

DOI: 10.11766/trxb201904090429

王腊红, 陈家赢, 汪善勤, 王天巍, 谭黄元. 基于本体的土壤系统分类逻辑表达与检索[J]. 土壤学报, 2020, 57(6): 1378–1386.

WANG Lahong, CHEN Jiaying, WANG Shanqin, WANG Tianwei, TAN Huangyuan. Logic Expression and Retrieval of Soil Taxonomy Based on Pedon[J]. Acta Pedologica Sinica, 2020, 57(6): 1378–1386.

## 基于本体的土壤系统分类逻辑表达与检索\*

王腊红<sup>1</sup>, 陈家赢<sup>1, 2†</sup>, 汪善勤<sup>1, 2</sup>, 王天巍<sup>1, 2</sup>, 谭黄元<sup>1</sup>

(1. 华中农业大学资源与环境学院, 武汉 430070; 2. 农业农村部长江中下游耕地保育重点实验室, 武汉 430070)

**摘要:** 中国土壤系统分类 (Chinese Soil Taxonomy, CST) 是建立在诊断层和诊断特性基础上的定量化土壤分类体系, 它的不断成熟为实现土壤类型自动化检索提供了理论基础。野外土壤描述与采样规范的形成成为土壤分类的语义规范提供了依据。目前, 我国已出现一系列基于 CST 的土壤类型检索系统, 但仍存在以下问题。首先, 现有的土壤类型检索系统仅注重分类规则的表达, 忽略了诊断对象、土壤类型与检索框架 (推理过程) 的耦合性对系统更新的影响。其次, 土壤信息的载体并不是单一的, 从空间结构上可分为土壤层次 (Horizon)、剖面 (Profile)、单个土体 (Pedon) 和聚合土体 (h), 但现有的检索系统并未将上述结构区分开来, 不利于土壤信息的管理。最后, 现有的检索系统均是通过传统计算机语言表达土壤特征的范围以及土壤特征之间复杂的逻辑关系, 表达方式繁琐且冗余。因此, 本文引入本体概念, 以土壤地理学和 CST 规则为理论基础, 分析土壤实体的空间结构及其与土壤类型、诊断对象之间的相互关系, 在此基础上建立了关于土壤实体、土壤特征和 CST 对象 (土壤类型与诊断对象) 的本体模型, 并定义了相应的谓词逻辑来表达三类本体模型的逻辑、隶属关系。本文采用 Python 语言实现了本体模型和谓词逻辑模型, 研发了 CST 中土纲到亚类的检索系统, 并采用湖北省土系调查数据完成系统测试。

**关键词:** 土壤系统分类; 本体; 谓词逻辑; 检索; 诊断对象

**中图分类号:** S155      **文献标志码:** A

## Logic Expression and Retrieval of Soil Taxonomy Based on Pedon

WANG Lahong<sup>1</sup>, CHEN Jiaying<sup>1, 2†</sup>, WANG Shanqin<sup>1, 2</sup>, WANG Tianwei<sup>1, 2</sup>, TAN Huangyuan<sup>1</sup>

(1. College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, China; 2. Key Laboratory of Arable Land Conservation (Middle and Lower Reaches of Yangtze River), Ministry of Agriculture and Rural Affairs of the People's Republic of China, Wuhan 430070, China)

**Abstract:** [Objective] At present, certain progress has been made in the research on a Chinese Soil Taxonomy (CST)-based Soil Type Retrieval System, but the research has come across the following problems: (1) Ignorance of the impact of the

\* 国家自然科学基金项目 (41877071, 41101192, 41471179)、国家科技基础性工作专项项目 (2014FY110200A16)、国家重点研发计划项目 (2016YFD0800907) 资助 Supported by the National Natural Science Foundation of China (Nos. 41877071, 41101192, 41471179), the Fundamental Research Funds for the Special Project (No. 2014FY110200A16) and National Key Research and Development Program of China (No. 2016YFD0800907)

† 通讯作者 Corresponding author, E-mail: chen.jiaying@mail.hzau.edu.cn

作者简介: 王腊红 (1994—), 女, 湖南湘潭人, 硕士研究生, 主要从事地理信息应用和土壤信息系统的研究。E-mail: wanglahong0201@qq.com

