

面向多源数据深度融合的农作物病虫害 本体构建研究*

李悦¹ 孙坦^{2,3} 鲜国建^{1,3} 赵瑞雪^{1,3} 李娇¹ 黄永文¹ 罗婷婷¹

(1. 中国农业科学院农业信息研究所, 北京 100081; 2. 中国农业科学院, 北京 100081;
3. 农业农村部农业大数据重点实验室, 北京 100081)

摘要: 我国农作物病虫害种类繁多, 每年因病虫害的影响导致农作物产量、农民收入的损失非常严重。由于农作物病虫害数据来源广泛, 数据表示、组织方式等不尽相同, 信息资源处于无序与相对混乱的状态, 导致领域知识难以被有效利用。本文借鉴融合多来源、多载体的专业化高质量知识资源, 尝试将农作物病虫害领域本体与通用的科技文献本体以及传统的知识组织体系进行深度融合, 并对本体进行形式化表示, 促进领域知识的共享与语义互操作, 为构建基于多源高质量数据的领域知识图谱奠定语义基础, 以期实现以科技为支撑助力农作物病虫害的科学防治, 为农业新型智能化知识服务实践提供相关思路。

关键词: 领域本体; 农作物病虫害; 数据融合

中图分类号: G353 **DOI:** 10.3772/j.issn.1673-2286.2021.02.001

引文格式: 李悦, 孙坦, 鲜国建, 等. 面向多源数据深度融合的农作物病虫害本体构建研究[J]. 数字图书馆论坛, 2021 (2) : 2-10.

农业在我国经济体系中占有重要比重, 农作物病虫害是制约“农业增效、粮食增产、农民增收”的重要灾害之一, 仅常见病虫害就在1 400种以上, 其爆发成灾往往直接影响社会经济的健康发展。农业农村部发文指出^[1], 2019年我国水稻病虫害将呈偏重发生态势, 发生面积11.9亿亩次, 其中, 稻飞虱、二化螟、纹枯病偏重发生, 局部大发生; 稻纵卷叶螟、稻瘟病、稻曲病中等发生, 局部偏重发生; 穗腐病、南方水稻黑条矮缩病等病毒病、白叶枯病、细菌性基腐病等次要病虫在局部稻区发生呈上升态势, 需采取及时有效的防控措施, 最大限度降低粮食风险。坚持农业科技自立自强, 形成科学有效的方法体系是防治农作物病虫害的关键环节。农作物病虫害防治体系的科学化、专业化对知识服务提出新的要求。同年, 面对草地贪夜蛾对粮食作物造成的危害, 中国农业科学院通过明确生态学规律, 确定害虫基因组类型、繁殖区域, 研发形成一整套针对草地贪夜蛾病虫害的防控技术, 彰显出强化农业科技, 健全农作物

病虫害防治体系的重要作用。

农作物病虫害数据广泛存在于农业知识在线资源、农业科学数据库和科学专著中, 数据表示、组织方式等不尽相同, 信息资源处于相对无序的分散状态。随着数据密集型科学研究第四范式的提出, 如何整合多源异构病虫害数据, 挖掘隐含在文本内的知识, 为农作物病虫害的科学防治提供服务成为当前重要的研究课题。

本体最初源于哲学, 20世纪80年代本体的概念被引入计算机领域用来刻画知识, 包含类、关系、公理、函数和实例5种要素^[2], 是揭示领域概念体系的重要手段。同时, 本体也是构成知识图谱语义模型的基础^[3], 在语义层面规范与约束了知识图谱的数据填充, 使得知识图谱能够大规模构建并支撑上层应用。近年来, 国内外学者开展了大量关于构建农作物病虫害本体进行知识组织与管理的研究, 并将其应用于专家诊断、领域知识管理、智能搜索、智能问答等场景, 进一步发挥了本

*本研究得到国家社会科学基金项目“科技论文全景式摘要知识图谱构建与应用研究”(编号: 19BTQ061)资助。

体的应用价值: Chaudhary等^[4]构建了包含农作物、病虫害、土壤、栽培过程等信息的棉花本体,并开发了农业咨询服务系统,该系统能够回答农民查询的领域知识,有助于缩小农民和农业领域专家之间的差距,为棉农的生产实践提供指导;郑颖等^[5]通过资源整合构建了五大类小麦病虫害本体,并将其应用于问答系统;Chougule等^[6]提出了一种虫害关键词提取算法构建作物虫害本体,使之作为农业专家系统的一部分;巩如悦^[7]采用基于文献的领域本体构建方法,构建包含苹果病虫害、病变部位、病变症状、发病时期、防治办法等核心概念的苹果病虫害本体,并将领域本体引入到垂直搜索引擎中,有效提升了苹果病虫害信息检索的准确率。Lagos-Ortiz^[8]提出了一个基于本体的甘蔗、水稻、大豆和可可作物虫害防治决策支持系统,为虫害诊断和防治提供指导;Chougule等^[9]通过将不同形式的葡萄作物病虫害知识转化为本体,并利用气象条件与病虫害发生的关系,开发了葡萄酒病虫害管理系统,该系统帮助葡萄种植者尽量减少杀虫剂的使用,从而降低化学药剂对环境的副作用;丁浩然等^[10]从不同知识维度构建了包含7个一级类、15个二级类的油茶本体,并基于本体构建了油茶领域知识图谱,实现智能搜索、智能问答等应用。

