

G

GUOJIA JICHUDILIXINXIBENTI
GUANJIANWENTI YANJIU

国家基础地理信息本体 关键问题研究

王 红 李 霖 朱海红 编著



科学出版社



国家基础地理信息中心编
中国地理信息产业协会编

国家基础地理信息本体 关键问题研究

王 德 明 王 德 明 编著



科学出版社

国家基础地理信息本体 关键问题研究

王 红 李 霖 朱海红 编著

国家高技术研究发展计划(863 计划,2007AA12Z219)

中国博士后科学基金(20070410963)

资助出版

科 学 出 版 社

北 京

版权所有,侵权必究

举报电话:010-64030229;010-64034315;13501151303

内 容 简 介

本书是作者依托国家 863 课题和中国博士后科学基金,在从事地理信息本体研究的基础上编写而成的。本书以国家基础地理信息概念为研究范畴,充分分析国家基础地理信息概念的特点,借助于形式本体,从语义表达的一般方法出发,确定地理信息的本体属性,提出一套建立基于形式本体的地理信息语义分析方法;从概念的内涵和外延出发,实现基于“格”的地理信息语义关系的构建;通过建立本体驱动的地理信息检索机制,提出一种基于领域知识的国家地理信息本体的构建方法,为地理信息领域及其他相关领域本体的构建提供参考和依据。

本书可作为计算机、地理学、地理信息系统及相关专业学习者的参考资料,也可作为相关领域有一定研究基础的专业人员的参考书。

图书在版编目(CIP)数据

国家基础地理信息本体关键问题研究/王红,李霖,朱海红编著. —北京:科学出版社,2011.6

ISBN 978-7-03-031202-0

I. ①国… II. ①王… ②李… ③朱… III. ①地理信息系统—研究
IV. ①P208

中国版本图书馆 CIP 数据核字(2011)第 098955 号

责任编辑:高 嵘 / 责任校对:董艳辉

责任印制:彭 超 / 封面设计:苏 波

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

武汉中科兴业印务有限公司印刷

科学出版社发行 各地新华书店经销

*

2011 年 6 月第 一 版 开本:B5(720×1000)

2011 年 6 月第一次印刷 印张:11 1/2

印数:1—1 500

字数:221 000

定价:38.00 元

(如有印装质量问题,我社负责调换)

随着国家信息化建设的推进,各种专题信息库和公共信息库的建设已初具成效,在这些信息库中,70%~80%是与地理位置有关的地理空间数据,具有地理空间参照作用的基础地理信息在各类专业和公共信息库的业务运行中起着关键作用。由于行业和应用的不同,对基础地理信息从数据源、数据格式、数据操作及语义的要求各不相同,因此地理信息共享与互操作问题一直是地理信息科学领域研究的热点。语义层面上互操作是语法层面上互操作的深入和实现地理信息共享最终目标的必经之路。由于地理现象的多样性和复杂性,用词汇和语义丰富的自然语言来描述,既增强人们对地理现象的认识,又增加了描述的歧义性。在语义层面上实现地理信息共享的关键是对地理信息所表达概念的本体表示,特别是对地理信息语义及关系的形式描述。

国家基础地理信息是国家信息化战略的重要资源,它作为国家地理信息的骨架,是国家地理信息的共享数据平台。一切与地理位置有关的人口、资源、国土、规划、建设、交通道路等行业和部门的海量信息库都以国家基础地理信息作为基本地理参考。地理信息系统在面向部门的专业应用不断拓展的同时,已开始向社会化、大众化应用发展,GIS从传统意义上的地理信息系统拓展为地理信息科学、地理信息系统和地理信息服务等多个方面。然而,对地理概念的不同理解将极大地妨碍行业和部门信息库的建设,基础地理信息中地理对象语义描述的微小差别将影响许多相关行业信息库的信息共享,并会导致大量的重复性工作和信息资源的浪费,基础地理信息的语义描述决定了国家地理信息共享工程的程度。

