、扩展信息、引用与负责机构信息。在元数据字典中,给出了每个元数据实体和元素的角色名称、短名称、定义、限制条件、最大出现次数、数据类型以及取值域信息。另外,按照所描述数据内容的差异,ISO DIS19115 把空间元数据分为数据集系列空间元数据、数据集空间元数据、要素类型和要素实例空间元数据、属性类型和属性实例空间元数据四个层次,分别用于描述数据集聚集、独立的数据集、单独的地理要素,以及构成一个地理要素的不同对象类。数据集空间元数据是整个空间元数据标准的核心。

2 空间元数据建库

在对具体的空间数据应用建立元数据库时,应当有一些原则性的考虑,以保证建成的元数据库不仅满足当前应用的需要,而且能够较好地适应与其他相关系统进行资源共享和将来升级扩展的需要。下面是本文认为应当考虑的一些基本原则:

- (1) 先进性原则。元数据库的设计和建设应当尽量借鉴国际国内先进的元数据标准。ISO 19115 是目前最新最具权威的空间元数据标准,元数据的设计应当遵循这个标准或根据它制定的专用标准,使元数据库具有良好的互操作性和兼容性。
- (2) 管理高效原则。元数据存储可以采用文件系统或者商用数据库系统,但元数据库的组织要便于提高元数据的存储、访问效率。使用文件系统组织存储元数据文档简单,但不便于查询和安全管理,一般用于小型或试验性系统;目前较多利用大型商用关系型数据库如 Oracle 或 SQL Server,这样可以充分利用 DBMS 的数据管理能力,但在解决结构化元数据与关系型数据库模式之间的相互映射上需要花费一定的代价。
- (3) 可扩展性原则。由于系统的持续性,元数据拥有者必然会对原有的元数据添加一些新的元素和数据子集,因此元数据库必须具有良好的可扩展性,以适应系统发展的需要。
- (4) 分层分级原则。空间元数据是高度结构化的复杂数据,元数据的组织应当根据不同层次的应用需要和描述的详细程度进行分层分级。元数据可能偏重于管理,或偏重于描述,也可能两者兼顾;另外,一级元数据集中了最基本、最主要的实体和元素,用于了解数据集的总体,而二级元数据可以按不同的数据类别详细描述数据集信息。

3 空间元数据管理与共享框架

空间元数据标准的制定为空间元数据的实施规定了操作对象的内容、组织形式,以及建模技术和方法。空间元数据实施的基本目的就是要使用户通过客户端工具实现对分布的空间元数据库进行查询、检索访问,根据查询的元数据结果可以定位和检索相应的空间数据。数据拥有者和管理员可以通过元数据库管理系统实现对空间元数据库的管理和维护。一个典型的空间元数据管理与共享系统的基本框架如图 2 所示。

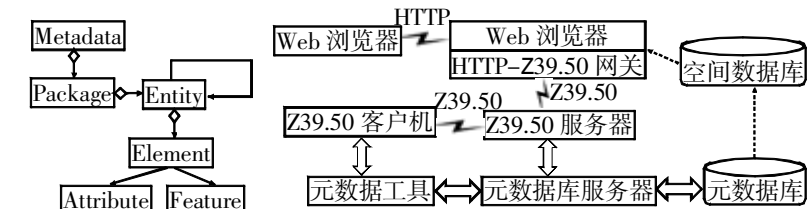


图 1 ISO19115.3 元数据内容组成结构图

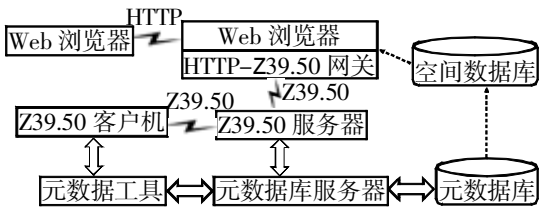


图 2 空间元数据管理与共享系统基本框架

在框架中,Z39.50 客户机/服务器是 Z39.50 信息检索协议的实现,它提供了对分布的多数据库进行一致访问的机制。HTTP-Z39.50 网关负责 HTTP 协议与 Z39.50 协议之间的通信和数据交换。元数据库服务器针对具体的空间元数据库,提供进行数据访问、存储管理、安全管理、系统维护等服务功能。

元数据与空间数据库之间的连接可以是离线或在线的形式,可能处于安全保密的原因,空间数据并不对外发布,但元数据指明了获取空间数据的方式。

4 关键问题与解决方案

4.1 元数据的关系化存储

空间元数据管理系统的一个关键问题是存储。XML 技术所具备的一些优点,如(半)结构化的数据表达、语义明确、丰富的查询支持等,使它很自然地成为解决元数据表达与存储的第一选择。在建立空间元数据库时,一般首先根据用 XML Schema 表示的元数据标准,创建元数据库,然后将 XML 格式的元数据文档导入数据库。这个过程包括两个问题,即元数据模式和元数据文档与数据库的映射。目前已经有多种 XML 与关系型数据库的映射解决方案,但往往导致数据库结构非常复杂,数据存取的效率也不高。因此,笔者提出了一种数据类型驱动的映射方法,具有表结构与数量固定、算法实现简单、较高的存取效率等特点。