然而,上述研究涉及的领域要素相对比较单一,对于覆盖农作物病虫害全领域所构建的本体较少,存在研究粒度较粗的问题,本体概念体系的全面性、可扩展性与泛化能力有待进一步提升。因此,本文通过对农作物病虫害全领域的信息资源进行深入分析与加工处理,总结与抽象领域概念体系,并与科技文献知识以及现有的知识组织体系深度融合,大幅拓展现有农作物病虫害本体中知识单元的广度、粒度与深度,使得所构建的领域本体在应用场景上能够满足精准问答、精准推荐等精细化多场景^[11]的农业专业知识服务需求,进而为构建基于多源高质量数据的领域知识图谱提供语义框架基础,以期促进农作物病虫害领域的知识共享与应用。

1 本体构建目标及流程

农作物病虫害本体的构建目标是依照植物保护领域的专业知识,对农作物病虫害知识体系进行建模,并与通用的科技文献本体以及传统的知识组织体系进行深度融合,形成客观且明确的描述规范,大幅度拓展当前农作物病虫害知识库中知识单元的广度、粒度与深

度,在结构化、形式化和语义化层面提高其描述与表达知识的能力,并通过本体描述语言使知识能够被计算机理解 and 处理,为构建基于多源高质量数据的农作物病虫害领域知识图谱奠定基础。

构建领域本体的常用方法是七步法^[12]。然而,七步法没有运用软件工程生命周期管理的思想,也缺少对本体质量的评估过程,导致将其运用于领域本体构建会存在本体的更新迭代受限以及无法满足特定领域业务需求的问题。因此,本文结合农作物病虫害领域特点,遵循本体构建原则,在七步法的基础上改进构建步骤,将本体验证与评估、编程语言描述本体及迭代优化的环节补充进领域本体的构建流程中,如图1所示。

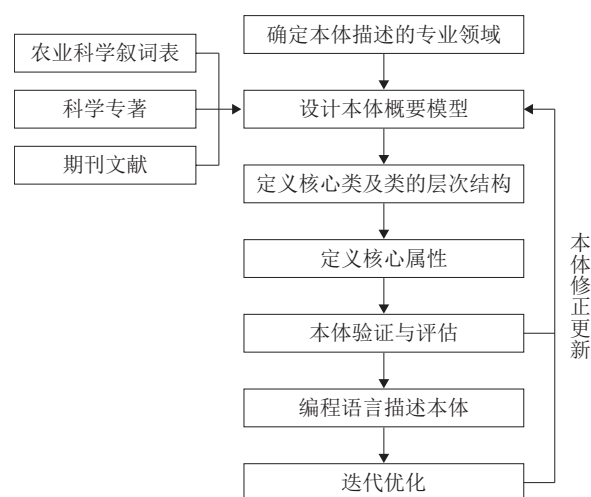


图1 农作物病虫害本体构建流程

(1) 确定本体描述的专业领域。全领域所包含的数据或者知识太过广泛,即使是同一领域,因侧重点不同所建立的本体也有较大差异,因此,在构建本体之前首先需要确定本体所描述的领域范畴以及边界。本文构建的农作物病虫害领域本体的范围覆盖农作物病虫害资源中的基础知识、相关的科技文献知识以及现有的知识组织体系,以实现多元化语义关联,解决各类资源间的语义互操作问题。

(2) 基于参考的数据源设计本体概要模型。该阶段的任务是尽可能全面地搜集领域内的信息资源,以充分了解相关的知识体系,对概念进行总结、抽象从而建立本体概要模型,形成农作物病虫害本体的框架。本文参照的专业资源包括系统划分并概括组织了农业学科主题概念的《农业科学叙词表》^[13]、蕴含丰富语义信息的病虫害科学专著《中国农作物病虫害(第三版)》^[14]以及学术科研领域相关的期刊文献。

(3) 定义核心类及类的层次结构。对本体概要模型进行细化和补充,围绕需要解决的问题和应用场景,梳理、精炼核心概念及概念类间的层次结构。

(4) 定义核心属性。通过定义本体类的属性关系,可将病虫害自身的特征信息,以及病虫害与科技文献、现有的知识组织体系的语义关联引入本体模型,以完整地展现农作物病虫害的知识关联体系。

(5) 本体验证与评估。初步构建领域本体后,为了保证其能够充分描述农作物病虫害知识体系,满足业务需求,需要通过创建实例来验证本体的可用性,并从本体结构、应用场景等方面对所构建的本体进行评估,实现本体的修正与完善。

(6) 编程语言描述本体。验证与评估通过后,运用OWL+SKOS语言对本体模型进行形式化表示以增强本体的表达能力。

(7) 本体迭代优化。领域本体的构建是一个动态的过程,后续出现的新知识经过上述流程综合分析后,可扩充到农作物病虫害本体中,以进一步迭代和完善本体。

2 农作物病虫害本体构建

农作物病虫害本体模型的设计和构建应与领域特

点紧密结合,且该领域本体模型应具有可复用性。本节根据提出的农作物病虫害本体的构建流程,遵循领域本体独立性、共享性特征,设计本体概要模型、定义核心类及层次结构、定义核心属性、本体验证与评估,并使用本体开发工具Protégé对农作物病虫害本体进行构建和管理。

