DOI: 10.5846/stxb202206091637

高梅香,朱家祺,刘爽,程鑫,刘冬,李彦胜.土壤动物知识图谱构建理论、方法与技术——以浙江天目山土壤螨类为例.生态学报,2023,43(16):6862-6877.

Gao M X, Zhu J Q, Liu S, Cheng X, Liu D, Li Y S.Theory, method and technique of soil animal knowledge graph construction: a case study of soil mites in Tianmu Mountain, Zhejiang Province.Acta Ecologica Sinica,2023,43(16):6862-6877.

土壤动物知识图谱构建理论、方法与技术
——以浙江天目山土壤螨类为例

高梅香[1,2],朱家祺[1,2],刘 爽[3],程 鑫[4],刘 冬[5],李彦胜[4,*]

1 宁波大学地理与空间信息技术系,宁波 315211

2 宁波市高等学校协同创新中心“宁波陆海国土空间利用与治理协同创新中心”,宁波 315211

3 北京师范大学密云实验中学,北京 101500

4 武汉大学遥感信息工程学院,武汉 430079

5 中国科学院东北地理与农业生态研究所,湿地生态与环境重点实验室,黑土地保护与利用全国重点实验室,长春 130102

摘要:土壤动物学面临以全新知识体系为科学研究框架的变革时期,其核心内容是以数据驱动为主要特征的人工智能技术方法。目前广泛应用的基于数据库的数据处理分析方法,面临着数据多源异构、快速增长和处理能力不足之间的矛盾。基于快速发展的大数据科学和人工智能技术的数据挖掘方法在解决前述矛盾中有突出优势,但需要依赖一个强大的领域知识库,然而土壤动物领域知识图谱的研究十分匮乏。土壤动物知识图谱是一个具有有向图结构的知识库,其中图的节点代表与土壤动物相关的实体或概念,图的边代表实体或概念之间的各种语义关系。提出了土壤动物知识图谱的定义、内涵、理论模型和构建方法,以浙江天目山土壤螨类多样性为例,分析了构建山地土壤动物知识图谱的技术方法;以土壤动物多样性研究关注的物种分布、物种共存、环境条件对物种的影响作用为例,探讨了基于山地土壤动物知识图谱可以解决的相关科学问题。研究表明,土壤动物知识图谱在解决生物多样性重要科学问题方面具有独特的潜力和优势,有力推动了土壤动物学、信息科学和数据科学交叉的土壤动物信息学的发展。

关键词:土壤动物;知识图谱;大数据;土壤动物信息学;天目山

Theory, method and technique of soil animal knowledge graph construction: a case study of soil mites in Tianmu Mountain, Zhejiang Province

GAO Meixiang[1,2], ZHU Jiaqi[1,2], LIU Shuang[3], CHENG Xin[4], LIU Dong[5], LI Yansheng[4,*]

1 *Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo* 315211, *China*

2 *Ningbo Universities Collaborative Innovation Center for Land and Marine Spatial Utilization and Governance Research at Ningbo University, Ningbo* 315211, *China*

3 *Miyun Experimental Middle School Attached to Beijing Normal University, Beijing* 101500, *China*

4 *School of Remote Sensing and Information Engineering, Wuhan University, Wuhan* 430079, *China*

5 *Key Laboratory of Wetland Ecology and Environment, State Key Laboratory of Black Soils Conservation and Utilization, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun* 130102, *China*

Abstract: Soil zoology is undergoing a period of development with a knowledge system at its core and data-driven. The current widely used data processing and analysis methods based on databases are facing the contradiction between the rapid

基金项目:国家自然科学基金项目(42271051, 41871042);浙江省公益技术应用研究项目(LGN22D010006);宁波市自然科学基金项目(2021J129)

收稿日期:2022-06-09; **网络出版日期**:2023-04-17

* 通讯作者 Corresponding author.E-mail: yansheng.li@ whu.edu.cn

growth of data and the lack of data processing capacity. Data mining methods based on rapidly developing big data science and artificial intelligence techniques have the outstanding advantages in solving the aforementioned contradictions, but they need to rely on a strong domain knowledge base. Yet there is a paucity of research on knowledge graph in the soil animal domain. The soil animal knowledge graph is a knowledge base with the directed graph structure, where the nodes of the graph represent entities or concepts related to soil animals, and the edges of the graph represent various semantic relationships between the entities or concepts. This paper introduced the concept, connotation and construction method of the soil animal knowledge graph. The soil animal knowledge graph is composed of triples (entity, relation, entity) to describe the relationships among entities. Take the soil mite diversity across an altitudinal gradient in Tianmu Mountain, Zhejiang Province as an example, this paper provided the method and process of constructing a mountain soil animal knowledge graph. In the process of constructing a soil animal knowledge graph, the data source was identified firstly, and the ontology construction objective and process of the soil animal knowledge graph were designed. The process of ontology construction included designing the soil animal knowledge graph model, dividing the core classes and class hierarchies, defining ontology attributes, and evaluating the ontologies. Finally, a preliminary graph of the mountain soil animal knowledge graph was displayed with Neo4j platform. To show the process of a data mining for a soil animal knowledge graph, this paper answers three important scientific questions based on the constructed mountain soil animal knowledge graph: that is where are the soil animals distributed, what species live together, and how environmental conditions affect the soil animal distribution. Finally, this paper pointed out that the soil animal knowledge graph has friendly portability and excellent scalability and indicated that some important scientific issues would be solved quantitatively for soil zoology in the future. In total, this study shows that the soil animal knowledge graph has unique potential and advantages in addressing importantly scientific questions about biodiversity, and has promoted the development of soil animal informatics at the intersection of soil zoology, information science and data science.

Key Words: soil animal; knowledge graph; big data; soil animal informatics; Tianmu Mountain

土壤动物多样性研究目前主要以多样性格局、维持机制和功能作用等为研究重点,相较于其他生物类群,土壤动物研究在这些领域的进展相对缓慢,缺少智能化、定量化的数据分析和深度挖掘方法是其重要障碍之一。线粒体基因组[1]、环境 DNA[2]、数字全息显微镜[3—4]、高清相机[5]、深度学习[6—7]等智能化监测仪器和现代方法技术的推广应用,实现了小时级甚至秒级的自动化监测,促进了土壤动物研究数据量的快速增加。目前土壤动物研究的数据管理、分析仍主要采用相对简单的数据库[8],在面对如此大规模的数据量时,产生了数据快速增长和数据处理能力不足之间的矛盾[9—10]。土壤动物相关研究经历了实验科学、理论科学、计算科学这三个科学研究范式阶段,正面临着从第三研究范式到数据驱动的第四研究范式转变的快速发展时期[11—13],表现为第一、二、三研究范式和第四研究范式并存和相互促进的科学研究阶段[7, 14—15]。推动基于海量数据的定量化、智能化研究模式,是解决土壤动物学面临难题的重大机遇,也是当代土壤动物多样性研究的前沿领域。

知识图谱(Knowledge Graph)是解决前述难题和挑战的有效途径。知识图谱是一种用图结构建模事物及事物间联系的数据表示形式,其背后是庞大的信息库,可以将信息表达成更接近人类认知的方式,提供了一种组织管理和认知理解海量信息的能力[16]。知识图谱受到学术界的广泛关注[17—18],已被应用在动物学、农学、林学等领域[19—22],但土壤动物领域未见相关报道。知识图谱在定量刻画土壤动物及相关事物之间联系和作用方面具有突出优势,土壤动物知识图谱(Soil animal knowledge graph, SAKG)的研究与发展,将进一步拓展数据与知识双重驱动的研究新方向[23],促进土壤动物相关领域的定量化、智能化研究。

土壤动物知识图谱是土壤动物与知识图谱的交叉研究,体现了土壤动物学与信息科学、计算机科学、数据库科学等多学科交叉优势,有助于推动土壤动物信息学这一全新学科的发展。本文首先分析土壤动物知识图

谱的基础理论知识;然后以“山地生物多样性联合研究网络”的阶段性成果为例,阐述土壤螨类知识图谱的构建方法;利用基于土壤动物知识图谱的查询、推理和知识表示等数据挖掘方法,探讨可以定量化解决的重要科学问题,并进一步分析土壤动物知识图谱的发展趋势、可以解决的重要科学问题与前沿方向。旨在通过山地土壤动物知识图谱的构建和实践,推动土壤动物信息学的发展。

1 土壤动物知识图谱理论方法

1.1 知识图谱定义

谷歌于2012年引入知识图谱技术[24],通过在搜索查询中找到概念之间的关联来获取知识[25]。知识图谱是很多相关技术相互影响和集成发展的产物,包括语义网络、知识表示、本体论、自然语言处理等[16]。其本质是一种结构化的大规模语义知识库,用符号形式描述物理世界中的概念及其相互关系[26],既包含丰富的语义信息,又具有图的各种特征。

知识图谱被定义为 $G=\{E, R, F\}$,其中 E,R 和 F 分别代表实体、关系和事实的集合,事实被定义为一个三元组 $(h, r, t) \in F$,h 和 t 分别代表头实体和尾实体,r 代表头实体和尾实体之间的关系。图1是土壤动物知识图谱的一个片段,图中节点表示实体,有向边表示实体之间的关系,边的方向表明了实体是作为主体还是对象出现。三元组(端三甲螨属,包含,姬端三甲螨)表达了端三甲螨属包含姬端三甲螨这一物种的系统分类事实;三元组(姬端三甲螨,具有,特征)、(特征,具有,体色)和(体色,表现,棕黄色)表达了姬端三甲螨体色是棕黄色的功能性状事实;三元组(姬端三甲螨,分布,地理区域)、(地理区域,省,浙江省),表达了姬端三甲螨分布在浙江省的地理分布事实。这些三元组是表达土壤动物知识图谱的片段或基本形式,基于这些实体、实体间关系,可以实现土壤动物基础知识、科学问题的基本搜索和深度挖掘。