收稿日期: 2019-04-09; 收到修改稿日期: 2020-06-15; 网络首发日期 (www.cnki.net): 2020-10-12

coupling of diagnosis object, soil type and retrieval framework (Inference Process) on updating of the system; (2) Lack of consideration of division of the spatial structure of soil information carriers, which is un conducive to management of the soil information; and (3) Scope matching of the soil characteristics always expressed in traditional conditional nested statement and logical relationships between soil characteristics, making retrieval language cumbersome and redundant. 【 Method 】 In order to improve the above situations, this study introduced the concept of ontology, set soil geography and CST classification rules as its theoretic bases, analyzed spatial structure of soil entities and their relationship with CST objects (soil types and diagnostic objects) and eventually on such a basis, established an ontological model for relationship between soil entities and CST objects. To standardize the expression of soil characteristics, this paper defined soil attribute models, and divided soil characteristics into two categories, i.e. ordinary soil characteristics and complex soil characteristics. Moreover, this paper also defined corresponding predicate logics to express the relationships between types or models in logic and membership. 【 Result 】 This paper used Python language to realize construction of the ontology models and definition of the predicate logics, developed a type retrieval system covering all the four levels (Order, Suborder, group and subgroup), and cited the data of some single pedons representative of the Soil Series of China (Hubei volume) for test. And the test not only helped determine types of the pedons at all the four levels, but also recorded the entire retrieval processes, which facilitated analysis of the result later. 【 Conclusion 】 To compare with other existing retrieval models, this ontological model divides the carrier of soil information into four categories, i.e. Horizon, Profile, Pedon, and Polypedon, which facilitates management of the soil information and reduces complexity of the classification rules. Besides, the retrieval model separates the rules from the framework and can be characterized by high cohesion and low coupling, so that the model can better support the retrieval system in updating and expanding. The retrieval framework no longer takes soil characteristics as retrieval object, but encapsulates soil characteristics into soil entities, diagnostic objects, and soil types, thus elevating retrieval objects up to the level of category, which tallies more with human cognition. In brief, a method of constructing a retrieval model is proposed from the perspective of ontology in this paper. It can solve some problems of the existing retrieval models to a certain extent. However, the issue of semantics of soil characteristics and quantification of soil morphological characteristics remains to be a topic to work on. Therefore, to further improve the model, further efforts should be made in the next phase of the research to establish a soil characteristics semantic dictionary with complete metadata and information and introduce uncertain information as system auxiliary information.

**Key words:** Soil taxonomy; Ontology; Predicate logic; Retrieval; Diagnosis object

土壤分类是土壤科学的基础, 是土壤科学发展水平的标志, 是信息交流的媒介, 是土壤调查与制图的前提, 也是土壤科学应用推广的动力。以诊断层和诊断特性为基础的, 定量化和标准化的土壤分类方案得到国内外广泛认可, 国际上主流的分类系统均采用量化分类, 如欧洲学派代表的世界土壤资源参比基础 (World Reference Base for Soil Resources, WRB)、美国土壤系统分类 (Soil Taxonomy, ST) 和联合国粮食与农业组织 (Food and Agriculture Organization, FAO) 的土壤图图例单元。我国土壤系统分类经多年发展, 形成《中国土壤系统分类检索》(第三版)<sup>[1]</sup>, 分类定义了诊断现象、诊断特性和诊断层, 并在此基础上建立层次结构分明的土壤类型检索体系。中国土壤系统分类 (CST) 的不断完善使得计算机实现土壤类型的自动化检索成为可

能, 从而推动土壤学发展。Fisher 和 Balachandran<sup>[2]</sup> 研发了用于教学的土壤类型检索系统, 其主要针对土壤剖面 and 特征的类型检索。Galbraith 等<sup>[3]</sup> 在人工智库的基础上, 整合数据驱动和目标驱动两个推导过程, 基于确定性检索过程补充了专家知识, 一定程度上弥补了土壤类型检索的数据缺失问题。钟骏平等<sup>[4]</sup> 首次在我国系统分类的基础上, 建立了基于诊断层和诊断特性的 CST 检索系统。刘光等<sup>[5]</sup> 采用数据库管理、地理信息系统 (Geographic Information System, GIS)、面向对象编程技术、浏览器/服务器模式 (Browser/Server) 和多层数据库技术等研究方法, 开发了基于 Browser/Server 体系的 CST 检索系统。CST 检索系统要求输入的土壤剖面特征信息全面且准确, 但在实际应用中常因数据不完整而无法实施, 现有的系统均补充专家知识或环境发生信息

来辅助检索过程,该类检索为不确定过程。本文分析人工检索过程基础上,建立本体数据模型和通用逻辑谓词来实现量化的土壤类型检索。

汪善勤等<sup>[6-7]</sup>将诊断层和诊断特性作为一种专家知识,结合 GIS 等土壤发生环境信息,提出了土壤分类决策树的构造方式和推理方法。但该种土壤类型识别方式忽视了分类体系本身处于不断完善中,导致专家知识更新困难。土壤类型边界的更新是基于整个推理过程的更新,邱琳和李安波<sup>[8]</sup>基于 CST 的分类规则,利用规则引擎和 RuleML 规则描述语言降低检索规则和运行引擎之间的耦合。该研究实现了土壤类型自动检索的规则库与推理过程分离,增强了土壤类型自动检索系统的可拓展性,降低其因分类规则变化而引起的维护和升级的难度。上述检索系统的研究将土壤类型判别规则与数据分离开,能适应土壤分类的不断发展和更新。

