

# 基于形式化本体的基础地理信息分类

李霖<sup>1,2</sup> 王红<sup>1,2</sup>

(1 武汉大学资源与环境科学学院, 武汉市珞喻路 129 号, 430079)

(2 武汉大学教育部地理信息系统重点实验室, 武汉市珞喻路 129 号, 430079)

**摘要:** 从本体概念出发, 以认知分类和现有的本体分类方法为基础, 将地理信息概念分为元概念和复合概念两大类, 并基于一定的约束条件, 根据地理信息概念的元属性, 以形式化方法实现基础地理信息的分类。

**关键词:** 地理信息; 形式化本体; 分类

**中图法分类号:** P208; P28

GIS 理论和技术的发展, 使得基础地理信息生产和地形图制图技术在近年来得到迅速发展。然而, 数据库中的数据分类存在同一要素称谓不同、分类不一致、分类系统多样性的情形。原有的分类系统是具有思维行为的人, 人可以通过视觉自动识别分类体系, 但是计算机却不能理解。如何让计算机根据一定的规则, 实现概念的自动分类, 并解决分类中存在的问题, 是目前迫切需要解决的问题之一。近几年来, 国内外许多学者<sup>[1-5]</sup>从地理本体概念、语义及分类等多个角度将本体论应用于地理信息科学研究, 这些均为地理信息的分类开辟了一条新的途径。

## 1 地理信息概念的形式化分类

### 1.1 分类依据

Rosch<sup>[6]</sup>等人从认知的角度出发, 要求分类时从物体的垂直和水平结构出发, 确定基本层、上一层和下一层的属性分类标准对物体进行分类。Guarino<sup>[2]</sup>等人对概念分类深入研究的基础上, 从概念的本体出发, 归纳出概念的元属性, 并以公式形式给出元属性的严格定义。在此基础上, 又讨论了元属性之间的关系和约束, 最终将研究结果作为概念分类的基本理论工具并提出一套完整的概念分类体系结构。

### 1.2 地理信息概念库的确定

地理信息概念库是建立概念分类的基础, 在

基础地理信息分类中, 概念主要来源于专家领域所制定的地形图图式、国土基础信息分类、全国土地分类、公路工程技术标准等与基础地理信息相关的标准。确定了地理信息概念库, 还需根据汉语词语的特点, 结合地理信息领域概念的特点, 将地理信息概念分为元概念和复合概念两大类。其中, 元概念是指地理信息概念从汉语构词结构上不可分或者拆分后概念所表达的涵义发生变化, 如植被、水系、水鹤等; 复合概念是指地理信息概念从汉语构词结构上可拆分并且拆分所表达含义不发生变化, 如常年湖、高速公路、地下管道等。

### 1.3 地理信息元属性

在基础地理信息分类中, 参照 Guarino 所提出的形式化本体分类方法, 根据地理信息概念的基本特性, 引入地理信息元属性, 主要包括抽象性、存在性、统一性、依赖性和拆分性, 分别用大写字母 *A*、*E*、*U*、*D* 和 *P* 表示, 通过在字母前加“+”和“-”分别表示概念“承载”和“不承载”此类元属性。其中, 抽象性是指地理信息概念与所指代的实体之间不存在一一对应关系; 存在性是指地理信息概念与所指代的实体在真实自然界中有具体的实物相对应; 统一性是指概念所表达的实体的每一个组成部分都以某种方式与实体的其他组成部分相连接, 同时在这种连接方式下不与任何其他的事物连接; 依赖性主要从以下两个方面来描述: 包含在部分-整体关系的整体和部件之间的关系, 其中部件依赖于整体而存在; 依赖物

体本身不是其一部分,但却因其存在;拆分性则是从地理信息概念库出发,根据概念库中的定义确定词的拆分性,如果概念  $\Phi$  是地理信息元概念,则概念  $\Phi$  不可拆分,反之,则可拆分。

1.4 约束和假设

元属性的定义为分类关系实施规定了一些具

有操作性的约束条件,就地理信息概念而言,元属性之间又存在着相互约束关系。表1中详细列举了地理信息概念元属性之间的包含关系,其中,单元格中“|”左侧的“√”表示所在行的元属性可以包含所在列的元属性,右侧的“√”表示所在列的元属性可以包含所在行的元属性,“×”则表示不能包含。

