计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2021,30(6):286−292 [doi: 10.15888/j.cnki.csa.007986]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

小麦品种知识图谱构建与可视化研究①

许　鑫[1,2,3], 岳金钊[1], 赵锦鹏[1], 王亚坤[1], 马新明[1,2,3], 钱学霖[1]

[1](河南农业大学 信息与管理科学学院, 郑州 450002)
[2](河南粮食作物协同创新中心, 郑州 450002)
[3](河南农业大学 农学院, 郑州 450002)
通讯作者: 马新明, E-mail: wheatdoctor@163.com

摘　要: 为探索知识图谱技术在农业智能生产中应用与落地, 解决复杂多样的农业生产数据的精准查询与可视化问题, 本研究以小麦品种知识为例, 利用爬虫技术, 爬取 1852 个小麦品种信息、735 个微百科、102 349 个词条; 基于知识图谱技术, 设计品种知识图谱实体与关系, 对抓取数据进行清洗、抽取与融合, 经过实体识别、关系构造等处理, 构建实体 258 484 个, 关系 328 933 个. 在此基础上, 设计了小麦品种知识存储方式, 结构化数据存储在 MySQL 中, 非结构化数据存储在 MongoDB 中, 使用 Neo4j 图数据库存储知识图谱来提高知识的查询性能, 在此基础上实现小麦品种关系查询与实体识别, 提供品种数据精确表达与可视化, 表明利用知识图谱技术实现品种等信息的可视化是可行的, 该研究可以为知识图谱在农业中的应用提供技术参考和理论支撑。

关键词: 小麦; 品种; 知识图谱; NLP; Neo4j

引用格式: 许鑫,岳金钊,赵锦鹏,王亚坤,马新明,钱学霖.小麦品种知识图谱构建与可视化研究.计算机系统应用,2021,30(6):286–292. http://www.c-s-a.org.cn/1003-3254/7986.html

Construction and Visualization of Knowledge Map of Wheat Varieties

XU Xin[1,2,3], YUE Jin-Zhao[1], ZHAO Jin-Peng[1], WANG Ya-Kun[1], MA Xin-Ming[1,2,3], QIAN Xue-Lin[1]

[1](College of Information and Management Science, Henan Agricultural University, Zhengzhou 450002, China)
[2](Henan Grain Crops Collaborative Innovation Center, Zhengzhou 450002, China)
[3](College of Agronomy, Henan Agricultural University, Zhengzhou 450002, China)

Abstract: In order to explore the application and implementation of knowledge mapping technology in intelligent agricultural production and realize the accurate query and visualization of complex and diverse agricultural production data, this study took wheat varieties as an example and collected the information of 1852 wheat varieties, 735 micro encyclopedias, and 102 349 entries by a crawler. Through knowledge mapping technology, this study designed the entities of variety knowledge graphs and their relationships, with data cleaned, extracted, and fused. A total of 258 484 entities were recognized and 328 933 relationships built. On this basis, the approach to storing wheat variety knowledge was worked out, with structured data stored in a MySQL, unstructured data in the MongoDB. Neo4j was employed to optimize knowledge query. In this way, the query about relationships between wheat varieties and entity recognition was made possible with variety data expressed accurately and visualized, proving the feasibility of knowledge mapping in visualization of information such as variety. This research can provide technical reference and theoretical support for the application of knowledge mapping in agriculture.

Key words: wheat; variety; knowledge graph; NLP; Neo4j

① 基金项目: 十三五国家重点研发计划 (2016YFD0300609); 河南省科技创新杰出人才 (184200510008); 河南省现代农业产业技术体系 (S2010-01-G04)
Foundation item: National Key Research and Development Program of China during Thirteenth Five-Year Plan (2016YFD0300609); Outstanding Talents of Scientific and Technological Innovation, Henan Province (184200510008); Modern Agricultural Industrial Technology System (S2010-01-G04)
收稿时间: 2020-10-16; 修改时间: 2020-11-18; 采用时间: 2020-12-12; csa 在线出版时间: 2021-06-01

信息化已成为农业现代化的重要组成部分[1], 生产数据结构复杂且类型多样, 数据可视化技术可以实现复杂的数据直观化、量化和简化, 能大力的推动农业信息化的发展[2].

