

面向数字城市的空间数据库设计与实现^{*}

李琦 常磊 王凌云

(北京大学数字地球工作室 北京100871)

摘要 空间数据库是数字城市空间数据基础设施的核心技术平台。本文在分析面向数字城市空间数据库的特点及功能定位的基础上,提出其核心内容,包括元数据库、遥感影像数据库、电子地图数据库、GPS数据库、地理编码数据库和社会经济基本单元数据库等。同时,对其数据组织模型和接口体系进行设计,并基于J2EE规范和Oracle Spatial进行了初步实现。

关键词 空间数据库,数字城市,J2EE,Oracle Spatial

The Design and Implementation of Spatial Database for Digital City

LI Qi CHANG Lei WANG Ling-Yun

(The CyberGIS Studio, Peking University, Beijing 100871)

Abstract Spatial Database is the core technique platform of digital city data infrastructure. This paper discusses the characteristic of spatial database for digital city, and figures out it's function should be data share and interoperability. It mainly consists of metadata database, remote sensing image database, electron vector map database, GPS database, Geocoding database and unit database. At last we design and implement Spatial Database based J2EE and Oracle Spatial for digital city. The main technique includes JPEG2000 compress arithmetic based Jasper and JJ2000 source code, and data update schema.

Keywords Spatial database, Digital city, J2EE, Oracle spatial

1 前言

空间数据库是数字城市空间数据基础设施(Urban Spatial Data Infrastructure, USDI)的核心技术平台。在建设空间数据库的过程中,如何对其进行功能定位和内容定位是设计者必须回答的两个基本问题。

从功能上看,作为数字城市的基础性设施,空间数据库的服务对象是社会各个部门,实现基础空间数据共享是一个基本要求。数据共享包括数据交换、数据在线访问和互操作等几个层次。回顾地理信息系统(GIS)的发展历史可以看出,应用系统与数据正在从紧密联系向相对分离发展,但就现状而言,数据共享多数仍然局限于数据交换等低层次。究其原因,与GIS的体系结构和空间数据的组织管理模式有直接的关系。定位在数据共享的哪个层次,将对数字城市的体系结构和数据组织模式产生重要影响。因此,在数字城市的建设中,不但要设计好的体系结构,更要突破传统GIS领域的空间数据管理模式,才能提高数据的共享层次。

从内容上看,空间数据库的管理内容以城市地理空间基础数据为主。数字城市与数字国家、数字地球的尺度不同,其核心框架数据也不尽相同。数字城市核心框架数据模型是空间数据库的设计依据。

本文从数据共享的角度,依据数字城市核心框架数据模型对空间数据库进行了设计,并进行了初步实现,希望能够对各地数字城市建设提供借鉴作用。

2 系统设计

根据功能定位,空间数据库的设计应当考虑互操作。这就要求数据库必须遵循软件规范,对不同的平台和系统都开放。

2.1 总体结构

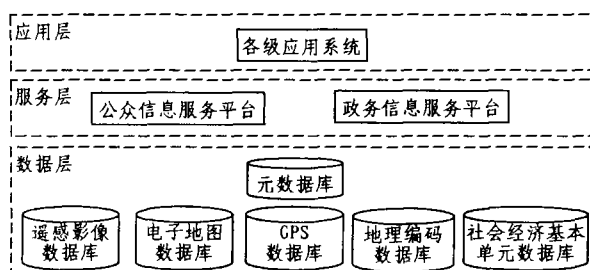


图1 数字城市总体结构图

数字城市的总体结构图如下^[1],分为数据层、服务层和应用层。数据层是空间数据库的建设内容,数据层之上是向社会和政府提供服务的服务平台,服务层之上是包括政府决策支持在内的各级应用系统。根据郝力等人的研究成果,数字城市核心框架数据集包括地理位置基准数据(影像、高程、坐标、水系)和社会经济公用数据(区划、道路、地籍房籍)七层^[2]。数据库设计中,遥感影像数据库包括遥感影像数据和DEM数据。电子地图库包括水系、区划、道路、地籍房籍数据。GPS数据库存储定位信息。地理编码库包括地址及其空间位置,是空间

^{*} 本文得到国家863课题“基于SIG框架的数字城市服务与示范研究”的资助,课题编号2002AA134030。李琦 教授,博士生导师,中国图形图像学会技术委员会主任委员,主要研究领域为空间信息科学、数字地球、数字城市理论等。常磊 硕士生,主要研究内容为空间数据库。王凌云 硕士生,主要研究内容为空间数据库。

