

面向 GIS 数据共享的概念模型设计研究^{*}

陈常松

(中国测绘科学研究院 北京 100039)

何建邦

(中国科学院地理研究所资源与环境信息系统国家重点实验室 北京 100101)

摘 要 GIS 数据共享的主要问题是数据的语义共享。在数据共享的保障基础——技术基础、标准化体系及政策法规体系中,标准化工作最为直接,而其中面向数据共享的概念模型的设计又是最基础性的工作之一。面向数据共享的概念模型设计的主要任务是寻求一种规范化、标准化的约束法则来保障数据的共享,为了保障数据共享的安全性,所设计的概念模型应能提供用于共享的抽象概念,这些抽象概念所描述的是地理概念而非纯空间概念,并能提供语义抽象机制。采用语义数据模型的方法,提出利用地理特征这一概念作为建模的基本概念。通过一个实例对其应用进行了初步探讨。

关键词 数据共享, 语义共享, 概念模型

实际的数据共享是一种在语义层次上的数据共享,远非仅限于数据格式的转换。为 GIS 数据共享所设计的概念模型,其出发点是:(1) 提出一组地理概念,使数据交换、数据互操作等 GIS 共享行为在这些抽象概念基础之上进行,克服传统 GIS 体系对 GIS 空间数据和非空间数据进行分离管理的现状;(2) 使不同的用户对同一数据集具有相同的理解,为不同用户之间数据集的共同使用提供基础;(3) 生成一系列的约束法则,保障安全的 GIS 数据共享行为,特别是保障数据共享过程中语义的正确性。由于其特定的应用目的,面向数据共享的概念模型必须遵守 ISO 为概念模型设计所规定的 100% 原则,即对问题域的结构与动态的描述达 100% 的准确。

1 GIS 数据共享和面向数据共享的概念模型设计

GIS 技术发展至今,地理数据的处理工具——各种 GIS 软件工具已十分成熟,采用这些软件工具建立的数据库及提供的数据产品已遍布全球,因此,现在讨论数据共享问题应立足于上述的基础。从实现上讲,GIS 数据共享存在下列技术途径¹⁾: 第一,数据转换,包括有语义约束的数据格式转换和没有

语义约束的数据格式转换;第二,基于元数据的 GIS 数据网络查询和应用,指在网络环境下在元数据的支持下对地理数据的查询、下载和应用;第三, GIS 互操作,以消息机制为实现基础的数据共享行为,共享的不但包括数据,还包括处理资源;第四,空间数据基础设施(SDI)^[1],SDI 是一种机制,它允许多于 3 个的用户同时对一个数据集进行操作。

面向数据共享模型设计过程中所关注的已不再是数据如何存储及管理,如何为用户提供各种空间操作符及如何建立数据库的问题,焦点转移到各种不同软件工具之数据寻求共同的基础并进而实现共享。从共享的技术途径上看,概念模型设计的主要任务应当是建立一套规范化的抽象概念,这些抽象概念要满足:(1) 保证数据共享行为的完成。现有 GIS 体系是一种将空间数据和非空间数据分离管理的体系,在这种体系中对地理概念的定义,例如对“河流”的定义是通过图形数据“线”和有关这条线的非空间属性共同定义的,尽管在图形数据和非空间数据之间存在语义联系,但这种语义联系的维护要通过用户完成,可见在传统 GIS 体系中的“所见”并不是用户的“所得”,面向数据共享模型中的抽象概念要提供对这种语义联系的维护机制,使得所有对

^{*} 国家“九五”重点攻关项目支持,编号是 97-759-01-01-04 和 97-925-01-01。

1) 黄裕霞,陈常松,何建邦. GIS 的互操作. 中国地理信息系统协会 1998 年年会论文集. 1998, 108-112.

收稿日期: 1998-05-29; 收到修改稿日期: 1999-05-04

该数据集的使用与访问建立在抽象概念之上。(2) 这些抽象概念应当支持所有的GIS软件工具的实现,即要求在所有GIS软件工具中具有实现这些抽象概念的机制。因此,面向数据共享的概念模型是一种高层次的数据模型,具有以下特点:(1) 概念模型中的抽象概念必须提供维护空间数据和非空间数据之间语义联系的机制,同时利用这些抽象概念能方便地形成地理应用。例如“河流”概念是一种独立的地理概念,方便于地理应用的形成,如果在概念模型中具有了描述“河流”概念的空间数据和非空间数据之间语义联系的维护机制,则“河流”概念满足这一条件。(2) 概念模型中必须具有语义抽象能力。由于抽象概念必须在各种GIS软件工具中实现,因此要求面向数据共享的概念模型必须具有语义抽象能力。