表1 地理信息元属性之间的包含关系

Tab.1 All Inclusion Relation Among the Geographic Information Meta-properties

包含关系	+A	-A	+E	-E	+U	-U	+D	-D	+P	-P
+A	√ √	√ ×	√ √	√ √	√ √	× ×	× ×	√ √	√ √	√ √
-A	× √	√ √	√ √	× ×	√ √	√ √	√ √	√ √	√ √	√ √
+E	√ √	√ √	√ √	× √	√ √	√ √	√ √	√ √	√ √	√ √
-E	√ √	× ×	√ ×	√ √	√ √	× ×	× ×	√ √	√ √	√ √
+U	√ √	√ √	√ √	√ √	√ √	× ×	× ×	√ √	√ √	√ √
-U	× ×	√ √	√ √	× ×	× √	√ √	√ √	× ×	√ √	√ √
+D	× ×	√ √	√ √	× ×	× ×	√ √	√ √	× √	√ √	√ √
-D	√ √	√ √	√ √	√ √	√ √	× ×	√ ×	√ √	√ √	√ √
+P	√ √	√ √	√ √	√ √	√ √	√ √	√ √	√ √	√ √	√ √
-P	√ √	√ √	√ √	√ √	√ √	√ √	√ √	√ √	√ √	√ √

1.5 形式化分析方法

根据以上关于元属性不同组合方案的意义,在地理信息元属性中,A、E、U、D、P分别取布尔值,共有32种可能的组合,根据表1中地理信息元属性之间的约束限制,实际的组合为8种。分析元属性之间的组合不难发现,不同的组合方案对应于地理信息分类中不同的分类层次,如表2所示。在地理信息概念分类中,根据地理信息概念各自的元属性,分析得到概念元属性所在的层,即可实现地理信息概念的形式化分类。

表2 地理信息元属性之间的结合

Tab.2 All Possible Combination Among the Geographic Information Meta-properties

+A	-A	+E	-E	+U	-U	+D	-D	+P	-P	Tier
								+	-	Tier 1
								-	+	Tier 2
+		+		+				-		Tier 3
+		+		+				+		Tier 4
-		+		+				-		Tier 5
-		+		+				+		Tier 6
-		+		-				-		Tier 7
-		+		-				+		Tier 8

1.6 复合概念约束条件

与其他概念库一样,地理信息概念库中存在大量的复合词,利用形式化分析方法可以得到地理信息的基本分类框架,但是通过形式化方法得到的同一级词汇中可能很多是基于不同角度对同一要素的划分,例如常年湖、淡水湖,这种情况下需要确定哪一类要素是基于概念本体的分类。

裴炳镇<sup>[8]</sup>提出对预分类好的领域词汇利用

“ $x + A$ ”规则建立概念之间的分类关系。通过“ $x + A$ ”方法可以得到概念和其父概念之间的关系,但当两个子类同属一父类时,仍存在分类结果不一致现象。

地理信息概念一般均为名词,因此复合概念的结构一般为形容词+名词,简称  $A + N$ ,前面的形容词主要用来限定后面名词的空间位置、功能、性质等。不难发现,名词前的限定形容词在概念分类中起重要作用,限定形容词与地理信息是自然要素或人工要素存在直接关系。自然地理要素是人所认识的外部世界,是自然界真实存在的,是人类利用其为人类服务的实体。因此,空间性和时间性两大特征为其最重要的特征,同一自然地理要素在不同的空间和时间,其性质可能会发生本质变化。在进行分类时,从不同的角度出发对要素分类会有不同结果,自然要素则优先考虑其空间性和时间性。人工要素是指人们依据对自然规律的认识,利用自然的物质条件创造出来的、自然界中原来没有的东西,即所谓“物化智能化”。因此,从一定意义上讲,人工要素是人类为了满足人类的某种需求而制造的,这些要素的存在并不因空间和时间而改变,因此,功能和用途成为人工要素分类的主要依据。根据以上分析,在分类中制定以下两个规则。

