

第九章 本体论驱动的地理空间推理策略

第一节 本体论的基本概念

本体论 (Ontology) 原本是一个哲学概念, 用来表示真实世界实体的客观存在。由于本体注重实体本身固有的属性和关系的表达, 可以解决仅仅依靠数学模型所不能解决的问题, 因此, 近年来本体论研究在信息科学和计算机科学中日益受到重视, 目前在人工智能、计算机语言学、地理信息科学和数据库理论中获得了广泛的应用。

一、哲学本体论与科学本体论

本体论是一个哲学概念, 也常翻译为“存在论”、“万有论”等, 被认为是哲学的本源。该词最早来自拉丁文 *Ontologia* 一词, 而拉丁文又源自希腊文, 它是指古希腊哲学术语, 意为世界的规律性。作为哲学的一门分支学科, “本体论”最早出现在亚里士多德的“第一哲学”(即“形而上学”)中, 而作为一个哲学术语, 是在 18 世纪才由沃尔夫给出明确定义的。就本质而言, 古代西方本体论强调的是纯粹的理念, 是一种先验哲学。它使用了现实中的语言词汇, 并对其赋予了本体论固有的语义, 即它超脱于现实世界, 无法被现实检验, 因此本体论难以按常识推断, 这正是它超脱于现实世界之外的具体表现, 也是它被引入科学领域用于规范概念术语的原因。

科学本体论是将哲学本体论的思想引入科学领域而形成的一种知识表示方法, 它沿用了哲学本体论的核心思想, 即概念的尽可能完备的描述和建立在这类概念之上的推理。科学本体的定义目前有很多种, 最早的定义是 Neches 等 (1991 年) 给出的, 即“本体定义了组成主题领域的词汇的基本术语和关系, 以及用于组合这些术语和关系、以及定义词汇的外延的规则”。这个定义明确指出, 要建立一个本体, 必须确定该领域的基本术语和这些术语之间的关系, 以及组合这些术语和关系的规则, 并提供这些术语和关系的定义。Gruber (1995 年) 的本体定义为“本体是概念化的详细描述”。

科学本体是指工程上的人造物, 其目标是确定领域内共同认可的词汇, 并从不同层次的形式化模型上给出这些词汇和词汇间相互关系的明确定义, 从而获取相关领域的知识, 提供对该领域知识的共同理解; 同时, 它是通过描述概念及概念之间的关系反映现实世界的概念模型, 因此本体中的概念必须真实地反映现实世界。科学本体的作用主要体现在: 建立本体的分析过程, 理清领域知识的结构, 有利于知识的计算机表示; 重组可交互的大众化的词汇, 为交流协作提供基础; 使用统一的术语和概念使知识的共享成为可能; 可以重用 (reuse) 和继承, 避免重复的领域知识分析, 也为知识的获取提供方便; 增强系统的逻辑一致性、规范性和可靠性。

本体论是以思维和存在这一哲学基本矛盾为前提的, 而整个哲学的发展就是围绕哲学基本矛盾而展开。其中, 所有以“存在”问题为核心的研究构成了“本体论”, 而以“思维”问题为核心的研究则构成了“认识论”。“本体论”和“认识论”是相互依存的, “本体论”是“认识论”的基础, “认识论”问题的研究有助于“本体论”问题的解决。我们不能在“本体论”和“认识论”之间划出一条明确的界限。

本体论层次的信息是事物所呈现的运动状态和变化方式；认识论层次的信息是认识主体所感受到的（或所表达出来的）事物的运动状态和变化方式。具体来说，本体论是关于“存在是什么？”、“构成世界的根本要素是什么？”的哲学，它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。而认识论是关于“我们知道什么？”、“我们有什么样的认识能力？”的哲学，它主要关注的焦点在于人们对现实世界的认识和把握，即认识主体所感受到的或表述出来的事物的运动状态和变化方式。