可以采用语义数据模型的方法设计这种概念模型,语义数据模型提供语义对象作为构造抽象概念的工具^[2],利用概括、聚集、联合、分类等形成语义抽象概念体系。在设计建立面向数据共享目的的概念模型时,首先要根据前面所述的原则选取合适的空间概念,由对选择的空間概念本身所进行的描述、空间概念之间关系的描述(包括抽象语义关系)等共同构成作为GIS数据共享基础的概念模型。

2 一个初步的概念模型框架

E·Lynn Usery, Robert D·Rugg 和 Agatha Y·Tang 等,提出基于地理特征的方法作为组织地理数据和进行地理应用系统的开发方法^[3-6],其中最主要的是地理特征的概念。定义如下:

地理特征是真实世界实体或对象(数字的和/或图形的)的表达,具有两个方面的内容。第一,是关于现实世界实体集的实例,构成实体集的实体具有一个限定条件,即实体作为一种现实世界的现象,其不可以被进一步分解为同一类实体;第二,是关于现实世界特征的描述。按照面向对象技术的要求,对地理特征的描述必须:

- 具有唯一的标识符;
- 具有一个或多个名称;
- 有一组属性唯一确定该特征;
- 描述时间、空间和非空间属性等。

可以看出地理特征概念在真实地理系统中的表现是组成地理系统的地理实体集,是对地理实体素所进行的规范化描述。从应用的角度讲,地理特征具有应用多维性,所谓应用多维性是指由同一特征模式可以形成不同的应用模式,例如同为“河流”特征,既可以与一定的描述性信息共同构成资源应用模式,也可以构成环境应用模式,地理特征的应用多维性是我们选择它作为面向数据共享概念模型设计基本空间概念的主要原因。

根据前述,将地理特征作为语义对象,根据语义数据模型的一般原理,下面给出一个初步的概念模型框架。

2.1 地理特征的属性描述

根据 Agatha Y·Tang, Tersea M·Adams 等,一个描述地理特征的对象应当包含6个方面的内容,即唯一标识符、位置信息、非空间属性、拓扑关系、非拓扑关系及方法^[6]。这6个方面是对一个地理特征进行描述的6个必需的方面,陈常松曾讨论过一个地理特征的、全面的属性描述问题¹⁾。本文着重解决的问题有:(1) 地理特征空间数据和非空间数据之间的语义联系的维护;(2) 为保证数据共享,地理特征应当具有的概念模式等。

空间数据和非空间数据之间的语义联系的维护

从上一节,一个地理特征的模式可以表达为:

〈唯一标识符,名称,空间属性,非空间属性,⋯〉 (1)

为了独立于比例尺定义一个地理特征的空间属性,采用元组表达¹⁾。

〈比例尺(分母),投影方式,坐标系,空间形态〉 (2)

其中,前三项可取值为标识它们的数值或字符串,空间形态指点、线、面等空间目标,由各种原子空间目标经过聚集生成,每一个空间目标具有一个唯一标识。基于地理特征的GIS语义数据模型中,空间数据和非空间数据之间的语义联系的维护实际上是要维护地理特征与对应空间数据的参照关系,称为use关系。不同的比例尺、投影、坐标系下具有不同的参照关系。

以某一“河流”实体为例,假设这一河流实体由弧段 a_1 , a_2 和 a_3 组成,如果河流的空间参考系统(包括比例尺等)均已知,则在模型中可以省去对它们的描述,在这种情况下,可以将 a_1 , a_2 , a_3 作为地

1) 陈常松. 地理特征的规范化描述及其编码问题. 中国地理信息系统协会, 中国海外地理信息系统协会 1998 年年会论文集, 1998, 192.

理特征“河流”的取值, a_1, a_2, a_3 本身是一语义对象的实例的标识, 代表不具有任何地理意义的、纯粹的空间目标。在 GIS 语义数据模型中作为域对象类型^[7]对待。河流特征的模式如下:

- 地理特征: 河流
- 唯一标识符: 21000
- 标准名称: 河流
- 别名 1: ...
- 别名 2: ...
- 空间属性: line (取值为 a_1, a_2, a_3 的聚集)
- 非空间属性: ...

