

卜伟琼,方 遼,张晓玲,等. 基于本体的柑橘病虫害知识模型构建[J]. 江苏农业科学,2013,41(10):363-366.

基于本体的柑橘病虫害知识模型构建

卜伟琼,方 遼,张晓玲,陈益能

(湖南农业大学信息科学技术学院,湖南长沙 410128)

摘要:针对现有农业病虫害知识系统的领域局限,以及知识重用与共享的困难,提出了构建柑橘病虫害知识本体的一套工程方法。在对柑橘病虫害领域知识分析的基础上,给出了柑橘病虫害知识本体模型的形式化定义及构建元数据语义关系定义,确定了本体构建框架及核心本体,并运用本体开发工具 Protégé 实现了柑橘病虫害知识本体模型的构建,为后续的本体学习及应用打下基础,并为进一步的本体规则推理与检索研究提供参考。

关键词:本体论;领域本体;本体建模;规则推理;柑橘病虫害

中图分类号: S436.66;S431.9 **文献标志码:** A **文章编号:** 1002-1302(2013)10-0363-04

农作物病虫害是我国的主要农业灾害之一,它具有种类多、影响大、时常暴发成灾的特点,发生范围和严重程度给农业生产常造成重大损失。柑橘作为我国主要经济作物之一,被广泛种植于全国 20 个省、市、自治区,1000 多个县,已经成为很多地区发展农村经济的支柱产业,为加强农民和农技人员对柑橘病虫害的认识,有效指导病虫害的防治工作,柑橘病虫害知识和信息的表示与共享技术显得尤为重要。相关学者对农作物病虫害知识开展了研究,唐惠丽等采用基于模糊推理的知识表示方法^[1];王霓虹等基于产生式规则设计知识库^[2];刘鹤等采用基于案例推理的知识库,根据目标案例提示在以前源案例中提取信息,稍加修正后进行应用^[3]。但这些方法都存在表示知识有限、不能很好的共享和重用以及结构单一等弊端。在这种情况下,引入了本体概念。2000 年,联合国粮农组织提出了 AOS(农业本体服务)研究计划,极大促进了农业本体论的发展,并开发了渔业本体、作物-有害生物本体;2001 中国农业科学院农业信息研究所开始了农业本体论的研究,开展了土壤学、花卉学等领域的本体模型研究;戴才萍等构建了水稻病虫害本体^[4];姜大庆等基于本体构建了蔬菜病虫害知识库^[5]。但目前大多数相关研究都存在

领域的局限性。本研究在对柑橘病虫害知识模型分析的基础上,引入本体概念,给出了柑橘病虫害知识本体模型的形式化定义及构建元数据语义关系定义,运用本体开发工具构建了柑橘病虫害知识核心本体,为促进柑橘病虫害防治知识共享和应用打下了基础,并为下一步的本体推理和检索等工作提供参考。

1 本体论

本体最早起源于哲学,随着人工智能的发展,逐渐被用于知识工程和信息科学等领域,用于知识的表示和组织。其定义和内涵也随之改变和不断完善。最普遍被人接受的一种定义认为本体是“共享概念模型的明确的形式化规范说明^[6]。”本体具有概念化、明确、形式化、共享四大特征,因而被广泛应用于知识的表示和组织。本体包含 5 个基本建模元语^[7]:类、关系、函数、公理和实例。类通常也被称作概念,可以指任何事物;关系是描述概念相互之间的交互关系;函数是一类特殊关系,在这种关系中,前 $n-1$ 个元素可以唯一决定第 n 个元素;公理代表永真断言;实例代表元素。随着本体在各个领域的广泛应用,出现了多种本体构建方法,主要有骨架法、TOVE 本体、G&Fox 方法、KACTUS 及 Bernaras 方法、XHEMICALS 本体与 METHONTOLOGY 方法、SENSUS 本体及方法、IDEF5 方法与 7 步法等^[8]。经典的本体描述语言有 RDF 和 RDF(S),OLL,DAML,OWL,KIF,SHOE,XOL,Cycl 及 Ontolingua 等^[9]。同时,也出现了很多构建本体的工具^[10],较为成熟、常用的主要有 OntoLingua、OntoSaurus、WebOnto、Protégé、OntoEdit 等。本研究将用 Protégé 来实现对柑橘病虫害本体的构建。