2.1 设计本体概要模型

本体概要模型描述了在抽象层次由概念及其间关系建立的逻辑模型,借助概要模型可以梳理本体构建的概念范围及核心属性的取值范围。本文以农作物病虫害为主要研究对象,在充分参考主要数据源的基础上,通过设计全面细粒度揭示病虫害资源的本体概要模型,使资源充分发挥价值,为进一步本体模型的详细设计奠定基础。

概要模型中各要素的收集,包括抽象层面自顶向下的要素列举,及数据层面自底向上的要素总结、抽象,如枚举《农业科学叙词表》中叙词等级体系下的相关概念、查阅文献数据库中表的字段及每条记录,总结并提炼《中国农作物病虫害(第三版)》等科学专著中隐性知识的知识单元、语义关系等。最终得到的农作物病虫害本体的概要模型如图2所示。

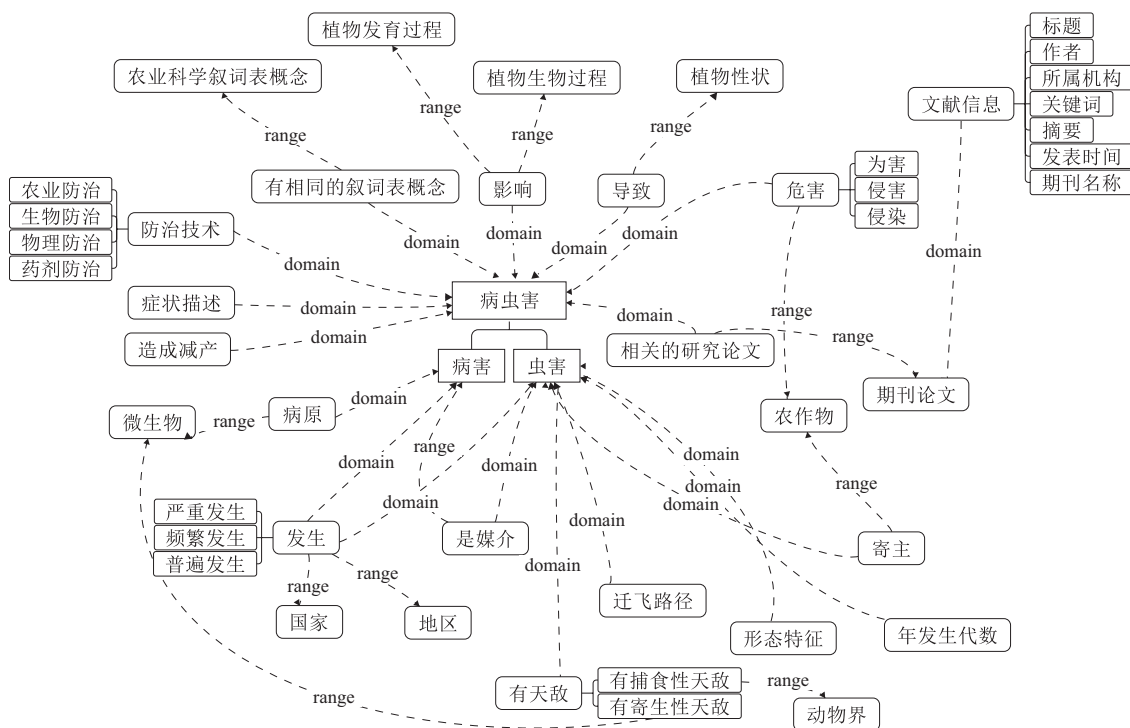


图2 农作物病虫害本体概要模型

模型从核心描述对象农作物病虫害出发, 建立病虫害类, 将其分为病害和虫害两类。病虫害类通过对象属性危害与农作物类建立关联, 描述病虫害危害的农作物; 通过对象属性发生分别与国家类、地区类建立关联, 描述不同的病虫害发生的国家和地区; 通过对象属性影响分别与植物发育过程类、植物生物过程类建立关联, 通过对象属性导致与植物性状类建立关联, 描述病虫害影响农作物的发育过程和生物过程, 其产生叶片干枯等植物性状, 造成粮食减产。同时, 添加了描述病虫害类自身的数据属性: 防治技术、症状描述、造成减产。对于病害类, 通过对象属性病原与微生物类建立关联, 描述病害存在病原微生物, 气候和地理环境的差异会使病原毒性有所不同。对于虫害类, 通过对象属性是媒介与病害类建立关联, 描述虫害会成为病害传播的媒介; 通过对象属性有捕食性天敌、有寄生性天敌分别与动物界类、微生物类建立关联, 描述虫害在动物界及微生物中有天敌, 可采用释放虫害天敌的方法进行生物防治。描述虫害自身的数据属性主要有形态特征、年发生代数、迁飞路径。