在具体实现上, 地理特征与其空间表达之间的 use 关系通过特征标识码和空间目标唯一标识码之间的相互参照维护, 在本例中通过 21000 和 a_1, a_2, a_3 之间的参照维护, 空间参考系统、比例尺等作为这种参照关系的语义约束加入模型之中。

非空间属性

根据研究表明, 影响数据共享的因素除了数据质量说明、管理属性说明等外, 还有一个很重要的方面是语义交叉。数据质量说明、管理属性说明等通过两个途径解决, 第一, 通过元数据说明¹⁾; 第二, 通过编码的方法²⁾。语义交叉一般表现为以下几种情况: ①同名异义; ②同义异名; ③一词多义; ④意义交叉等^[8]。可以看出名词在语义交叉中具有关键作用, 其中关于地理特征的名称是最重要的, 因此在设计面向数据共享的概念模型时, 必须考虑名称的处理方法。

美国在制定其转化标准 SDTS^[9]时, 也遇到了相似的问题, 主要表现在所使用的名词术语的同义现象和含义相互覆盖现象。SDTS 所采取的措施是: (1) 通过在各个部门广泛征求意见来消除自然语言含义的不确定性, 最终的特征定义基本保持了定义的唯一性; (2) 一些词汇之间不易分辨的细微差别通过其属性值反映出来; (3) 通过对地理名词实行标准化, 非标准化名词采用“Included Terms”机制解决。

在基于地理特征的概念模型中, 可以采用相类似的方法解决这一问题, 如上述关于河流要素的模式所述。在模型中首先定义地理特征的标准名称,

由于在实际应用中, 应用部门并不总是采用这一标准的名称, 因此, 将各个部门使用的名称作为别名, 加入到模式中, 这样无论在数据交换中, 还是在互操作中, 都可以通过标准名称保证在数据共享中语义的正确性, 而应用部门可以通过别名(这一部门所通常采用的名称)来访问数据, 从而使参与共享的数据库具有继承性。

2.2 地理特征之间的相互参照关系

为了使所设计的概念模型满足 GIS 数据共享的需求, 除了在作为语义对象的地理特征模式中处理诸如语义交叉等问题外, 还要解决地理特征的应用多维性与具体应用之间的关系问题, 地理特征的应用多维性的意义是基于同一地理特征可以形成多种具体的应用数据库。在概念模型设计中, 这种应用多维性表现为两种地理特征之间的相互参照关系, 它们是 use 关系和 reference 关系。

use 关系我们在前面已有所涉及, 其含义是一种语义对象可以成为另一种语义对象的属性取值。图 1 为作为语义对象“井”的模式和另一对象“水”之间的 use 关系表达。

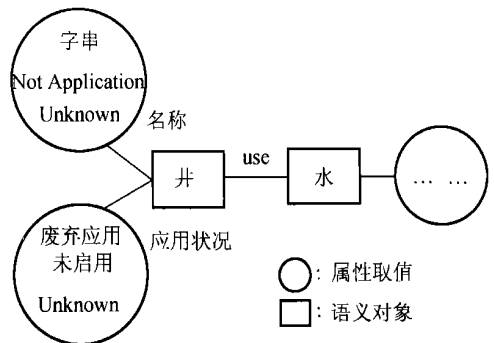


图 1 “井”特征和“水”特征之间的 use 关系

Fig. 1 Use relations between features “Well” and “Water”

实际上, use 关系是对地理特征之间的组成关系的一种在模式构成方面的说明, 它提供了在建立一种新的地理特征模式的过程中对已有地理特征模式的引用机制。例如在图 1 中, 如果“水”特征的模式是已知的, 那么在建立“井”特征模式的过程中, 可以直接引用“水”特征的模式。

1) 刘若梅, 蒋景瞳, 贾云鹏. 国家基础地理信息中心, 中国 21 世纪议程管理中心. 中国可持续发展信息共享(metadata 标准实施. 中国可持续发展信息共享示范 metadata 研究, 标准和实施. 1998. Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

2) 何建邦, 柯正谊, 毛 锋等. 对我国若干资源环境数据库的现状分析及其集成的初步探讨. 中国地理信息系统协会, 中国海外地理信息系统协会 1998 年年会论文集. 1998, 63-70.

