《农业知识图谱构建技术规程》

地方标准（征求意见稿）编制说明

一、工作概况

**（一）任务来源**

本标准根据陕西省市场监督管理局《关于下达2022年地方标准计划的通知》(陕市监函〔2022〕380号)立项，项目编号SDBXM120-2022，项目名称《农业知识图谱构建技术规程》。

**（二）本标准制订的必要性和意义**

根据国家战略发展需求，21世纪以来共发布了20个指导“三农”工作的中央一号文件。2023年2月13日《中共中央 国务院关于做好2023年全面推进乡村振兴重点工作的意见》进一步指出，深入实施数字乡村发展行动，推动数字化应用场景研发推广。加快农业农村大数据应用，推进智慧农业发展。农业农村部《数字农业农村发展规划（2019-2025年）》提出构建重要农业种质资源大数据，为品种选育、产业发展、行业监管提供大数据支持。

陕西是我国重要的农业产区，地形复杂、海拔高度、雨热资源和生态条件差异大，形成了超过20个各具特色的种植业气候类型区。随着农业产业的不断发展，借助信息技术快速、有效地获取农业产业全链条生产、销售等信息变得越发重要，基于人工智能技术的农业自动问答、推荐、溯源系统快速发展，为农户、合作社、农技人员等农业从业者在一定程度上提供了相关的信息服务。但当前不同种植业气候类型区作业方式、作业时间、农资农具等需求各不相同，使得农业生产数据呈现种类繁多、多源异构等特点，且在计算机中存储与表达标准不统一，导致农业信息搜索、推荐、问答的有效性、准确性无法进一步提高。

知识图谱是一种新的多关系图，通过知识图谱可以快速推测隐含的业务逻辑，构建小而轻的存储载体。农业知识图谱可以更直观地挖掘、分析、构建、绘制和显示农业知识。基于知识图谱的农业知识问答、搜索、推荐等系统能够高效查询、匹配信息，为农业从业人员提供快速敏捷的信息服务。然而，由于农业领域知识名称与术语没有统一规范、各地作物管理过程和知识获取途径差异较大，知识融合与知识推理方法也不尽相同，造成农业领域知识图谱的构建缺乏相应的标准。因此，为了推动陕西智慧农业、数字乡村乃至乡村振兴的快速发展，有必要制定符合陕西农业特色的知识图谱构建标准。

本标准充分考虑了陕西省农业地域特性，规定了农业知识图谱相关术语和基本概念；适用于指导农业领域知识图谱的构建，规范农业知识抽取、推理、问答和推荐过程，方便相关从业人员敏捷高效地获取信息，为智慧农业应用奠定数据基础和技术支撑，从而促进陕西数字化农业农村的发展。

**（三）起草单位、主要起草人及任务分工**

**1、起草单位**

本文件起草单位：西北农林科技大学、中国电信股份有限公司陕西分公司、西安交通大学、西北工业大学、西安电子科技大学、西安农链互联网科技有限公司、杨凌现代农业产业标准化研究推广服务中心。

**2、主要起草人**

本文件主要起草人：张宏鸣、耿耀君、朱珊娜、吕志明、杜宝红、李亚佳、孙鹤立、李 青、冯志玺、陈 欢、李书琴、宋荣杰、牛当当、刘 斌、蒲 攀、黄小星、张 婷、周永明。