数字城市电子地图库逻辑结构

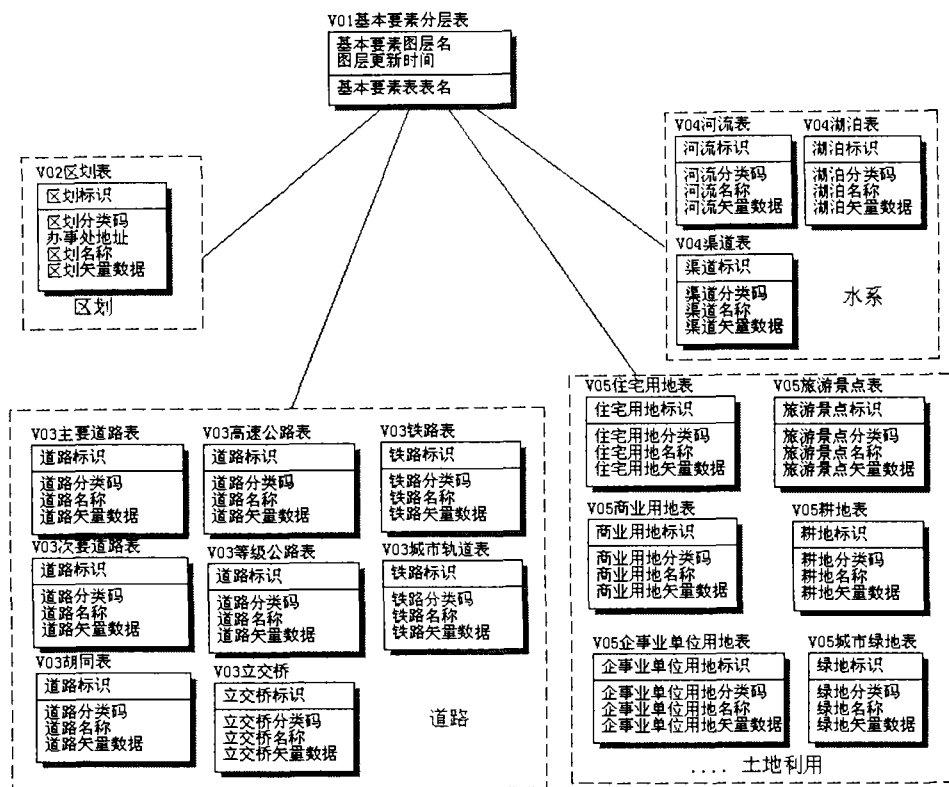


图4 数字城市电子地图数据库逻辑结构图

2.4 数据访问模式

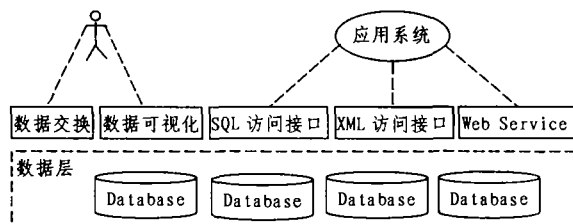


图5 数字城市空间数据库访问模式

空间数据库针对不同用户提供多种访问模式，分别对应数据共享的不同层次。数据交换和数据可视化的服务对象是人，功能是使社会各部门可以获取和浏览城市基础数据。SQL接口、XML接口和Web Service的服务对象是应用系统。其中SQL接口是底层接口，需要开放数据库结构等底层信息，灵活性大；XML接口是互操作的较高级形式，它屏蔽了数据库结构等信息，灵活性小。Web Service是互操作的高级形式，从数据级互操作进入到信息级的互操作。

这几类接口实质构成了数据库访问的若干底层的中间件。这些中间件必须是跨平台、开放式的才能满足这些访问需求。我们采用Java技术，按照J2EE规范进行设计^[5]。这几类接口需要在一定安全机制下实现，这里不再详述。

3 系统实现

数据库采用Oracle 9i Spatial^[6]，数据库设计工具为Erwin 4.0，软件设计工具为Rational Rose。

数据预处理和数据库维护采用Client/Server结构，用Borland C++ 5.0开发。

用于数据交换的元数据查询、在线数据浏览均为

Browse/Server结构，采用java技术开发。影像浏览的中间件自主开发。矢量可视化的中间件采用第三方软件二次开发，目前支持Oracle mapviewer和Mapxtreme for java几类产品。影像与矢量叠加也是自主开发。