XML 文档是由标记和内容所组成的文本文档,标记包括元素和属性,内容则是这些元素和属性的值。XML Schema 是文档有效性验证的机制,符合同一个模式的所有 XML 文档,应当具有相似的结构。因此,可以把 XML 文档分为结构和内容两个部分,分别用不同的数据库表来存储。为了便于支持数据的比较查询,具有不同数据类型的元素又分别保存在不同的表中;因为 XML 所支持的数据类型是有限的,对应一个模式的数据表也就比较简单。按模式建库时,对模式进行分析,并将其中的元素和属性用 GID 唯一标志,生成模式表。在 XML 文档的映射时,按照广度优先的算法遍历文档树,一方面结合模式表中的标志号,记下所有节点的出现顺序,获取文档结构信息;另一方面取出每个节点的属性值和叶节点的值,根据取值类型的不同存入相应的数据表。这样得到的 XML 文档的存储结构如图 3 所示。

4.2 Z39.50 协议服务查找引擎

Z39.50 协议是一个基于客户机/服务器模型的互连网络协议,在客户机与服务器建立 Z 连接(Z_Association)之后,两者之间可以进行多次通信,客户机可以在上一次查询的结果集中继续请求提取或再查询等服务,直到 Z 连接被关闭。在空间元数据的网络共享发布系统中,一般要实现 Z39.50 协议地理空间元数据专用标准(“GEO”),它要求至少提供 Init, Search 和 Present 三种 Z39.50 服务。

Z39.50 协议服务器在逻辑上可分为两个部分: Z39.50 协议核心服务器,负责与客户机的数据通信、多线程任务处理以及应用协议数据单元处理; 查找服务器,负责在元数据库中的查询检索任务和结果集的管理维护。

结合 Z39.50 协议查找服务器,空间元数据库管理系统中的一个基于 XML 的元数据存取模型示意如图 4 所示。



图 3 XML 文档的关系存储结构关系图

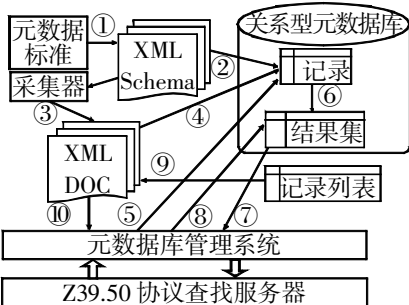


图 4 基于 XML 的元数据存取模型

图 4 中各步骤表示: 将元数据标准用 XML Schema 文件表达; 标准模式文件注册用于数据库建立和数据文档的自适应采集; 采集生成 XML 格式的元数据文档; XML 格式的元数据文档经关系化后入库存储; 提出查询请求; 生成查询结果集(抽象集), 用视图等形式表示; 根据元素集名称、记录语法生成检索结果(如 ESN = “B”, 简单列表记录); 提出检索提取记录的请求; 生成检索结果(如 ESN = “F”, 完整记录); 将检索结果返回请求者。

5 结论

元数据在空间数据处理领域正在发挥越来越重要的作用。元数据的组织、存储、表达、传输、查询、管理等是元数据系统实现时必须考虑的一些问题。按照本文所提出的设计思路开发了一个原型系统, 初步实现了空间元数据的数据库存储管理和网络发布。实际上, 完善的元数据库机制还要考虑很多问题, 如元数据的安全, 对高级别安全的元数据可以考虑加密传输与数字签名服务; 另外, 对分布式数据的检索, 可以运用智能代理技术。因此, 在将来的工作中, 要结合信息技术领域的新进展, 对元数据库机制进一步完善, 并在面向网络的大型地理信息系统中发挥支柱作用。

(上接第 21 页)

[12] adia S K, C S Yeung. A Generalized Model for a Relational Temporal Database[C] . Proceedings of ACM SIGMOD International Conference on Management of Data, 1988. 251-259.

[13] Goodchild M F. Geographical Information Science[J] . International Journal of Geographical Information Systems, 1992, 6(1) : 31-45.

[14] Goodchild M F. Geographical Data Modeling[J] . Computers & Geosciences, 1992, 18(4) : 401-408.

[15] Bishr Y. Overcoming the Semantic and Other Barriers to GIS Interoperability[J] . International Journal of Geographical Information Science, 1998, 12(4) : 299-314.

[16] David J Wilcon, *et al.* Modeling Dynamic Polygon Object in Space and Time: A New Graph-based Technique[J] . Cartography and Geography Information Science, 2000, 27(2) : 153-164.