此外,现有的检索系统均依赖大量的土壤数据,有效地组织与管理土壤数据是降低检索系统复杂度的关键。为了保证数据的通用性,多家土壤学组织联合制定了 SoilML 作为全球土壤数据交换的模型,针对 SOTER 数据库而提出 SoTerML 数据交互规范<sup>[9]</sup>。澳大利亚定义了 OzSoilML 用于土壤和地形数据的交换<sup>[10]</sup>。ANZSoilML 则是澳大利亚和新西兰共同发布的兼容 OGC 地理信息服务规范的土壤和地形数据交换数据模型<sup>[11]</sup>。但这些土壤信息交换的抽象模型均将土壤类型作为土体的基本属性特征,未将土壤类型的边界作为土壤交换模型的组成部分。

综上所述,基于 CST 的土壤类型自动检索的研究已经取得了一定的成果并在不断改进中,但这些检索模型均是以土壤特征为检索对象,通过传统计算机语言表达土壤特征的范围以及土壤特征之间复杂的逻辑关系,检索语言繁琐且冗余。其次,土壤信息的载体依据土壤空间结构可分为土壤层次(Horizon)、剖面(Profile)、单个土体(Pedon)和聚合土体(Polypedon),但现有的检索模型忽略了这一特点,并未将上述结构区分开来,不利于土壤信息的管理,并增大了分类规则表达的复杂度。因此,本文以土壤发生学,土壤地理学为理论基础建立关于土壤实体的本体模型表达土壤的空间结构;总结归纳土壤特征值的特点,建立土壤特征模型来统一土壤特征的类型,使得土壤特征在检索规则中被规范化表达;将 CST 检索规则中土壤特征的条件

约束以及特征之间的逻辑关系进行封装,将土壤类型与诊断对象视为规则边界,建立关于土壤类型和诊断对象的本体模型,将检索对象从土壤特征值的范围匹配转换为土壤实体与土壤类型、诊断对象之间隶属关系的判别,并通过定义的谓词逻辑来表达。这种判别方法使得规则表达的语言更为精简直观,与传统的条件嵌套语句相比更符合人类的认知。

## 1 土壤本体模型

### 1.1 土壤本体模型结构

土壤是覆盖地球表面的一层不连续疏松物质,因此具有空间结构。土壤学中,土壤空间结构分别从聚合土体、单个土体、剖面和土层进行表达,这四个基本概念同时具备空间上的逻辑关系。土壤剖面是指从地表向下挖掘至母质层所裸露的一段垂直切面;土层是土壤剖面上表现出的水平层状构造,它反映了土壤形成过程中物质的迁移、转化和累积的特点<sup>[12]</sup>;单个土体是土壤剖面的立体化形式,是体积最小的土壤三维实体<sup>[13]</sup>;聚合土体是两个以上的单个土体组成的群体,在土壤制图上为最小制图单位,在分类上则为基本分类单位。本文以土壤地理学与土壤分类学为理论基础,将土壤作为可解剖的实体,建立上述四种空间结构的土壤本体模型。进一步,明确结构之间的逻辑关系并定义相应的谓词逻辑来表达这些关系。

本文在四种土壤空间结构的基础上抽象出土壤结构的顶级本体概念——抽象土壤对象,该对象作为所有空间结构本体概念实例化的基础,并包含土壤的特征属性。土层根据其物质组成进一步划分出土壤基质、新生体、填充物和侵入体等实体对象。此外,模型还定义了地理信息类型和多媒体类型与单个土体关联,地理信息类描述土体的经纬度、高程等空间信息;多媒体类描述土体周围的自然景观照等多媒体信息。

上述土壤空间结构之间的关系包括继承(泛化)、实现、聚合和组成(强聚合)。其中,四种基本土壤空间结构均继承自抽象土壤对象,且每个土壤结构通过各自具体特征进行实例化。组成关系在空间上表现为连续性,是一种强聚合,如:聚合土体必然由连续的单个土体组成,剖面由若干连续的土层组成。聚合关系则表现为较弱的依赖性,如土

层中可能包含也可能不包含新生体。

## 1.2 土壤本体的谓词逻辑

本文基于土壤空间结构之间的组成与聚合关系定义了一组一阶谓词 INCLUDE 与 INCLUDED, 语义上理解为“包含”与“被包含”。组成和聚合关系为一对多的映射, 即 INCLUDE 返回被包含对象集合, INCLUDED 返回一个包含对象。例如, INCLUDE (p) 返回被剖面 p (下文均以 p 代表剖面) 包含的所有土层, INCLUDED (h) 返回包含土层 h (下文均以 h 代表土壤层次) 的一个剖面。该组谓词的定义使得检索系统既在独立表达土壤结构概念的同时又实现了它们之间的组合关系。