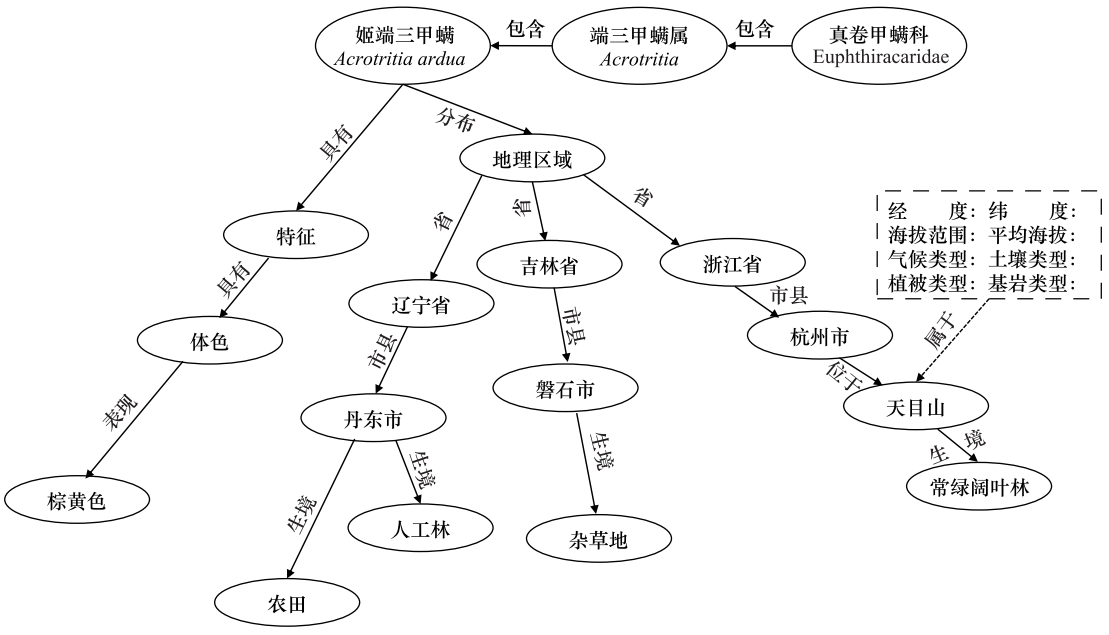


图1 土壤动物知识图谱案例

Fig.1 Example of soil animal knowledge graph

虚线框内所示为实体对应的属性

按照数据来源,可将知识图谱分为基于网络百科资源和基于结构化、半结构化、非结构化数据的知识图谱[27—28];按照覆盖范围,可分为通用和领域知识图谱[29]。通用知识图谱覆盖内容广,融合实体多,但准确度不高,难以借助标准知识库来规范实体、属性和关系。领域知识图谱针对特定领域数据来构建,实体属性和数

值详细规范,包含更复杂更全面的领域知识。本文采用基于野外调查和文献资料获取的数据来构建土壤动物知识图谱,属于基于结构化数据构建的领域知识图谱。

知识图谱主要有自顶向下和自底向上两种构建方式[16]。自顶向下方式首先构建顶层关系本体,然后将抽取到的实体匹配更新到所构建的顶层本体中[30]。自底向上的方式直接将抽取数据中发现的类别、实体、属性及关系合并到知识图谱中[31]。采用自顶向下的方式构建土壤动物知识图谱。

以大规模语义网络为基础的语义知识图谱(Semantic Knowledge Graph),不同于基于文献计量学的科学知识图谱(Mapping Knowledge Domain)。科学知识图谱起源于 Garfield 提出的将引文索引应用于检索文献的思想[32],于 2005 年被我国学者引入到相关研究中[33],目前已成为国内文献研究的重要方法。依托科研文献发现实体关系,实现从科学知识图谱计量关联到语义知识图谱语义关联的转换,有助于构建更专业化、标准化的领域知识图谱,这将是土壤动物学领域一个重要的研究方向[34—35]。

1.2　土壤动物知识图谱的基本定义、理论模型

土壤动物知识图谱是一种语义网络知识库,是一个具有有向图结构的知识库,其中图的节点代表与土壤动物相关的实体或概念,图的边代表实体或概念之间的各种语义关系。构建土壤动物知识图谱的目的在于建模、识别、发现和推理土壤动物学相关概念、事物之间的复杂网络关系,基于土壤动物概念、事物关系的可计算模型,采用基于知识图谱的数据查询、推理、知识表示等数据挖掘方法,解决土壤动物学面临的基础理论和实践应用难题,促进数据科学范式时代土壤动物学的快速发展。土壤动物知识图谱涉及如下基本概念(图 1):

实体:也被称为实例,每一条实际的数据或信息被称为实体,如姬端三甲螨、人工林等。

概念:也被称为类型,用于管理土壤动物知识图谱中的节点,通过配置属性可结构化实体,实体与概念的关系是多对多,如果一个实体属于多个概念,则其结构为这些概念属性的集合,如姬端三甲螨属于物种名称、甲螨亚目、疥螨目、蜱螨亚纲、蛛形纲、节肢动物门、动物界这些概念。

关系:用于管理土壤动物知识图谱中的边,通过配置属性可结构化边实体,一个边实体只能属于一个关系,如包含、分布等。

属性:每个概念或关系可以设置多个属性,如天目山的属性包括经度、纬度、气候类型等。

1.3　土壤动物知识图谱构建的基础理论方法

土壤动物知识图谱的方法论涉及多个方面。一般流程为:首先确定知识表示模型,然后根据数据来源选择不同的知识获取手段导入知识,接着综合利用知识抽取、知识融合、知识加工等技术对构建的知识图谱进行质量提升,最后根据需求设计不同的知识访问、呈现和挖掘方法,如图谱可视化分析等(图 2)。

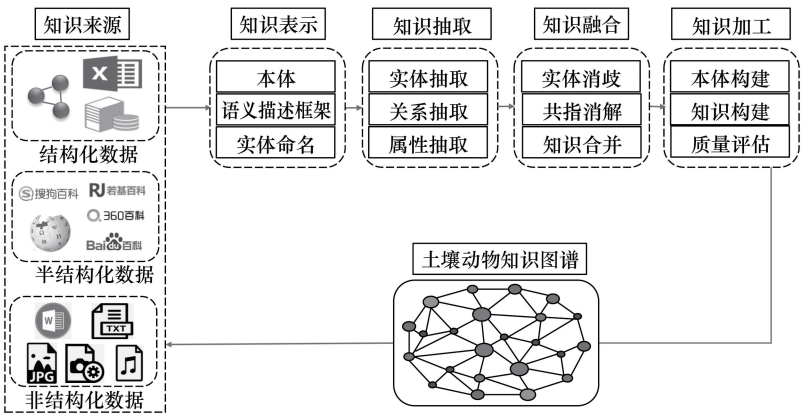


图 2　土壤动物知识图谱构建技术流程

Fig.2　Technical process of constructing soil animal knowledge graph

知识来源。可以从多种来源获取土壤动物知识图谱数据,包括文本、结构化数据库、多媒体数据、传感器

数据和人工众包等。每一种数据源的知识化都需要综合各种不同的技术手段,结构化、半结构化和文本数据是目前土壤动物知识图谱最主要的知识来源[16, 36]。

知识表示。是指用计算机符号描述和表示人脑中的知识,以支持机器模拟人的心智进行推理的方法和技术。知识表示决定了土壤动物知识图谱构建的产出目标,即土壤动物知识图谱的语义描述框架、本体、知识交换语法、实体命名及 ID 体系[16, 37—38]。

知识抽取。是从不同来源、不同结构的数据中提取知识,形成结构化数据存入土壤动物知识图谱[5],其任务一般包括概念抽取、实体识别、关系抽取(实体属性抽取、实体关系抽取)、事件抽取和规则抽取等[16, 37]。

知识融合。是将不同知识库得到的同一实体或概念的描述信息融合起来。描述信息可以是同种类型,也可以是不同类型,如文字、图像、音频、视频等[16, 37]。

知识加工。主要包括基于土壤动物知识图谱的知识检索、知识推理、知识挖掘等,实现形式包括语义检索、智能问答、知识挖掘等[16, 37]。

土壤动物知识图谱属于交叉领域,除了土壤动物学之外,涉及土壤学、生态学、人工智能、数据库、自然语言处理、机器学习、分布式系统等,具体涉及数据库系统、智能问答、推理、推荐系统、区块链与去中心化的知识图谱等技术[16],相关研究算法涉及传统模型(LaSIE-II、Bootstrap)、基于神经网络的模型(CNN-CRF、ED-LAN)和基于图神经网络的模型(CGN、GNED)等[36, 39]。

2 土壤动物知识图谱构建方法

2.1 数据来源

数据来源为“山地生物多样性联合研究网络”(Biodiversity along Elevational Gradients: Shifts and Transitions, BEST)调查的山地土壤动物数据。截至 2022 年 6 月 1 日,收集了 183 个垂直梯度、732 个样方、1512 个土壤动物样品数据。本文以浙江天目山为例说明土壤动物知识图谱的构建和数据挖掘方法。

天目山位于浙江省西北部的杭州市临安区(30°18′—30°21′ N, 119°24′—119°27′ E),主峰仙人顶海拔 1506m,最低海拔约 300m。地处中亚热带北缘,受海洋性季风气候影响,年均气温 8.8—14.8℃,年均降水量 1390—1870mm。海拔 600m 以下为红壤带,600—1200m 为黄壤带,1200m 以上为棕黄壤带[40—41]。

2021 年 6 月,在天目山 12 个 20m×20m 的海拔梯度样地中分别设置 4 个 2m×2m 样方,在每个样方内随机采集 3 个 15cm×15cm 的凋落物层样品,再使用土钻随机采集 3 个 0—15cm 的土壤层样品。按照 Krantz 和 Walter[42]分类系统将土壤螨类鉴定到种,共获取 85 个物种、6918 只个体。

