

基于地理本体的空间数据库概念建模

苏里¹, 朱庆伟^{1,2}, 陈宜金¹, 周丹卉³

(1. 中国矿业大学(北京)资源与安全工程学院, 北京 100083; 2. 西安科技大学测量工程系, 西安 710054;

3. 中国科学院沈阳应用生态研究所, 沈阳 110016)

摘要: 针对用 E-R 模型建立空间数据库概念模型存在的不足之处, 将本体论的思想和方法引入到地理信息领域。讨论了地理本体的意义及其表达语言。阐述了空间 E-R 模型的特点, 分析了地理本体和空间 E-R 模型之间的关系。提出了以地理本体为语义基础设计空间 ER 模型的思想, 并用扩展 OWL 表示的灌溉本体为例, 介绍了从地理本体出发, 设计空间 E-R 模型来建立空间数据库概念模型的方法。

关键词: 地理本体; 空间 E-R 模型; OWL; 空间数据库

Conceptual Modeling of Spatial Database Based on Geographic Ontology

SU Li¹, ZHU Qingwei^{1,2}, CHEN Yijin¹, ZHOU Danhui³

(1. School of Resource and Safety Engineering, China University of Mining and Technology, Beijing 100083; 2. Department of Surveying, Xi'an

University of Science and Technology, Xi'an 710054; 3. Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110016)

【Abstract】 Aiming at deficiency of building conceptual model of spatial database with E-R model, the ontological ideology and methodology are introduced into the geographic information domain. The exact meaning and its formal represented language of geographic ontology are discussed. Characteristics of spatial E-R model are clarified and relationship between geographic ontology and spatial E-R model is analyzed. The thinking of designing spatial E-R model based on geographic ontology as semantic foundation is presented. And a method of establishing conceptual model of spatial database by designing spatial E-R model from geographic ontology is introduced with an example about irrigation ontology represented by expanded OWL.

【Key words】 Geographic ontology; Spatial E-R model; OWL; Spatial database

在数据库概念模型的设计中, ER 模型的建立是一件费时又枯燥的工作, 需要分析人员不断与用户交流, 了解应用的需求, 而且开发人员对应用域不熟悉常常会出现对应用错误理解现象, 这给以后的设计工作带来困难, 并且, 这些理解上的错误在开发周期里, 越晚发现, 修改越困难, 耗费越大。另外, 相同领域的 ER 模型中存在着大量的共同之处, 如果能充分利用这些共性, 将大大减少工作量, 并提高 ER 模型的质量。改进的 ER 模型虽然可用于空间数据库的概念建模, 但是 ER 模型中实体或联系的语义是隐含的, 依赖于人为的解释和利用 ER 模型的应用程序, 这样就给不同 ER 模型之间的集成带来很大的困难。

针对上述的问题, 本文引入本体作为解决问题的一种有效方法, 探讨地理本体在空间数据库概念建模中的作用。

1 本体及地理本体

在哲学上, 一般认为本体论就是关于存在的理论。在信息科学领域, 本体论最有代表性的定义是: 本体(Ontology)是共享概念模型的明确的形式化规范说明^[1]。这个定义体现了本体的 4 层含义:

(1)概念模型: 通过抽象出客观世界中一些现象的相关概念而得到的模型, 其表示的含义独立于具体的环境状态。

(2)明确: 所使用的概念及使用这些概念的约束都有明确的定义。

(3)形式化: 本体是计算机可读的。

(4)共享: 本体中体现的是共同认可的知识, 反映的是相

关领域中公认的概念集, 它所针对的是团体而不是个体。

不同于哲学中的本体论, 信息科学中的本体论的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。

地理信息科学中的本体则兼具哲学本体和信息本体的双重含义。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, 其目的是产生有关地理世界结构的更好理解, 为空间数据库的建立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。而信息本体的含义体现在通过对共享的地理概念进行明确的形式化定义, 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。

2 地理本体的形式化语言表达

描述本体的语言不仅要有描述能力, 同时也应具备推理能力, 目前开发的本体语言主要是基于一阶逻辑和描述逻辑的。一阶逻辑是一种形式语言系统, 具有很强的表达能力。

作者简介: 苏里(1978-), 男, 博士生, 主研方向: 地理信息系统应用, 地图语义模型; 朱庆伟, 博士生、讲师; 陈宜金, 博士、教授; 周丹卉, 博士生