二、地理本体

将本体论引入地理信息科学，源于对朴素地理学（Naive Geography）的研究。朴素地理学的研究旨在为建立大众化和智能化地理信息系统提供基础（Egenhofer 和 Mark, 1995 年），Mark（1996 年）首次提出了一个基于常识的地理本体框架，自此以后，关于地理空间信息本体的研究受到广泛的关注，取得了很多成果。2002 年 9 月，在英国 Ilkley 召开的关于地理信息本体的会议上，与会学者提出了地理信息本体的一个术语——地理本体（Geo-ontology）。地理本体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。

人工智能领域引入本体的主要目的是知识共享与重用，但是，地理本体兼具哲学本体和科学本体的双重含义。哲学本体突出表现在对地理目标域本身的关注，主要涉及地理概念、类别、关系和过程的研究。地理时空本体、不确定性本体、尺度本体也是哲学本体的重要体现。通过哲学本体的研究，尤其是对地理目标域中地理种类、实体类型的本体设计，可以产生对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。地理本体的含义体现在通过对共享的地理概念的明确形式化定义，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。地理本体的研究是跨学科的交叉研究，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，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。地理本体的研究会成为我们加深对地理信息和地理信息系统认识的重要途径。

第二节 基于地理本体的空间推理策略

本体可以对表达空间实体（或者事件、过程、行为等）的概念进行详细解释，包括概念的属性、关系、约束等，并且这些属性、关系、约束等可以用形式化方式表达。因此，用地理本体来表示地理知识，不仅可以实现知识共享和重用等目的，而且有助于地理空间推理的实施。

一、地理本体的组成

一个本体是一个包含概念、对象、属性、关系及约束的定义集合，可以将地理本体形式化定义为一个五元组形式： $O = \langle U, C, R, A, F \rangle$ ，其中， U 代表给定的论域， C 代表论域中的概念， R 代表概念之间的关系， A 代表概念的属性集合， F 代表概念和概念之间

的所有公理的集合。

概念原本是思维科学的一个术语，是“思维的基本形式之一，反映客观事物的一般的和本质的特征”。概念的内涵所反映的是事物的本质特征；概念的外延所反映的是事物所包含的范围。在信息科学领域，我们将概念理解为物体或事件的模型知识。在本体工程中，概念是人类对现实世界理解的表意符号，是机器操作的主要对象，是数学模型的主要构成元素。简言之，概念是人与机器交互的桥梁。

关系在本体中用来表示概念之间的一种关联，常见的概念之间的二元关系有 *is-a* 关系，它表示逻辑层次不同的概念之间的一种关系。

属性是指对概念的定性或者定量的描述。属性可以分为基本属性和角色属性。基本属性是指概念的本质属性，如“高度”；而角色属性是指基本属性在不同领域中的应用，例如，与基本属性“高度”相对应的角色属性有“海拔、深度、高程”等。

实例也是一种概念，是逻辑层次最低的概念，因此称为概念的个体实例，简称实例。例如，“黄河”是“河流”这一概念的一个实例，它本身也是一个概念。

性质是概念的所有实例所共同具有的一般抽象特征，用于区分概念集合中不同的概念。性质是一种特殊的属性，或者是属性和属性值的组合。

公理是指在本体中公认的事实（或者推理规则），用于知识推理。

用地理本体进行描述的地理概念，包括地理实体、地理现象、地理行为、地理过程等。建立地理本体必须考虑空间问题，例如空间关系、空间分布特征、实体占据空间的类型等；还要考虑边界问题（真实边界和人为边界）。

地理本体的组成主要包括地理概念及其属性、关系、以及公理和规则。表 9-1 是“国家”本体的一个组成框架。“中国”本体是“国家”本体的一个实例，其组成框架见表 9-2。

表 9-1 “国家”本体的一个组成框架

属性	关系	公理和规则
首都	东濒	总人口数=男性人口数+女性人口数
洲属	西邻	夏季平均气温>冬季平均气温
人口数	南接	月最高气温≤年最高气温
官方语言	北连	东邻(A, B)→西邻(A, B)
陆地面积	…	$\forall R \in \{\text{位于东部, 位于西部, 位于南部, …}\},$
领海面积		$R(A,B) \wedge R(B,C) \rightarrow R(A,C)$
…		…