基于此,本书通过对国内外相关资料的分析与整理,系统地介绍了本体、地理本体的相关概念;基于形式本体,通过概念化手段实现国家基础地理信息的语义分析;结合地理信息的特点,提取地理信息概念的本体属性,形成基于形式本体的基础地理信息语义分析方法;借助于“格”,实现地理信息语义关系的自动构建;结合知识工程、计算机科学等,建立国家基础地理信息本体的逻辑模型,设计并实现了国家基础地理信息本体原型系统;基于已有的地理数据集,就其应用进行了简单的探讨。本书以国家基础地理信息为研究内容,系统地介绍了领域本体的构建过程,对于学习或了解本体,特别是地理本体的读者,具有较好的参考价值。

本书是中国博士后科学基金“基于形式本体的基础地理信息语义表达”和国家高技术研究发展计划(863计划)“地球空间信息系统技术专题”子课题“国家基础

地理信息本体原型研究(2007AA12Z219)”的研究成果之一。

本书的完成离不开许多人的贡献,正是他们的努力工作,丰富了本书的内容。借此机会,特别感谢参与项目开发的优秀青年工作者:宋平超、普帆、蒯希、苏凯、钟璇、周晶、曾微微等。在此,对他们表示衷心的感谢!

由于作者水平有限,难免出现疏漏和不足,诚恳希望读者批评指正。

编 者

2010 年 12 月

第 1 章 概论	1
1.1 研究背景及研究意义	1
1.2 地理本体的概念及研究现状	6
1.3 国家基础地理信息本体	13
第 2 章 本体理论基础	17
2.1 本体的含义	17
2.2 本体描述语言	24
2.3 本体的构建	37
2.4 几种典型的本体	47
2.5 国内外典型的本体知识库	53
第 3 章 国家基础地理信息概念分析	58
3.1 地理信息的概念化	58
3.2 地理信息概念的组织和来源	61
3.3 国家基础地理信息概念域	69
第 4 章 基于形式本体的地理信息语义分析	74
4.1 基于本体属性的地理信息概念的语义表达	74
4.2 地理信息本体属性的建立	86
4.3 地理信息本体属性的形式化表达	89
第 5 章 基于格的地理信息语义关系的构建	92
5.1 形式概念分析原理	92
5.2 地理信息概念之间的语义关系	98
5.3 地理信息本体的形式化描述	105
5.4 基于格的地理信息概念语义关系的建立	107
第 6 章 本体驱动的地理信息检索	112
6.1 本体在信息检索中的作用	112
6.2 基于本体的地理信息检索	114
6.3 本体推理的设计	122

第 7 章 国家基础地理信息本体的设计与实现.....	131
7.1 国家基础地理信息本体的创建	131
7.2 国家基础地理信息本体系统的实现	137
第 8 章 基于本体的地理空间数据集的设计.....	149
8.1 数据准备与分析	149
8.2 基于本体的地理数据语义关系及表达	151
8.3 基于本体的地理空间数据推理	161
参考文献.....	166

进入 21 世纪以来,人类社会正从工业经济迈向知识经济时代,一场以信息技术为核心的革命正在深刻改变着人类生活与社会面貌,作为全球信息化浪潮重要组成部分的地理信息系统的建设与应用,日益引起科技界、企业界和政府部门的广泛关注。地理信息系统、遥感技术和全球定位技术三者有机结合,构成科学地理学日臻完善的技术体系,引起世界各国普遍的重视。

——黄杏元

1.1 研究背景及研究意义

1.1.1 研究背景

作为空间数据的有效管理工具,地理信息系统目前逐步成为空间信息管理与应用的主要平台,因此,地理信息系统面临着对不同来源、不同数据组织形式的空间信息进行有效管理和综合应用的难点。目前的空间数据已经不再是简单的图形数据,它还包括卫星遥感影像数据、数字地面模型数据等,空间数据的管理已经明显扩大;在应用层次上,地理信息系统与其他专业密切相关,这给当前空间数据共享带来了很大的困难;另外,随着地理信息系统应用的不断发展,各个部门存在将本部门的专业系统与地理信息系统相结合的问题,不同部门之间的数据协作也变得日益重要。在处理某项工作的时候,往往需要不同部门的数据,如用于城市规划管理的地理信息系统,它不仅需要城市基础地理信息,还需要城市道路、土地利用、管网以及规划等数据,这些工作的完成需要不同部门、不同空间数据库的协同工作。但是,由于这些不同空间数据库建立在不同的时间,或者适合于不同的部门行业,所以往往存在很大的差异性,也就是所说的异构性特点(龚健雅,2004)。