收稿日期:2013-03-17

基金项目:国家科技支撑计划(编号:2011BAD21B03);湖南省科技重大专项(编号:2010FJ1006)。

作者简介:卜伟琼(1987—),女,湖南常德人,硕士研究生,主要从事农业信息化技术研究。E-mail:buwq_molice@163.com。

通信作者:方 遼,博士,教授,博士生导师,研究方向为农业信息工程。

[9]叶五梅,杨超珍. 静电喷雾中的雾滴荷电特性分析[J]. 农机化研究,2009,31(12):15-18.

[10]余泳昌,史国栋,赵朝会,等. 风送弥雾机施加静电场喷雾技术的研究[J]. 河南农业大学学报,1995,29(2):152-155.

[11]闻建龙,王军锋,罗乾乾. 荷电改善燃油雾化的实验研究[J]. 中国公路学报,2002,15(3):108-110.

[12]杨超珍,叶五梅,赵伟敏. 电极位置影响荷电量的试验研究[J]. 高电压技术,2007,33(10):79-82.

[13]陈志刚,孙英琨,储金字,等. 网状目标法测量雾滴或粉尘荷质

比的精度分析[J]. 排灌机械,2006,24(4):40-43.

[14]Edward L S. Agricultural electrostatic spray application;a review of significant research and development during the 20th century[J]. Journal of Electrostatic,2001,51/52:25-42.

[15]王贞涛,闻建龙,陈 燕,等. 静电雾化理论及应用技术研究进展[J]. 排灌机械,2004,22(6):41-44.

[16]王贞涛,闻建龙,王晓英,等. 高压静电液体雾化技术[J]. 高电压技术,2008,34(5):1067-1072.

2 柑橘病虫害本体知识库的构建

2.1 本体模型

柑橘病虫害本体是柑橘病虫害知识概念化的显式规范说明,它把柑橘病虫害知识以一种明确的、形式化的方式表示出来。本体可以表示为一个五元组 $O = \langle V, F, C, H, Root \rangle$ 。其中, V 表示术语集(词汇集), C 表示概念集。将一个术语集 $V_i \subset V$ 映射到一个概念集,一般来说,术语和概念是多对多的关系,即一个术语可以映射到多个概念,而一个概念也可以被多个术语映射。 F 表示参照函数。 H 表示层次关系,即概念所构成的树状结构, $H \subseteq C \times C, H(c_1, c_2), (c_1 \neq c_2)$ 表示的是 c_1 是 c_2 的一个子概念, H 具有自反性、有向性、无环性和传递性。 $Root$ 表示根概念,在一个本体中,根概念有且只有一个。

2.2 元数据定义

柑橘病虫害本体是柑橘病虫害知识 DK 的概念化描述,包括一系列的基本概念及其关系,可表示为 $OD = \{CS, RS\}$, CS 表示的是概念集合, RS 表示关系集合。(1)概念:领域中具有相同属性的对象的集合,即 $AC_i = \{O_{i1}, O_{i2}, \dots, O_{ik}\}$,其中, O_{ik} 是柑橘病虫害领域中的对象;(2)知识原子:由原子概念和原子关系与公理构成的事实知识断言;(3)知识单元:由知识原子经过交或并操作后而产生的较大粒度的知识实体本

成为柑橘病虫害;(4)概括:即 $general - of$ 关系,其逆关系是 $kind - of$,是概念之间的一般和特殊的关系;(5)聚合:即 $part - of$ 关系,由简单对象构造复杂的实体,描述了知识的封闭结构,如柑橘由叶、花、果、枝干、根和芽等部位组成;(6)角色:即 $role - of$ 或 $property - of$ 关系,概念与领域中其他概念的关系;(7)属性:即 $attribute - of$,概念自身的一些信息与特征;(8)实例:个体对象,个体与其所属概念之间是 $instance - of$ 关系。