此外, 土壤实体的描述是建立在定量化的土壤特征的基础之上, 而土壤特征也依附于土壤实体而存在, 因此本文通过一组二元一阶谓词 GET\_ATTRIBUTE (soil, attribute) 和 SET\_ATTRIBUTE (soil, attribute) 表达土壤实体与土壤特征的关系, 其中 soil 为土壤实体, attribute 为土壤特征, 谓词 GET\_ATTRIBUTE (soil, attribute) 的语义为“获取土壤实体对象的某一特征值”, 其返回值为特征 attribute 的属性值, 谓词 HAS\_ATTRIBUTE (soil, attribute) 的语义为“判断某一特征值是否属于土壤实体对象”, 其返回值为布尔值。

本文除了在空间结构拓扑关系的描述中引入 GML 表达剖面、单个土体和聚合土体之间的空间关系, 还定义了基于深度的土层位置关系 RELATION ( $h_i, h_j$ ), 采用九交矩阵实现定量化描述。其定义为: 设  $h_i, h_j$  是同一剖面中任意两个土层,  $B$  表示土层的水平边界,  $I$  表示土层的内部,  $E$  表示土层的外部。若  $h_i, h_j$  交集为空, 则值为 0, 非空则为 1 (式 (1))。针对两个土层之间的 9 种位置关系: 上部非接触、上部接触、上部嵌入、下部非接触、下部接触、下部嵌入、包含、被包含以及重合, 矩阵 RELATION ( $h_i, h_j$ ) 具有 9 个有效值, 并为每个矩阵值定义了相应的二元一阶谓词表达其位置关系的语义: ABOVE ( $h_i, h_j$ )、ABOVE\_TOUCH ( $h_i, h_j$ )、TOP\_INTERSECT ( $h_i, h_j$ )、BELOW ( $h_i, h_j$ )、BELOW\_TOUCH ( $h_i, h_j$ )、BOTTOM\_INTERSECT ( $h_i, h_j$ )、CONTAIN ( $h_i, h_j$ )、CONTAINED ( $h_i, h_j$ )、OVERLAP ( $h_i, h_j$ ), 谓词的返回值为布尔型。基于土层的位置关系, 模型能表现土壤特征随深度的变化, 反应成土过程中物质的迁移、转化和累积

的特点。

$$\text{RELATION} \left( \begin{matrix} h_i h_j \\ \left[ \begin{array}{ccc} I(h_i) \cap I(h_j) & I(h_i) \cap B(h_j) & I(h_i) \cap E(h_j) \\ B(h_i) \cap I(h_j) & B(h_i) \cap B(h_j) & B(h_i) \cap E(h_j) \\ E(h_i) \cap I(h_j) & E(h_i) \cap B(h_j) & E(h_i) \cap E(h_j) \end{array} \right] \end{matrix} \right) \quad (1)$$

## 1.3 土壤特征模型及其谓词逻辑

土壤分类是依据土壤性状质与量的差异, 系统地划分土壤类型及其相应的分类级别, 从而拟定土壤分类系统<sup>[14]</sup>。土壤系统分类的分类规则核心是土壤特征的定性 with 定量化描述, 但土壤特征的类型并不是单一的, 本文将其归为两大类: 普通土壤特征和复合土壤特征。

普通土壤特征不具有子属性, 特征值单一; 复合土壤特征则包含多个子属性, 特征值由其多个子属性值组合而成。基于此, 本文通过建立土壤特征模型来描述上述两种类型的土壤特征, 并定义了土壤特征的谓词逻辑来表达类型之间的关系。土壤特征类型具有特征名称、特征值和量纲三个元素, 其中特征名称和量纲为字符型, 特征值的类型则抽象为一个顶级的特征值类型, 并基于该类型派生两个子类: 普通值和复合值类型, 分别描述普通土壤特征和复合土壤特征。特别地, 本文定义了二元一阶谓词 COMPOSITE\_TO (c\_attribute, attribute) 表达复合特征类型与普通特征类型的组合关系, c\_attribute 为复合土壤特征名称, attribute 为 c\_attribute 的子特征名称, 谓词返回 attribute 的特征值。

普通土壤特征值类型即简单值类型, 包括数值型、字符型和枚举型等。数值型土壤特征的谓词逻辑包括基本的四则运算、不等式运算、一元函数运算、集合运算等; 字符型土壤特征的谓词逻辑包括字符的连接、索引、截取、匹配、查找、大小写转换等; 枚举型的特征值描述是建立在数值型与字符型谓词逻辑的基础上, 通过一组离散值来限定特征的值域范围, 且取值唯一。

复合土壤特征的谓词逻辑是建立在普通土壤特征的谓词逻辑的基础上, 进一步处理多个属性值之间的逻辑关系。复合土壤特征类型的典型实例包括土壤的机械组成与土壤颜色。

机械组成是土壤中各粒级占土壤重量的百分比组合, 包括黏粒 (<0.002 mm)、粉粒 (0.002~0.05 mm) 和砂粒 (0.05~2 mm)。本文基于普通值类型定义了