地理信息系统在取得巨大发展的同时,其缺陷也越来越明显,突出表现在:传统的地理信息系统是封闭的、孤立的系统,没有统一的标准,各自采用不同的数据格式、数据存储和数据处理方法;地理信息系统及其应用系统的开发都是基于具体、相互独立和封闭的平台,并且在数据语义表达上往往存在不可调和的矛盾,从而无法直接进行应用系统之间的数据共享。随着信息社会的发展,如何使不同的

地理信息系统软件能够迅速快捷地获取这些来源不同的数据,并将它们集成起来进行分析,使这些集成数据能够在不同的系统下相互可操作以及在异构分布数据库中获取所需要的数据信息就变得非常重要。

地理信息系统技术经过几十年的发展,已经进入以时空调控、全球变化、区域可持续发展和战略决策支持系统为主要目标的时代。为适应这一时代发展目标的要求,全球范围、区域范围及国家范围内的地理信息必须实现共享。信息共享已经成为现代信息社会发展的一个重要标志,而地理信息互操作的产生则是信息共享的必然产物,地理信息系统共享和互操作将成为 21 世纪地理信息系统研究领域的一个重要组成部分(龚健雅,2004)。

空间数据共享的方法异常复杂,目前最常用的数据共享主要有外部数据交换、空间数据共享平台和空间数据互操作三种方式。外部数据交换仅仅是从数据转换角度考虑共享,它是基于文件级的共享,仅能用于数据的集成,不能达到要素级的实时共享。空间数据共享要求所有相关应用部门都采用一个空间数据库管理系统,该部门所有的空间数据及各个应用软件模块都共享一个平台,但是目前实现这种共享方式比较困难,因此当前数据共享的重点是研究空间数据互操作技术。

GIS 互操作的核心是用统一的标准协议通信,解决对象跨平台的连接和交互问题,从而实现在不同分布式计算平台、不同数据模型、不同数据结构、不同空间数据表示,不同数据定义语言和数据操作语言之间自由地进行数据交换与协同处理。OGC(OpenGIS Consortium,开放地理信息系统协会)把地理信息系统互操作分为数据层、语法层和语义层三个方面。数据层和语法层两种互操作主要由数据结构、概念模型和软硬件环境的差别引起,可以通过统一数据结构标准(如 OpenGIS)和服务功能原子化、模块化以及开放接口标准来解决,而语义层的互操作就相对复杂得多。

在语言学中,语义是语言反映人类思维过程和客观实际的方向,是人们对客观事物的反映,但并不完全等同于客观事物。在计算机中,语义是指对数据库中形式化语言表达的数学解释。语义互操作是实现互操作中的一个难题。地理信息共享是一个交流过程,在交流中不仅涉及数据值本身,还涉及数据的语义或含义。同样,成功的 GIS 互操作不仅要求技术上获取分布式地理信息和服务,还需理解这些信息和服务的文档。由于地理信息语义的复杂性,使语义互操作成为 GIS 互操作中的一个难题。

GIS 语义互操作(黄裕霞,2001)是面向应用层的 GIS 互操作。在某些特定的 GIS 应用中,可能会用不同术语来描述相同概念,或用相同术语来描述不同概念,因而导致语义不一致。GIS 语义互操作强调的是处理语义不一致的能力,它着重处理三类问题:因不同应用分类定义导致的属性含义差异性;因不同几何描述导致的空间定位差异性;因不同管理要求导致的信息质量和共享限制的差异性。通过语义互操作,能使 GIS 语义在转换和使用过程中保持一定程度的准确性、完整性

和一致性,以达到最为有效的 GIS 语义共享。

国家基础地理信息是表述国家主权、自然资源、社会经济各要素分布的重要信息,是国民经济建设各专业部门专题信息定位和定量分析的空间基础,是国家信息化建设的重要组成部分(李莉,2003)。它作为国家地理信息的骨架,不仅是国家信息化战略的重要资源,而且是国家地理信息的共享数据平台。一切与地理位置有关的人口、资源、国土、规划、建设、交通道路等行业和部门的海量信息库都以国家基础地理信息为基本地理参考。