2.3 元关系定义

语义关系主要考虑 2 类关系^[11]:一类是处于不同逻辑层次上的概念之间的关系,包括种属关系($is - a$)和实例关系($instance - of$);另一类是反映对象组成结构的关系,它是部分和整体($part - whole$)之间的关系($part - of$)。

定义 1(概念,实例):设概念集为 SC ,对于概念集 SC 中的任意概念 C ,概念 C 的外延集 $E(C) = \{x | x \in C\}$;对于 $E(C)$ 中的任一元素 $C_i \in E_C$,如果 C_i 的外延集 $E(C_i) = \{C_i\}$,则称 C_i 为概念 C 的实例。

定义 2(部分关系,part):部分关系是指对象个体之间的部分关系,将 x 是 y 的一部分记作 $part(x, y)$ 。

实例关系($instance - of$)、部分关系($part - of$)和种属关系($is - a$)的概念和性质如表 1 所示。

表 1 关系的概念与性质

关系种类	概念	性质			
		自反性	对称性	传递性	反对称性
实例关系	对于概念(或类) C 及其实例集 S_{ic} ,则实例集中的元素 $e(e \in S_{ic})$ 和概念 C 之间的关系称为实例关系,记作 $inst(e, C)$ 。	-	-	-	-
部分关系	A 与 B 是部分关系,当且仅当:对 A 的任意实例 x ,存在 B 的某些实例 y 在实例级与 x 为部分关系;反之亦然。	-	-	√	-
种属关系	对于概念集 SC 中的概念 $C_1, C_2 \in SC$,如果有(1)概念 C_1 的内涵包含 C_2 的内涵,即 $I(C_1) \supset I(C_2)$;(2)概念 C_1 的外延包含 C_2 的外延,即 $I(C_1) \subset I(C_2)$;则将概念 C_1 和 C_2 之间的关系称为种属关系,记作 $is - a(C_1, C_2)$ 。	√	-	√	√

2.4 本体构建

根据以上对柑橘病虫害知识本体元语的定义,结合柑橘病虫害知识的特点,对本体进行构建^[12-15]。首先将柑橘病虫害本体表示为一个三元组: $CO = \{C_{co}, R_{co}, H_{co}\}$,其中, $C_{co} = \{\{病因\}, \{疾病名称\}, \{病症\}, \{病理\}, \{患病部位\}, \dots\}$ 为概念集合; $R_{co} = \{part - of, instance - of, kind - of, \dots\}$ 为关系集合; $H_{co} = \{\{炭疽病, 溃疡病, 疮痂病, 黄龙病, \dots\}, \{叶, 果, 花, \dots\}, \{叶片脱落, 枝梢枯死, 茎部黑色霉变, 果实腐烂, \dots\}, \dots\}$ 为实例集合。

2.4.1 类的定义 知识采集、知识抽取之后,得到一系列的概念和术语,在领域专家的指导下,对这些概念和术语进行细致的分析、提炼和归类,确定类的特性和等级,以及类之间的关系。经过反复讨论与协商得到柑橘病虫害知识本体中的类包括柑橘疾病类(Disease)、柑橘患病症状类(Symptom)、柑橘病因类(Cause)、柑橘部位类(Part)、柑橘生长环境类(Environment)、防治方法类(Prevention)等多个类树。