此外, 将具有标题、作者、关键词等文献信息属性的科技文献本体, 以及知识组织领域的农业科学叙词表叙词概念引入领域本体中, 分别形成期刊论文类和农业科学叙词表概念类。病虫害类通过对象属性相关的研究论文与期刊论文类建立关联, 通过对象属性有相同的叙词表概念与农业科学叙词表概念类建立关联, 实现病虫害领域资源与通用的科技文献本体以及传统的知识组织体系的深度融合。

2.2 定义核心类及层次结构

在本体中, 类(Class)用于描述具有相同属性的一类概念的集合, 是构成本体的重要成分, 属性(Property)和实例(Individual)都是对类的明确和说明。类的上下位关系表达了类之间的基本关系, 主要根据一级类对其他类进行子类的划分, 并通过SubClassOf描述。

基于上文设计的本体概要模型进行本体中类及层次结构的详细设计, 最终定义了农作物病虫害本体包含11个一级类、8个二级类。同时, 为了使本体的构建、管理过程更加规范, 便于后续的编程工作, 本文对本体中的类和属性运用英文名称进行定义, 并为其设置了带有释义的中文和英文标签(rdfs:label)以方便相关人员查看。本体中的核心类具体为以下11类。①病虫

害类(DiseaseAndPest)。本文研究的是农作物病虫害领域本体构建并将其进行形式化表示, 因此在本体的构建中以病虫害类作为核心类。病虫害描述为害各类农作物的病害和害虫, 因此将其分为病害(Disease)和虫害(Pest)2个子类。②文献类(Literature)。文献的基本信息可以描述农作物病虫害相关文献的外部特征和内容特征, 为了将其融入领域本体中, 创建了文献类并添加期刊论文(JournalArticle)子类。③农业科学叙词表概念类(CAT)。农业科学叙词表的病虫害叙词概念具有上下位关系, 如“块茎病”有上位词“茎病”和下位词“块茎腐病”等上下位语义关系。将其融入领域本体中并与病虫害类实例建立连接, 可提升所建本体的专业性与关联性, 实现领域本体与外部知识的集成、关联与共享。④植物界类(Plant)。根据生物分界系统的划分, 植物界是生物的一界。在本体中创建植物界一级类, 为其添加农作物类(Crop)子类, 描述植物界下包含的农作物, 也是病虫害为害的对象。⑤动物界类(Animal)。根据生物分界系统的划分, 动物界也是生物的一界。在本体中创建动物界用来描述虫害在动物界的天敌。⑥植物生物过程类(PlantBiologicalProcess)。植物生物过程描述植物发生光合作用等生物过程。⑦植物发育过程类(PlantDevelopmentProcess)。植物发育过程描述萌芽期、幼苗期等各个植物生长发育时期。⑧植物性状类(PlantTraits)。植物性状描述植物的表型特征, 分为外源物刺激性状(ExogenousStimulatingTraits)和症状类型(SymptomType)2个子类。⑨微生物类(Microorganism)。微生物描述病害的病原微生物, 分为真菌(Fungus)、细菌(Germ)、病毒(Virus)3个子类。⑩国家类(Nation)。其表述病虫害最早被发现的国家, 以及主要分布的国家。⑪地区类(Region)。其表述病虫害发生的地区。

此外, 本文将病害类(Disease)、虫害类(Pest)、植物界类(Plant)和动物界类(Animal)与SKOS概念体系中具体的概念(skos:Concept)建立等价关系(Equivalent To), 形成农作物病虫害核心本体与领域本体的结合。

2.3 定义核心属性

类本身不足以描述专业领域的知识体系, 故有必要通过定义类的内部属性信息和外部关系信息来描述概

念的具体内涵。本体中,类的属性用于描述类具有的特征,其分为两种。其中,对象属性揭示了类之间的语义关系,是在本体中进行逻辑推理的基础,决定了知识图谱的丰富程度和应用效果。数据属性描述了类自身的特点,通常使用具体的数据类型进行表示,不同数据属性的取值类型可能有所差异,一般包括文本、数值、日期

等多种类型。

通过参照数据源分别为核心类填充具体实例,实例间多次出现的语义关系构成类的对象属性,实例自身的共性描述构成类的数据属性。最终定义了16个一级对象属性,同时为了增强语义严谨性设置对象属性的定义域、值域的取值范围(见表1)。