2006 年,国家基础地理信息系统 1:50 000 数据库通过了验收,它标志着我国国家基础地理信息系统初步建成,真正实现以“数字化”形式再现神州大地。结合已建成的 1:400 万、1:100 万及 1:25 万数据库和各地正在建设的 1:1 万以及更大比例尺基础地理信息数据库,在未来几年我国将建成数字中国地理空间框架的数据体系。地理信息共享的关键是要实现地理信息的标准化及地理信息处理的通用性,只凭地理数据转换永远不能实现信息的共享,只有在更高层次上实现不同应用系统之间的相互合作,才能真正达到地理信息共享的目的。目前对于地理信息服务的研究,主要依据 OGC、ISOTC211 等制定的一些规范(WMS, WFS, WCS 等),它们在实现空间信息共享与互操作方面起到了一定的作用,但语义的描述、海量分布式数据处理等仍未给出明确的方法,将网络服务技术、语义技术引入地理信息领域的研究仍处于发展阶段。时空信息服务的大众化代表了当前和未来的时代特征,是空间信息行业能否产业化运转的关键(李德仁,2004)。如何在数据生产者、管理者和应用者之间搭建一座桥梁,为用户透明地提供地理信息服务,已成为当前一项十分重要的、紧迫需要研究的课题。

1.1.2 研究意义

地理信息体是人们进行空间信息认知的基本单元,在 OpenGIS 指南中,地理信息体被表述如下(宋炜,2004)。

① 地理信息体是一群人的集合(同一学科的研究者,一个工程中的合作伙伴等),这些人至少在部分时间内共享一种数字化的地理信息语言并共享相同的空间要素定义。

② 地理信息体是一个系统或者个人的集合,这些人能成功共享数字化的地理空间信息,也就是要素。

③ 地理信息体根据信息体成员的元数据进行形式化定义。

第①个定义没有区分两个截然不同的世界,即工程世界和真实世界。一个工程世界是一个专门的地理空间应用领域的计算机表达。而其定义中信息体是由工程世界外的人、组织等所构成的,尽管他们不能在两个信息体中共享一个相同的数字信息语言或共同的空间要素定义,但仍有可能共享地理信息。

第②个定义是有问题的,这是因为它假定地理空间信息仅仅只能在信息体内

被共享。这个限制并不能回答为什么一个国土资源部的地籍地理信息体与建设部的土地利用地理信息体能够共享信息,尽管它们是不同的信息体。

地理信息体的定义不应该基于计算机中被表达的要素,而是应该基于人,这些人属于一个特定的领域并概念化真实世界。这个特定领域的概念化是特定地理信息体的论域,概念被其外延和内涵所描述,它与地理信息体相关联。因此,在一个地理信息体中,人们可以在相同的论域中以不同的方式进行概念抽象,这也创造了不同的工程世界。一个工程世界主要是由一个固定的词汇集合、一个固定的模式元素集合以及一个要素的实例集合组成。

第③个定义表明地理信息体是由有元数据的地理空间应用形成的。元数据是描述一个数据库的信息而不是一个地理信息体。概念建模和元数据限制了数据库中对象的说明。一个信息体的定义应该没有这种限制,这是因为我们根本不可能为一个特殊的信息体提供一个单一的标准数据模型。因此,信息体要求由一个高层的语言,也就是在概念模型之上形式化定义基本地理信息体的论域而不限制它的表达方式。建议用本体作为一种更高层次的语言。

OGC 认为,在一个具有公共语言、公共定义集和稳定概念模型的信息体内,可以实现数据的无损转换。但是由于信息体之间的语言、定义和概念模型不同,除非采取特殊的步骤进行控制,否则信息共享是不完美的而且会有信息损失。信息体之间的数据共享由于语义不匹配,会导致数据丢失或者数据被错误翻译。理论上,至少在以下三种情况下会出现语义冲突问题,在不同的信息团体之间通信时也会有损失。