收稿日期: 2006-06-29 **E-mail:** bud_hui@sohu.com

描述逻辑是一种基于对象的知识表示的形式化,它是一阶逻辑的一个可判定的子集。描述逻辑的表达能力不如一阶逻辑,但它能提供可判定的推理服务,它保证推理算法总能停止,并返回正确的结果。一阶逻辑虽然具有很强的表达能力,但其推理过程复杂,不利于本体的检验;而描述逻辑的推理复杂度可知,更适用于本体检验。并且,描述逻辑的语法容易转换成 XML/RDF 形式,因此,基于描述逻辑的本体模型更适合于网络环境下的概念建模与知识共享。

OWL本身设计用于网络环境,其逻辑基础是描述逻辑,按照XML语言格式实现。OWL能够清晰地描述领域内概念的含义以及这些概念之间的关系。它遵循面向对象的思想,按类和属性的形式描述领域知识所包含的结构,即一个本体由类和属性构成。尽管OWL是表示本体的理想语言,因为对于一般本体来说,要表达的语义关系主要是part-of、kind-of、instance-of、attribute-of 4种关系,而OWL提供了丰富的建模原语,因而是足够的。但对于地理本体来说,除了表示地理概念的属性特征外,更为重要的是要表示地理概念的空间特征,尤其是空间关系。OWL语言在处理地理本体的空间特征时就无能为力,因为它本来就是面向一般本体的。属于地理本体所特有的问题,必须要扩展OWL才能表示^[2]。

下面是矿区中灌溉领域本体的部分OWL表示:

```
<owl:Class rdf:ID="&灌溉;渠道">
  <owl:cross rdf:resource="&灌溉;倒虹吸" />
  <owl:within rdf:resource="&灌溉;水闸" />
</owl:Class>
<owl:Class rdf:ID="&灌溉;干渠">
  <rdfs:subClassOf rdf:resource="&灌溉;渠道"/>
</owl:Class>
...
<owl:Class rdf:ID="&灌溉;农渠">
  <rdfs:subClassOf rdf:resource="&灌溉;渠道"/>
  <owl:cross rdf:resource="&灌溉;田块" />
</owl:Class>
<owl:ObjectProperty rdf:ID="&灌溉;灌溉">
  <rdfs:domain rdf:resource="&灌溉;农渠" />
  <rdfs:range rdf:resource="&灌溉;田块" />
  <owl:minCardinality>1</owl:minCardinality>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="&灌溉;控制水量">
  <rdfs:domain rdf:resource="&灌溉;水闸" />
  <rdfs:range rdf:resource="&灌溉;渠道" />
</owl:ObjectProperty>
...
```

在OWL里省略了概念的数据类型属性。这里共有渠道、倒虹吸、水闸、田块概念和控制水量、灌溉对象属性。渠道概念又可分为干渠和农渠等,它们的关系类似于面向对象模型中的继承。OWL中的属性分为两种:数据类型属性和对象属性。数据类型属性每个具体概念的属性,而对对象属性反映的是概念之间的关系。控制水量对象属性的domain是水闸,range是渠道,这说明水闸和渠道之间是控制水量关系。由于灌溉领域本体属于地理本体的范畴,因此每个概念之间除了语义关系外还有空间关系。例如水闸位于渠道中,渠道穿越倒虹吸,农渠穿越田块。

3 地理本体与空间 E-R 模型

3.1 空间 E-R 模型

E-R模型大多作为概念设计的数据模型应用在关系数据

库。许多学者改进了ER模型用以空间数据库的概念建模^[3,4],本文称之为空间ER模型,它与传统的ER模型的最大区别是地理实体和地理实体空间关系的表达。

地理数据描述的实体具有空间特性。地理实体不同于一般数据库的实体,一般实体是物理实体,而地理实体除了具有物理实体,还具有与物理实体相对应的空间实体。空间实体一般用点、线、面或 grid-cell、tin、image 像元表示。在传统的 E-R 模型中,用矩形表示实体,只能描述和表达地理实体的一个层面,即只能表达物理实体或只能表达空间实体。在空间 E-R 模型中要做的就是将物理实体和空间实体同时表示出来。与地理实体一样,地理实体间的关系也具有双重性,既具有物理实体间的关系,如拥有/属于关系、集/子集关系、父/子关系等,也具有空间实体间的关系,如拓扑关系等。这两种关系也需要在空间 E-R 模型中表达出来。

除此之外,空间 E-R 模型还具有 EER(extended E-R data model)的特征,用来表示地理实体间的分化与综合、聚集、范畴/类等关系,这些关系是由地理实体本身的特点决定的。