表1 农作物病虫害本体的核心对象属性

属性名称	英文标识	定义域	值域	释义
最早被发现的国家	earliest site be discovered	病害 (Disease)	国家 (Nation)	描述病害最早被发现的国家
病原	pathogeny	病害 (Disease)	微生物 (Microorganism)	描述病害的病原微生物
有症状类型	has symptom type	病害 (Disease)	症状类型 (SymptomType)	描述病害的症状类型
有寄主	has host	虫害 (Pest)	农作物 (Crop)	描述虫害在农作物中的寄主
有捕食性天敌	has predaceous natural enemy	虫害 (Pest)	动物界 (Animal)	描述虫害在动物界的捕食性天敌
有寄生性天敌	has parasitic natural enemy	虫害 (Pest)	微生物 (Microorganism)	描述虫害在微生物中的寄生性天敌
传播	spread	虫害 (Pest)	微生物 (Microorganism)	描述虫害传播的微生物
完成世代在	complete generation in	虫害 (Pest)	农作物 (Crop)	描述虫害完成世代所在的农作物
是媒介	is vehicle	虫害 (Pest)	病害 (Disease)	描述虫害是病害的传播媒介
分布的国家	distributed in	病虫害 (DiseaseAndPest)	国家 (Nation)	描述病虫害分布的地区
发生的地区	happen	病虫害 (DiseaseAndPest)	地区 (Region)	描述病虫害发生的地区
危害	harm	病虫害 (DiseaseAndPest)	农作物 (Crop)	描述病虫害危害的农作物
影响	affect	病虫害 (DiseaseAndPest)	植物生物过程 (PlantBiologicalProcess)	描述病虫害对植物生物过程/ 植物发育过程的影响
			植物发育过程 (PlantDevelopmentProcess)	
导致	lead to	病虫害 (DiseaseAndPest)	外源物刺激性状 (ExogenousStimulatingTraits)	描述病虫害导致植物产生的 外源物刺激性状
相关的研究论文	be researched by	病虫害 (DiseaseAndPest)	期刊论文 (JournalArticle)	描述研究病虫害的期刊论文
相同的叙词表概念	same as	病虫害 (DiseaseAndPest)	农业科学叙词表概念类 (CAT)	描述农业科学叙词表中与 病虫害名相同的概念

为描述类自身的特征,在本体中为农作物病虫害类 (DiseaseAndPest)、期刊论文类 (JournalArticle) 和农业科学叙词表概念类 (CAT) 复用、定义了共计31个一级数据属性,4个二级数据属性。

对于农作物病虫害类 (DiseaseAndPest), 复用、定义了12个一级数据属性、4个二级数据属性,包括复用SKOS本体模型中词法标签下的概念skos:prefLabel来标识病害的名称和虫害的名称; skos:altLabel来标识病害的别名和虫害的异名; 定义了标识符 (identifier) 用来唯一标识类的实例; 最早被发现的时间 (earliest time be discovered) 描述每种病虫害最早被发现的时间;

造成减产 (cause crop failure) 描述病虫害会造成农作物减产的数量; 年发生代数 (annual generation) 描述虫害在不同环境下的年发生代数; 世代及发生期 (generation and emergence period) 描述虫害的世代发生期; 寿命 (lifetime) 描述虫害各代数的寿命; 形态特征 (morphological characteristics) 描述虫害的体长及生长发育各阶段的形态特征; 迁飞路径 (moved way) 描述虫害的迁飞路径; 症状描述 (symptom description) 是指农作物遭遇病虫害后的症状表现; 防治技术 (control technology) 描述病虫害的防治方法, 包括农业防治 (agricultural measure)、生物防治

(biological measure)、物理防治 (physical measure) 和药剂防治 (potion measure) 4个二级数据属性。

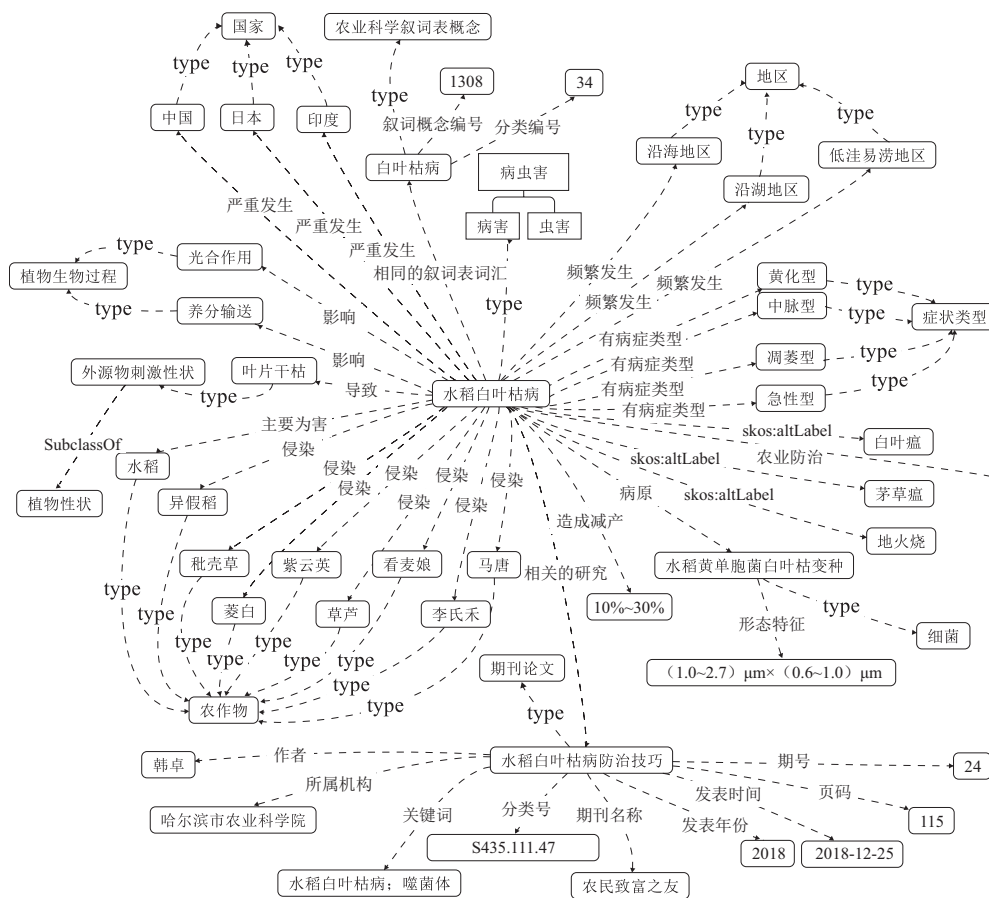
对于期刊论文类 (JournalArticle), 复用、定义了17个一级数据属性, 包括复用都柏林核心元素集中的语种 (dc:language) 描述期刊论文的语种; 定义了标识符 (gid) 对期刊论文类资源进行唯一的标识、标题 (title)、作者 (author)、所属机构 (affiliate institution)、摘要 (abstract)、关键词 (keywords)、分类号 (classification)、doi码 (doi)、发表年份 (year)、发表时间 (create time)、期刊名称 (journal name)、是否为核心期刊 (isCore)、出版商 (publisher)、期号 (issue)、卷 (volume)、页码 (page)。