第一种情况,两个信息体之间共享概念和定义,但没有公共的语言;或者两个信息体之间虽然有公共的语言却使用不同的行话。这个问题可以在两个语言之间进行双向映射,通过简单的翻译转换解决。只要语言本身是稳定的,相关的术语间会有一个 1:1 的关系,因此该方案可以支持有效的通信。

第二种情况,两个信息体之间没有共享一个稳定的术语定义集合。要解决这个问题,需要一个对公共定义的直接映射,另外加一个不能被映射的术语的解释集合。只要信息体间存在着 1:M 的映射,则将一个专门定义的概念映射到一个通用的概念时就必须概括归并,这时信息损失可能发生。

第三种情况,信息体之间没有共享的概念。也就是说,两个团体有着完全不同的世界观,没有公共的集合。

因此,信息体之间地理信息的共享首先要建立信息体的概念模型,并在它们之间建立一种语义转换机制,使得有着共享需求的信息体间能以最小的语义损失最大限度地充分利用已有的数据,以减少重复劳动和节省费用,达到信息共享和互操作。

在数据共享过程中,应有一种对地理认知的概念模型作为不同部门之间数据共享和应用的基础,面向地理系统过程语义的数据共享的概念模型包括一系列约

定法则,如地理实体几何属性的标准定义和表达,地理属性数据的标准定义和表达,元数据定义和表达等。这种模型中的内容和描述方法,有别于面向地理信息系统软件设计或地理信息系统数据库建立的面向计算机操作的概念建模方法。地理信息本体的出现,使一组空间数据生产者或使用者,可以共享表达真实世界现象的一个 Geo-Ontology。

Geo-Ontology 可以看做是地理数据模型之上的一个元语言。要求每一个地理信息体提出一个 Geo-Ontology 以减少在信息体之内或者跨信息体之间的语义损失。如果两个信息体完全或者部分地共享相同的 Geo-Ontology,那么这两个信息体就有可能共享它们的数据。例如,如果每一个信息有自己的 Geo-Ontology,那么两个信息体必须共享一个共同的 Geo-Ontology 以便能够共享某种信息(宋炜,2004)。

可以看出,GIS 技术下的地理信息共享不单纯是技术问题,而是一项系统工程,需要解决管理、政策、标准和技术等问题。基于这种认识上的提高,已经形成相对完善的 GIS 标准体系、相对成熟的技术体系(如 WebGIS 技术、互操作),正在建立和健全法规保障体系,为 GIS 技术下的地理信息共享创造了一个良好的外部环境。但是所有这些工作着力要解决的问题仍然停留在 GIS 数据的“共享”层次上,即在保证 GIS 数据不失真、不损失精度的情况下,一个 GIS 系统中的数据怎样与另一个系统中的数据相互集成、融合、重组和转换,或者多个组织(个人)如何获得一个 GIS 系统中的数据。地理数据经过分析加工形成地理信息,该如何通过实现数据共享来真正实现地理信息共享,尽管也有所涉及,但绝非重点。不同部门、单位、个人所持有的空间概念之间的相互融合,称之为语义共享。数据的共享和语义的共享两者合起来,称之为 GIS 语义共享。

要真正实现 GIS 语义共享,需从概念的本体属性出发,明确要素之间的各种语义关系。本体的引入给 GIS 学科发展开辟了一片崭新的研究领域。因此,从本体出发,研究地理信息共享和互操作,是 GIS 学科发展的迫切要求。

任何一门科学在其发展过程中都是不断学习借鉴其他学科而不断深化的,一般的规律是:学科要实用化就不得不借助于数学方法,而要夯实理论根基则需要求助于哲学。地理信息系统的发展也符合这一规律。人地关系学说、系统哲学、地理信息认知哲学等已成为地理信息系统的理论基础。但是它们却并不能为地理信息系统领域的人类活动提供支持,地理信息系统领域的专家们便转而寻找一种能够表示地理本体中更本质的东西,于是本体论进入了人们的视野。