颗粒类型并派生三个子类：黏粒类型、砂粒类型和粉粒类型；在机械组成基础上，根据不同的颗粒组成划分的土壤质地。基于土壤质地和土壤的机械组成之间存在映射关系，建立不同质地之间的相互比较谓词。针对 CST 土壤质地划分的 12 种类型，定义了 12 个枚举值来限定 12 种质地名称，并通过一元一阶谓词 DECIDED\_TO( $m$ ) 表达质地( $t$ ，下文均以  $t$  代表质地)与机械组成( $m$ ，下文均以  $m$  表示机械组成)的单向依赖关系，语义即根据机械组成获得质地类型，谓词返回质地的属性值。同时在检索过程中存在质地类型之间的对比，如“CaCO<sub>3</sub> 相当物颗粒大小较壤质更黏”。针对比较关系，定义三元一阶谓词 TEXTURE\_GREATER( $h, t, \text{particle}$ )，其中  $\text{particle}$  为枚举型的颗粒级别列表，即[‘clay’, ‘silt’, ‘sand’]，TEXTURE\_GREATER( $h, \text{‘壤质’, ‘clay’}$ ) 语义为“土层  $h$  的质地较壤质更黏”。该谓词根据质地划分规则识别并返回一个存储了较“壤质”更黏的所有质地类型的集合，然后判断  $h$  的质地是否属于该集合中，最后返回一个布尔型表示结果的真假。

CST 采用孟塞尔颜色系统描述土壤颜色，其是建立在色相、明度、彩度三个特征上的颜色立体<sup>[15]</sup>。与机械组成类似，孟塞尔颜色类型由孟塞尔色调类型、孟塞尔明度类型和孟塞尔彩度类型三种类型组合而成。土壤颜色的量化描述可视为对颜色特征值范围的匹配。其中，明度与彩度为数值型，其范围的匹配即数值类型的不等式运算。色调是一个被均分为 40 等分的圆盘且随顺时针方向具有一定的排列顺序，因此色调的描述可视为角度值范围的匹配。本文将所有色调赋予角度值，从 5Y 开始顺时针方向递增，相邻色调之差为 9 度，即 5Y=0, 7.5Y=9, 10R=18，以此类推。

针对上述颜色属性的描述形式，本文定义了 MUNSELL\_VALUE( $h, a, b$ )、MUNSELL\_CHROMA( $h, a, b$ ) 和 MUNSELL\_HUE( $h, a, b$ ) 三个三元一阶谓词分别表达明度、彩度和色调的范畴匹配，其语义即判断土层  $h$  的值是否在  $a \sim b$  的范围内，返回值为布尔型。色调的量化描述包括比较形式和范围形式，MUNSELL\_HUE( $h, a, b$ ) 具有 2 种语义：MUNSELL\_HUE( $h, \text{‘2.5YR’, ‘5Y’}$ ) 即判断土层  $h$  色调是否在 5Y~2.5YR 范围内，MUNSELL\_HUE( $h_i, h_j, \text{‘5Y’}$ ) 即判断土层  $h_i$  的色

调是否在 5Y 与  $h_j$  的色调值范围内。此外，CST 关于明度和彩度的描述除了两个土层的对比之外，还存在同一土层中干态与润态对比的情况。本文定义了一元一阶谓词 VALUE\_MOISTURER( $h$ ) 和 CHROMA\_MOISTURER( $h$ ) 表示土层  $h$  的明度与彩度的干润态比较，其语义即判别土层  $h$  的润态明度是否大于干态明度或润态彩度是否大于干态彩度，返回值为布尔型。

## 2 CST 本体模型

### 2.1 CST 本体模型结构

基于 CST 的本体模型是对 CST 分类体系的结构化描述，它表达了诊断对象和土壤类型及其之间的逻辑关系，系统基于该模型建立了分类规则框架。土壤诊断层和诊断特性及其与各种成土因素空间变异的关系是土壤类型划分的唯一标准<sup>[1]</sup>，从土纲到亚类的所有土壤类型的划分均依赖于诊断对象，在现有的检索模型中其两者存在高度耦合关系。因此，本文的 CST 本体模型将诊断对象与土壤类型剥离。基于 CST 对象的顶级本体概念——抽象 CST 对象类型派生出诊断对象类型和土壤类型，土壤类型结构化表达分类体系中土纲到亚类的所有土壤类别，诊断对象类型派生了诊断层类型、诊断特性类型和诊断现象类型，分别描述诊断层、诊断特性和诊断现象。在 CST 检索系统中，土壤本体模型的实例作为被检索对象，CST 本体模型作为类型概念规则边界，两者通过两个抽象类型（抽象土壤对象类型与抽象 CST 对象类型）之间的谓词逻辑 IS\_INSTANCE( $\text{soil}, \text{cst}$ ) 连接，其语义为：“土壤对象  $\text{soil}$  满足 CST 对象  $\text{cst}$  的规则边界”，返回值为布尔型。通过 IS\_INSTANCE 建立土壤个体模型与 CST 类型模型之间的关系判别，表达了土壤个体与诊断对象，土壤类型之间的隶属关系。

