

文章编号 1000-5013(2004)01-0087-04

GML 的空间信息映射模式

谭玉敏 池天河 唐中实

(中国科学院地理科学与资源研究所, 北京 100101; 清华大学土木工程系, 北京 100084)

摘要 对几种空间数据集成模式进行总结分析, 提出基于 GML 的空间信息映射并用于空间信息集成, 实现空间信息的共享和互操作. 文中以 Shapefile 为例, 说明基于 GML 编码的空间数据信息映射模式. 同时, 基于 .net 平台, 开发了在 SQL Server 2000 数据库中对 GML 文档进行存储和解析的程序实例.

关键词 GML, 信息映射, 互操作, 空间数据集成, Shapefile

中图分类号 TP 311.132.4 TP 391

文献标识码 A

GIS 无论作为一门学科, 还是一项技术, 它关心的应该是描述和表达现实世界的的数据, 而数据是独立于生产和管理数据的软件. “GIS 数据”不应该看作某类特殊类型的数据, 不应该用数据格式来区分. 用于生产、存储和管理其它数据的软件工具, 应该也可以用于 GIS 数据. GIS 软件不应该要求专用的、折衷的 GIS 编程语言, 应该可以通过满足通用工业标准的开发工具来进行. 目前, GIS 平台都采用自己封闭的数据格式, 数据通常以二进制的形式来存储, 整合这些数据是非常困难的. 另一方面, GIS 数据的生产成本普遍很高, 从经济学的角度考虑, 应该“一次生产, 多次利用”. 由于各 GIS 平台都采用自己的数据模型和数据格式, 通过相互间的格式转换不可避免会产生信息损失. 这在一定程度上降低了数据的利用价值. 本文提出了一种基于 GML 的空间信息映射机制, 用于解决空间数据的共享和互操作问题.

1 现有空间数据集成模式

多源空间数据的集成, 大致可以分为数据格式转换、数据互操作、直接数据访问、制定数据转换标准等模式^[1]. (1) 数据格式转换模式. 它是传统 GIS 数据集成方法, 也是目前地理空间数据集成的主要方法. 多年实践证明, 该模式存在一些问题. 由于不同数据格式描述空间对象时采用的数据模型不同, 使得转换后不能完全准确地表达源数据的信息, 难免丢失信息. (2) 直接数据访问模式. 它是指在一个 GIS 软件中, 实现对其它软件数据格式的直接访问. 其典型代表是 Intergraph 的 GeoMedia 及中国超图公司的 SuperMap. 但该模式存在的问题主要是, 必须保证对被访问的数据格式破译完全正确, 而且需要及时跟踪宿主软件数据格式升级的变化. (3) 数据互操作模式. 它是 OGC 制定的规范, 旨在满足用户在异构分布式环境中, 能透明地获取操作所需的地理空间数据. OGC 标准是将计算机软件领域的非空间数据处理标准成功地应用到空间数据上 (ODBC 接口), 为空间数据集中式管理和分布存储与共享提供了操作的依据. 该模式在应用中的局限性, 主要表现在需要每格式的宿主软件都按统一的规范实现数据访问接口, 涉及 GIS 系统底层编码的修改, 在一定时期内还很难实现. (4) 数据转换标准. 很多国家或组织都制定了本国或本领域的统一空间数据格式规范, 如美国的 SDTS 和我国的 VCT 系列格式等. 它们都包括几何坐标、投影、拓扑关系、属性数据、数据字典, 以及栅格和矢量等不同空间数据格式的转换标准. 但目前还很不完善 (以 SDTS 为例). 它不能完全概括空间对象的不同描述方法, 没有为数据的集中和分布式处理提供解决方案. 所有的数据仍需要经过格式转换复制到系统中, 不能自动同步更新. 通过

收稿日期 2003-09-25

作者简介 谭玉敏 (1977-), 女, 博士研究生, 主要从事地理信息共享和 3S 应用的研究. E-mail: tanyym@igsr.cn

基金项目 科技部“十五”国家科技攻关计划重点课题基金资助项目 (2001BA608B)

对上述多种空间数据集成模式的比较研究,我们认为比较理想的是多源数据集成模式.它符合今后网络化数据共享和互操作的发展方向,而且要在短期内有可能实现.鉴于此,我们提出了基于 GML 的空间信息映射模式,用于解决多源空间数据的集成问题.

2 GML 的空间信息映射模式

Internet 环境中空间信息集成问题,可以从空间信息在语法和语义差异的两个不同层次研究.语法上的差异,是指不同的空间信息资源可以采用不同存储格式,而同一类存储格式也可能有版本的差异.语义上的差异,是指不同的空间信息资源可以采用不同的概念体系表示,而同一个概念体系中的概念也可能有同型异义或同义异型的现象.本文主要讨论基于 GML 的空间信息映射模式,及其在空间信息语法结构层次(即空间实体层次)上的数据集成方面的应用. GML 是 OGC 推出的面向 GIS 的 XML 标记集,是一个基于文本的地理特征编码标准.它基于 OGC 创建的公共地理模型(OGC 抽象规范),可以对很复杂的地理实体进行编码.