规则1 自然地理要素作为运动着的物质,空间性和时间性属性是其两个重要的且优于其他所有属性的属性。

规则2 人工要素是人为了满足自己的需求而创造的实体,其功能和用途是两个重要的且优

于其他所有属性的属性。

通过规则 1 和规则 2, 可以实现分类中地理信息复合概念的选取问题。另外, 对复合概念进行分析, 在分类中, 对于抽象不存在概念, 一般复合概念均位于元概念的上一层; 抽象存在概念和非抽象概念则恰恰相反, 复合概念位于元概念的下一层, 如表 2 所示。这也是概念的内涵与外延之间存在的反比关系, 即内涵越丰富, 外延越小, 内涵越贫乏, 外延越大。

### 1.7 概念之间关系的确定

地理信息概念库中, 概念之间的关系主要有同义关系、上下位关系、整体部分关系、属性宿主关系和实例关系等。利用概念的定义确定同义关系, 借助于语义通过形式化分析方法可以得到概念的上下位关系、整体部分关系, 但是属性宿主关系和实例关系则需要地理信息本体创建过程中通过其他定义得到。

## 2 分类实例

基础地理信息<sup>[9]</sup> 主要由自然地理信息中的地貌、水系、土地覆盖和社会地理信息中的居民地、交通、境界、特殊地物和地名以及用于地理信息定位的地理坐标系格网等基本信息构成。因此, 本文基于形式化本体的思想对自然地理信息中的水系及社会地理信息中的管线中要素的分类予以详细说明。

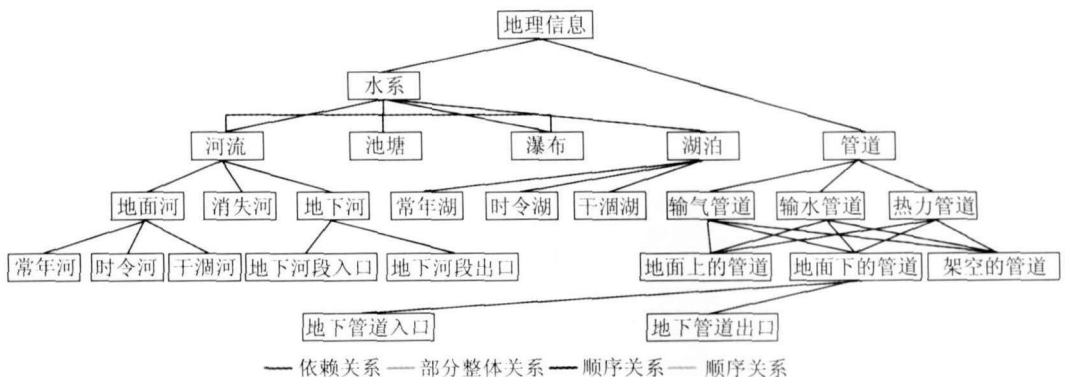
假设给定以下地理信息概念: 地理信息、管道、热力管道、输气管道、输水管道、地面上的管道、架空的管道、地面下的管道; 水系、河流、湖泊、池塘、瀑布、常年湖、时令湖、干涸湖、地面河、消失河、地下河、地下河段出入口、常年河、时令河、干涸河、地下河段出入口、地下河段出口

口、地下河段入口。对这些概念进行分析, 其中, 管道、水系、河流、湖泊、池塘、瀑布为元概念, 其他概念均为复合概念; 水系为一个抽象不存在的概念; 管道、河流、池塘、湖泊、瀑布为非抽象存在的概念; 复合概念输气管道、输水管道和热力管道是一组描述管道功能的概念; 地面上的管道、地面下的管道和架空管道是一组描述管道空间特性的概念; 常年河、时令河和干涸河是一组描述河流时间特性的概念, 地面河、地下河和消失河则是一组描述河流空间特性的概念; 常年湖、时令湖和干涸湖为一组描述湖泊时间特性的概念。同时, 根据复合概念的约束条件, 对于管道这一人工要素而言, 其限定形容词优先级为“功能> 空间性”; 对于河流和湖泊这两类自然要素而言, 其限定形容词优先级为“空间性> 时间性> 功能”, 分别用“+”和“-”表示限定形容词优先或落后于同层的形容词。根据以上分析可得到表 3 所示的分类结果表。在具体实施过程中, 结合概念的语义结构, 根据表 3 所提供的元属性, 结合概念的语义特征, 得到图 1 所示的所有要素分类结果图。