2.4.2 确定类之间的关系 确定类之后,就进行类之间的关系的定义,部分类与类之间的关系如图 1 所示,其中,矩形表示类,单向箭头表示关系,箭头上是关系的名称。

2.4.3 定义类的属性 定义属性时包括名称的定义、描述、

数据类型的定义以及关于值的限定等多个方面。数据类型有很多可以选择,如 any 、 $boolean$ 、 $class$ 、 $float$ 、 $instance$ 、 $integer$ 、 $string$ 、 $smbol$ 。属性进行取值时,可以是本体中已经定义的某个类,也可以是某个类的实例。Protégé 提供了很多可选的设置来进行具体的取值,包括允许值的数量、固定值、缺省值、逆转属性等。比较特殊的是其中的逆转属性,比如,我们定义的疾病和病因这 2 个类,他们互为属性,疾病是病因所引发的结果,他们之间是 $result - in$ (果)的关系,而病因是疾病产生的原因,他们之间又是 $caused - by$ (因)的关系,我们就可以将该疾病类和病因设置为逆转属性,实例化时,当为病因的疾病属性新建一个实例时,该实例就会自动的出现在疾病的病因属性之下^[16]。

2.4.4 创建类实例及形式化表示 前面的内容将柑橘的病虫害知识用本体的形式构建了出来,但是还不能被我们所应用,需要进一步的创建类实例和进行形式化表示,这都可以通过本体构建工具 Protégé 来实现。Protégé 是在 1987 年由斯坦福大学开发的一种开放源码的本体编辑和知识获取软件,目前已成为使用最广泛的本体论编辑器之一。

实例是类的成员,类是实例的集合。类的实例就是类层次中最底层的叶子节点。利用 Protégé 4.1 构建的柑橘病虫

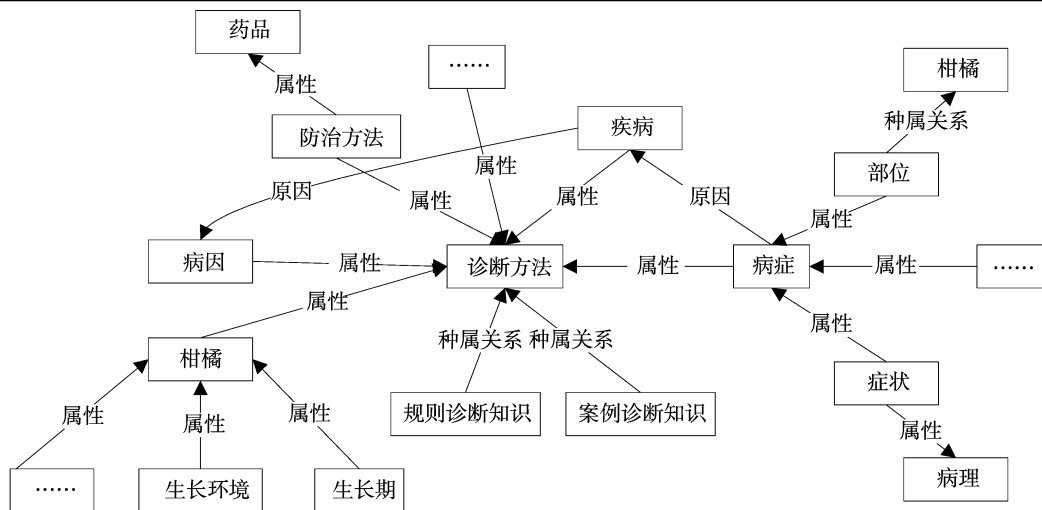


图1 类之间的关系 (部分)