对于农业科学叙词表概念类 (CAT), 复用、定义了4个一级数据属性, 包括同样复用了SKOS本体模型中的 skos:prefLabel 来标识叙词表中的概念名称, skos:altLabel 标识叙词概念的别名, 定义了叙词概念编号 (termcode) 用来唯一标识叙词概念, 分类编号 (catid) 描述叙词概念所在的学科领域分类。

2.4 本体验证与评估

本节根据病虫害信息资源为本体增加实例进行验证, 并基于本体评估标准对初步构建的农作物病虫害本体模型进行评估和迭代修正, 以保证所建本体的规范性、专业性和适用性, 使其能够充分描述领域的概念体系, 满足实际的知识服务需求。

农作物病虫害本体为病虫害相关的信息资源提供了一套语义表示框架, 可对多来源、多载体的专业资源进行结构化的知识描述。白叶枯病是我国水稻生产中的“三大病害”之一, 严重影响水稻的产量和品质^[15]。选取水稻白叶枯病这一典型实例对农作物病虫害领域本体模型进行验证与评估, 采用的主要数据来源于《中国农作物病虫害 (第三版)》中水稻白叶枯病相关的文本数据, 该数据涉及的知识面广, 知识单元类别多样, 详细介绍了水稻白叶枯病的分布和危害、症状、病原、防治技术等信息。图3为对农作物病虫害本体进行实例填充后的示意, 展示了以“水稻白叶枯病”为核心



- (1) 严格执行检疫制度, 防止带菌种子传入无病区。引种时, 应小面积种植, 证明无病后再大面积种植。
- (2) 尽量使用无病种子, 必要时进行种子消毒。可用20%叶枯唑可湿性粉剂500~600倍液浸种48~72h, 或链霉素200IU浸种24h。
- (3) 及时清理病稻草、病稻桩及病谷, 以减少菌源。病区强调不用病稻草扎秧把; 不用病草作草套围秧畦; 不用病草作浸种催芽覆盖物; 不用病草堵塞水口和涵洞; 不用病草铺垫拖拉机道路等。打谷场及村庄附近应开截水沟, 防止菌水流入进水渠污染秧田和大田。在华南稻区, 宜在春耕及复种前结合放养绿萍, 提早整田, 消除田边再生稻株和杂草。

图3 农作物病虫害本体实例示意

的农作物病虫害领域语义网络,说明了水稻白叶枯病相关的研究论文,该论文的作者、所属机构、关键词、分类号、期刊名称、发表年份、发表时间、页码、期号等外部特征和内容特征信息,水稻白叶枯病与农业科学叙词表相同的概念,该叙词概念含有唯一标识编号、分类编号,水稻白叶枯病的病原、症状类型、发生的国家和地区、危害对象,影响光合作用、养分输送等植物生物过程,导致叶片干枯等外源物刺激性状,及水稻白叶枯病自身的别名、农业防治措施等属性特征。由此可验证,农作物病虫害领域本体实现了多种资源的整合与关联,为下一步基于多源数据构建领域知识图谱提供语义基础。

本体的评估标准一般包括清晰性、一致性、完善性和可扩展性^[16]。清晰性指所定义的类和属性是明确的,不存在歧义;一致性指类间的关系在逻辑上是一致的;完善性指所定义的领域内类和属性是完整的,可以用于描述主要参考数据源中的知识体系;可扩展性指当农作物病虫害领域内出现新的概念时,本体可规模化地扩展。笔者邀请植物保护领域专家对本体进行评估,经过讨论,专家认为农作物病虫害科学著作描述了病虫害的各方面信息,却缺乏病虫害领域概念体系间的语义关联,农业科学叙词表仅使用“用、代、属、分、族参”描述叙词概念的层级关系,科技文献本体同样缺乏对于病虫害的清晰针对性描述。因此,总的来说,本文构建的领域本体相较于单一资源包含了较为完整的农作物病虫害概念体系,类的层次结构及属性可以清晰、有逻辑、细粒度地揭示水稻白叶枯病的相关知识以及知识间的语义关联,通过迭代优化可实现本体的扩充与优化。因此,本文所构建领域本体符合上述评估标准,同时通过填充水稻病虫害实例验证了本体的有效性,可适用于大规模的农作物病虫害领域本体构建,为进一步实现农业智能化知识服务奠定基础。