2.2 本体构建目标及流程

本体构建的目标是依照土壤动物领域的专业知识,对山地土壤动物知识体系进行建模,形成客观明确的描述规范,在结构化、形式化和语义化层面实现描述、表达、挖掘、推理和联想知识的能力,并通过本体描述语言使土壤动物专业知识可以被计算机理解和处理,构建流程如下(图 3):

(1)确定本体描述的专业领域。在构建本体之前确定本体所描述的领域范畴以及边界,包括山地不同垂直带土壤动物的基础知识、与土壤动物生存有紧密关系的土壤、植被数据等。

(2)设计本体概要模型。以土壤动物专业领域知识为基础,搜集该领域的信息资源,掌握相关的知识体系,总结、抽象概念并建立本体概要模型,形成山地土壤

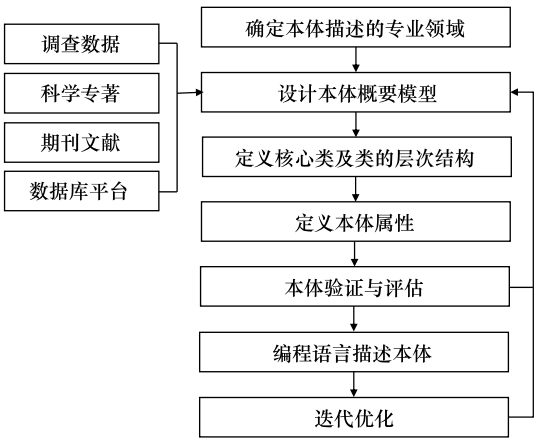


图 3 山地土壤动物本体的构建流程

Fig.3 Construction workflow of mountain soil animal ontology

动物本体框架。

(3)定义核心类及类的层次结构。对山地土壤动物概要模型进行细化和补充,围绕科学问题和实践应用需求,梳理、精炼核心概念及概念类间的层次结构。

(4)定义本体属性。通过定义本体类的对象属性和数据属性,将土壤动物自身的特征信息,以及土壤动物与样地、环境因子、现有的知识组织体系的语义关联引入本体模型,体现完整的土壤动物知识关联体系。

(5)本体验证与评估。构建领域本体后,采用专家咨询法,请领域专家从本体结构、实践应用场景等方面对所构建的本体模型进行评估,进一步修正与完善本体。

(6)编程语言描述本体。通过验证和评估后,对本体模型进行形式化表示,增强本体的表达能力。

(7)本体迭代优化。概念模型的构建是一个动态过程,本文以土壤螨类数据为例构建知识图谱,后续获得土壤跳虫、线虫、蚯蚓等类群数据后,可通过上述流程综合分析,将其它类群的数据扩充到本体模型中,进一步迭代和完善山地土壤动物知识图谱的本体。

2.3　土壤动物本体构建

2.3.1　土壤动物本体模型

本体概要模型描述了在抽象层次由概念及概念间关系建立的逻辑模型,借助概要模型可梳理本体构建的概念范围及核心属性的取值范围。本文采用人工构建的方法,由领域专家协作和指导,使用 Protégé 构建本体库,将山地森林土壤动物本体设置为 7 层,包括土壤动物类、类群类、山地类、样地类、样方类、样品类、物种类。以天目山土壤螨类为主要实例,用抽象层面自顶向下的要素列举方法,设计的细粒度土壤动物本体概要模型见图 4。

2.3.2　核心类及类的层次结构

在山地土壤动物知识图谱的本体中,类是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,实例和属性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,本研究主要根据一级类对其他类进行类的划分,并进行了描述。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,对本体中类和属性运用英文名称进行了定义。本体中的类具体如下:

(1) 1 个一级类,即土壤动物类。研究的是山地森林土壤动物领域本体构建并将其进行形式化表示,因此以土壤动物作为最高层次类。

(2) 8 个二级类,即类群类 (Group)。设置了螨 (Mite)、跳虫 (Collembola)、线虫(Nematode)、蚯蚓(Earthworm)、甲虫 (Beetle)、马陆 (Millipede)、蜈蚣 (Centipede)、蜘蛛 (Spider)用于描述山地森林主要土壤动物类群。本研究主要添加了螨类的数据属性,后续可以补充其他类群的数据属性,进一步拓展类群类及其之间的关联。

(3) 12 个三级类,即山地类 (Mountain)。包括天目山 (TMS)、天山野核桃沟 (TS1)、天山库尔德宁(TS2)、大别山 (DBS)、峨眉山 (EMS)、金佛山 (JFS)、官山 (GS)、玉龙雪山 (YLXS)、戴云山 (DYS)、大明山(DMS)、高黎贡山 (GLGS)和霸王岭 (BWL)。

(4) 183 个四级类,即样地类 (Plot)。以天目山为例,设置了样地 1 (TMS1)、样地 2 (TMS2)、样地 3 (TMS3) 等 12 个位于不同海拔梯度的样地。根据事实,不同山地设置的样地数量不同,例如高黎贡山和玉龙雪山分别设置了 31 和 19 个样地类。

(5) 732 个五级类,即样方类 (Quadrat)。以天目山的样地 1 (TMS1) 为例,设置了样方 1 (TMS1_1)、样方 2 (TMS1_2)、样方 3 (TMS1_3) 和样方 4 (TMS1_4) 4 个本体。

(6) 1512 个六级类,即样品类 (Sample)。以天目山样地 1 的样方 1 (TMS1_1) 为例,设置了凋落物层样品 (TMS1_1LL) 和土壤层样品 (TMS1_1SL) 2 个本体。