本体是认知形式的基础,它是横贯科学和信息系统分支的理论。本体主要关注的焦点在于实体本身,即事物本身所呈现的运动状态和变化方式,具体到地理信息系统领域就是对地理空间实体目标域的关注。地理信息系统领域中的本体是解决地理概念分类以及地理概念之间的相互关系,它注重的是实体和实体之间的关系,而不是对实体的操作。

本体引入地理信息系统领域的最大意义在于对空间信息语义理论的丰富,主要反映了地理信息学科研究重心的转移,即从过去过分强调计算模型的形式化到目前对空间目标域本身的关注。本体的基本目的是把类特征化,得出实体类型间的区别。本体代表了 GIS 科学研究的新方向,在 GIS 中扮演了一个重要的角色,它允许人们去捕获独立于数据表达的数据语义。本体作为共享概念模型的明确的形式化规范说明,其优点在于建立一套共享的术语和信息表示结构,多数据源上的异构信息通过共享的术语和信息表示结构,成为同构的信息,从而实现多数据源信息的集成,利用它可以实现空间数据的详细定义和说明。

本体在领域内的共享保证了领域内信息在语义上的统一和规范,而显式和形式化的特性一方面为知识的机器阅读和理解奠定了基础;另一方面保证了具有不同知识背景的用户仍可以借助本体进行基于语义的交流,从而使各个领域的应用研究不仅遵从本领域特有的规律健康发展,同时也可以与其他领域的应用研究相辅相成,共同进步。在这个过程中,本体保证了对知识理解和运用的一致性、精确性、可重用性和共享性。

1.2 地理本体的概念及研究现状

1.2.1 地理本体的含义

为了便于建模,本体通常被用作定义领域知识和提供标准词汇表及其应用规则工作中确定分级、概念、关系和规则的一种方法(Smith,1998)。地理信息科学领域应用本体的最初目的就是定义一套能促进不同系统和系统与用户之间的互操作和解决数据集成问题的公共词汇表。地理信息科学领域广泛引用的本体定义是“规范化概念模型”,并把本体用作“能够用作反映客观世界的指代物(referent)的实体的定义或标识”(Mark,1999)。地理信息科学领域认可的其他本体定义还有“存在的科学”、“所有客观存在的分类分级的定义”(Smith,2001)、“客观世界及其行为规则和客观世界同概念世界的联系的抽象定义”、“决定一个领域的实体、关系和规则”。地理信息科学领域研究本体的方法可分为两大类:从形式化本体的观点出发,确定顶级分类的哲学方法;面向具体领域和具体任务的一个具体分类的行为、术语和关系的明确化方法,可以从自然语言表达达到严格的形式化规范。Sowa(2000)把本体定义为“领域内可能存在的事物的分级的全面研究”,Timpf(1992)则持更具有目的驱动的观点,认为本体是人们为了“某些目的”表达的简化的世界模型,只包括领域内与某一特定任务相关的概念。

Mark 等人(1999)认为,地理信息科学中的本体论研究是高度跨学科的交叉研究,与地理信息的认知、表达、互操作、尺度和不确定性密切相关,其最重要的一点是研究空间信息的语义理论,或者更一般地说,就是研究人类思维、信息系统与

地理现实世界之间的关系。

国内也有不少学者相继提出了地理本体的概念。

孙敏等人(2004)给出了地理本体论的定义:地理本体论是研究地理信息科学领域内不同层次和不同应用方向上的地理空间信息概念的详细内涵和层次关系,并给出概念的语义标识。并且他认为地理信息本体论的研究对解决地理信息的建模、语义互操作、空间数据重用及共享、数据挖掘等地理信息科学理论的多方面问题以及地理信息系统的智能化、网络化和大众化发展有着变革性的意义。

崔巍(2004)认为:一个地理本体系统是空间信息科学中具体应用领域里共享的一个概念化的知识系统的形式化和显示的说明规范。地理本体的概念构成层次结构,该层次结构来源于地理领域已有的概念分类体系,本体的概念由属性集描述,属性集组成一个多维的特征向量空间。