**3、任务分工**

任务分工如表1所示。

表1 任务分工表

| **序号** | **姓名** | **单位** | **任务分工** |
| --- | --- | --- | --- |
| 1 | 张宏鸣 | 西北农林科技大学 | 规程总体设计与最终审定，工作方案制定。 |
| 2 | 耿耀君 | 西北农林科技大学 | 知识图谱架构设计。 |
| 3 | 朱珊娜 | 西北农林科技大学 | 规程术语和定义整理。 |
| 4 | 吕志明 | 西北农林科技大学 | 知识图谱构建流程设计。 |
| 5 | 杜宝红 | 中国电信股份有限公司陕西分公司 | 陕西省农业相关领域的发展状况分析与总结。 |
| 6 | 李亚佳 | 中国电信股份有限公司陕西分公司 | 知识图谱构建环节面临的主要问题和未来趋势分析。 |
| 7 | 孙鹤立 | 西安交通大学 | 规程技术要求制定。 |
| 8 | 李 青 | 西北工业大学 | 规程技术要求制定。 |
| 9 | 冯志玺 | 西安电子科技大学 | 规程相关国际标准、国家标准查阅、收集、整理与分析。 |
| 10 | 陈 欢 | 西北农林科技大学 | 规程相关行业标准、地方标准查阅、收集、整理与分析。 |
| 11 | 李书琴 | 西北农林科技大学 | 国内外有关智慧农业、农业知识图谱构建等其它文献资料整理与分析。 |
| 12 | 宋荣杰 | 西北农林科技大学 | 规程的验证试验。 |
| 13 | 牛当当 | 西北农林科技大学 | 规程的验证试验。 |
| 14 | 刘 斌 | 西北农林科技大学 | 规程的验证试验。 |
| 15 | 蒲 攀 | 西北农林科技大学 | 规程实地调研环节资料整理与分析。 |
| 16 | 黄小星 | 西安农链互联网科技  有限公司 | 规程技术相关实践经验、存在问题及解决方案的提供。 |
| 17 | 张 婷 | 西安农链互联网科技  有限公司 | 规程实地调研环节资料整理与分析。 |
| 18 | 周永明 | 杨凌现代农业产业标准化研究推广服务中心 | 规程相关法律法规查阅、收集，标准内容形式把关。 |

**（四）起草过程**

**1、成立标准起草工作组**

《农业知识图谱构建技术规程》地方标准任务下达后，由西北农林科技大学牵头，联合中国电信股份有限公司陕西分公司、西安交通大学、西北工业大学、西安电子科技大学、西安农链互联网科技有限公司、杨凌现代农业产业标准化研究推广服务中心等单位，成立了《农业知识图谱构建技术规程》陕西省地方标准起草工作组，讨论确定标准编制思路和工作方案。

**2、调查研究和标准起草**

标准起草工作组通过查阅文献资料和实地走访等形式，开展标准需求调研，深入了解陕西省农业相关领域的发展状况，分析研究知识图谱构建环节面临的问题和未来趋势，并根据GB/T1.1-2020《标准化工作导则 第1部分：标准化文件的结构及起草规则》的要求，开始规程的编制起草，经过内部多次讨论，形成标准草案。

**3、标准草案提升完善**

标准起草工作组对西安农链互联网科技有限公司、西北农林科技大学猕猴桃示范站、西北农林科技大学合阳葡萄试验示范站、陕西省汉中市镇巴县中药材园区进行实地调研，就标准编制思路和主要内容与农业相关从业者、管理、信息化等专业技术人员进行讨论，记录一线技术人员反馈的意见；同时积极征求听取农业信息化领域专家建议。根据多方意见对标准草案进行修改，重点围绕农业知识图谱构建技术要求，对技术规程进行细化调整，不断完善标准草案。

**4、形成标准征求意见稿**

为了做好标准的起草工作，根据编制团队多年来在农业知识图谱构建方面的实践经验，同时广泛收集和参阅国内外有关智慧农业、农业知识图谱构建等方面的文献资料，确定了本规程的编写内容。邀请监管部门、企业、高校、科研院所等相关单位以及示范站、园区经营用户、标准化领域专家等召开标准研讨会，对标准框架和内容逐条逐项进行认真讨论。工作组根据各方意见和建议，对标准草案进行修订，于2023年4月形成标准征求意见稿。

二、标准编制原则和标准主要内容

**（一）标准编制的原则**

**1、科学性原则**

本标准的制定严格按《中华人民共和国标准化法》、《地方标准制定工作规范》（DB14/T1665-2018）、《标准化工作导则第1部分：标准化文件的结构和起草规则》（GB/T 1.1-2020）的要求进行。

**2、适用性原则**

“规范”编写充分考虑了陕西省地方特色和农业知识图谱构建的当前现状，既要保持技术上规范性、合理性，也必须兼顾适应新形势下从事农业生产、研究人员的实际需求，体现对未来需求的适应性。