(7) 85 个七级类,即物种类 (Species)。设置的物种本体是天目山的 85 个土壤螨类物种。

2.3.3　本体属性定义

类本身不足以描述土壤动物专业领域的知识体系,需要通过定义类的内部属性信息和外部关系信息来描

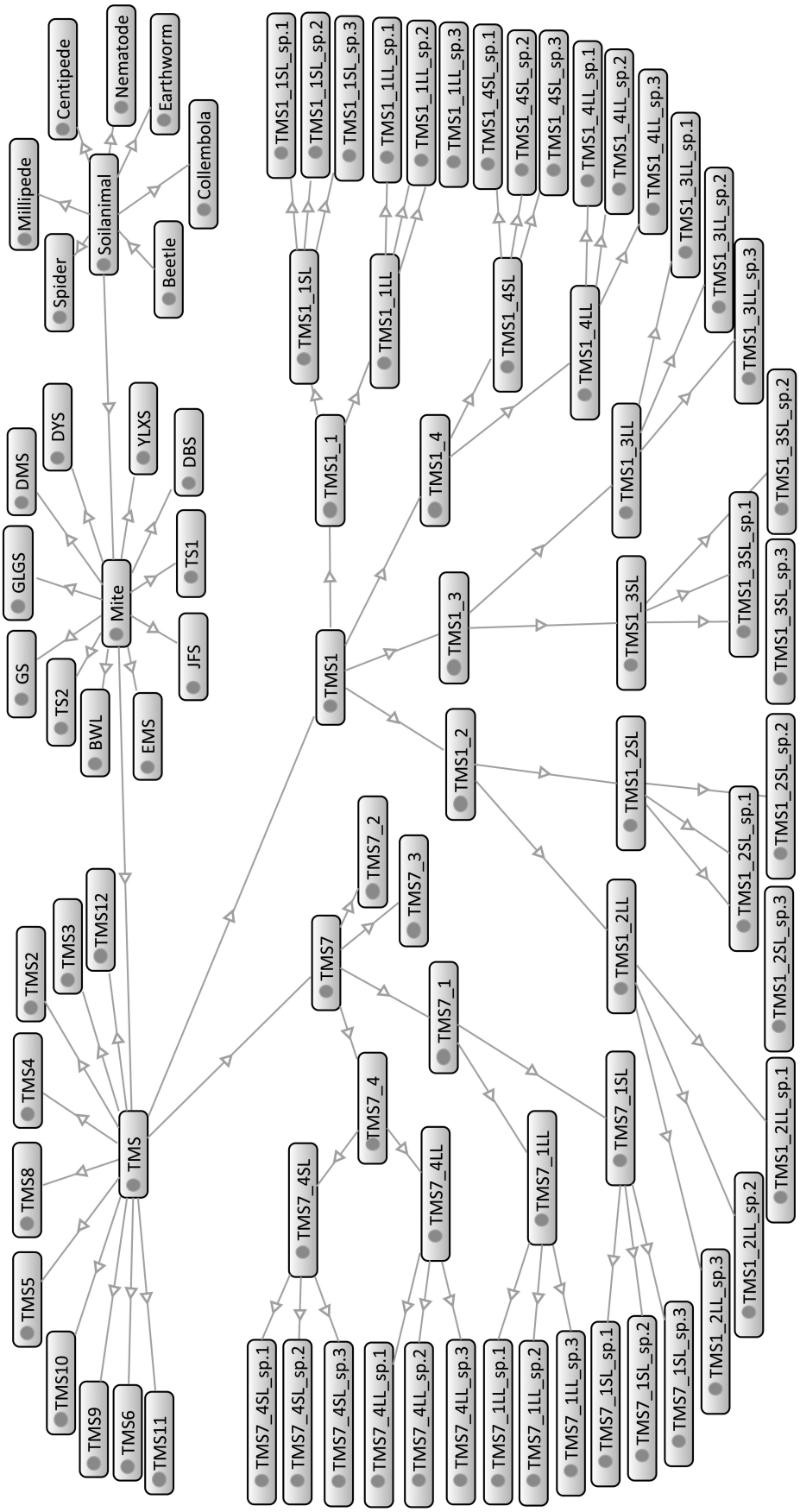


图4 土壤动物知识图谱部分本体模型

Fig. 4 Ontology model for soil animal knowledge graph

TMS：天目山；DYS：戴云山；DMS：大明山；GLGS：高黎贡山；GS：官山；TS2：天山库尔德宁；BWL：霸王岭；EMS：峨眉山；JFS：金佛山；TS1：天山野核桃沟；DBS：大别山；YLXS：玉龙雪山；TMS1：天目山1号样地；TMS2：天目山2号样地；TMS3：天目山3号样地；TMS4：天目山4号样地；TMS5：天目山5号样地；TMS6：天目山6号样地；TMS7：天目山7号样地；TMS8：天目山8号样地；TMS9：天目山9号样地；TMS10：天目山10号样地；TMS11：天目山11号样地；TMS12：天目山12号样地；TMS1_1：天目山1号样地的第1个样方；TMS1_2：天目山1号样地的第2个样方；TMS1_3：天目山1号样地的第3个样方；TMS1_4：天目山1号样地的第4个样方；TMS7_1：天目山7号样地的第1个样方；TMS7_2：天目山7号样地的第2个样方；TMS7_3：天目山7号样地的第3个样方；TMS7_4：天目山7号样地的第4个样方；TMS1_1LL：天目山1号样地的第1个样方的土壤层样品；TMS1_1SL：天目山1号样地的第1个样方的凋落物层样品；TMS1_2LL：天目山1号样地的第2个样方的土壤层样品；TMS1_2SL：天目山1号样地的第2个样方的凋落物层样品；TMS1_3LL：天目山1号样地的第3个样方的土壤层样品；TMS1_3SL：天目山1号样地的第3个样方的凋落物层样品；TMS1_4LL：天目山1号样地的第4个样方的土壤层样品；TMS1_4SL：天目山1号样地的第4个样方的凋落物层样品；TMS7_1LL：天目山7号样地的第1个样方的土壤层样品；TMS7_1SL：天目山7号样地的第1个样方的凋落物层样品；TMS7_2LL：天目山7号样地的第2个样方的土壤层样品；TMS7_2SL：天目山7号样地的第2个样方的凋落物层样品；TMS7_3LL：天目山7号样地的第3个样方的土壤层样品；TMS7_3SL：天目山7号样地的第3个样方的凋落物层样品；TMS7_4LL：天目山7号样地的第4个样方的土壤层样品；TMS7_4SL：天目山7号样地的第4个样方的凋落物层样品；sp.1：物种1；sp.2：物种2；sp.3：物种3，下同

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,分为对象和数据属性。对象属性揭示类之间的语义关系,是在本体中进行逻辑推理的基础,决定了土壤动物知识图谱的丰富程度和应用效果。数据属性描述类自身的特点,使用具体的数据类型进行表示,包括文本、数值、日期等多种类型。

(1)定义本体对象属性。定义了 6 个对象属性,为了增强语义严谨性,设置了对象属性的英文标识、定义域和值域(表 1)。

表 1　山地森林土壤动物知识图谱本体的对象属性

Table 1　Object property of ontologies for mountain soil animal knowledge graph

属性名称 Property name	英文标识 English label	定义域 Domain of definition	值域 Range	释义 Description
包含 Include	include	土壤动物类	类群类	描述土壤动物的类群
分布于 Be distributed in	distribute	类群类	山地类	描述土壤动物类群分布的山地
设置样地 Setting plot	set_plot	山地类	样地类	描述在山地设置的样地
设置样方 Setting quadrat	set_quadrat	样地类	样方类	描述在样地内设置的样方
采集样品 Collecting sample	set_sample	样方类	样品类	描述在样方内采集的样品
生存着 Exist	exist	样品类	物种类	描述生存在样品内的物种

(2)定义本体数据属性。用来描述本体自身的特征,详见表 2:

土壤动物类,定义了 1 个数据属性,描述山地土壤动物知识图谱的主要对象。

类群类,定义了 6 个数据属性,描述类群的生物学分类地位。

山地类,定义了 14 个数据属性,描述山地的基本特征,其中最低和最高纬度、最低和最高经度,是为了数据挖掘时分析该山地土壤动物所在位置、分布范围、生存界限等。

样地类,定义了 21 个数据属性,涉及微气候、地形和植被等对土壤动物有重要影响的因子。

样方类,设置了 5 个数据属性,其中长边、短边、半径用于数据挖掘时计算和分析样方面积。

样品类,设置了 12 个数据属性,形状、长边、半径和深度等用于计算和分析样品的面积和体积。采集年份、采集月份和采集日,可用于计算和分析土壤动物多样性动态特征。

物种类,设置了 8 个数据属性,用于计算分析物种多样性和功能多样性特征。

表 2　山地森林土壤动物知识图谱本体的数据属性

Table 2　Data property of ontologies for mountain soil animal knowledge graph

本体 Ontology	数据属性及对应的英文标识 Data attribute and corresponding English label
土壤动物类 Soilanimal	名称(title)
类群类 Group	名称(title),中文名称(Chinese name),界(kingdom),门(phylum),纲(class),目(order)
山地类 Mountain	名称(title),省(province),最低纬度(lat_low),最高纬度(lat_high),最低经度(lon_low),最高经度(lon_high),年均温度(mean_T),年均最低温(min_T),年均最高温(max_T),年均降雨量(mean_P),土壤类型(soil_type),植物物种数量(plant_species),气候类型(climate),最高海拔(max_elevation)
样地类 Plot	名称(title),海拔高度(elevation),纬度(lat),经度(lon),坡度(slope),坡向(aspect_cos),植物物种数量(richness),胸径(BA_perHectare),土壤 pH(pH),土壤全氮(TN),土壤全磷(TP),土壤有机碳(OC),土壤含水量(WC),土壤有效磷(AP),土壤有效钾(AK),铵根离子(NH_4^+),硝酸根离子(NO_3^-),土壤表层温度(tempSoil6cm),土壤表面温度(tempSurface),近地面大气温度(temAir12cm),土壤湿度(moistSoil)
样方类 Quadrat	名称(title),形状(shape),长边(long_edge),短边(short_edge),半径(radius)
样品类 Sample	名称(title),形状(shape),长边(long_edge),短边(short_edge),半径(radius),深度(depth),采集年份(year),采集月份(month),采集日(day),采集单位(unite),采集人(collector),采集时天气(weather)
物种类 Species	名称(title),科(family),属(genus),数量(number),体长(body_length),体宽(body_width),体色(colour),DNA 条形码(DNA barcoding)

2.3.4　本体验证与评估

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。清晰性指所定义的山地土壤动物知识图

谱的类和属性必须明确,有科学依据和意义,不存在歧义。一致性指土壤动物知识图谱类间的关系在逻辑上必须是一致的,保证数据推理和挖掘的科学性、严谨性。完整性指所定义的山地土壤动物知识图谱领域内类和属性是完整的,可以用于描述数据来源中的知识体系。可扩展性指当领域内出现新的概念时,山地土壤动物知识图谱的本体可规模化扩展。本体初步构建完成后采用专家咨询的方法,邀请土壤动物学和山地森林生态学领域多名专家进行评估。专家认为本文所构建的山地土壤动物领域本体包含了较完整的山地土壤动物概念体系,符合上述评估标准和预期要求,可进行本体的语言描述。

2.4 山地土壤动物本体的形式化表示

图数据库将知识图谱的实体和概念作为图节点,实体属性和关系作为边,以图的形式进行存储,比较直观地反映知识图谱的内部结果,有利于进行图查询以及知识推理,并具有很强的扩展性。Neo4j 是一个开源的图数据库系统,底层使用图数据结构进行存储,大幅度提升数据检索的性能,是目前用于知识图谱存储的主要途径。将山地土壤动物知识图谱存储于 Neo4j 图数据库中[43]。

首先将获取的实体节点和关系数据分别保存为.csv 文件,然后通过 Cypher 语言的 LOAD CSV 语句导入节点和关系。采用 Cypher 语句将实体和实体间的关系存储到 Neo4j 图数据库中[44],形成山地土壤动物知识图谱(图 5)。