下面,我们具体介绍以 GML 的信息映射模式,用于空间数据集成.利用基于 GML 的信息映射模式

进行空间数据集成,实际上是对不同格式的空间数据进行 GML 格式转换.即将 GML 作为一种标准的数据转换文件格式来看待,如图 1 所示.其信息映射基本原理是,将复杂符号构造若干一般符号的组合,通过语义转换将复杂符号解析若干一般符号^[2].按形态特征,可将地理实体分为 4 类.(1)点类.具有一定大小、颜色、真北方向的点状符号.(2)线类.各种不同宽度和颜色的实线.(3)填充类(即面类).按一定的填充规则,在范围线内均匀地填充图元、文字、斜线而生成的面状符号.(4)标注类.一般嵌套在其它类中使用.下面分别给出各类地理实体基于 GML 的编码模式.()基本点.有

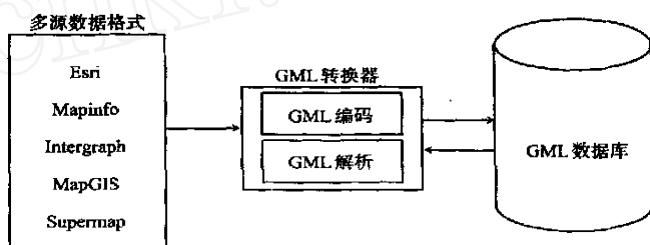


图 1 基于 GML 的信息映射模式

进行空间数据集成,实际上是对不同格式的空间数据进行 GML 格式转换.即将 GML 作为一种标准的数据转换文件格式来看待,如图 1 所示.其信息映射基本原理是,将复杂符号构造若干一般符号的组合,通过语义转换将复杂符号解析若干一般符号^[2].按形态特征,可将地理实体分为 4 类.(1)点类.具有一定大小、颜色、真北方向的点状符号.(2)线类.各种不同宽度和颜色的实线.(3)填充类(即面类).按一定的填充规则,在范围线内均匀地填充图元、文字、斜线而生成的面状符号.(4)标注类.一般嵌套在其它类中使用.下面分别给出各类地理实体基于 GML 的编码模式.()基本点.有

```
< Primitive code = " ... " LayerName = " ... " Color = " ... " LineWidth = " ... ">
```

```
< Coord > xxx. xxx, yyyy. yyy, zzzz. zz </ Coord >
```

```
</ Primitive >
```

()基本线.有

```
< Line code = " ... " LayerName = " ... " Color = " ... " LineWidth = " ... ">
```

```
< Coord > xxx. xxx, yyyy. yyy, zzzz. zz </ Coord >
```

```
< Coord > xxx. xxx, yyyy. yyy, zzzz. zz </ Coord >
```

```
⋮
```

```
< Coord > xxx. xxx, yyyy. yyy, zzzz. zz </ Coord >
```

```
< /Line >
```

()基本面类.有

```
< Area code = " ... " LayerName = " ... " Color = " ... " LineWidth = " ... ">
```

```
< Coord > xxx. xxx, yyyy. yyy, zzzz. zz </ Coord >
```

```
< Coord > xxx. xxx, yyyy. yyy, zzzz. zz </ Coord >
```

```
⋮
```

```
< Coord > xxx. xxx, yyyy. yyy, zzzz. zz </ Coord >
```

```
< /Area >
```

基于 GML 多源空间数据集成的基本思想是,通过将不同格式的空间数据文件中的空间实体基于上述映射基本原理,生成对应的 GML 文档.由于 GML 文档本质上是文本的,因此可以很方便地进行集成.下面以数据格式公开的 Shapefile 为例,说明 GML 的编码模式. Shapefile 是 ESRI 提供的存储地理数据的矢量格式,地图元素以 X, Y 形式出现,其坐标系是笛卡尔坐标.图 2 是 Shapefile 格式的福建某地区行政