在基于地理特征方法支持下的语义数据模型的框架内,语义对象之间的 reference 关系,也就是地理特征之间的 reference 关系。其含义可以被直观地理解为不同应用模式中,地理特征模式的相对稳定性。为了更加清楚地说明其含义,我们通常称这种关系为特征相等关系。特征相等关系存在于不同部门之间的分类体系之间。在文献[8]中我们已给出了两种关于地理特征相等的定义及其判别标准,并给出了一种应用——基于地理特征如何组织资源与环境数据。在已有资源环境数据库之间的GIS数据转换或互操作中,特征相等关系还可以用来解决上述语义交叉问题,主要是名词不一致问题。假如有两个数据库A、B需要共享,如果可以判断出在A、B中,拥有不同名词标识的特征 a 、 b 存在相等的特征模式,则可以认为 a 、 b 代表的是同一种特征,从而可以统一两个数据库中关于 a 、 b 的标准名称,并采用上节对名词的处理方法进一步处理。

2.3 数据抽象

面向数据共享的概念模型必须支持语义抽象,根据语义数据模型所提供的语义抽象工具,利用概括、聚集、联合、分类等表达数据之间的抽象关系^{[10],1)}。需要强调的是,在面向数据共享的GIS概念模型设计中,语义抽象包括下列几方面的内容:

纯空间语义抽象

欧氏空间中,连续的点组成线段或弧段,由线段或弧段组成线状目标,由线状目标组成面状目标,并进而形成更为复杂的几何空间。这一系列的语义抽象过程是高层次GIS概念模型中语义抽象机制的重要组成部分。

利用空间目标和非空间属性形成地理概念

以点、线、面等为原子空间目标所形成的欧氏空间在一定的地球参考系统下,成为关于地球表面的度量空间,要使这一空间成为地理空间,还要建立这些空间目标与地理描述性概念之间的关系,形成地理概念,这部分内容构成语义抽象的一部分内容组成。

地理概念语义抽象

地理宏观概念与微观概念之间的相互关系,宏观概念通过继承可形成微观概念;微观概念通过聚集、联合或概括形成宏观概念,地理概念的分类体系部分地表达了这种关系。

利用地理概念形成资源与环境等应用模式

如文献[8]所述,基于地理特征的数据模型可以形成不同的资源环境应用,从一定意义上讲,也属于一种语义抽象的机制。

一般地,借助于语义数据模拟方法设计概念模型,由于语义数据模型可以提供形式化的语言支持,因此,可以自动形成概念模式,但是由于GIS领域中仍然没有一种成熟的语义模拟方法,也就不可能具有这种形式化的数据描述语言,国际标准化组织ISO及CEN为了寻找一种描述这种概念模型的方法,对已出现的所有数据描述语言进行了全面的对比分析,发现已出现的数据描述语言均不能准确地表达地理概念模型的全部语义,建议暂时采用 express 语言(ISO10303-11, 1994和ISO10303-G)作为数据描述语言。适合于描述GIS数据,并进而支持GIS语义模式生成的GIS数据描述语言仍然是一个重要的研究方向。

3 关于应用的讨论及研究实例

在实现GIS数据共享的4种技术途径中,高层次的概念模型具有控制的作用,本文就概念模型在数据交换、数据互操作方面的作用进行讨论,最后结合我们正在进行的中国资源环境与地区经济信息系统(NREDIS)设计的工作对概念模型在部门之间数据共享的作用进行了讨论。

3.1 数据转换

在高层次的概念模型支持下的数据转换实际上首先是模式的转换。模式转换的含义不只是数据格式的转换,更强调语义的转换。因为概念模式是在一定的语法规则下的语义表达,因此模式转换也就意味着语义的转换。

3.2 GIS互操作

图2是C/S体系结构下的GIS互操作示意图,在这一过程中,由于对一个GIS数据集的请求必须同时涉及空间数据和非空间数据,在将这两种数据分离管理的传统GIS体系中,满足这一请求的处理必须直接建立在这两种数据结构之上,因此很难保证对这一请求的响应不出现意外结果。必须在现有GIS属性数据和空间数据彼此独立管理的体系上增加一层抽象概念,这一层抽象概念能准确地反映地

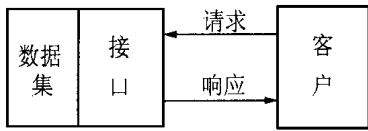


图 2 一个 GIS 互操作的进程
Fig. 2 A GIS interoperability procedure

理现实,同时又利于计算机的实现。我们所设计的概念模型显然是设计这种接口的基础,由于它能提供比矢量、栅格等模型层次高的抽象概念,为抽象概念定义标准模式后,客户端的请求和服务端所做出的响应都将与抽象概念直接打交道。由于这种模型内部提供了语义约束机制。因此服务器发出的响应结果是可预料的。