3 农作物病虫害本体的形式化表示

本体的描述语言可以使本体的表示更加形式化,并且可以使其被计算机所理解 and 处理。其中,OWL语言具有较好的语义表达能力,病害、虫害的名称可以作为SKOS^[17]词表中概念(skos:Concept)下的实例进行声明,概念在传统的知识组织体系中也大量存在,运用SKOS语言可以继承其中的知识,形成领域概念体系,以提高与叙词表等知识组织体系的共享和互操作,因

此本文采用OWL+SKOS语言对农作物病虫害领域本体进行形式化的表示,具体如下。

(1) 类的OWL+SKOS表示。农作物病虫害本体中的类用OWL中的类Class进行表示,类间的层次关系用subClassOf来体现。如创建病虫害类(DiseaseAndPest)的子类虫害(Pest),设置其中文标签为“虫害”,英文标签为“Pest”,同时用skos:Concept将其定义为领域内的一类概念,OWL+SKOS代码如下。

```
<owl:Class rdf:about="http://linked.aginfra.cn/onts/cdp#Pest">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <rdfs:subClassOf rdf:resource="http://linked.aginfra.cn/onts/cdp#DiseaseAndPest"/>
  <rdfs:label xml:lang="en">Pest</rdfs:label>
  <rdfs:label xml:lang="zh-cn">虫害</rdfs:label>
</owl:Class>
```

(2) 对象属性的OWL表示。对象属性即Object Property,对象属性用于创建类之间的关系。如创建对象属性影响(affect)并设置其定义域为病害类(Disease),值域为植物生物过程类(PlantBiologicalProcess),英文标签为“affect”,中文标签为“影响”,OWL代码如下。

```
<owl:ObjectProperty rdf:about="http://linked.aginfra.cn/onts/cdp#affect">
  <rdfs:domain rdf:resource="http://linked.aginfra.cn/onts/cdp#Disease"/>
  <rdfs:range rdf:resource="http://linked.aginfra.cn/onts/cdp#PlantBiologicalProcess"/>
  <rdfs:label xml:lang="en">affect</rdfs:label>
  <rdfs:label xml:lang="zh-cn">影响</rdfs:label>
</owl:ObjectProperty>
```

(3) 数据属性的OWL表示。数据属性即Datatype Property,数据属性可以为实体赋予一个具体的属性值。如建立在数据属性防治技术(control technology)下的子数据属性药剂防治(potion measure),并设置其英文标签为“potion measure”,中文标签为“药剂防治”OWL代码如下。

上述基于OWL+SKOS的农作物病虫害本体形式化表示,通过将语义信息存储为结构化的形式化代码,实

```
<owl:DatatypeProperty rdf:about="http://linked.aginfra.cn/
onts/cdp#potion_measure">
  <rdfs:subPropertyOf rdf:resource="http://linked.aginfra.
cn/onts/cdp#control_technology"/>
  <rdfs:label xml:lang="en">potion measure</rdfs:label>
  <rdfs:label xml:lang="zh-cn">药剂防治</rdfs:label>
</owl:DatatypeProperty>
```

现了领域知识从计算机可读向计算机可理解、可推理发展,为农作物病虫害语义共享与互操作、语义推理和语义检索提供支撑。

4 总结与展望

中国农作物病虫害种类繁多,每年因病虫害导致农作物减产的现象屡见不鲜。本体建模是有效揭示农作物病虫害概念体系的重要手段。本文针对农业植物保护领域的特点提出了农作物病虫害本体的构建目标与构建流程,详细介绍了本体构建流程中的各个步骤。完成了本体概要模型设计,运用Protégé工具完成了核心类及属性的定义,并将领域本体与通用的科技文献本体及现有的农业科学叙词表叙词概念进行深度融合,以水稻白叶枯病实例开展了本体的验证与评估,最后基于本体描述语言对农作物病虫害本体进行了形式化表示,以为农业植物保护领域提供可复用的本体构建流程以及规范化的语义表示模型,促进领域知识与外部知识的深度融合与共享,提升知识的利用效率。

知识图谱是人工智能的重要组成部分,也是机器实现认知智能的基石。未来,在应用层面,基于本文所构建的面向多源数据深度融合的农作物病虫害本体,可结合深度学习算法应用于领域知识图谱构建中的实体识别、关系抽取、属性抽取等场景,进而支撑病虫害辅助分析、决策支持、知识推理、智能问答等应用。