黄茂军(2005)提出了一个基本观点:地理本体涵盖了哲学本体、信息本体以及空间本体三个层面的含义。哲学本体即从地理真实世界这个角度去研究地理本体,其目的是产生有关地理世界结构的更好理解,为地理信息系统的发展提供更合理的概念模型,从而避免现有的数据模型与人类空间认知机制的巨大反差。信息本体主要通过对共享的地理概念的明确的形式化定义,应用于地理信息共享与互操作、基于语义的地理信息集成以及地理信息服务等方面。空间本体主要表现为与地理信息空间特征相关的本体,具体说来也就是与空间位置、空间形状和大小等几何特征,以及空间关系等相关的本体。

王敬贵(2005)提出:地理本体是指特定地理空间信息领域概念化模型(conceptualization)或学科感知世界的明确的形式化规范说明。他认为,与一般信息系统所研究的“封闭”世界不同,地理世界是一个开放世界,地理事实存在模糊性和不确定性、动态性等复杂特征,因而人类对地理事实的认知即地理实体和现象也存在模糊性、时态特征及身份识别等复杂问题,这为地理本体的构建提出了一些困难而且是根本性的问题,因此地理本体并不能简单地看做有别于其他领域的一般本体。相反地,地理本体具有区别于一般本体的重要特征,即空间特征和时态特征。因此,地理本体必须发展方法表达这些一般本体中不具备的特征,而地理本体理论研究也应该成为一般本体理论研究的一个重要分支,而不是一般本体理论在地理领域的简单应用。

宋炜(2004)提出:地理信息本体是一个特定研究论域的概念模型的形式化表达机制。参照本体的定义,一个地理信息本体是对地理领域被共享的概念化的形式说明。其中,地理领域的概念就是地理信息概念模型,它通过对地理现象的认识和抽象形成地理现象的抽象模型,显式地指出所用概念的类型以及约束,并通过形式化表达使人 and 机器可读。Geo-Ontology获取了领域内的一致知识,它不是某个个体私有的,而是可以被一个群体所接受和共享的。Geo-Ontology为需要共享地理领域的信息研究者定义了一个公共的词汇集。它包括地理领域概念的定义和概念之间的

关系,这些都是机器可以理解的。

1.2.2 地理本体的国内外研究现状

本体作为一种信息科学研究的方法论被引入地理信息科学,其研究源于 Egenhofer 和 Mark(1995)对朴素地理学(*naive geography*)的研究。1996 年 11 月初,在美国国家地理信息分析中心(NCGIA)第 21 号研究动议的专家讨论会上,首次拟定了一个地理本体草案。随后,本体概念及思想引起了众多地理信息研究者的注意,开展了一系列的研究(Mark, 1999; Kuhn, 2001—2003; Kavouras, 2002—2005; Frank, 1997—2003)。这些研究就地理本体的基础理论、方法及其应用问题进行了卓有成效的探索,并取得了一定的进展。在地理本体基础理论(包括方法论)研究方面,主要对地理本体理论构成、地理本体的人文调查与试验研究、地理本体的信息学研究特别是地理本体的形式化描述语言和工具及本体创建和集成方法等方面进行了广泛而深入的思索与探讨,并将地理本体理论与方法研究成果应用于地理信息集成、地理信息检索、道路寻找和智能体系统设计与开发、社会及生态环境研究等多个应用领域(陈建军, 2006)。

一些专门的研究团队对地理本体进行了深入的研究。希腊雅典国家技术大学乡村与测量工程学院成立了地理本体研究组——OntoGeo,该研究组的研究与教学活动涉及地理概念与分类理论、时空建模、地理知识表达、语义互操作、本体集成、语义理论、空间语言、语义网、本体工程、地理空间实体的相似性与粒度、运动对象(*moving objects*)、概念图(*conceptual graphs*)、地理空间本体、形式化概念分析、认知语义、位置的语义、形式化本体、本体语言等领域。美国地理信息科学大学联盟(UCGIS)(Mark, 2000)是由进行地理空间信息科学教学研究的 80 多所北美高校、专业机构和私营部门组成。大学地理信息联盟已发展成为地理信息系统学术研究界的一个中心网络,尤其是在确定短期、中期和长期的关键研究领域方面。大学地理信息联盟将优先研究重点放在与获取、处理、判读、分布和保存地理空间信息有关的一些难题的基础研究上。年度优先研究计划以白皮书的形式发布。德国 Munster 大学地理信息科学研究院成立了语义互操作实验室——MUSIL,展开了对地理本体与互操作方面的研究。欧盟专门组织了一个网上的空间信息检索查询(*spatial-aware information retrieval on the Internet, SPIRIT*),用来研究地理本体的构建及其查询机制以求解决欧盟各国的地名智能匹配和查询问题(Jones, 2002)。