[17] Allen J F. Maintaining Knowledge about Temporal Intervals[J] . Communications of the ACM, 1983, 26(11) : 832-843

[18] Date C J. An Introduction to Database Systems[M] . The 6th Edition, Massachusetts, Reading: Addison-Wesley, 1990.

[19] Codd E F. The Relational Model for Database Management[M] . Massachusetts, Reading: Addison-Wesley, 1990.

[20] Armstrong M P. Temporality in Spatial Database[C] . Proceedings: GIS/LIS 88, 1988. 880-889

[21] Barrera R, Frank A, Al-Taha K. Temporal Relations in Geographic Information Systems[R] . NCGIA Technical Report, 1991.

[22] Beller A, Giblin T, *et al.* A Temporal GIS Prototype for Global Change Research[C] . Proceedings: GIS/LIS 91, 1991. 752-765.

[23] Frank A U, Mark D M. Languages Issues for GIS[A] . In Geographic Information Systems: Principles and Applications, Maguire D J, Goodchild M F, Rhind D W, (Esses: Longman Scientific & Technical Inc.) [C] . 1991. 147 -163.

[24] Freksa C. Temporal Reasoning Based on Semi-intervals[J] . Artificial Intelligence, 1992, 54: 199-227.

参考文献:

[1] 中国 21 世纪议程管理中心. 中国地理信息元数据标准研究[C] . 北京: 科学出版社, 1999.

[2] FGDC-STD-001-1998. Content Standard for Digital Geospatial Metadata v2[EB/OL] . <http://www.fgdc.gov/>, 1998. 6.

[3] ISO DIS19115 Geographic Information/Geomatics - Metadata[EB/OL] . <http://www.isotc211.org/>.

[4] Information Retrieval (Z39. 50) : Application Service Definition and Protocol Specification[EB/OL] . <http://www.loc.gov/z39.50/agency>, 2001.

[5] Z39. 50 Application Profile for Geo-spatial Metadata (" GEO") . version 2. 2[EB/OL] . <http://www.blueangeltech.com/standards/Geo-Profile>, 2000-05.

[6] Fernandez M; Amer-Yahia S. Techniques for Storing XML[C] . International Conference on Data Engineering, 2002.

[7] Banerjee, *et al.* Oracle8i: The XML Enabled Data Management System[C] . International Conference on Data Engineering, 2000.

作者简介:

高睿(1975-), 男, 湖北鄂州人, 讲师, 博士, 研究方向为 GIS 理论及应用研究; 刘修国(1968-), 男, 河南信阳人, 教授, 博士, 研究方向为 GIS 理论及应用研究。

[25] Jensen C S. Temporal Database Management[EB/OL] . <http://www.cs.auc.dk/~csj/Tesis/>, 2000.

[26] Temporal Database Links[EB/OL] . <http://www.db.informatik.uni-rostock.de/~bg/tdb-res.html>, 2001.

[27] Jensen C S. Introduction to Temporal Database Research Outline [EB/OL] . <http://infolab.usc.edu/csci599/Fall2001/presentations/session16.pdf>, Fall, 2001.

[28] Tang Chang-jie. The Background, Characteristics and Representative Researchers of Temporal Database[J] . Computer Science, 1999, 26(2) .

[29] Tang Chang-jie. The Achievements, Deficiency and Future Work in Temporal Database[J] . Computer Science, 1999, 26(3) .

[30] Gong Jian-ya. An Object-Oriented Spatio-temporal Data Model in GIS [J] . Acta Geodaetica et Cartographica Sinica, 1997, 26(4) : 289-298.

[31] Tang Yong, Tang Na, Ye Xiao-ping. Temporal Information Processing Research Survey[J] . Acta Scientiarum Naturlium Universitatis Sunyatseni, 2003, 42(4) .

[32] Chen jun, Cui bingliang. Use Voronoi to Extend Topology Function for MapInfo[J] . Journal of Wuhan Technical University of Surveying and Mapping, 1997, 22(3) : 195-200.

[33] Liu Ren-yi, Liu Nan. Extension of Spatial-temporal Data Models of Base State with Amendments and Its Implementation in Land Registration Management Systems[J] . Acta Geodaetica et Cartographica Sinica, 2001, 30(2) : 168-172.

[34] LU Feng, LI Xiao-juan, *et al.* Feature-based Temporal-Spatial Data Modeling: State of the Art and Problem Discussion[J] . Journal of Image and Graphics, 2001, 6(A) (9) : 830-835.

作者简介:

黄照强(1973-), 男, 江西赣州人, 博士生, 主要从事遥感与地理信息系统和时空数据模型研究; 冯学智(1953-), 男, 兰州人, 教授, 博士生导师, 主要研究方向为遥感与地理信息系统。