3.2 地理本体与空间 E-R 模型

地理本体是面向地理领域的概念模型,它包含领域内通用的、普遍的概念,并且规定了领域级别上的约束,这些约束可以被用来进行知识级别上的推理,因此,地理本体表达的是更高级别的信息需求。而空间 E-R 模型是面向数据的概念模型,和领域内特定的应用相关,设计时要考虑过程优化和系统性能等方面的因素,而且,在这个层次上,约束主要是数据操作上的约束,许多领域级别上的约束并没有被显示出来,所以空间 ER 模型表达的是较低级别的信息需求。可以根据地理本体中的概念设计空间 ER 模型,并生成最终的数据库模式,空间 ER 模型可以看作地理本体的视图,同一个地理本体可以有不同的视图。

因此,可以通过地理本体显示的、形式化的表达某一领域的概念模型,捕捉同一领域内空间 ER 模型的共性,提高同一领域内数据库设计的可重用性;并且通过把空间 ER 模型和地理本体结合起来,将地理本体作为空间 ER 模型的语义基础,使空间 ER 模型中实体、属性和联系隐含的语义显式的表达出来,方便空间 ER 模型间的集成。

4 基于地理本体的空间 E-R 模型

空间 ER 模型的实体在地理本体中均能找到相对应的概念,地理本体中的概念形式化地表达了实体的语义信息。本节以前面提到的矿区中的灌溉本体为例,说明如何根据地理本体设计空间 ER 模型。现在灌溉领域的专家已经设计好了灌溉本体,系统设计者正在着手设计空间数据库的概念模型,设计者可以按以下的步骤设计:

(1)参考灌溉领域本体中的概念,根据需求产生空间 ER 模型中的实体。若概念之间具有一般和特殊的关系,则相应的实体之间具有“is-a”关系。本例中,可以从本体中的概念“渠道”、“倒虹吸”、“水闸”、“干渠”、“农渠”、“田块”,确定空间 ER 模型对应的实体“渠道”、“倒虹吸”、“水闸”、“干渠”、“农渠”、“田块”。其中灌溉本体中的“渠道”和“干渠”、“农渠”之间存在一般和特殊的关系,在OWL中用 rdfs:subClassOf 表示。因此在空间 ER 模型中“渠道”和“干渠”、“农渠”之间要存在“is-a”的关系。

(2)参考灌溉本体中的对象属性,确定空间 ER 模型中对应实体之间的联系。例如对象属性“灌溉”,它的 domain 是“农渠”概念,range 是“田块”概念,表达了“农渠”灌溉“田块”的关系,因此,在空间 E-R 模型中实体“农渠”和“田块”之间应当具有灌溉的联系。同样根据对象属性“控制水量”,可以得出“水闸”实体和“渠

道”实体是控制水量的联系。另外参考灌溉本体中对对象属性的约束，可在空间 ER 模型中确定联系的基数。如对象属性“灌溉”上的约束表示一个农渠至少可以灌溉一个田块，用 owl:minCardinality 表示。因此，空间 ER 模型中对应的联系“灌溉”是 1:N 类型的。

(3)参考 OWL 中的数据类型属性，根据需要加入到空间 E-R 模型中对应的实体上作为其属性。数据类型属性利用 XML Schema datatype 定义的简单类型完成定义，实际上它们反映的是概念元素和 XML datatype 之间的关系。这些属性最终映射到数据库表格的字段。

(4)找出 OWL 语言构建的灌溉本体当中表示概念空间关系的建模原语，并把它们表示在空间 E-R 模型的联系中。空间 E-R 模型的联系包含语义和空间关系信息。本例中“渠道”概念的 owl:cross 表达“渠道”穿越“倒虹吸”，二者的空间关系“穿越”可表示在它们的联系中。同理“渠道”的 owl:within 表示“水闸”位于“渠道”中。“农渠”概念的 owl:cross 表示“农渠”穿越“田块”。OWL 中概念空间关系的表达是地理本体区别于一般本体的标志。

根据上述的步骤和方法得到最终的空间 ER 图，如图 1。

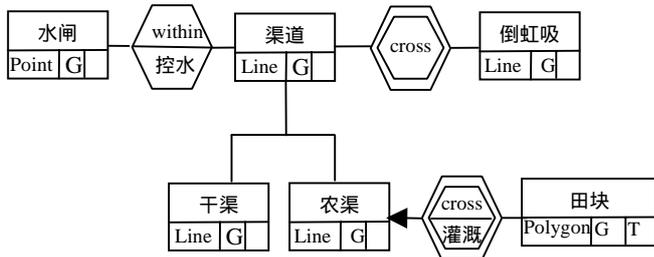


图 1 空间 E-R 模型

基于灌溉本体的空间 E-R 图中的实体不同于传统 E-R 模型中的实体，在矩形框中表示出它们的空间表现形式，例如渠道是线状的，水闸是点状的，田块是面状等。空间 E-R