**3、规范性原则**

本次编写采用章、节、条、款四级的条文编排方式，各自用于不同表达：

章——划分规程章节，例如以1-6表示。

节——划分技术模块，例如以1.1-1.3表示。

条——罗列具体技术步骤，例如以1.1.1-1.1.3表示

款——罗列具体技术要求，例如以a)-e)表示。

其中，款是具体要求，所以规定不能模糊，必须明确具体要求和效果，本次编写通过增加款的方式来细化调查、分析及实施的技术要求。

**（二）标准主要内容**

**1、缩略语**

**关键技术要求：**对适用于本文件的缩略语，RDF、NER、RE、LSTM等进行说明。

**确定依据：**根据农业知识图谱构建技术所包含的关键步骤和关键模型提出。

**2、架构与流程**

**（1）关键技术要求**

农业知识图谱架构包括：基础数据采集层、图谱构建层、信息解析层和服务层。基础数据采集层：农业知识图谱数据来源于用于构建初步农业知识图谱的术语集和规范资料，以及用于扩展农业知识图谱的各种数据源，包括各种网络资源、农业知识库以及智能感知设备采集的生产数据。采集的数据需要进行去噪处理。图谱构建层：从数据中抽取命名实体和关系，并进行知识融合和加工；信息解析层：借助模板匹配和机器学习方法，以及分词、词性分析、句法分析、图搜索、排序和分类组织等自然语言处理技术，通过信息解析和意图理解，准确理解文字含义及指令意图；服务层：以农业知识图谱为基础为农业相干从业者、企业、研究院所提供全产业链、全方位的产业发展现状和趋势，并给出产业发展相关的决策与指导。

农业知识图谱构建流程包括：数据准备、知识表示与知识建模、知识获取与知识融合、知识加工、知识应用服务。数据准备：通过农业知识库、术语集、规范资料和设备感知等方式获取结构化、半结构化和非结构化数据，构建农业领域内的领域知识和重要术语。知识表示与知识建模：定义知识表示的方式，确定知识图谱构建的框架和工具，设计知识图谱的数据结构，构建本体模型描述农业领域的知识体系。知识获取与知识融合：利用LSTM、Transformer和Bert等深度学习模型，结合人工众包完成实体抽取、关系抽取、属性抽取。通过知识体系映射，去除冗余知识数据，形成统一的知识结构存储。知识加工：基于已有知识图谱构建算法模型，预测未知的隐含知识，验证知识的正确性后，通过知识补全完善现有知识图谱。随着新知识的获取，完成新的实体、关系、属性和规则等知识单元的更新。知识应用服务：根据农业领域的应用场景部署知识图谱，通过知识图谱数据库的查询，提供知识检索、知识推理、知识可视化等接口和服务。

**（2）确定依据**

根据《知识图谱白皮书》对知识图谱构建和服务的相关描述及农业邻域实际需求，确定了农业知识图谱的架构和构建流程。架构同时也符合大数据处理及服务平台的常规要求。构建流程中核心步骤依据知识图谱核心要点而确定。

**3、技术要求**

**（1）关键技术要求**

知识抽取：a)支持从结构化数据中直接提取资源描述框架（RDF）三元组。b)支持基于农业领域半结构化、结构化数据的概念抽取、实体抽取、关系抽取和属性抽取。c)抽取模型应基于隐马尔科夫模型(HMM)、条件随机场(CRF)等统计学习或长短期记忆网络（LSTM）、预训练语言模型等深度学习（DL）的知识抽取方法。d)知识加工应支持基于术语匹配、结构特征、知识分块的本体层融合和匹配和基于表示学习的实 体对齐。e)知识抽取模型的性能应使用准确率、召回率和F1分数（F1-score）等公认的评价指标评价。

知识推理：a)支持本体公理的知识图谱推理。b)支持图结构和规则学习的知识图谱推理。c)支持表示学习的知识图谱推理。d)支持图神经网络的知识图谱推理。e)支持分布式表示学习的知识图谱推理。

知识问答：a)支持问句模板的知识图谱问答。b)支持基于语义解析的知识图谱问答。c)支持基于检索排序的知识图谱问答。d)支持基于深度学习的知识问答。e)支持通过PC端网站、APP和小程序等方式进行知识问答。

知识推荐：a)支持基于特征的推荐方法和基于路径的推荐方法。b)支持依次学习、联合学习和交替学习的方法，将知识图谱特征应用到推荐系统中。c)支持通过PC端网站、APP、小程序和短信等方式推荐。

**（2）确定依据**

知识抽取技术要求依据在可能是结构、非结构和半结构的特点知识图谱语料中提取信息的方式确定；知识推理技术、知识问答技术和知识推荐技术依据各自邻域典型技术分类确定，保证知识背景有差异的研发人员均可以依据本规程构建农业知识图谱。

三、实证研究

依托国家重点研发计划“绿色宜居村镇技术创新”重点专项“村镇数字化科技信息服务综合示范（2020YFD1100600）”课题一“农业先进适用技术社会化服务平台研发与示范”，在示范基地收集相关数据，构建葡萄、猕猴桃、苹果、红枣、中草药、谷子等6种农作物知识图谱。部分可视化展示如图1-图2所示。