区界数据. 图中 Shape 格式数据对应的 GML 编码文档(限于篇幅,模式文件忽略)为

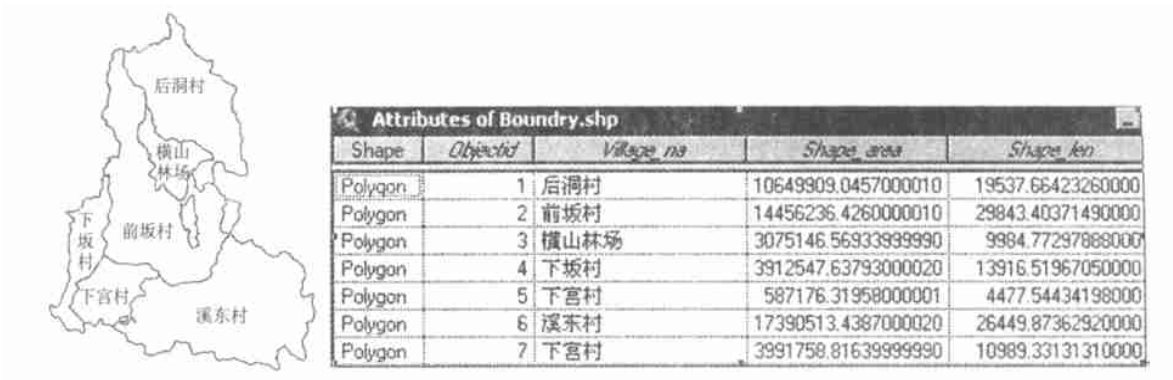


图 2 Shapefile 模式数据实例

```

< ? xml version = " 1.0 " encoding = " UTF-8 " ? >
< dataset xmlns = " f : / namespaces / shp2gml "
  :
< schemaFeatures >
< gml :featureMember >
< Feature >
< featureType > Boundry < / featureType >
< property name = " VILLAGE - NA " > char (20) < / property >
< / Feature >
< / gml :featureMember >
< / schemaFeatures >
< dataFeatures >
< gml :featureMember >
< Feature >
< featureType > Boundry < / featureType >
< property name = " VILLAGE - NA " > 后洞村 < / property >
  :
< property name = " gml2 - coordsys " / >
< gml :polygonProperty >
< gml :Polygon gml :srsName = " 0 ">
< gml :outerBoundaryIs >
< gml :LinearRing >
< gml :coordinates >
576255. 2980114007 ,2680473. 3690133323
  :
< / gml :coordinates >
< / gml :LinearRing >
< / gml :outerBoundaryIs >
< / gml :Polygon >
< / gml :polygonProperty >
< / Feature >
< / gml :featureMember >
< / dataFeatures >

```

</ dataset >

3 GML 文档的数据库存储

GML 文档本质上也是 XML 文档. XML 数据库可分为 3 类^[3]. (1) 本原(自然) XML 数据库(Native XML Database ,NXD). NXD 最基本特点是为 XML 文档定义了(逻辑)模型. 它不同于文档中的数据模型, 可根据该模型进行存储和检索. 它用 XML 文档作为基本的(逻辑)存储单元, 无需特定的物理存储模型. 但不足之处在于没有利用现有 ORDB, 而且无法让用户对 XML 文档和存储在 ORDB 中的数据实施无缝查询. (2) XML 使能数据库(XML Enabled Database ,XEDB). 此类数据库有一个附加的 XML 映射层(通常由数据库生产商或第三方提供), 映射层管理 XML 数据的存储和检索. 对数据的操作可用 XML 方法(如 Xpath, XSLT, DOM 或 SAX), 也可用数据库方法(如 SQL). 数据库中基本的存储单元独立于程序实现. (3) 混合 XML 数据库(Hybrid XML Database ,HXD). 该数据库既可看作是本原 XML 数据库, 也可看作 XML 使能数据库. 取决于应用的需求. 本文采用第 2 种数据库存储模式. 即将 GML 文件及其相关文档以字符串的形式, 存储到 SQL Server 2000 数据库的 Varchar 字段.

4 结束语

本文介绍了地理实体基于 GML 的信息映射规则, 说明通过 GML 进行空间数据集成是切实可行的. 它实现了对 GML 文档进行数据库存储和解析. 但是, 要真正实现基于 GML 的多源空间数据集成和互操作, 需要每种格式的宿主软件都按照统一的 XML 规范实现数据访问接口. 这在一定时期内还不现实. 文中仅考虑了格式公开的数据类型(如 Shapefile, mif 等), 至于其它类型的数据, 留待以后集成其宿主软件公司提供的接口工具. 另外, 对地理实体的表现(通常指符号化)没有提及, 而如何保证空间实体表现形态的一致性(包括数据正确显示和满足 GIS 制图要求), 是多源数据集成不能回避的问题.

参 考 文 献

- 1 宋关福, 钟耳顺, 刘纪远等. 多源空间数据无缝集成研究[J]. 地理科学进展, 2000, (8): 110 ~ 115
- 2 李进强. 基于 .NET 框架的分布式政务空间信息平台的设计与开展[D]. [学位论文]. 北京: 清华大学土木工程系, 2003. 24 ~ 42
- 3 廖述梅, 陶 皖. 对象关系数据库中 XML 文档的存储技术[J]. 计算机与现代化, 2003, (3): 41 ~ 50

GML-Based Spatial Information Reflection

Tan Yumin Chi Tianhe Tang Zhongshi

(Inst. of Geograph. Sci. & Natural Resource Res., CAS, 100101, Beijing, China;

Dept. of Civil Eng., Tsinghua Univ., 100084, Beijing, China)

Abstract By summarizing and analysing several modes of spatial data integration, the authors suggest to use GML-based spatial information mapping in spatial information integration so as to realize sharing and interoperability of spatial information. An example of shape file is given to illustrate the mode of GML encode-based spatial data information mapping. Moreover, a net platform based prototype program about storing GML documents in SQL Server 2000 data base and analysing them is developed.

Keywords GML (graphic machine language), information mapping, interoperability, spatial data integration, shape file