3.3 NREDIS 设计——基于高层次概念模型的部门之间资源环境信息共享研究

在中国各个部门之间的数据共享研究中,同一地理实体往往具有多种资源与环境应用^[8]。图 3 以

某一河流实体 D 作为例子,整个实体的基本属性存储于 A、B 两个文件之中,从 A 中可知,该条河流由 a_1, a_2, \dots, a_n 条弧段共同构成。分类码 21000 表明这一实体所属的特征为一个河流特征,该特征的标准名称为“河流”。这一特征的基本非空间属性由 B 中给出;包括它的标准名称和长度等,C 是 D 的一个应用,应用码 43030 表明其应用属性是“国内(内河、沿河)航线”,“应用别名”是在应用部门中对其专门称呼(它的标准的名称则在其特征文件 A 中“特征名”中给出),从图 3 中还可看出具有“国内(内河、沿河)航线”应用属性的并不是河流的整个实体,而是一部分(弧段 a_1, a_2),它们之间的 part-of 关系由具体的 GIS 技术加以解决。

实际上,可以形成许多类似于 C 的具体应用,这些具体应用形成之后再通过进一步的聚集,可以形成满足部门应用的资源或环境信息分类体系,经过模式集成,可形成综合性资源环境信息系统。

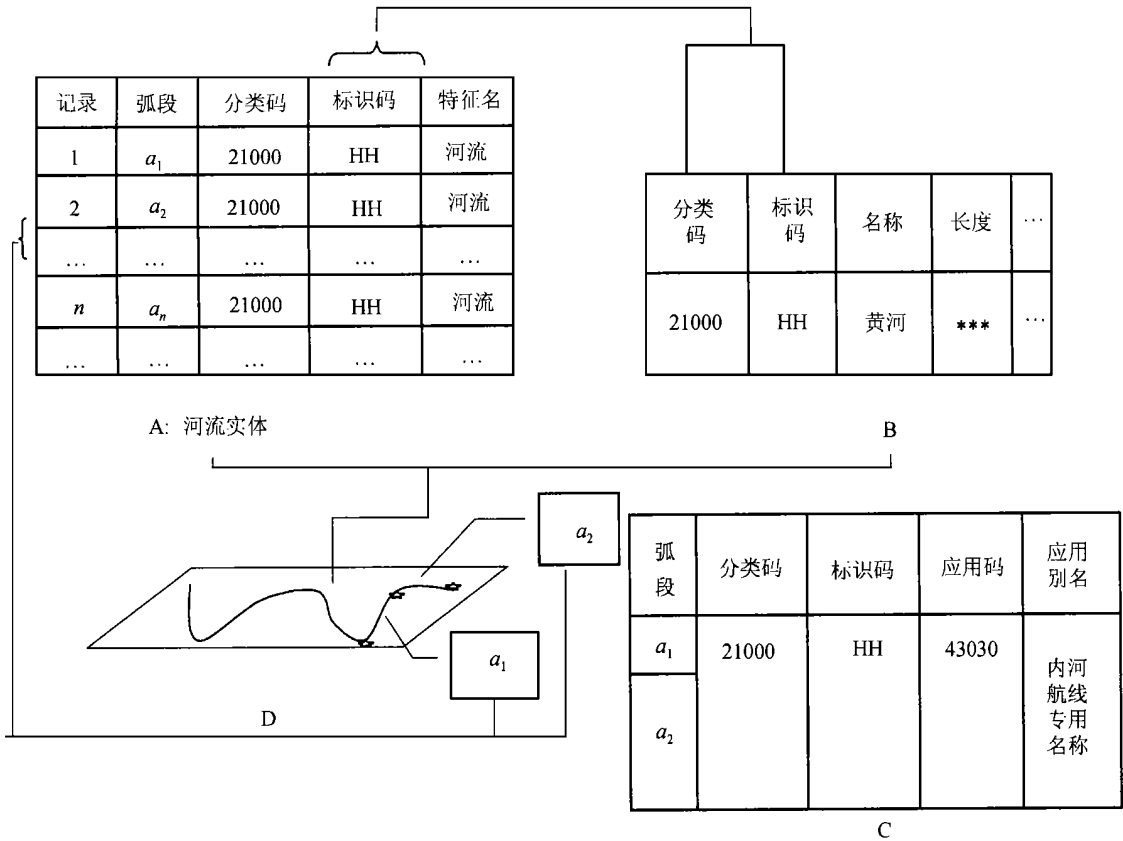


图 3 河流特征及其应用
Fig. 3 An example river feature and its applications

4 结 语

对 GIS 技术的广泛应用来讲,数据共享问题是

一个核心问题。数据格式转换问题一度成为该领域研究的核心问题,但是本文认为,保证数据集语义正确传输的概念模型在 GIS 数据共享中扮演着更为重要的角色。面向数据共享的概念模型的设计属于高