表 9-2 “中国”本体

中华人民共和国：国家本体		
简称：中国	东濒：太平洋	官方语言：汉语
首都：北京	西邻：	气候类型：亚热带季风气候
洲属：亚洲	南接：	陆地面积：960 万平方公里
人口数：13 亿（2001 年）	北连：内蒙、俄罗斯等	领海面积：300 多万平方公里

二、基于地理本体的空间推理

1、地理本体的属性继承推理

由于本体中的概念是以种概念、属概念、实例的层次树结构来呈现（如图 9-1、图 9-2 所示），因此不同层次的概念之间的属性存在一定的关联关系，可以根据这些关系实现地理本体的知识推理。

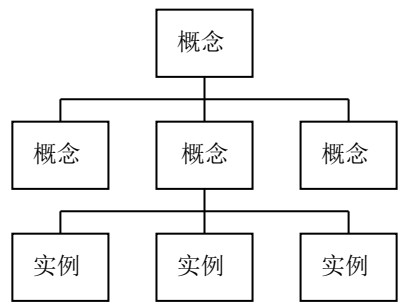


图 9-1 概念的层次结构

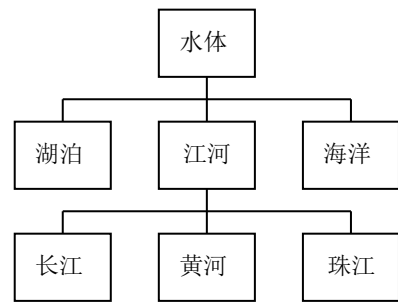


图 9-2 概念“水体”的层次

对于概念 C 及其实例集 S_{IC} ，实例集 S_{IC} 中的元素 e ($e \in S_{IC}$) 和概念 C 之间的关系称为实例关系 (Instance-of Relation)，记为 Instance-of (e, C)，表示的含义为“实例 e 是概念 C 的一个实例”。实例和概念之间具有很好的性质和属性的继承性。通过继承规则，可以进行基于实例关系的知识推理。

对于概念集 S_C 中的概念 $C_1, C_2 \in S_C$ ，如果有概念 C_1 的内涵包含 C_2 的内涵，且概念 C_1 的外延包含于 C_2 的外延，则将概念 C_1 和 C_2 之间的关系称为种属关系 (Is-a Relation)，记为 Is-a(C_1, C_2)，表示的含义为“概念 C_1 是概念 C_2 的子概念；概念 C_2 是概念 C_1 的父概念（或类概念）。”基于种属关系的知识推理规则如下：

传递性：($Is-a(C_1, C_2) \wedge Is-a(C_2, C_3)$) $\rightarrow Is-a(C_1, C_3)$ ；

属性继承：($Is-a(C_1, C_2) \wedge HasAttribute(C_2, A)$) $\rightarrow HasAttribute(C_1, A)$ ；

性质继承：($Is-a(C_1, C_2) \wedge HasProperty(C_2, A)$) $\rightarrow HasProperty(C_1, A)$ ；

实例归属：($Is-a(C_1, C_2) \wedge Instance-of(e, C_1)$) $\rightarrow HasAttribute(e, C_2)$ ；

2、地理本体的空间关系推理

我们可以借助于部分学 (Mereology)、拓扑学 (Topology) 和位置论 (Location Theory) 这三个理论工具来建立地理本体 (Casati 等, 1998 年)。利用这几种工具，可以对本体概念的空间位置和空间关系进行形式化表达，并建立一套公理体系。

部分学用来描述部分与整体之间的关系，其核心关系表示为 Part-of (A, B)，表示的含义为“ A 是 B 的一部分”。与此关系有关的公理可以用一阶谓词的形式表示如下：