互操作接口的体系结构遵循J2EE规范，中间件和各类应用程序之间通过HTTP协议收发XML请求与响应数据包。

限于政策影响，目前数据分发没有采用Internet方式。但是这套体系提供了网上数据分发的实现机制，为今后电子政务的深入应用奠定了基础。

4 关键技术

4.1 数据压缩

高质量的压缩算法一直是遥感影像领域的关键技术之一。基于小波压缩的算法已经在该领域得到广泛应用且效果显著，代表性产品是ECW和Mrsid，这二者都是商业产品，除了价格因素外，它们没有提供数据库访问接口是我们没有采用的主要原因。JPEG2000也采用小波压缩算法，其中Part I已于2000年成为国际标准^[7]。基于JPEG2000的开放源代码有两大类，即基于C语言的Jasper和基于java的jj2000。我们在研究这两类代码的基础上进行改进，应用到我们的软件开发中。就影像管理而言，数据库中存储的是JPEG2000格式的数据(jp2格式)。中间件从数据库中读取数据后有两种解码模式。一是在中间件进行解码，以常用格式发送到客户端；二是直接发送到客户端，由浏览器的Applet或插件进行解码。

4.2 数据更新

数据更新是空间数据库设计的重要内容，也是难点。下面分别从影像和矢量两个方面进行说明。

4.2.1 遥感影像数据更新 包括整体更新和局部更新。

(下转第105页)

Else IF C_1 和 C_2 是 A_i -属性值混合联通,
 Then $\text{ind}(C_1, C_2, A_i) = D(V_1, V_2)$;
 Else IF C_1 和 C_2 是 A_i -属性值上位联通,
 Then $\text{ind}(C_1, C_2, A_i) = D^+(V_1, V_2)$;
 Else IF C_1 和 C_2 是 A_i -属性值下位联通,
 Then $\text{ind}(C_1, C_2, A_i) = D^-(V_1, V_2)$;
 Else IF C_1 和 C_2 是 A_i -属性值序联通,
 Then IF A_i 服从正态分布 $N(m, \sigma)$
 Then $\text{ind}(C_1, C_2, A_i) = 1 - \min(P(|V_1 - m| > 3\sigma), P(|V_2 - m| > 3\sigma)) + \text{常数}(>1)$
 Else $\text{ind}(C_1, C_2, A_i) = \{4, 4\}$
 Else $\text{ind}(C_1, C_2, A_i) = 1$
 $\text{ind}(C_1, C_2) = \text{ind}(C_1, C_2) + \text{ind}(C_1, C_2, A_i)$

现在,分析该算法的复杂性。设 $|C| = m, \text{Max}\{|D_{A_i}| | n \geq i \geq 1\} = k$ 。在最坏情况下,该算法需执行指令数当 $n \geq 3$ 时为 $3n[(k^2)(m-2)+4]/4$, 当 $2 \geq n$ 时为 $2k^2mn$; 在最好情况下,该算法需执行指令数为 $5n$ 。

总结 本文利用本体这一手段重点分析了 A-属性联通极其测度。主要内容是:第2部分讨论了领域本体模型(SOM)和概念本体模型(COM),并给出了命题2.1-命题2.4;第3部分讨论了基于属性取值域的属性划分,将属性划分为结构属性、分类型属性和非结构属性,并证明了命题3.1-命题3.2;第4部分探讨了 A-属性联通以及联通测度。

基于属性的联通还有很多值得进一步讨论的问题。例如两个概念在不同属性上的联通问题;两个概念在两个相同属性上的取值以及取值之间的各种关系等。总之,我们认为基于属性的联通是一类很有意义的联通,可以运用于自然语言查询系统中,提高查询质量。

致谢: 本论文的思想形成于在中科院计算所大规模知识

处理课题组学习期间。

参考文献

- 1 史忠植著. 高级人工智能. 北京: 科学出版社, 1998. 1~14
- 2 雷玉霞, 睦跃飞. 基于 Concept-Relation 模型的知识联通. 计算机科学, 2002, 29(11): 53~58
- 3 雷玉霞, 睦跃飞, 曹宝香. 基于固定概念分类 H 上的概念联通以及联通测度. 计算机科学(录用)
- 4 Lei Yuxia, Cao Cungen, Sui Yuefei. Based-CR Knowledge Interconnection in National Knowledge Infrastructure. ICII'2001, 2001, C: 134~139
- 5 Gallant S I. Connectionist expert systems. Communications of the ACM, 1988, 31: 152~169
- 6 Richardson S D, Dolan W B, Vanderwende L. MindNet: acquiring and structuring semantic information from text. Microsoft Research Technical Publications: [MSR-TR-98-23]. 1998
- 7 Richardson S. Determining similarity and inferring relations in a lexical knowledge base: [PhD. dissertation]. City University of New York, 1997
- 8 Guarino N, Giaretta P. Ontologies and knowledge bases towards a terminological clarification. In: Proc. of KB&KS'95, 1995. 25~32
- 9 Guarino N, Carrara M, Giaretta P. An Ontology of Meta-Level Categories. In: D. J., E. Sandewall, P. Torasso, eds. Principles of Knowledge Representation and Reasoning: Proc. of the Fourth Intl. Conf. (KR94). Morgan Kaufmann, San Mateo, 1994. CA: 270~280
- 10 Guarino N, Welty C. A Formal Ontology of Properties. In: Proc. of 12th Int. Conf. on Knowledge Engineering and Knowledge Management, 2000b
- 11 曹生根. 知识分析中的若干概念与方法: [中科院计算所学术报告]. 2001