层次的数据建模技术,要求这种数据模型提供方便于各种应用模式建立的抽象概念,并提供语义抽象机制,GIS语义数据模拟方法是建立这种概念模型的最佳方法。建立面向数据共享的、高层次的概念模型在中国仍然是一项被忽视的工作,许多问题,例如语义抽象、应用等还需进一步研究。

参 考 文 献 (References)

- 1 David J. Coleman, John D. McLaughlin. Defining global geospatial data infrastructure (GCDI): components, stakeholders and interfaces. *GEO-MATA*, 1998, **52**(2): 129-143.
- 2 Edited by David M. Kroenke, translated by Shi Bole *et al.* Database Processing: Basic, Design and Realization. Beijing: Publishing House of Electronics Industry, 1998. [David M. Kroenke, 施伯乐等译校. 数据库处理: 基础, 设计与实现. 北京: 电子工业出版社, 1998.]
- 3 Lynn Usery. Category Theory and the Structure of Features in Geographic Information Systems. *Cartography and Geographic Information Systems*, 1993, **20**(1): 5-12.
- 4 E. Lynn Usery. A Feature-Based Geographic Information System Model. *PE&RS*, 1996, **62**(7): 833-838.
- 5 Robert D. Rugg. Defining Standard Features for Land Use Application. *Cartography and Geographic Information Systems*, 1995, **22**(3): 195-204.
- 6 Agatha Y. Tang, Teresa M. Adams, E. Lynn Usery. A spatial data model design for feature-based geographical information systems. *INT. J. Geographical Information Systems*, 1996, **10**(5): 643-659.
- 7 Zhu Xinyan, Xu Yuntao, Zhang Yinzhou *et al.* Application of object-oriented semantic data model to the spatial database. *Journal of Wuhan Technical University of Surveying and Mapping*, 1993, **18**(4): 76-81. [朱欣焰, 许云涛, 张银州等. 面向对象语义数据模型及其在空间数据库中的应用. 武汉测绘科技大学学报, 1993, **18**(4): 76-81.]
- 8 Chen Changsong, Huang Yuxia. The logic management of the resources and environmental information for the design of the national resources, environment and district-economic information system (NREDIS). Proceedings of International Conference on Modelling Geographic and Environmental System with Geographical Information Systems. Hong Kong, 1998, 21-30.
- 9 Robin G. Fegeas, Janette L. Cascio, Robert A. Lazar. An overview of FIPS 173, The Spatial Data Transfer Standard. *Cartography and Geographic Information System*, 1992, **19**(5): 278-293.
- 10 Chen Changsong. The Application of the geographic information classification system to the design of a GIS semantic data model. *Bulletin of Surveying and Mapping*, 1998, (8): 17-20. [陈常松, 地理信息分类体系在GIS语义数据模型设计中的作用. 测绘通报, 1998, (8): 17-20.]

作 者 简 介

陈常松,男,1966年6月生,副研究员,1991年北京师范大学获硕士学位。主要从事地理信息理论, GIS数据共享及立法, GIS标准化规范化等方面的研究, 1995年入中国科学院地理研究所资源与环境信息系统国家重点实验室攻读在职博士学位, 发表论文30多篇。

Study of GIS Conceptual Model Design for GIS Data Sharing

CHEN Chang-Song

(Chinese Academy of Surveying and Mapping, Beijing, 100039)

HE Jian-Bang

(The Institute of Geography, CAS, Beijing, 100101)

Abstract The main challenge in GIS data sharing is semantics sharing of the data. Of the main basis for data sharing including technological development, data standardization and legal basis, the standardization is the most direct one. The design of one standard conceptual data model, which is a data sharing related fundamental task, is the most important strategy to make a precise semantics sharing. The authors' perception of standard conceptual model design is intended to find a suitable, systematic set of rules to ensure the safety of data sharing. The model designed must provide abstract concepts that can be interfaced for data sharing and therefore the concepts must be geographic concepts instead of spatial concepts. The abstract mechanism must also be designed in the process of modeling. The GIS semantic data modeling approach has been used as a tool to design the needed conceptual model. As a basic geographic concept, geographic feature has been used as a semantic object.

Key words Data sharing, Semantics sharing, Conceptual data model