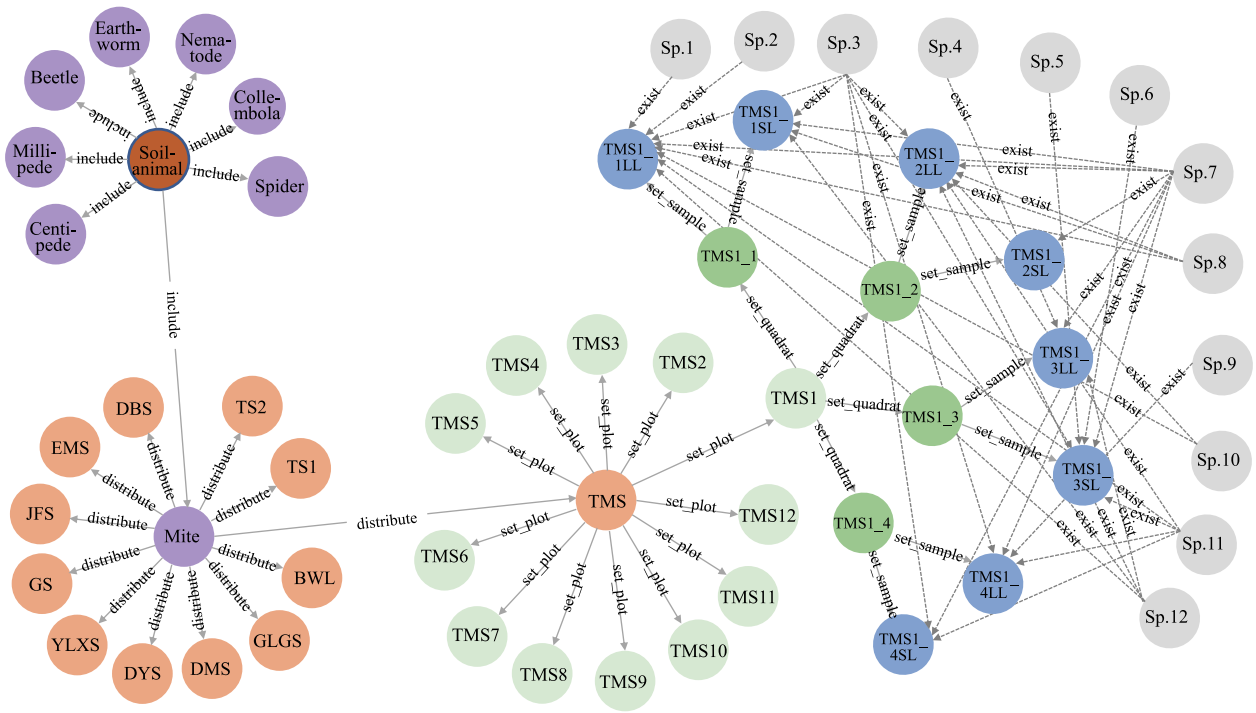


图 5 山地土壤动物知识图谱的可视化

Fig.5 Visualization of mountain soil animal knowledge graph

图中圆圈为实体节点,带箭头的边为关系,缩写的含义见图 4 和表 1;Sp.1:*Damaeus spiniger* 矩刺珠甲螨;*Sp.*2:Platynothrus peltifer;*Sp.*3:Hypochthonius rufulus 淡红缝甲螨;*Sp.*4:Hypochthoniella minutissima;*Sp.*5:Ceratozetella *sp.*1;*Sp.*6:Hypochthonius luteus 金黄缝甲螨;*Sp.*7:*Oribatida sp.*7 甲螨 *sp.*7;*Sp.* 8:Acrotritia ardua 姬端三甲螨;*Sp.* 9:Eremulus avenifer 阿沙甲螨;*Sp.* 10:*Oribatida sp.* 10 甲螨 *sp.* 10;*Sp.* 11:Suctobelbella *sp.*1;*Sp.*12:Atropacarus striculus

3 基于土壤动物知识图谱的分析

基于构建的土壤动物知识图谱,本文围绕生物多样性研究的三个重要科学问题“分布在哪里”、“哪些物种共存在一起”和“环境条件如何影响物种分布”,举例说明知识图谱如何帮助我们定量化地回答这些科学问题。

3.1　土壤动物的分布生境

头长单翼甲螨 *Protoribates capucinus* 分布在哪些乔木丰富度大于 20 种、坡度大于 20°、土壤全氮值大于 0.25的生境中？数据挖掘过程如下：

PROFILE

MATCH (a:Species{title:'Protoribates capucinus'})-[*3..5]-(b:Plot)

WHERE b. richness > 20 AND b. slope > 15 AND b.TN>0.25

RETURN a.title, b.title

可以发现头长单翼甲螨 *Protoribates capucinus* 分布在 6 个满足前述条件的生境中,包括 TMS1、TMS3、TMS7、TMS8、TMS10、TMS12 样地 (图 6)。可以进一步推理,如果补充了 12 个山地所有的土壤螨类数据,便可以推理从新疆天山到海南霸王岭跨越 25 个纬度的大空间尺度上,某个或某些土壤动物分布在哪些符合特定条件的地点或生境。

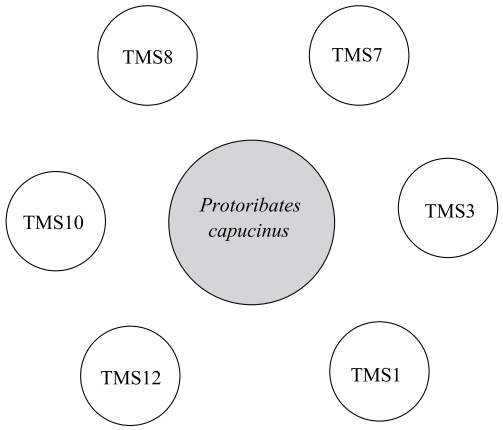


图 6　案例 1 的数据挖掘结果

Fig.6　Data mining result of case 1

缩写的含义见图 4;*Protoribates capucinus*:头长单翼甲螨

3.2　特定生境中的共存物种分析

头长单翼甲螨 *Protoribates capucinus* 和土库曼罗甲螨 *Lohmannia turcmenica* 共同生存在哪些个体数量大于 50 只、物种数量大于 10 种、时间节点为 5 月的生境中？数据挖掘过程如下：

PROFILE

MATCH (a:Species{title:'Protoribates capucinus'})-[*1..5]->(b:Sample)<-[*1..5]-(c:Species{title:'Lohmannia turcmenica'})

WHERE b.ind_num>50 AND b.spe_num >10 AND b.month=5

RETURN a.title,b.title,c.title

结果发现头长单翼甲螨 *Protoribates capucinus* 和土库曼罗甲螨 *Lohmannia turcmenica* 共存在 12 个生境中,主要为海拔高度为 500—1000m 的凋落物层(图 7)。该案例表明土壤动物知识图谱可用于共存物种之间、共存物种与生存环境因子之间的查询与推理,有助于揭示土壤动物物种共存特征、共存物种与环境因子之间的关系对共存物种的影响等科学问题,在生物间作用、环境过滤对多样性维持的贡献等关键科学问题中将发挥重要作用。

3.3　复杂环境条件对土壤动物分布的影响

海拔高度大于 300m、坡度大于 15°、乔木物种丰富度大于 15 种、土壤有机碳含量大于 10%、土壤 pH 值大于 3.7、土壤有效磷含量大于 7mg/kg 的复杂环境条件,影响哪些个体数量大于 50 头、在 5 月份活跃在天目山的土壤螨类的物种分布。数据挖掘过程如下：

PROFILE

MATCH (a:Plot)-[*1..5]-(b:Species), (c:Sample)

WHERE b.num>50 AND a.elevation>300 AND a.slope>15 AND a.richness >15 AND a.OC>10 AND a.pH>3.7 AND a.AP>7 AND c.month=5

RETURN a.title,b.title

结果表明,前述综合的复杂环境条件共同影响 20 个物种在天目山的分布,包括头长单翼甲螨 *Protoribates capucinus*、*Eremulus avenifer*、*Tectocepheus velatus* 等 (图 8)。

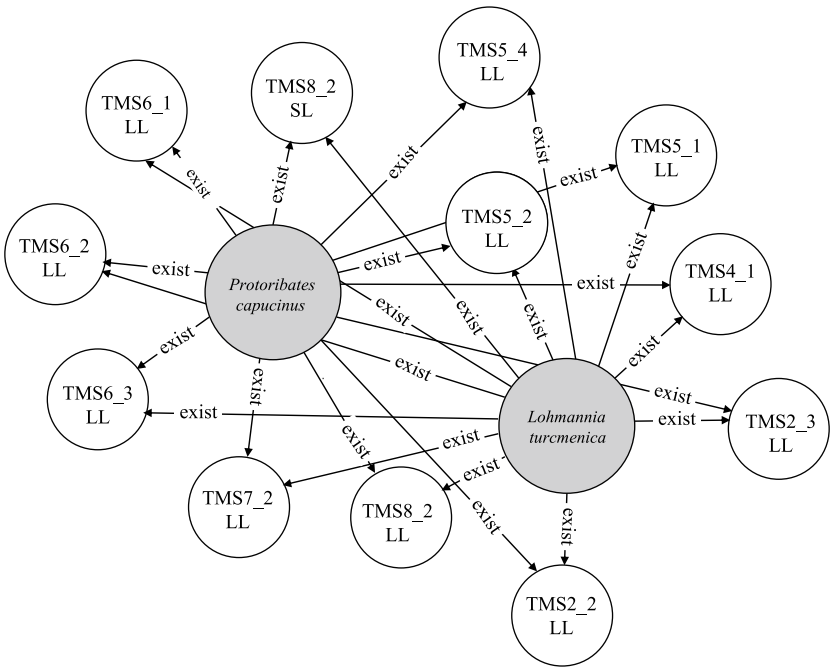


图 7 案例 2 的数据挖掘结果

Fig.7 Data mining result of Case 2

缩写见图 4 和表 1;*Protoribates capucinus*:头长单翼甲螨;*Lohmannia turcmenica*:土库曼罗甲螨

前述 3 个案例仅是基于天目山土壤螨类数据的实践和应用,基于山地土壤动物知识图谱的查询、推理和知识表示等方法,可以进一步挖掘更多科学问题。如:(1) 哪些土壤动物可以同时生活在土壤 pH 值在 4—7 的凋落物层和土壤层中。(2) 哪些土壤动物是特定山地的特有种,至少在 3 次数据调查中发现其只分布在该特定山地。(3) 哪些土壤动物是所有海拔高度的广布种,广泛分布在从 200—3000m 的山地中。(4) 哪些土壤动物是跨纬度山地生态系统的广布种,同时分布在从温带到热带的山地中。(5) 土壤有机碳含量大于 6%、山地坡度大于 15°、每公顷树木胸径大于 30mm 的生境中共存土壤动物的体长、体宽、生物量范围。(6) 是否存在着在所有山地都是优势种的土壤动物,其个体数量在所有山地的占比都大于 10%。(7) 什么性状的土壤动物更喜欢生活在凋落物层而不会生活在土壤层。

4 讨论

构建土壤动物知识图谱是一项系统性知识工程。相较于其他研究方法,土壤动物知识图谱在解决土壤动物学的基础理论、应用实践方面具有独特的优势和作用。

4.1 土壤动物知识图谱的特色与优势

(1) 友好的可移植性

依据生物地理学和生态学的基础理论知识,可以实现土壤动物知识图谱的移植、共享和完善。以“土壤动物-包含->类群->分布->山地-包含->样地-包含->样方-包含->样品-包含->群落-包含->物种”的 15 元组为核心对知识进行表达,通过一个完整的知识链路,将土壤动物数量和性状信息与其生存环境中植被、土壤、地形、气候等信息表达出来,实现了基于土壤动物知识图谱对科学问题的解释和挖掘。该知识表达方式和知识链路可以横向移植到其他生态系统中,如农田土壤动物知识图谱、湿地土壤动物知识图谱等。

