ICS 点击此处添加 ICS 号

点击此处添加中国标准文献分类号

**DB61**

陕 西 省 地 方 标 准

DB 61/ T XXXXX—XXXX

农业知识图谱构建技术规程

Technical specification for the construction of agricultural knowledge graph

点击此处添加与国际标准一致性程度的标识

（征求意见稿）

XXXX - XX - XX 发布

XXXX - XX - XX 实施

陕西省市场监督管理局 发 布

前 言

本文件依据GB/T 1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由陕西省农业农村厅提出并归口。

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。

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。

本文件首次发布。

农业知识图谱构建技术规程

1 范围

本文件规定了农业知识图谱的相关术语和定义、缩略语、架构与流程及技术要求。

本文件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。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 23703.2 知识管理 第2部分：术语

GB/T 35295 信息技术 大数据 术语

3 术语和定义

GB/T 23703.2、GB/T 35295界定的以及下列术语和定义适用于本文件。

3.1

现代农业 modern agriculture

以保障农产品供给、增加农民收入、促进可持续发展为目标，以提高劳动生产率、资源产出率和商品率为途径，以现代科技和装备为支撑，以家庭经营为基础，在市场机制与政府调控的综合作用下，农工贸紧密衔接，产加销融为一体，多元化的产业形态和多功能的产业体系。

3.2

知识图谱 knowledge graph

将互联网的信息表达成更接近人类认知世界的形式，以结构化的形式描述客观世界中概念、实体及其关系，是一种更好地组织、管理和理解互联网海量信息的方式。

3.3

农业知识图谱 agricultural knowledge graph

融合现代农业的地域性、季节性和多样性等特征后，利用农业领域的实体与关系，挖掘出农业潜在价值的智能系统。

3.4

实体 entity

指具有独特的标识并且可以与其他对象区分的对象或事物。其客观存在且相互区别，是知识图谱的最基本元素。

3.5

关系 relation

知识图谱中不同实体间的联系。

3.6

属性 attribute

与真实或抽象的事物相关联的一种受关注的特征，表示类的实例通常共有的某些特征。

3.7

实体抽取 entity extraction

也称为命名实体学习或命名实体识别。指从自然语言文本中自动识别出专有名称和有意义的数量短语并加以归类的过程。

3.8

关系抽取 relation extraction

从一段文本中抽取出（主体，关系，客体）这样的三元组。

3.9

属性抽取 attribute extraction

**从非结构化文本中抽取出实体的属性以及属性值**。

3.10

实体链接 entity linking

将文本中的实体与知识图谱中相应的实体链接起来，以解决实体间存在的歧义性问题。

3.11

命名实体消歧 entity disambiguation

一个实体可对应多个真实世界实体，根据上下文语境确定一个实体对应唯一真实世界实体的过程。

3.12

知识表示 knowledge representation

知识符号化的过程，即把知识用计算机可接受的符号以某种形式描述出来。

[来源：GB/T 23703.2-2010，4.4，有修改]

3.13

结构化数据 structured data

一种数据表示形式。按此种形式，由数据元素汇集而成的每个记录的结构都是一致的，并且可以使用关系模型予以有效描述。

[来源：GB/T 35295-2017，2.2.13]

3.14

非结构化数据 unstructured data

不具有预定义模型或未以预定义方式组织的数据。

[来源：GB/T 35295-2017，2.1.25]

3.15

本体 ontology

在大数据语境下，它是一些约束后续各种不同层次逻辑模型的语义模型。

注：本体，从本质上看，既可以是非常概括性的，也可以是极其专门化的。

[来源：GB/T 35295-2017，2.1.54]

3.16

知识推理 knowledge reasoning

通过从已知的知识出发，通过已经获取的知识，从中获取到所蕴含的新的事实，或者从大量的已有的知识中进行归纳，从个体知识推广到一般性的知识。

3.17

知识推荐 knowledge recommendation

根据用户的历史行为、兴趣点、等信息去判断用户当前需要或感兴趣知识，并提供给用户。

4 缩略语

下列缩略语适用于本文件。

RDF：资源描述框架(Resource Description Framework)

NER：命名实体识别(Named Entities Recognition)

RE：关系抽取(Relation Extraction)

HMM：隐马尔科夫模型(Hidden Markov Model)

CRF：条件随机场(Conditional Random Fields)

LSTM：长短期记忆网络(Long Short-Term Memory)

DL：深度学习(Deep Learning)

PC：个人计算机(Personal Computer)

APP：应用程序(Application)

5 架构与流程

5.1 概述

农业知识图谱的构建架构包括基础数据层、图谱构建层、信息解析查询层和服务层。如图1所示。

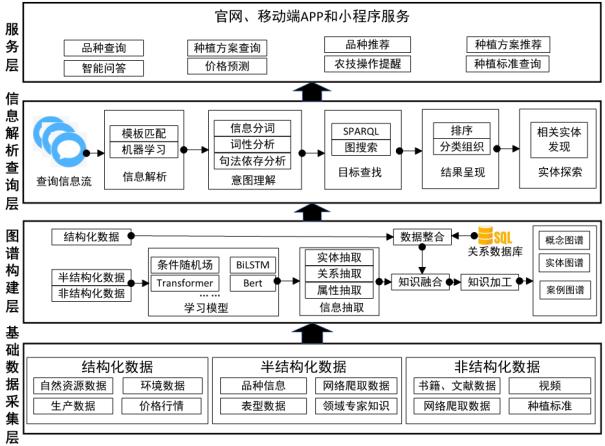


图1 农业知识图谱构建架构

5.2 知识图谱构建架构

5.2.1 基础数据采集层

是农业知识图谱数据的产生过程。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。

5.2.2 图谱构建层

5.2.2.1 信息抽取

基于BiLSTM、Transformer等模型从不同类型数据源中提取出实体、属性以及实体间的相互关系，形成本体化的知识表示。包括实体抽取、关系抽取(RE)/命名实体识别(NER)和属性抽取。

5.2.2.2 知识融合

通过实体链接、知识合并、命名实体消歧等方式整合数据，消除矛盾和歧义。

5.2.2.3 知识加工

经过本体构建、知识推理和质量评估后，将合格的知识加入到知识库中。

5.2.3 信息解析查询层

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，并根据相关意图做出对应的反馈，快速搭建问答场景应用或完成指定语音任务。

5.2.4 服务层

5.2.4.1 农业信息服务

为农业相关从业者提供信息检索及问答相关的服务。

5.2.4.2 产业指导

通过可视化分析为农业相关从业者、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。

5.2.4.3 农产品推荐

通过农产品知识图谱构建与机器视觉、自然语言处理等技术结合，根据客户需求推荐相应农产品。

5.3 知识图谱构建流程

5.3.1 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。

5.3.2 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。

5.3.3 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。

5.3.4 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。

5.3.5 知识应用服务

根据农业领域的应用场景部署知识图谱服务网站、APP或小程序等，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。

6 技术要求

6.1 知识抽取

从不同来源、不同结构的农业数据中抽取实体、关系和属性，并加工形成农业知识。知识抽取过程应满足如下要求：

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、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。

6.2 知识推理

依据已有农业知识，获取新的农业知识或结论的过程。知识推理模型应满足如下要求：

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。

6.3 知识问答

用准确、简洁的自然语言回答用户提出的农业问题。知识问答模型和接口应满足如下要求：

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。

6.4 知识推荐

根据用户的历史行为、兴趣点，判断并给用户提供当前需要或感兴趣的农业知识，其中知识推荐模型和接口应满足如下要求：

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。