### 2.2 特征边界的约束规则

系统论思想表明，结构必须依赖边界定义，任何一项科学研究提出的假设均不是无中生有，均要建立在先前科学发现的基础之上<sup>[16]</sup>，土壤分类同样如此。CST 的土壤类型、诊断对象等概念是建立在其本身的特征属性范围的基础之上，特征属性的范围实际上是对特征属性变幅的规定，即特征边界约束。每个特征属性的边界约束均可通过对应的特征

谓词逻辑进行表达，如“酸性湿润淋溶土”中 pH 的边界被约束为  $0 \sim 5.5$ ，该谓词逻辑基于不等式运算和集合运算表示为  $(\text{pH} \geq 0) \cap (\text{pH} \leq 5.5)$ ，又如“红色酸性湿润淋溶土”中孟塞尔色调的边界为  $5R \sim 5YR$ ，用孟塞尔特征类型的谓词逻辑表示为  $\text{MUNSELL\_HUE}(h, 5Y, 5YR)$ 。

但 CST 中的范畴对象的边界不是针对单个特征属性的约束，是建立在一系列的特征约束集合之上。它是由一系列条件（即特征属性的约束）通过组合和替代关系形成复合概念，采用逻辑运算关系组织条件集合。基于此，本文定义条件类型（Condition）规范表达 CST 的本体边界。条件类型具有三个属性：（1）‘description’：条件的文字描述；（2）‘sub\_condition’：子条件，条件类型，它使得条件类型（Condition）形成嵌套结构；（3）逻辑综合（synthesis）：将土体在‘sub\_condition’中的所有匹配结果根据组合与替代关系进行组合，返回值为布尔型或为空，若判断结果为空则说明土体中 CST 本体所必须的土壤特征值缺失；（4）‘result’：判断结果，当‘sub\_condition’不为空时属性值等于‘synthesis’的值，否则为单个条件的判断结果。至此，CST 中土壤类型和诊断类型的表达均是将其土壤特征的边界表达为 Condition 类型。

### 2.3 土壤类型的谓词逻辑

CST 检索表现为一个树状结构且检索有一定的检索顺序，该顺序使得土壤类型之间存在相应的关系，如上下位关系、前置条件关系和后置条件关系。给定概念  $c_1$  和  $c_2$ ，若  $c_2$  的外延包含  $c_1$  的外延，则认为  $c_1$  和  $c_2$  具有上下位关系，称  $c_2$  为  $c_1$  的上位概念（hypernym）， $c_1$  为  $c_2$  的下位概念（hyponym）<sup>[17]</sup>，本文通过二元一阶谓词  $\text{IS\_A}(c_1, c_2)$  描述上下位关系。例如“水耕人为土是一种人为土”，则“人为土”即为“水耕人为土”的上位概念，“水耕人为土”为“人为土”的下位概念。前置条件关系与后置条件关系表达基于同一上位概念的不同土壤类型之间的检索顺序，本文通过二元一阶谓词  $\text{PRE\_CONDITION}(c_1, c_2)$  和  $\text{POST\_CONDITION}(c_1, c_2)$  进行描述。例如“人为土”类型发生的前提条件是不为“有机土”，即  $\text{PRE\_CONDITION}(\text{Histosols}, \text{Anthrosols})$  返回的布尔值为假。又如，当土体满足“人为土”后继续分类，则结果一定不是“灰土”，即  $\text{POST\_CONDITION}(\text{Anthrosols},$

$\text{Spodosols})$  为假。

### 2.4 土壤实体和 CST 类型的谓词逻辑

在检索体系中，土壤实体作为输入对象传入 CST 类型中，并返回一个检索结果。如土体作为被分类对象，在土壤类型中得到一个分类结果；剖面 and 土层作为被诊断对象，分别在诊断特性类型和诊断层类型中得到一个诊断结果；因此，土壤实体与 CST 类型之间存在一个实例与类型的映射关系。本文定义了谓词  $\text{IS\_INSTANCE}(c_1, c_2)$  来定义这种映射关系，其中  $c_1$  表示土壤实体对象， $c_2$  表示 CST 类型，其语义为“ $c_1$  是否为  $c_2$  的实例”。当执行  $\text{IS\_INSTANCE}(c_1, c_2)$  时， $c_1$  根据  $c_2$  的特征边界（Condition），调用谓词  $\text{GET\_ATTRIBUTE}(c_1, \text{attribute})$  获取对应的特征值，并判断该特征值是否在类型边界内，结果返回 Condition 的‘result’值作为  $\text{IS\_INSTANCE}(c_1, c_2)$  的结果。

## 3 基于本体逻辑的土壤类型检索

### 3.1 土壤类型检索的过程与结果表达

遵照 CST 分类体系的树形结构，本文从土纲到亚类通过广度优先的遍历策略进行检索，当土体在某个土壤类型中的判别结果为真，则停止在当前级别的检索并转入该土壤类型的所有下位关系土壤类型中，按照前置条件与后置条件关系的顺序继续检索，直至检索到亚类结束。