(2) 优良的可扩展性

可扩展性是土壤动物知识图谱的一个基本特性,当专业领域出现新概念或需要添加新知识时,可以方便地实现知识本体的规模化扩展[45]。构建的山地土壤动物知识图谱,以土壤动物类为起点,可以进一步拓展土

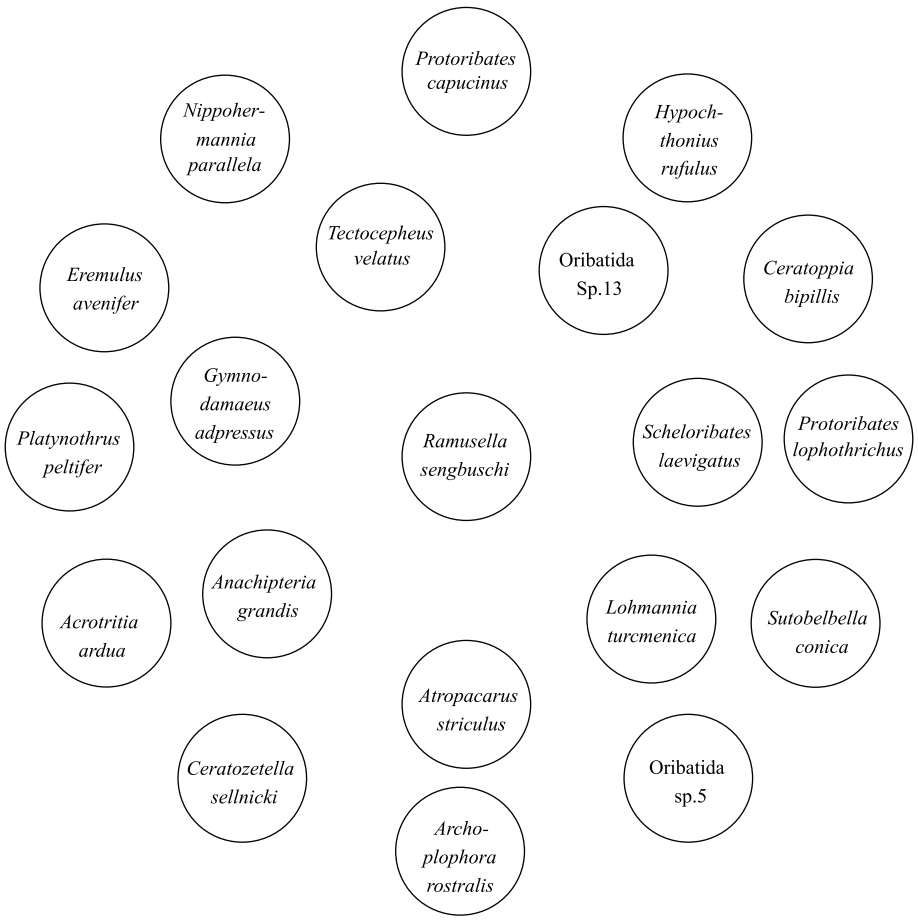


图 8　案例 3 的数据挖掘结果

Fig.8　Data mining result of Case 3

Nippohermannia parallela:平行日本汉甲螨;*Eremulus avenifer*:阿沙甲螨;*Acrotritia ardua*:姬端三甲螨;*Ceratozetella sellnicki*:塞氏小尖棱甲螨;*Archoplophora rostralis*:吻直卷甲螨;*Oribatida sp.*5:甲螨 *sp.*5;*Protoribates lophothrichus*:冠长单翼甲螨;*Hypochthonius rufulus*:淡红缝甲螨;*Tectocepheus velatus*:覆盖头甲螨;*Gymnodamaeus adpressus*:扁裸珠甲螨;*Anachipteria grandis*:大无前翼甲螨;*Scheloribates laevigatus*:滑菌甲螨;*Ramusella sengbuschi*:沈氏枝奥甲螨;Oribatida Sp.13:甲螨 Sp.13

壤动物与植物、微生物、哺乳动物、鸟类等知识概念之间的关联,构建土壤动物与其他生物之间彼此互联的复杂知识图谱。又如基于本研究中的山地类,后续可以进一步拓展长白山、武夷山、贡嘎山等本体类的实例和属性,构建中国乃至全球山地土壤动物知识图谱。

4.2　土壤动物知识图谱可以解决的科学问题

(1) 土壤动物学基础科学问题

相对于农作物害虫[46]、林业[47]、医学[48]等领域,土壤动物学对知识发现和知识梳理的表现性和科普性仍显不足。如姬端三甲螨分布在我国哪些省、市、县的哪些生境地点,这些分布地点年均温、年均降水量、土壤有机质含量分别相差多少。虽然已经积累了较多的数据和资料,由于缺乏有效的知识表达形式,仍然难以回答这些土壤动物学的基础科学问题。土壤动物知识图谱还可以精准地量化其他基础科学问题,如年均温小于等于 10℃的地区生活着哪些土壤动物;同种土壤动物在不同分布区的体长、体宽等功能性状相差多少;某个土壤动物物种喜好什么生境,喜好与哪些物种共存在一起等。

(2) 土壤动物多样性维持机制

土壤动物多样性维持机制是土壤动物学研究的核心内容之一[49]。目前解释土壤动物多样性维持机制的假说主要包括中性理论、生态位理论等[50—51],致力于解决的科学问题集中于扩散[52]、环境过滤[53]和生物间

作用[54]的影响。由于一直难以精准量化前述基本科学问题,导致相关研究进展缓慢。基于土壤动物知识图谱的分析与挖掘,有望量化前述相关科学问题。如量化并回答哪些 DNA 序列与特定范围的体长、体宽等功能性状、特定范围的土壤、植被、地形等环境条件之间存在紧密联系;在给定土壤、植被、地形、气候等特定环境条件下,有哪些土壤动物可以存活,这些存活的土壤动物在数量、性状、DNA 信息等方面有什么适应性特征和一般性规律;在给定土壤、植被、地形、气候等特定环境中,感兴趣的土壤动物之间是否可以共存,在什么环境范围内共存关系最显著。

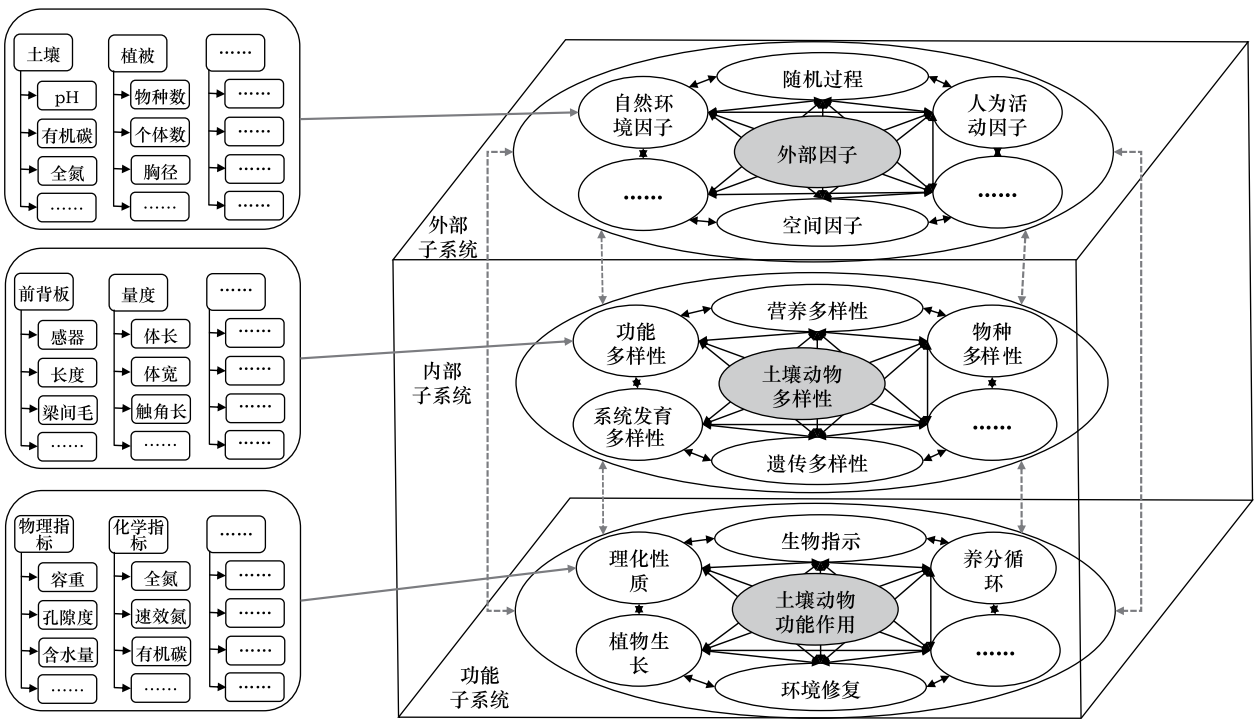
尺度是土壤动物多样性维持机制研究关注的一个重要问题[55],基于土壤动物知识图谱的数据挖掘可以帮助解决尺度相关的科学难题。尺度是连接格局与过程的桥梁和纽带[56—57],相关研究证明地理距离和环境因子对土壤动物分布格局的影响具有尺度依赖性[58—59],但尺度一直是土壤动物多样性维持机制研究的难点。基于严谨的知识体系和知识链路,土壤动物知识图谱有潜力将“种群-群落-样地-生境-局地-区域-全球尺度”的土壤动物多样性与其生存环境的关联表达出来,再通过数据挖掘等方式发现多尺度、跨尺度的科学规律。