（1）构建的葡萄知识图谱，包括中国葡萄品种、葡萄土肥水管理、葡萄树体管理、葡萄病害、葡萄虫害、农药使用方法等，总计4774条三元组数据。

（2）构建的猕猴桃知识图谱，包括猕猴桃形态特征、生长习性、种植管理流程、病虫害防止和加工运输等知识，总计3044条三元组数据。

（3）构建的苹果知识图谱，包括苹果种植资源、栽培技术、病虫害防治、采后加工等知识，总计2531条三元组。

（4）构建的红枣知识图谱，包括红枣品种、枣树生长环境、枣树育苗、枣树种植、枣树管理、枣树病虫害、红枣采收加工、枣树适用技术等，总计4184条三元组数据。

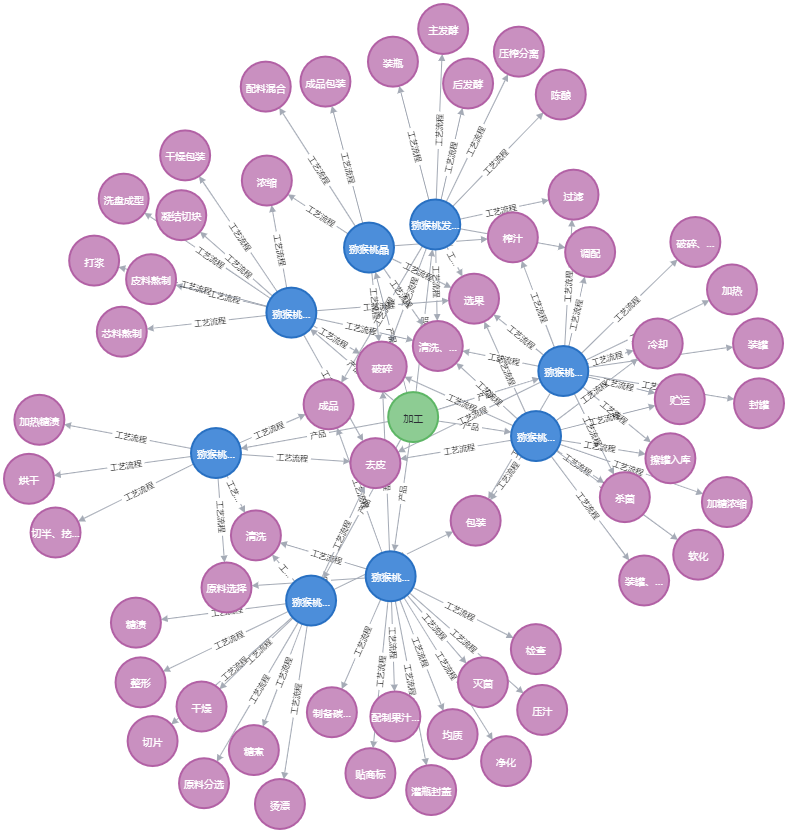


图1 猕猴桃知识图谱部分可视化展示

（5）构建的中草药知识图谱，包括药材名，功效，部位，数值，工具，科，病症，形态，种植方法，病害，属，化学成分，成长期，田间管理，病原，产地，气候，时间段，选种标准，症状，气象指标，地理环境，禁忌，时机，发病规律，药性，检测指标，实施者，种植密度，虫害，药味，标准，触发词，肥料，害虫成长阶段，经络，植物类别，操作对象，施肥量，防治方法，药剂，包装记录，储藏条件等类型的实体及关系，总计3765条三元组数据。



图2 中草药知识图谱部分可视化展示

（6）构建的谷子知识图谱，相关知识总计3424条三元组数据。

构建的农业知识图谱集成于农业先进适用技术社会化服务平台，在陕西省合阳县、镇巴县、佳县进行了成果应用，通过农业知识图谱优化农业社会化资源配置，运用知识搜索、推荐、问答等模块实现农业科技成果供需精准匹配和线上敏捷服务。相关成果服务于生产、经营和管理等主体，覆盖5万亩耕地，惠及农户6000多户。

四、知识产权说明

标准未涉及相关知识产权。

五、重大意见分歧的处理

无重大意见分歧。

六、采标情况

未采用国际标准或国外先进标准。

七、其他应说明的事项

无其他说明事项。

《农业知识图谱构建技术规程》起草组

2023年4月10日