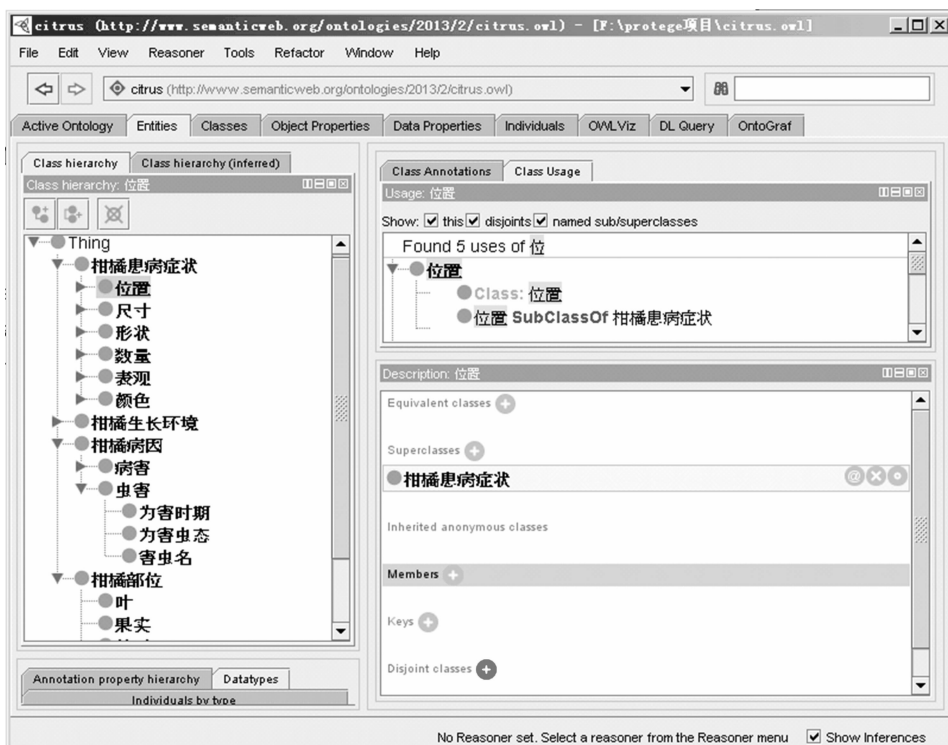


图2 柑橘病虫害知识本体

害本体如图2所示。本研究在构建本体之处,选择用OWL来描述本体,完成后的本体文件能导出以支持OWL浏览和编辑的工具中被保存、阅览和修改。

3 推理本体

本体知识库构建完成之后,利用Jena对本体进行解析,在知识本体的基础上构建自己推理机制,基于已有知识,依据推理规则库中的规则从已有的事实得出隐含知识,完成对问题的求解,并把结果返回给用户。其中,推理中的规则定义最为关键^[17]。规则库中的规则是根据柑橘病虫害的症状、发病条件、防治方法等概念层次和概念之间的关系,定义出来的一些类似于“if-then”关系的规则,这些规则能够对柑橘病虫害

知识的诊断、推理或判定提供依据。张伶俐等提出了一种改进的数学诊断方法来定义计算规则^[18]。该规则的思想是根据患病症状在不同疾病中所占的比例来进行加权。在进行计算时,将用户检索的症状与某疾病的症状进行匹配,匹配成功,则在该疾病上加分值,分别对每一种疾病进行匹配,最后所得分值与症状个数比值最大者为诊断结果。

目前,已出现了大量的本体表示语言,如OWL、DAML、RDFS/RDF等,他们都带有设计良好的推理机制,可以协助完成推理机制的实现。规则集在本体中被称为公理,他严格控制着本体的行为。Jena是由HP公司开发用于Java开发的工具包,它为推理机制提供了一个语义网开发平台,它不仅可以提供对OWL、RDFS/RDF模型的解析、创建和查询支持,而且