表 3 分类结果表

Tab. 3 Table of Classifying Result

概念名称	A	E	U	D	P	层次	约束规则
地理信息	+ A	- E	+ U	- D	+ P	1	1
水系	+ A	- E	+ U	- D	- P	2	2
管道	- A	+ E	+ U	- D	- P	5	5
功能 adj. + 管道	- A	+ E	+ U	- D	+ P	6	6+
空间性 adj. + 管道	- A	+ E	+ U	- D	+ P	6	6-
地下管道出入口	- A	+ E	- U	+ D	+ P	8	8
河流、湖泊、池塘、瀑布	- A	+ E	+ U	- D	- P	5	5
时间性 adj. + 湖泊	- A	+ E	+ U	- D	+ P	6	6
空间性 adj. + 河流	- A	+ E	+ U	- D	+ P	6	6+
时间性 adj. + 河流	- A	+ E	+ U	- D	+ P	6	6-
地下河段出入口	- A	+ E	- U	+ D	+ P	8	8



— 依赖关系 — 部分整体关系 — 顺序关系 — 顺序关系

图 1 分类结果图

Fig. 1 Figure of Classifying Result

### 3 结 语

本文通过分析地理信息概念的元属性,利用形式化方法将原本复杂的分类通过计算机完成,不仅节省了人力物力,而且有利于分类系统的更新。实验表明,分类的结果基本与人类认知结果相同,分类方法具有一定的可行性。但是,当分类中出现一些特殊情况,如果存在一个概念,从形式化分析方法得到这个概念和水系、植被同属第一层,但是从字面涵义理解却明显位于上一层,这种情况分类的形式化如何实现;同时,分类过程的细分,分类结果的完善,分类约束条件的限制,概念之间属性关系的定义,地理信息分类中汉字处理技术的改进,均有待于进一步研究。

#### 参 考 文 献

- [1] Smith B, Mark D M. Ontology and Geographic Kinds. The 8th International Symposium on Spatial Data Handling[C]. Vancouver, Canada, 1998
- [2] Guarino N, Welty C. A Formal Ontology of Properties[C]. The 12th International Conference on Knowledge Engineering and Knowledge Management, Berlin, Germany, 2000

- [3] 何建邦,李新通,毕建涛,等. 资源环境信息分类编码及其与地理本体关联的思考[J]. 地理信息世界, 2003, 1(5): 6-11
- [4] 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[J]. 测绘通报, 1998(8): 17-20
- [5] 宋炜,张铭. 语义网简明教程[M]. 北京: 高等教育出版社, 2004
- [6] Rosch E. Principles of Categorization, Cognition and Categorization[M]. Hillsdale, NJ: Lawrence Erlbaum, 1978
- [7] Guarino N, Welty C. Identity, Unity, and Individuality: Towards a Formal Toolkit for Ontological Analysis[C]. The European Conference on Artificial Intelligence, Amsterdam, Netherlands, 2000
- [8] 裴炳镇,陈晓明,胡熠,等. 一种建立中文概念分类关系的新算法[J]. 计算机工程与应用, 2004, 40(36): 18-21
- [9] 宋鸿运,王东华,商瑶玲,等. 1: 25 万国家基础地理数据的管理及应用[J]. 测绘通报, 2001(10): 34-35

第一作者简介:李霖,教授,博士生导师,研究方向为地理信息系统、本体论与数字地图制图工程等。

E-mail: lilin@whu.edu.cn

## Classification of Fundamental Geographic Information Based on Formal Ontology

LI Lin<sup>1,2</sup> WANG Hong<sup>1,2</sup>

(1 School of Resource and Environment Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(2 Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

**Abstract:** Setting about ontology concept, the paper defines the geographic information concept as meta-concept and compound concept based on the classification of cognitive and ontology taxonomy that exists. And according to meta-properties of geographic information concepts, it realizes the classification of fundamental geographic information with formal ontology method under some constraint conditions.

**Key words:** geographic information; formal ontology; classification

**About the first author:** Li Lin, professor, Ph. D supervisor. His main research orientations are GIS, ontology and knowledge engineering of digital cartography, etc.

E-mail: lilin@whu.edu.cn