模型的联系大体上分为 3 种：两个实体间只有语义联系，只有空间联系，或者二者都存在。例如农渠和田块之间的语义关系是灌溉，空间关系是穿越。渠道和倒虹吸只有穿越空间关系等。

5 结论

将地理本体作为空间 E-R 模型的语义基础，可以把空间 E-R 模型中实体、属性和联系隐含的语义显式的表达出来。当某一地理领域内已有成熟的本体存在，从该本体出发，不仅能简化空间 E-R 模型的设计，而且能提高空间 E-R 模型的质量，有利于不同空间 E-R 模型之间的集成。

参考文献

- 1 Gruber T R. Toward Principles for the Design Of Ontologies Used for Knowledge Sharing[J]. International Journal of Human and Computer Studies, 1995, 43 (5): 907- 928.
- 2 黄茂军. 地理本体空间特征的形式化表达机制研究[J]. 武汉大学学报(信息科学版), 2005, 30(4): 337-340.
- 3 Worboys M F, Hearnshaw H M, Magvire D J. Object-oriented Data Modeling for Spatial Databases[J]. Journal of Geographical Information System, 1990, 4(4): 369-383.
- 4 邵全琴, 周成虎. 海洋渔业数据建模的 E-R 方法研究[J]. 地理研究, 1998, 17(增刊): 108-115.
- 5 葛 亮, 谢祖铭. 基于本体的实体联系模型[J]. 计算机工程, 2004, 30(15): 53-54.
- 6 李生琦, 徐福缘, 徐 莹. 支持数据库 E-R 模型设计的本体论及其获取方法[J]. 计算机工程, 2004, 30(9): 7-8.

(上接第 83 页)

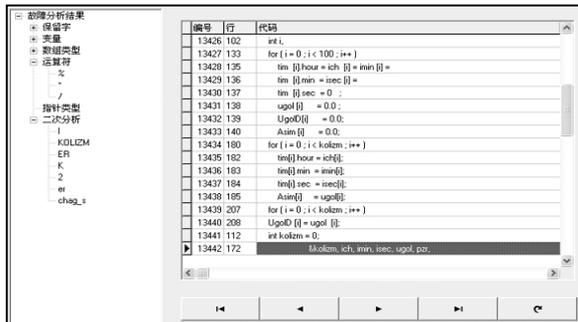


图 4 代码检测结果

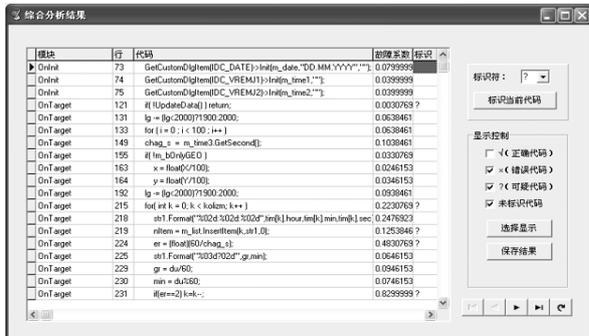


图 5 最终分析得到的故障代码集

4 结论

由于软件系统和软件故障的复杂性，因此针对软件故障进行的故障定位是软件故障诊断中研究的难点。本文针对目前软件故障定位方法的缺陷，提出了一种基于代码检测的软件故障定位方法，该方法已经成功运用于某软件密集型系统软件故障诊断，应用结果表明该方法能快速有效地定位软件故障，从而为软件密集型装备的软件故障诊断提供了一种新途径。

参考文献

- 1 朱小冬, 王小巍. 基于 CMM 的软件维护过程研究[J]. 计算机工程与应用, 2005, 41(29): 66-69.
- 2 Parikh G. Software Support, Management, and Evolution (SSME) in the Coming Decade and Beyond Opportunities & Challenges[C]// Proceedings of the 21st IEEE International Conference on Software Maintenance, 2005: 1-2.
- 3 吴 际. 软件故障定位和失效预测研究[D]. 北京: 北京航空航天大学, 2003.
- 4 李合平, 尚朝轩, 薛周成, 等. 新型电子装备软件故障诊断方法研究[J]. 微计算机信息, 2005, 21(3): 113-114.
- 5 宫云战, 高文玲, 李晓维. 软件故障暴露率的计算方法[J]. 计算机辅助设计与图形学学报, 2003, 15(6): 752-754.