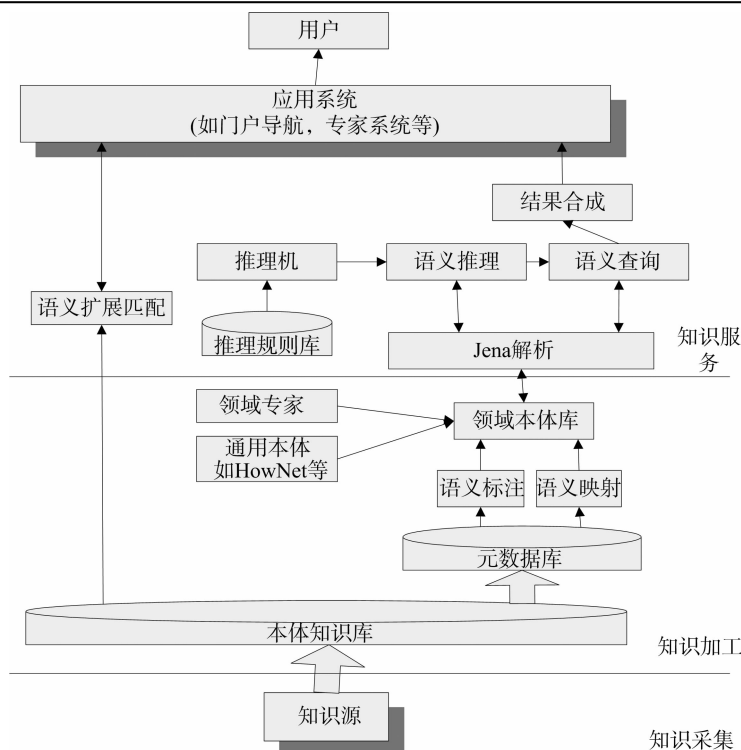


图3 柑橘病虫害知识的本体模型总体结构

它支持基于规则的简单推理。柑橘病虫害知识的本体模型总体结构如图3所示。

4 结语

首先对本体论的相关理论知识进行阐述,针对农业病虫害系统本体构建,指出了现有本体构建方法的不足之处,以柑橘为例,提出了一套完整的病虫害知识应用本体构建方法,研究了农业本体构建与应用中的关键技术,并进行了实现。

参考文献:

- [1]唐惠丽,周炼清,叶基瑶,等. 基于模糊推理的农业病虫害专家系统[J]. 浙江农业学报,2009,21(5):506-509.
- [2]王霓虹,窦智勇. 城市绿化林带病虫害专家系统知识库设计与实现[J]. 林业机械与木工设备,2009,37(8):36-37,40.
- [3]刘鹤,李东明,陈桂芬. 基于CBR的蔬菜病虫害诊治专家系统的研究[J]. 安徽农业科学,2010,38(27):15380-15381,15413.
- [4]戴才萍,黄义德,钱平,等. 水稻病虫害草本体的构建研究[J]. 广东农业科学,2011,38(1):191-194.
- [5]姜大庆,蔡银杰. 基于本体的蔬菜病虫害知识库构建[J]. 江苏农业科学,2012,40(7):368-370.
- [6]Studer B, Benjamins V R, Fensel D. Knowledge engineering: principles and methods[J]. Data and Knowledge Engineering, 1998, 25(1/2):161-197.
- [7]Perez A G, Benjamins V R. Overview of knowledge sharing and reuse components; ontologies and problem solving methods[C]//Conference on Artificial Intelligence, Belgium, 1999.
- [8]李景,孟连生. 构建知识本体方法体系的比较研究[J]. 现代图书情报技术,2004(7):17-22.
- [9]刘琳娜,薛建武,汪小梅. 领域本体构建方法的研究[J]. 情报杂志,2007,26(4):14-16.
- [10]杨秋芬,陈跃新. Ontology 方法学综述[J]. 计算机应用研究,2002,19(4):5-7.
- [11]方卫东,袁华,刘卫红. 基于Web挖掘的领域本体自动学习[J]. 清华大学学报:自然科学版,2005,45(增刊1):1729-1733.
- [12]陈刚,陆汝钤,金芝. 基于领域知识重用的虚拟领域本体构造[J]. 软件学报,2003,14(3):350-355.
- [13]郭宏亮. 玉米病虫害草害诊断系统的研究与实现[D]. 长春:吉林大学,2008.
- [14]范轶. 基于本体推理的心电图辅助诊断系统研究[D]. 长春:吉林大学,2010.
- [15]郭永洪. 基于本体的鱼病知识获取与诊断推理集成系统研究[D]. 北京:中国农业大学,2004:1-123.
- [16]姜海燕,朱艳,汤亮,等. 基于本体的作物系统模拟框架构建研究[J]. 中国农业科学,2009,42(4):1207-1214.
- [17]张信,马衍忠,皮会庆,等. 动物疾病数学诊断与防治[M]. 北京:中国农业出版社,2008:9-10.
- [18]张伶子,段青玲,李道亮. 玉米病虫害诊治本体构建技术研究[J]. 农机化研究,2012,34(1):41-45.