参考文献

- [1] 农业农村部. 2019年水稻重大病虫害防控技术方案 [EB/OL]. [2020-12-10]. http://www.moa.gov.cn/gk/nszd_1/2019/201904/t20190425_6212817.htm.
- [2] PÉREZ A G, BENJAMINS V R. Overview of knowledge sharing and reuse components: Ontologies and problem-solving methods [C] //Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving methods (KRR5), Stockholm, Sweden. 1999: 1-15.
- [3] 王昊奋, 漆桂林, 陈华钧. 知识图谱方法、实践与应用 [M]. 北京: 电子工业出版社, 2019.
- [4] CHAUDHARY S, BHISE M, BANERJEE A, et al. Agro advisory system for cotton crop [C] //7th International Conference on Communication Systems and Networks (COMSNETS). IEEE, Bangalore, India, 2015: 1-6.
- [5] 郑颖, 金松林, 张自阳, 等. 基于本体的小麦病虫害问答系统构建与实现 [J]. 河南农业科学, 2016, 45 (6): 143-146.
- [6] CHOUGULE A, JHA V K, MUKHOPADHYAY D. Adaptive ontology construction method for crop pest management [C] //Proceedings of the International Conference on Data Engineering and Communication Technology. Springer, Singapore: 2017, 665-674.
- [7] 巩如悦. 基于本体的苹果病虫害垂直搜索引擎研发 [D]. 杨凌: 西北农林科技大学, 2017.
- [8] LAGOS-ORTIZ K, MEDINA-MOREIRAL J, MORÁN-CASTROL C, et al. An ontology-based decision support system for insect pest control in crops [C] //International Conference on Technologies and Innovation. CITI 2018, Guayaquil, Ecuador. 2018: 3-14.
- [9] CHOUGULE A, JHA V K, MUKHOPADHYAY D. Decision support for grape crop protection using ontology [J]. International Journal of Reasoning-based Intelligent Systems, 2019, 11 (1): 24.
- [10] 丁浩宸, 王忠明. 基于本体的油茶中文知识图谱构建与应用 [J]. 世界林业研究, 2020, 33 (4): 50-55.
- [11] 赵瑞雪, 李娇, 张洁, 等. 多场景农业专业知识服务系统构建研究 [J]. 农业图书情报学报, 2020, 32 (1): 4-11.
- [12] NOY N F, MCGUINNESS D L. Ontology development 101: a guide to creating your first ontology [R/OL]. [2021-01-05]. https://protege.stanford.edu/publicatipub/ontology_development/ontology101.pdf.
- [13] 农业部情报研究所. 农业科学叙词表 [M]. 北京: 中国农业出版社, 1994.
- [14] 中国农业科学院植物保护研究所, 中国植物保护学会. 中国农作物病虫害: 第三版 [M]. 北京: 中国农业出版社, 2015.
- [15] 王剑, 朱燕, 赵黎宇, 等. 水稻白叶枯病的发生流行与防治技术 [J]. 四川农业科技, 2020 (10): 35-36, 39.
- [16] 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.

[17] W3C. SKOS Simple Knowledge Organization System

Namespace Document [EB/OL]. [2021-01-10]. [https://www.](https://www.w3.org/2009/08/skos-reference/skos.html#)

[w3.org/2009/08/skos-reference/skos.html#](https://www.w3.org/2009/08/skos-reference/skos.html#).

作者简介

李悦, 女, 1996年生, 硕士研究生, 研究方向: 知识组织、知识图谱。

孙坦, 男, 1970年生, 博士, 研究员, 通信作者, 研究方向: 数字信息描述与组织, E-mail: suntan@caas.cn。

鲜国建, 男, 1982年生, 博士, 研究员, 研究方向: 大数据融汇治理、知识组织、知识图谱。

赵瑞雪, 女, 1968年生, 博士, 研究员, 研究方向: 信息管理与信息系统、数字图书馆、知识组织与知识服务。

李娇, 女, 1989年生, 博士研究生, 助理研究员, 研究方向: 知识组织、知识图谱。

黄永文, 女, 1975年生, 博士, 副研究员, 研究方向: 知识组织与知识服务。

罗婷婷, 女, 1985年生, 硕士, 助理研究员, 研究方向: 知识组织、大数据融汇治理、信息管理与信息系统。

Research on Ontology Construction of Crop Diseases and Pests for Deep Fusion of Multi-source Data

LI Yue¹ SUN Tan^{2,3} XIAN GuoJian^{1,3} ZHAO RuiXue^{1,3} LI Jiao¹ HUANG YongWen¹ LUO TingTing¹

(1. Agricultural Information Institute of CAAS, Beijing 100081, China; 2. Chinese Academic of Agricultural Sciences, Beijing 100081, China;

3. Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, Beijing 100081, China)

Abstract: There are many kinds of crop diseases and pests in China, and the loss of crop yield and farmers' income is very serious every year due to the influence of diseases and pests. However, the data of crop diseases and pests come from a wide range of sources, and the data representation and organization are different. The information resources are in a state of disorder and relative chaos, which makes it difficult for domain knowledge to be effectively utilized. For this reason, this paper integrates professional and high-quality knowledge resources from multiple sources and carriers, try to deeply integrate the domain ontology of crop diseases and pests with the general scientific and technological literature ontology and the traditional knowledge organization system, and based on formalized representation of ontology, to promote the field of knowledge sharing and semantic interoperability. This lays a semantic foundation for the construction of field knowledge graph based on multi-source and high quality data, in order to realize the scientific control of crop diseases and pests supported by science and technology, and to provide relevant ideas for the practice of new intelligent agricultural knowledge service.

Keywords: Field Ontology; Crop Diseases and Pests; Data Fusion

(收稿日期: 2021-01-12)