(上接第91页)

整体更新指整个批次的数据进行更新。对于成果影像库, 导入新的批次数据, 迁移旧的批次数据到备份介质中, 同时更新总表中批次数据变动情况。对于应用影像库, 预处理后装载新数据, 一般不删除旧数据, 使用户可以在线浏览所有影像数据。

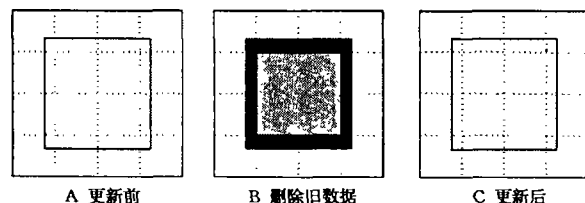


图6 遥感影像数据局部更新

局部更新指少量的标准分幅影像进行更新。对于成果影像库, 一条记录对应一个分幅的影像, 因此只需替换相应的数据即可。对于应用影像库则比较复杂。一个批次的影像构成一个完整的金字塔结构, 金字塔的每一层对应不同分辨率的影像。更新时需要分别更新每一层的影像。以第 i 层影像为例, 首先得到单幅影像相对该层影像的位置, 图 A。然后删除完全落入该范围的影像块, 把其他的影像块重叠部分的像元赋值为零, 见图 B。之后对单幅影像按照该层影像的分辨率进行重采样并分块, 把不重叠的影像块入库, 重叠部分的影像块进行异或运算, 见图 C。其他层的更新以此类似。

4.2.2 矢量数据更新 矢量数据的更新比较复杂, 矢量数据整体更新需要迁移旧的数据, 然后把新的数据按照相同的表结构导入库中。

矢量数据局部更新分为两种情况, 一种是通过客户端维护工具直接修改数据库中的矢量数据, 特点是保持了数据的现势性, 但丢失了历史数据。另一种是复制所要更新的图层

表, 对新表进行编辑修改, 最后把表的变化情况记录到总表中, 该方案的特点是数据的现状和历史都得到保存, 但是增加了数据冗余。

4.3 图例的管理

在传统的 GIS 应用中, 图例的制作管理是很费时的工作。在 Oracle Spatial 中, 图例用数据库管理, 用户可以任意添加自定义图例, 方便了开发和应用。

小结 国内外空间数据库产品也实现了空间数据的有效管理, 但是数据共享层次不够高, 一般不支持第三方厂商的中间件产品访问空间数据, 不能有效实现互操作。本文试图从这一角度回答这一问题。同时还存在很多问题需要研究。例如, JPEG2000 算法除了高压缩比, 还有渐进传输、局部解压以及兴趣点解压等优势没有很好地体现出来, 还需要进一步研究和开发。

参考文献

- 1 李琦. 数字城市服务系统与示范总体框架设计和关键技术研究. 863 技术报告. 基于 SIG 的数字城市体系结构和关键技术研究. 2003
- 2 郝力. 城市空间信息共享应用政策与协议研究. 863 技术报告. 基于 SIG 框架的数字城市服务与示范研究. 2003
- 3 Barclay T, Eberl R, Gray J. The Microsoft TerraServer™. 1998
- 4 李军. 影像数据库的研究、设计和实现: [中国地质大学博士学位论文]. 2000
- 5 Sun Microsystems, Inc. 2000. Java™ 2 Platform Enterprise Edition Specification, 1.4. 2002. 12
- 6 Murray C. Oracle(r) Spatial User's Guide and Reference. Release 9.2. 2002. 3
- 7 Adams M D. The JPEG-2000 Still Image Compression Standard, ISO/IEC JTC 1/SC 29/WG 1. 2002. 12