(3) 土壤动物多样性功能作用

土壤动物功能作用是土壤动物学、群落生态学等相关研究的重要问题[60—61],现代技术方法和数据科学的交叉与耦合有助于解决该问题[62]。土壤动物活跃于复杂系统中,目前的研究框架难以将多样性子系统、外部环境子系统、功能作用子系统有机地联系在一起[61, 63],难以有效地表达土壤动物功能作用与其他子系统所具有的相互依赖性和非平稳性特征[64],在处理海量数据和复杂关系过程中面临诸多困难和挑战。土壤动物知识图谱不仅可以建立子系统内部要素间的横向关联 (图 9a),还可以构建跨子系统的细粒度要素间的纵向关联 (图 9b),为研究土壤动物多样性功能作用提供细粒度的、可量化方向和关系的研究模式。

(4) 土壤动物学的实践应用

土壤动物知识图谱在土壤动物多样性保护、土壤质量提高、土壤健康管理、农产品质量提升等方面,都将有很大的应用潜力。在土壤动物知识图谱中构建与人类活动相关的类,如农药施用类、化肥施用类、农业灌溉类、耕作方式类、农产品产量类、农产品质量类等,便可以基于知识体系和知识链路,挖掘某些农药、化肥、灌溉、耕作等特定管理模式下会生存哪些土壤动物,以及这些土壤动物的性状、DNA 信息特征,这些预见成果将为土壤质量、土壤健康及人类健康评估等重要科学问题提供有效的数据和方法[65]。



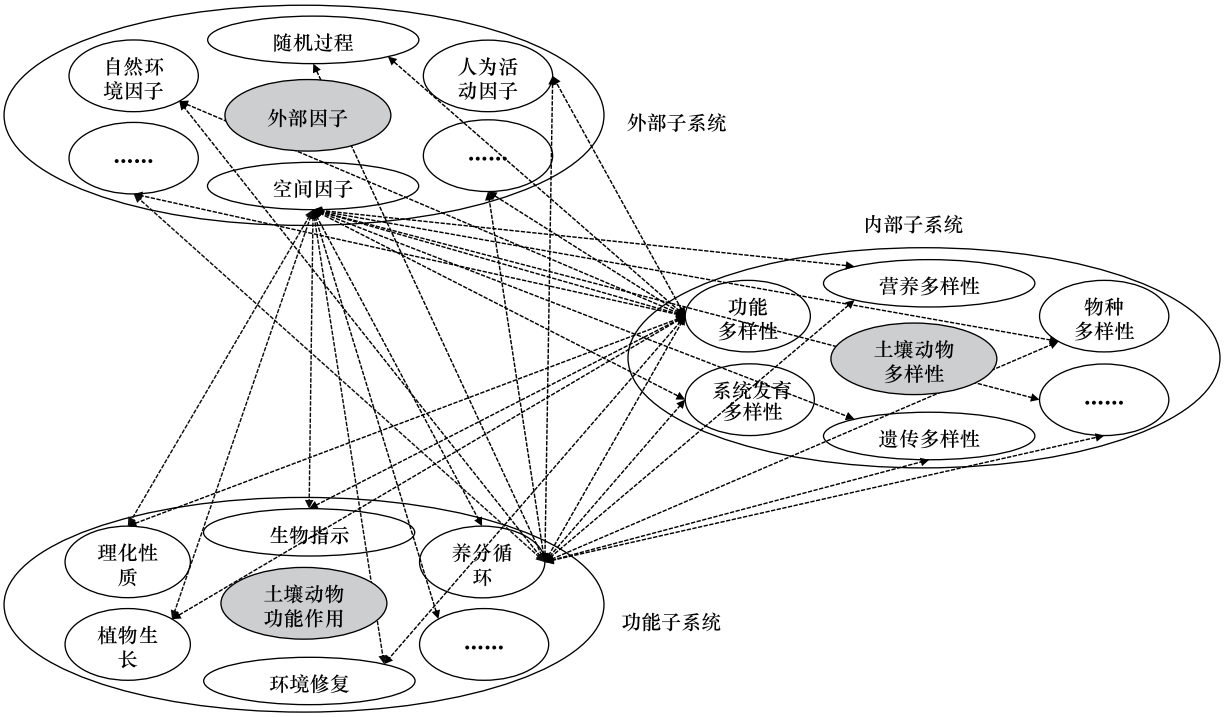


图 9　土壤动物复杂系统

Fig.9　Complex systems of soil animals

a 刻画了子系统内部要素间的横向关联;b 刻画了跨子系统的细粒度要素间纵向关联,以每个子系统中的单个要素为例

5 结论

土壤动物知识图谱正处于萌芽时期,即将迎来快速发展阶段。采用哪些模式和方法构建科学的土壤动物知识图谱,采用哪些数据挖掘方法有效发现科学规律,利用土壤动物知识图谱解决哪些关键科学难题,都是土壤动物知识图谱发展的前沿领域和重要方向。介绍了土壤动物知识图谱的基础理论和构建方法,以构建的山地土壤动物知识图谱为例,围绕分布在哪里、哪些物种共存在一起和环境条件如何影响物种分布这三个重要科学问题,探讨了土壤动物知识图谱的优势和可以解决的科学问题。研究结论将推动土壤动物学、信息科学和数据科学的交叉研究,促进土壤动物信息学新领域的发展。

参考文献(References):

[1] Godeiro N N, Bellini B C, Ding N F, Xu C, Ding Y H, Zhang F. A mitogenomic phylogeny of the Entomobryoidea (Collembola): a comparative perspective. Zoologica Scripta, 2021, 50(5): 658-666.

[2] Lunghi E, Valle B, Guerrieri A, Bonin A, Cianferoni F, Manenti R, Ficetola G F. Environmental DNA of insects and springtails from caves reveals complex processes of eDNA transfer in soils. Science of the Total Environment, 2022, 826: 154022.

[3] Sirico D G, Cavalletti E, Miccio L, Bianco V, Memmolo P, Sardo A, Ferraro P. Kinematic analysis and visualization of Tetraselmis microalgae 3D motility by digital holography. Applied Optics, 2022, 61(5): B331-B338.

[4] 朱晓艳, 韩苗, 韩天富, 韩峥, 王钢. 铜绿假单胞菌在液滴界面的二维和三维运动特征. 微生物学报, 2022, 62(8): 3124-3136.

[5] Flórián N, Grúnicz L, Gergócs V, Tóth F, Dombos M. Detecting soil microarthropods with a camera-supported trap. Insects, 2020, 11(4): 244.

[6] MacNeil L, Missan S, Luo J L, Trappenberg T, LaRoche J. Plankton classification with high-throughput submersible holographic microscopy and transfer learning. BMC Ecology and Evolution, 2021, 21(1): 123.

[7] Djerdj T, Hackenberger D K, Hackenberger D K, Hackenberger B K. Observing earthworm behavior using deep learning. Geoderma, 2020, 358: 113977.

[8] 傅声雷. 利用新方法和野外实验平台加强土壤动物多样性及其生态功能的研究. 生物多样性, 2018, 26(10): 1031-1033.