定义：O (x, y): $= \exists z$ (Part-of (z, x) \wedge Part-of (z, y))；

公理 P1: Part-of (x, x)；

公理 P2: Part-of (x, y) \wedge Part-of (y, x) $\rightarrow x=y$ ；

公理 P3: Part-of (x, y) \wedge Part-of (y, z) \rightarrow Part-of (x, z)；

公理 P4: $\forall z (\text{Part-of}(z, x) \rightarrow \text{O}(z, y)) \rightarrow \text{Part-of}(x, y)$;

公理 P1 和公理 2 表示“部分”关系具有自反性和反对称性;公理 P3 表示该关系具有传递性;公理 P4 保证了该关系是扩展的。

位置理论是建立在部分学基础之上、用来研究地理对象与地理对象所占据的空间之间的关系的一种理论(Casati 等, 1998 年)。位置理论的理论基础是基本关系“恰好位于”(Exact Location), 用 $L(x, y)$ 来表示“对象 x 恰好位于区域 y ”。该理论的基本定义如下:

定义 L1 (完全位于): $FL(x, y): = \exists z (\text{Part-of}(z, y) \wedge \text{Part-of}(x, z))$;

定义 L2 (部分位于): $PL(x, y): = \exists z (\text{Part-of}(z, x) \wedge \text{Part-of}(z, y))$;

定义 L3 (一致): $x \approx y: = \exists z (L(x, z) \wedge L(y, z))$;

这里, 定义 L1 是 L 的扩展, 定义 L2 为 L 的弱化, 定义 L3 为两个地理对象之间的“一致关系”。

拓扑学一般用来描述地理目标之间的相对位置关系。地理目标之间的相对位置关系可以用连通关系来描述, 而连通关系可以用边界来定义。

将这三个表达工具结合起来, 可以在地理本体中对地理位置、空间关系等进行形式化描述, 进而根据这些公理来实现空间推理操作。

3、地理本体语义关系推理

确定两个地理本体的语义关系, 实际上是确定两个地理本体概念之间的语义关系, 可以通过产生式规则推理的方式进行(崔巍, 2004 年)。推理过程所用的三个公理如下:

公理 1: 本体概念之间的关系是由概念之间的蕴涵或者同义关系组成, 概念的语义由其在属性集 $\{a_1, a_2, \dots, a_m\}$ 中的特征子集 $\{X_1, \dots, X_n\}$ 来描述。

公理 2: 如果两个概念 C_1 和 C_2 有共同的属性, 并且在所有属性上都有真值, 值域是相关的, 则两个概念 C_1 和 C_2 是同义关系。

公理 3: 对于两个概念 C_1 和 C_2 , 如果 C_1 的一个实例在 C_2 的所有属性上均有真值, 且值域是相关的, 反之不成立, 那么称 C_1 语义包含 C_2 。

推理过程为: 在两个本体系统中, 首先选择一个本体系统作为目标本体系统, 另一个是源本体系统; 然后, 从源本体系统中依次取出本体概念作为待转换概念, 把待转换概念在目标本体系统的体系中进行再分类, 确定待转换概念与目标概念的语义关系, 从而确定两个本体的语义关系。

第三节 基于地理本体的路径查询

路径查询问题是我们在日常生活中经常会遇到的一类问题, 也是地理信息科学中研究的基本问题之一。目前解决路径查询问题的途径主要有基于空间认知(Raubal, 2001 年)和基于地理本体这两种(曹菡, 2002 年), 并且大多数是研究在城市道路网络中的路径查找问题(曹菡, 2002 年; Timpf, 2002 年)。但在实际上我们经常需要在更大尺度范围(例如全国范围)内进行路径查询, 这必然会涉及到道路的不同等级以及对应的行政区域。

现在的问题是：我们要从某一个等级较低的行政区域（例如，某一个镇），查询到另一个省的某一个等级较低的行政区域（中的一个镇）的道路。由于在比例尺不同的空间数据库中对道路描述的详细程度不同，因此不能直接确定这样的两个地点之间的路径。我们可以通过建立道路本体系统和相应的行政区域本体系统，在此基础上通过地理空间推理的方法来解决这个问题。