国际上也有许多以地理信息本体为主题的研讨活动,其中比较有代表性的会议有:2000 年由 Schuurman 和 Mark 组织,在 Pittsburgh 召开关于本体的 AAG 会议分会;2000 年 9 月,由 Winter 组织,在法国召开“空间数据标准的本体论与方法论”会议,就空间本体、空间认知论、空间认识、空间不确定性、空间层次、空间一般

化、标准与标准化等主题进行了研讨;2000年和2002年,分别在美国佐治亚和科罗拉多召开地理信息科学会议,对地理信息本体进行了深入探讨;2002年9月,在英国召开关于地理信息本体(Geo-Ontology)的工作组会议,对地理信息本体的概念、关键问题以及研究的重点范围进行讨论,与会学者首次使用了Geo-Ontology一词。本次会议的报告指出,Geo-Ontology是地理信息本体的一般术语。地理信息本体是研究地理信息科学领域内不同层次和不同应用方向上的地理空间信息概念的详细内涵和层次关系,并给出概念的语义标识。

研究地理本体或利用本体方法研究地理概念语义被认为是地理信息集成和共享的有效途径(Bishr,1998),国内外许多学者在这方面进行了研究。Nunes(1991)指出建立下一代GIS的第一步是地理实体、属性和关系的系统整理和规范说明。这可以看做是建立地理本体的最原始思想。Frank(1997)研究了时空数据库的基础本体问题,认为本体将有利于信息系统更好地避免GIS中潜在的概念不一致问题。Guarino(1998)在详细阐述本体(ontology)含义的基础上,提出了本体论驱动的信息系统。Harvey(1999)认为将基本的语义处理方法引入GIS设计过程对克服语义差异具有特别重要意义。Uitermark(2001)研究了一种基于本体的地理数据集成方法。Kuhn(2001)提出了一种从自然语言文本中抽取地理领域本体的方法,该方法主要关注文本中的对象行为,并以从德国交通代码文本中抽取汽车导航领域本体为例说明了这种方法的可行性。Farshad(2001)引入逻辑描述语言,提出如何利用本体解决互操作中语义冲突问题。Kavouras(2000,2002,2005)认为本体问题对地理信息科学领域的很多研究方向具有根本性的影响,研究内容包括地理本体的概念映射、语义互操作、形式化概念分析及概念格等多个方面,他们应用形式化概念分析和概念格形式化表达地理类型及集成不同地理本体的具体方法。Fonseca(2001,2002)讨论了构建本体驱动的GIS的概念框架、构架ODGIS的方法学、本体融合等问题,提出了一种基于本体的地理信息集成框架。随后,Fonseca(2006)建议在使用地理对象的基础上进行概念的分类,并根据描述的物理世界和社会现实提出了地理本体的一套基本概念,同时还提出一个框架来衡量程度之间的互操作性地理本体。Visser等人(2002)对形式化本体及其在地理信息处理中的应用作了一个简洁综述,对各种本体语言的形式化程度进行了比较,包括描述层次的XML到一阶谓词逻辑层次的概念建模语言CML以及更复杂、形式化程度更高的知识交换格式KIF等,并简单介绍了一种基于语义的智能信息检索代理构架。Rodriguez(1999,2003)从语义出发,提出了基于本体的综合考虑特征和语义关系的混合模型。Tomai(2005)则提供了一个网上工具用于构建地理本体。Lutz等(2006)以水文学家在丰富的网络信息资源中查找给定时间点上,某特定水文控制点的水位为例,展示了不同的元数据描述给查询者带来的困惑。这个例子典型地表明地理信息语义共享需求在信息高速发展和发达的社会是何等重要和紧迫。