检索结果的结构与 CST 体系的树形结构相同，其节点的显示顺序依照土体在 CST 体系中的检索路径。当土体在土纲完成检索后，首先在结果树中返回一个根节点，该节点存储了检索级别以及检索结果，如“土纲判断-人为土”。基于根节点，系统依次返回土体在人为土及其之前的所有土纲类型的判别结果，作为第一层子节点，这些子节点是基于条件类型的对象，存储了土壤类型的所有条件描述、逻辑关系及判别结果。在第一层子节点的最后，系统返回土体在亚纲的检索结果，如“亚纲判断-水耕人为土”，并将该节点作为父节点，继续派生亚纲中土壤类型的判别结果，以此类推直至亚类，示例如图 1。

### 3.2 土壤类型检索的系统实现与测试

本文基于土壤本体模型、CST 本体模型和土壤特征模型及其谓词逻辑，通过 Python 语言实现土壤类型检索系统，模拟检索过程，并用湖北省的土系数



## 4 结论与展望

本文基于土壤地理学和 CST 分类知识,从领域本体的角度分析土壤的空间结构以及与诊断对象、土壤类型的隶属关系,建立了土壤实体的本体模型、土壤类型和诊断对象的本体模型和土壤特征模型,使用 Python 语言完成了模型的定义。总结在 CST 检索过程中存在的不同判断比较等结构,提取土壤分类检索中的基础谓词,定义相应的关系谓词表达模型内部类型之间的逻辑关系,表达土壤实体与 CST 对象之间的隶属关系,采用谓词逻辑的表达方式来实现土壤类型的检索,在本体概念的基础上建立更加符合自然语义土壤系统分类检索系统。

与其他已有的检索模型相比,本文的检索模型具有以下三方面的优点:首先,本文从土壤实体空间结构的角将土壤信息的载体区分为聚合土体、单个土体、剖面、土层和土壤基质等结构,建立相互之间的组合关系,建立土壤个体类型判断的表达方式,并能降低分类规则表达的复杂度。其次,模型在结构上具有高内聚低耦合的特性,将检索规则与检索框架剥离开来,能更好地支撑检索体系的更新与扩展;最后,从知识判别的角度来看,本文不再以土壤特征作为判别对象,而是将其封装在土壤实体、诊断对象以及土壤类型之中,将判别对象提升至范畴的高度,通过逻辑谓词直接表达对象之间的隶属关系,即“是”或“不是”、“属于”或“不属于”,与传统的条件嵌套语句相比,这种判别方式更符合人类的认知。

虽然本文的检索系统在模型构建上已基本完备,但系统在实现过程中仍存在不足之处,主要包括以下两方面:1)土壤特征的语义问题。土壤信息来源广泛,获取的方法也不同。例如铁的测定方法,包括邻菲罗啉比色法、原子吸收光谱法等。不同方法得到的土壤特征虽然名称相同,但语义上存在差异,因此本文系统中所有的土壤特征需要通过建立语义字典进行统一规范化表达,其字典包含特征的唯一标识(ID)、特征名称(Name)、语义描述(Description)、获取方法(Analysis method)和量纲(Unit)。2)形态学特征无法量化的问题。如水耕表层中的“糊泥化(Puddling)”缺乏具体定义,无法量化表达。此外,有些对属性值变化描述过泛,如干旱正常盐成土的条件中有“忽高”、“忽低”、“不

规则”缺少具体的变化界限,检索系统需要构造多种谓词表达式及其替代关系。

本文对土壤进行确定性的类型识别,分类结果的准确性依赖于土壤数据的精度。但在实际情况中土壤数据存在误差,如野外观测误差、专家经验性判断、实验室分析的仪器误差、人工误差等。土壤数据的缺失也会影响检索结果,因此在接下来的研究中可考虑添加补充替代信息,如用计量土壤学信息补充土壤形态学信息,借助 RS 和 DEM 信息补充土壤发生环境信息,通过现代传感器手段,如 X 射线荧光光谱分析(X-Ray Fluorescence spectrum method, XRF)获取的土壤重金属作为辅助信息等。此外,还可加入不确定信息,如模糊信息、未知信息和随机信息等,将系统扩展为不确定性分类系统,综合统计学理论,实现对土壤类型的预测并给出相应概率。