这里的假设条件是：（1）假定我们这里只考虑公路交通，因为公路更有代表性；（2）假定各级道路都是连通的，并且每一级道路在各级行政区域都有进出口，一般情况下，等级相同的行政区和公路满足该条件，但是这一条件在二者等级不同时可能并不成立，例如一条省级高速公路并不一定在经过的每一个乡镇都有进出口。

这种方法的主要思路是：根据道路和行政区的不同等级，首先确定始点和终点所在的省份之间的省级公路；然后分别在这两个省范围内（前面已经假设是两个省），查找与该省级公路连通的县级公路，继而寻找与该县级公路连通的乡镇级公路，从而完成始点和终点之间的路径查询。

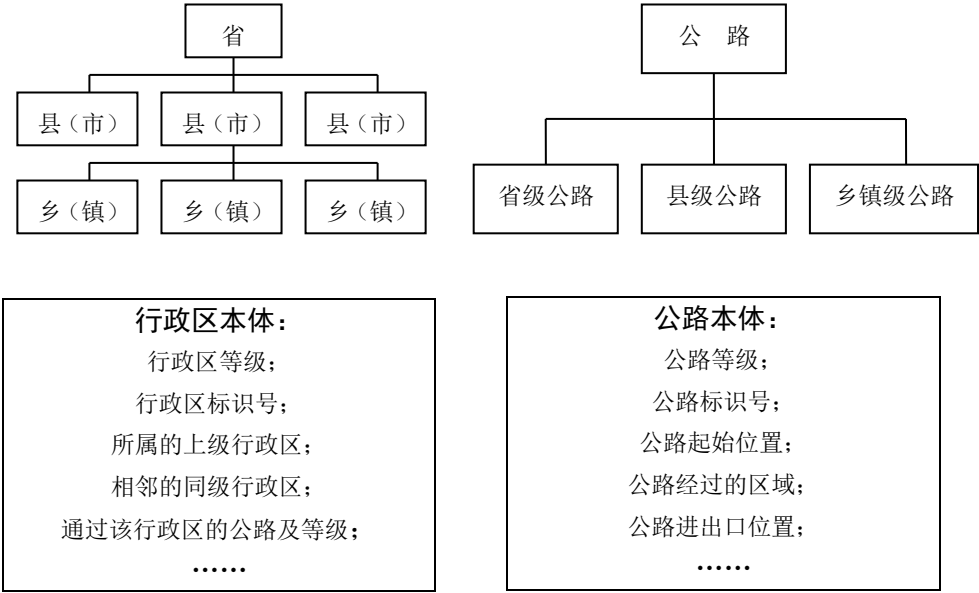


图 9-3 行政区本体与公路本体系统

基于本体的路径查询策略如下：

（1）建立道路本体系统和行政区本体系统，确定各行政区域的所属关系、各级公路与各行政区域之间的拓扑关系，以及各级公路进出口所在的行政区域（如图 9-3）。

（2）根据建立的行政区本体系统，确定路径查询中起始点以及目的地所在的乡镇、该乡镇所在的县市、及其所在的省份。这里假定起始点和目的地都与所在的乡镇级公路是连通的。

（3）若起始点和目的地都位于同一个乡镇，根据假设，它们都与乡镇公路连通，查询完成。

（4）若起始点和目的地位于同一个县的不同乡镇，查询同时通过这两个乡镇的乡镇公路，所查询到的该公路即为所求的路径。

（5）若起始点和目的地位于同一个省的不同县，则首先查询同时通过这两个县的县级公路。如果该县级公路恰好通过起始点和目的地所在的乡镇并且有进出口，则该县级公路即为查询的结果；如果该县级公路不经过起始点或目的地所在的乡镇，则分别查询与其连通的、并且通过起始点或目的地所在乡镇的乡镇公路，得到所查询的路径。

（6）若起始点和目的地位于不同的省份，则查询过程为：首先查询同时通过这两个省的省级公路，如果该省级公路恰好通过起始点和目的地所在的乡镇，并且有进出口，则该公路即为查询结果；否则，查询与该公路以及起始点和目的地所在县连通的县级公路，然后按照步骤（5）的方法得到最后的查询结果。