[9] 肖文宏, 周青松, 朱朝东, 吴东辉, 肖治术. 野生动物监测技术和方法应用进展与展望. 植物生态学报, 2020, 44(4): 409-417.
[10] Norouzzadeh M S, Morris D, Beery S, Joshi N, Jojic N, Clune J. A deep active learning system for species identification and counting in camera trap images. Methods in Ecology and Evolution, 2021, 12(1): 150-161.
[11] Akhtar M N, Shaikh A J, Khan A, Awais H, Bakar E A, Othman A R. Smart sensing with edge computing in precision agriculture for soil assessment and heavy metal monitoring: a review. Agriculture, 2021, 11(6): 475.
[12] Bui E N. Data-driven Critical Zone science: a new paradigm. Science of the Total Environment, 2016, 568: 587-593.
[13] Wadoux A M J C, Román-Dobarco M, McBratney A B. Perspectives on data-driven soil research. European Journal of Soil Science, 2021, 72(4): 1675-1689.
[14] Qing X, Wang Y H, Lu X Q, Li H B, Wang X, Li H M, Xie X J. NemaRec: a deep learning-based web application for nematode image identification and ecological indices calculation. European Journal of Soil Biology, 2022, 110: 103408.
[15] Akintayo A, Tylka G L, Singh A K, Ganapathysubramanian B, Singh A, Sarkar S. A deep learning framework to discern and count microscopic nematode eggs. Scientific Reports, 2018, 8: 9145.
[16] 王昊奋, 漆桂林, 陈华钧. 知识图谱: 方法、实践与应用. 北京: 电子工业出版社, 2019.
[17] Kejriwal M, Sequeda J, Lopez V, Kejriwal M, Lopez V, Sequeda J F. Knowledge graphs: construction, management and querying. Semantic Web, 2019, 10(6): 961-962.
[18] Ji S X, Pan S R, Cambria E, Marttinen P, Yu P S. A survey on knowledge graphs: representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(2): 494-514.
[19] Gharibi M, Zachariah A, Rao P. FoodKG: a tool to enrich knowledge graphs using machine learning techniques. Frontiers in Big Data, 2020, 3: 12.
[20] Xia Y, Sun N, Wang H, Yuan X, Gu L C, Wang C, Gao Q. Research on knowledge question answering system for agriculture disease and pests based on knowledge graph. Journal of Nonlinear and Convex Analysis, 2020, 21(7): 1487-1496.
[21] Rodríguez-García M Á, García-Sánchez F, Valencia-García R. Knowledge-based system for crop pests and diseases recognition. Electronics, 2021, 10(8): 905.
[22] Pedersen T, Purandare A, Kulkarni A. Computational linguistics and intelligent text processing. CICLing 2005. Lecture Notes in Computer Science. Name Discrimination by Clustering Similar Contexts, ed. A. Gelbukh. 2005, Springer, Berlin, HeidelBERG.
[23] 李彦胜, 张永军. 耦合知识图谱和深度学习的新一代遥感影像解译范式. 武汉大学学报: 信息科学版), 2022, 47(8): 1176-1190.
[24] Google, Introducing the knowledge graph: Things, not strings. May 16, 2012. https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html.
[25] van Luijt B, Verhagen M. Bringing semantic knowledge graph technology to your data. IEEE Software, 2020, 37(2): 89-94.
[26] Zou Q F, Lu C Z. An accurate matching query method of natural language knowledge graph based on hierarchical graph topological sequence. IEEE Access, 2022, 10: 24080-24094.
[27] Balabin H, Hoyt C T, Birkenbihl C, Gyori B M, Bachman J, Kodamullil A T, Plöger P G, Hofmann-Apitius M, Domingo-Fernández D. STonKGs: a sophisticated transformer trained on biomedical text and knowledge graphs. Bioinformatics, 2022, 38(6): 1648-1656.
[28] Lo Giudice P, Musarella L, Sofo G, Ursino D. An approach to extracting complex knowledge patterns among concepts belonging to structured, semi-structured and unstructured sources in a data lake. Information Sciences, 2019, 478: 606-626.
[29] Lin J J, Zhao Y Z, Huang W Y, Liu C F, Pu H T. Domain knowledge graph-based research progress of knowledge representation. Neural Computing and Applications, 2021, 33(2): 681-690.
[30] Zhao Y C, Zhang B K, Gao D. Construction of petrochemical knowledge graph based on deep learning. Journal of Loss Prevention in the Process Industries, 2022, 76: 104736.
[31] Liang Z W, Yang J N, Liu H, Huang K J, Cui L, Qu L Z, Li X. HRER: a new bottom-up rule learning for knowledge graph completion. Electronics, 2022, 11(6): 908.
[32] Garfield E. Citation indexes for science. A new dimension in documentation through association of ideas. 1955. International Journal of Epidemiology, 2006, 35(5): 1123-1127; discussion 1127-1128.
[33] 陈悦, 刘则渊. 悄然兴起的科学知识图谱. 科学学研究, 2005, 23(2): 149-154.
[34] Su N, Wang Z B. Visual analysis of global carbon mitigation research based on scientific knowledge graphs. International Journal of Environmental Research and Public Health, 2022, 19(9): 5766.
[35] 蓝婷, 唐立娜, 徐智邦, 贾玉秋. 基于文献计量的城市空间紧凑度研究知识图谱分析. 生态学报, 2022, 42(4): 1645-1654.
[36] Zhao J, Harmelen F V, Tang J, Han X P, Wang Q, Li X Y. Knowledge graph and semantic computing. knowledge computing and language understanding. Tianjin: Springer, 2018.
[37] Kejriwal M. What is a knowledge graph?. Domain-Specific Knowledge Graph Construction. Cham: Springer International Publishing, 2019: 1-7.
[38] Chein M, Mugnier M. Graph-based knowledge representation - computational foundations of conceptual graphs. London: Springer, 2008
[39] 马昂, 于艳华, 杨胜利, 石川, 李劼, 蔡修秀. 基于强化学习的知识图谱综述. 计算机研究与发展, 2022, 59(8): 1694-1722.

[40] 楼一恺, 范忆, 戴其林, 王铮屹, 库伟鹏, 赵明水, 余树全. 天目山常绿落叶阔叶林群落垂直结构与群落整体物种多样性的关系. 生态学报, 2021, 41(21): 8568-8577.
[41] 蒋文伟, 郭运雪, 杨淑贞, 赵明水. 天目山柳杉树干液流动态及其与环境因子的关系. 江西农业大学学报, 2011, 33(5): 899-905.
[42] Krantz G W, Walter D E. A Manual of Acarology. Third edition. Texas: Texas Tech University Press, 2009.
[43] Yin Y T, Zhang L, Wang Y G, Wang M Q, Zhang Q M, Li G Z. Question answering system based on knowledge graph in traditional Chinese medicine diagnosis and treatment of viral hepatitis B. BioMed Research International, 2022, 2022: 7139904.
[44] Fecho K, Bizon C, Miller F, Schurman S, Schmitt C, Xue W, Morton K, Wang P, Tropsha A. A biomedical knowledge graph system to propose mechanistic hypotheses for real-world environmental health observations: cohort study and informatics application. JMIR Medical Informatics, 2021, 9(7): e26714.
[45] Qiao B, Fang K, Chen Y M, Zhu X H. Building thesaurus-based knowledge graph based on schema layer. Cluster Computing, 2017, 20(1): 81-91.
[46] Liu X X, Bai X S, Wang L H, Ren B Y, Lu S H, Li L. Review and trend analysis of knowledge graphs for crop pest and diseases. IEEE Access, 2019, 7: 62251-62264.
[47] Zhao M X, Li D, Long Y S. Forestry big data platform by Knowledge Graph. Journal of Forestry Research, 2021, 32(3): 1305-1314.
[48] Lan N, Yang S Q, Yin L, Gao Y B. Research on knowledge graphs with concept lattice constraints. Symmetry, 2021, 13(12): 2363.
[49] Thakur M P, Phillips H R P, Brose U, De Vries F T, Lavelle P, Loreau M, Mathieu J, Mulder C, Van der Putten W H, Rillig M C, Wardle D A, Bach E M, Bartz M L C, Bennett J M, Briones M J I, Brown G, Decaëns T, Eisenhauer N, Ferlian O, Guerra C A, König-Ries B, Orgiazzi A, Ramirez K S, Russell D J, Rutgers M, Wall D H, Cameron E K. Towards an integrative understanding of soil biodiversity. Biological Reviews, 2020, 95(2): 350-364.
[50] Hacala A, Lafage D, Prinzing A, Sawtschuk J, Pétillon J. Drivers of taxonomic, functional and phylogenetic diversities in dominant ground-dwelling arthropods of coastal heathlands. Oecologia, 2021, 197(2): 511-522.
[51] Tsafack N, Borges P A V, Xie Y Z, Wang X P, Fattorini S. Emergent rarity properties in carabid communities from Chinese steppes with different climatic conditions. Frontiers in Ecology and Evolution, 2021, 9: 603436.
[52] Chen S Y, Hsu C H, Soong K. How to cross the sea: testing the dispersal mechanisms of the cosmopolitan earthworm Pontodrilus litoralis. Royal Society Open Science, 2021, 8(8): 202297.
[53] Xie Z J, Chen T W, Potapov M, Zhang F, Wu D H, Scheu S, Sun X. Ecological and evolutionary processes shape below-ground springtail communities along an elevational gradient. Journal of Biogeography, 2022, 49(3): 469-482.
[54] Pocock M J O, Schmucki R, Bohan D A. Inferring species interactions from ecological survey data: a mechanistic approach to predict quantitative food webs of seed feeding by carabid beetles. Ecology and Evolution, 2021, 11(18): 12858-12871.
[55] Bellino A, Baldantoni D, Milano V, Santorufo L, Cortet J, Maisto G. Spatial patterns and scales of Collembola taxonomic and functional diversity in urban parks. Sustainability, 2021, 13(23): 13029.
[56] Peterson D L, Parker V T. Ecological Scale: Theory and Applications. New York: Columbia University Press, 1998.
[57] Xiong D, Wei C Z, Wang X G, Lü X T, Fang S, Li Y B, Wang X B, Liang W J, Han X G, Bezemer T M, Li Q. Spatial patterns and ecological drivers of soil nematode β-diversity in natural grasslands vary among vegetation types and trophic position. Journal of Animal Ecology, 2021, 90(5): 1367-1378.
[58] Dong C X, Gao M X, Guo C W, Lin L, Wu D H, Zhang L M. The underlying processes of a soil mite metacommunity on a small scale. PLoS One, 2017, 12(5): e0176828.
[59] Gao M X, Guo Y X, Liu J, Liu J W, Adl S, Wu D H, Lu T Y. Contrasting beta diversity of spiders, carabids, and ants at local and regional scales in a black soil region, northeast China. Soil Ecology Letters, 2021, 3(2): 103-114.
[60] Bardgett R D, van der Putten W H. Belowground biodiversity and ecosystem functioning. Nature, 2014, 515(7528): 505-511.
[61] Brussaard L. Biodiversity and ecosystem functioning in soil: the dark side of nature and the bright side of life. Ambio, 2021, 50(7): 1286-1288.
[62] Chen X D, Dunfield K E, Fraser T D, Wakelin S A, Richardson A E, Condron L M. Soil biodiversity and biogeochemical function in managed ecosystems. Soil Research, 2020, 58(1): 1-20.
[63] Geisen S, Wall D H, van der Putten W H. Challenges and opportunities for soil biodiversity in the anthropocene. Current Biology, 2019, 29(19): R1036-R1044.
[64] Guerra C A, Heintz-Buschart A, Sikorski J, Chatzinotas A, Guerrero-Ramírez N, Cesarz S, Beaumelle L, Rillig M C, Maestre F T, Delgado-Baquerizo M, Buscot F, Overmann J, Patoine G, Phillips H R P, Winter M, Wubet T, Küsel K, Bardgett R D, Cameron E K, Cowan D, Grebenc T, Marín C, Orgiazzi A, Singh B K, Wall D H, Eisenhauer N. Blind spots in global soil biodiversity and ecosystem function research. Nature Communications, 2020, 11: 3870.
[65] 孙新, 李琪, 姚海凤, 刘满强, 吴东辉, 朱冬, 朱永官. 土壤动物与土壤健康. 土壤学报, 2021, 58(5): 1073-1083.