## 参考文献 (References)

- [1] Chinese Soil Taxonomy Research Group, Institute of Soil Science, Chinese Academy of Sciences, Cooperative Research Group on Chinese Soil Taxonomy. Keys to Chinese Soil Taxonomy[M]. 3rd ed. Hefei: University of Science and Technology of China Press, 2001. [中国科学院南京土壤研究所土壤系统分类课题组, 中国土壤系统分类课题研究协作组. 中国土壤系统分类检索[M]. 第3版. 合肥: 中国科学技术大学出版社, 2001.]
- [2] Fisher P F, Balachandran C S. Stax: A turbo prolog rule-based system for soil taxonomy[J]. Computers & Geosciences, 1989, 15 (3): 295—324.
- [3] Galbraith J M, Bryant R B, Ahrens R J. An expert system for soil taxonomy[J]. Soil Science, 1998, 163 (9): 748—758.
- [4] Zhong J P, Min Y, Jiang P A. Keys and data base of Chinese Soil Taxonomy[J]. Journal of Xinjiang Agricultural University, 1999, 22 (3): 215—218. [钟骏平, 闵勇, 蒋平安. 中国土壤系统分类检索及数据库系统[J]. 新疆农业大学学报, 1999, 22 (3): 215—218.]
- [5] Liu G, Li S D, Zhao Y, et al. The query system of Chinese Soil Taxonomy based on browser/server model[J]. Research of Soil and Water Conservation, 2002, 9 (4): 45—48. [刘光, 李树德, 赵焯, 等. 基于 Browser/Server 开发的中国土壤系统分类检索系统[J]. 水土保持研究, 2002, 9 (4): 45—48.]
- [6] Wang S Q, Zhou Y, Yu L, et al. XML-based metadata model for retrieval system of Chinese Soil Taxonomy[J]. Acta Pedologica Sinica, 2005, 42 (6): 889—895. [汪善勤, 周勇, 于雷, 等. 基于 XML 的土壤系统分类检索系统的元数据模型[J]. 土壤学报, 2005, 42 (6): 889—895.]

- [ 7 ] Wang S Q, Zhou Y, Zhang G L. Designing of soil classification expert system based on geographical information system[J]. *Acta Pedologica Sinica*, 2005, 42 ( 5 ): 705—711. [汪善勤, 周勇, 张甘霖. 基于 GIS 的中国土壤分类专家系统设计[J]. *土壤学报*, 2005, 42( 5 ): 705—711.]
- [ 8 ] Qiu L, Li A B. The study on automatic retrieval for soil classification based on rule engine[J]. *Chinese Journal of Soil Science*, 2011, 42 ( 5 ): 1031—1034. [邱琳, 李安波. 基于规则引擎的土壤类型自动检索的研究[J]. *土壤通报*, 2011, 42 ( 5 ): 1031—1034.]
- [ 9 ] Pourabdollah A, Leibovici D G, Simms D M, et al. Towards a standard for soil and terrain data exchange: SoTerML[J]. *Computers & Geosciences*, 2012, 45 ( 6 ): 270—283.
- [ 10 ] Simons B A, Wilson P L, Jacquier D. Ozsoilm1: An Australian standard for exchange of soil data. Joint SSA and NZSSS Soil Science Conference 2012.
- [ 11 ] Simons B, Wilson P, Ritchie A, et al. Anzsoilm1: An Australian-New Zealand standard for exchange of soil data. *Geophysical Research Abstracts*, Proceedings EGU General Assembly, 2013: 6802.
- [ 12 ] Zheng D. Dictionary of geographical regionalization and planning[Z]. Beijing: China Water and Power Press, 2012. [郑度. 地理区划与规划词典[Z]. 北京: 中国水利水电出版社, 2012.]
- [ 13 ] Liu F Z, Li Y J. Technology of soil monitoring and analysis[M]. Beijing: Chemical Industry Press, 2015. [刘凤枝, 李玉浸. 土壤监测分析技术[M]. 北京: 化学工业出版社, 2015.]
- [ 14 ] Wang G. *Agrology*[M]. Beijing: Higher Education Press, 2009. [王果. 土壤学[M]. 北京: 高等教育出版社, 2009.]
- [ 15 ] Xu F S, Xu H S, Wang Y. A new conversion algorithm between munsell color order system and CIE1931 standard colorimetric system[J]. *Acta Photonica Sinica*, 2007, 36 ( 4 ): 650—654. [徐芙珊, 徐海松, 王勇. 孟塞尔色序系统与 CIE1931 标准色度系统转换新算法[J]. *光子学报*, 2007, 36 ( 4 ): 650—654.]
- [ 16 ] Liu C D. An ontological attribute of linguistic structure: Structure-boundary unity[J]. *Foreign Language and Literature*, 2009, 25 ( 4 ): 9—14. [刘辰诞. 认知语言学视角下的语言结构本体属性: 结构—边界统一体[J]. *外国语文*, 2009, 25 ( 4 ): 9—14.]
- [ 17 ] Liu L, Cao C G, Zhang C X, et al. Sense recognition research of hyponymy based on concept space[J]. *Chinese Journal of Computers*, 2009, 32( 8 ): 1651—1661. [刘磊, 曹存根, 张春霞, 等. 概念空间中上下位关系的意义识别研究[J]. *计算机学报*, 2009, 32 ( 8 ): 1651—1661.]
- [ 18 ] Zhang G L, Wang T W. *Soil series of China: Hubei*[M]. Beijing: Science Press, 2017. [张甘霖, 王天巍. 中国土系志 湖北卷[M]. 北京: 科学出版社, 2017.]

( 责任编辑: 檀满枝 )