知识图谱作为大数据可视化和人工智能的重要组成部分被广泛应用[3]. Google 将知识图谱应用在搜索引擎上[4], 百度和搜狗相继推出了“知心”和“知立方”[5], 苏宁易购发布金融企业知识图谱系统. 蒋秉川等[6] 利用地理知识图谱结合交互式可视化分析 COVID-19 疫情态势; 车金立等[7] 构建了军事装备知识图谱, 实现了军事装备领域的知识问答; 李晓雪等[8] 利用领域知识图谱技术进行了农作物病虫害分析和分类; 张善文等[9] 提出了一种基于知识图谱与 Bi-LSTM 结合的小麦条锈病预测方法; 华东师范大学[10] 利用深度学习和自然语言处理构建了农业知识图谱; 叶帅[11] 将知识图谱引入到煤矿领域. 知识图谱在各个领域都有应用, 但在农业领域的应用和技术体系尚待研究[12].

目前的农业数据分散化、种类多、连贯性差, 挖掘有价值的信息是未来研究的重点[13]. 知识图谱技术可以将离散的、不集中的信息与可视语义网络关联[14], 便于通过图的形式直观地掌握和分析关系错综复杂的领域知识, 实现精确查询[12].

本研究以小麦生产知识为研究对象, 获取网络中现存的凌乱复杂的知识, 探索农业领域知识图谱的构建方法, 设计小麦品种图谱实体和关系, 通过知识图谱直观、清晰地展示错综复杂的品种知识, 以期为小麦生产知识的精准推荐, 农业知识图谱的构建提供技术方案依据.

1 小麦品种知识图谱框架设计

知识图谱可分为通用知识图谱和行业知识图谱[15]. 通用知识图谱都是常识性的知识, 面向全领域, 覆盖面较广, 但深度不足, 主要应用于互联网的搜索、推荐等业务场景, 如: FreeBase[16]、DBpedia[17]. 行业知识图谱覆盖特定领域的知识, 知识的深度相比通用知识图谱较深, 行业知识图谱需要收集特定领域的数据, 结合业务流程在领域专家的指导下来构建知识图谱模式之后构建数据层[18]. 本研究结合互动百科通用知识图谱和小麦生产行业知识图谱, 通过获取小麦品种等生产数据, 经过清洗、整理、知识抽取等步骤, 构建小麦生产领域知识图谱, 如图 1 所示.

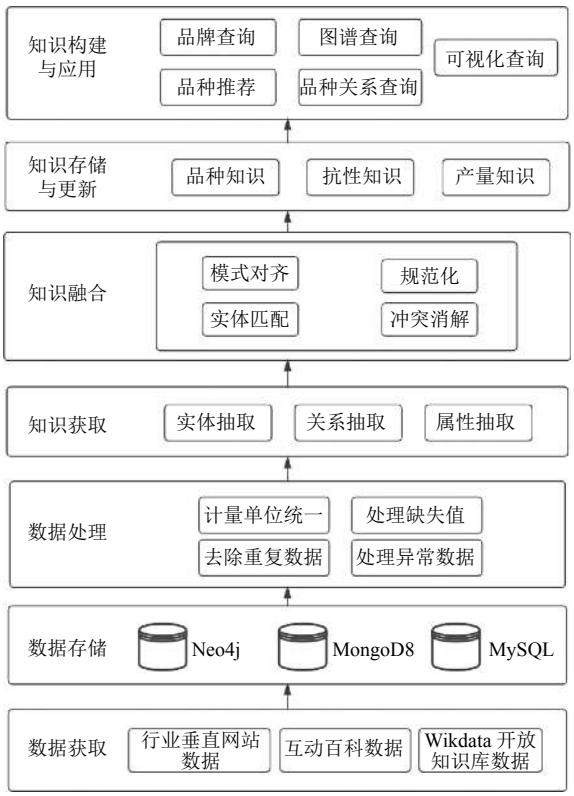

图 1 图谱构建流程图

(1) 数据获取、存储与处理: 数据获取之后需要对数据进行清洗、预处理, 提高数据的利用率, 增强知识图谱的准确性. 本研究选取行业垂直网站、在线百科、开放知识库等多个源头获取数据, 提升知识图谱的丰富性和有效性. 对于不同源头的不同类型数据, 进行分别存储. 结构化数据存储在 MySQL 数据库中, 非结构化数据存储在 MongoDB 数据库中. 获取到的数据往往会存在残缺、错误、重复等问题, 需要对数据进行计量单位统一、处理缺失值等处理.

(2) 知识获取: 针对不同类型数据采用不同的知识获取方式, 对于结构化数据, 各项之间存在明确的对应关系, 可以直接构建三元组; 而半结构化数据, 存在一定的结构, 需要进一步提取, 将半结构化数据转化为结构化数据. 非结构化数据, 利用自然语言处理 (Natural Language Processing, NLP) 技术对文本进行分段、分句、分词、去除停用词等处理, 进而进行命名实体识别和关系抽取.

(3) 知识融合: 不同来源数据会导致整体数据格式复杂, 出现实体属性名称不一致, 数据类型冲突等情况. 所以需要把将要抽取的知识和知识图谱现有的知识做

融合处理, 以消除矛盾和歧义. 选取实体的属性作为特征, 构建特征向量, 利用相似度计算, 将新的实体与知识图谱中现有的实体进行链接[19].

(4) 知识存储与更新: 在传统的关系型数据库存储中, 存储大量关系复杂的数据之后, 难以直观的描述实体与实体之间的关系, 每次查询都需要联结大量表, 造成查询效率低. 而基于属性图形模型的 Neo4j 数据库不仅能够直观的反应实体之间的关系, 还能够大大地提高查询效率[18]. 利用 Cypher 图数据库查询语言来解决知识更新问题, 易于理解, 方便用户对不合理的图数据进行更新操作.

(5) 小麦领域知识图谱的构建与应用. 将收集和整理好的数据, 结合小麦领域知识的特点, 构建知识图谱. 利用 Neo4j 来负责小麦知识图谱的存储, 将构造好的三元组——“实体-关系-实体”, 利用 Cypher 语言存储到数据库中. 从用户自然语句中提取实体和属性, 将实体和属性注入到 Cypher 查询模板中, 实现在小麦知识图谱中进行查询, 在此基础上, 研究开发小麦知识图谱查询系统, 实现了品种推荐、实体查询、关系查询、可视化查询等功能.

2 关键技术设计

2.1 多源异构数据的获取与处理

数据来源主要包括 3 个部分: 从小麦行业垂直网站上得到小麦品种数据、在线百科获取百科数据、开放知识库获取领域实体及实体之间的关系数据.

品种数据作为小麦生产行业知识主要针对于某一特定领域的专业性网站或数据库, 内容集中, 专一, 内容数据多偏半结构化数据, 但在数据一致性和完整性方面与通用的知识库相比更加完善, 通常需要先分析数据结构, 获取数据后按照其结构解析[15]; 利用互动百科[20] 中的微百科 (category system) 和词条信息模块构建本体; 目前已有很多开放知识库, 如德国马普研究所开发的 Yago[21]、复旦大学开发的 CN-DBPedia[22]、多语言并存的 DBpedia[17] 等. 也有垂直领域的知识库, 如浙江大学维护的新冠开放知识图谱、清华大学的影视双语知识图谱[23]. 本研究利用 Wikidata[24] 完善本地知识库中节点关系, 以便构造“实体-关系-实体”三元组.

获取到的数据往往会存在残缺、错误、重复等问题. 需要对数据进行清洗, 剔除无用数据. 数据清洗融合主要包含数据中含有干扰字符、字段冗余、非结构化文本处理、计量单位不统一等, 按照不同的类型进行单独的处理与转换.

2.2 知识图谱的表示和存储

知识图谱的表示和存储是将学术实体以及实体之间的关系按照一定的数据描述模型, 进行存储的过程[25]. 知识图谱中的知识表示方法是以本体为核心, 以 RDF 的三元组模式为基础框架, 但更多的体现实体、类别、属性、关系等多颗粒度多层次的语义关系.

知识图谱的表示和存储方法使用较广泛的有 RDF 存储、图数据库存储、关系型数据库存储 3 种. 国内的一些学者已将其成功的用于医学领域知识图谱的存储中[26,27]. 但由于 RDF 存储模型设计上不够灵活, 且查询时间复杂度高, 所以不适合作为知识图谱的表示工具. Neo4j 是一个图数据库, 属于非关系型数据库, 它具有高性能、嵌入式、轻量级的优势. Neo4j 以边、节点或属性的形式存储, 而不是以表的形式存储, 对于处理具有复杂关系的海量的知识数据来说是一个利器[28]. Fatima 等[29] 在社交网络场景下, 比较了 Neo4j 图数据库和 MySQL 数据库的表现力. Neo4j 数据库的关系模型可以表达面向网络的数据, 与关系数据库相比, Neo4j 可以在存储数据时连接数据, 使其能够更快地遍历关联数据, 从而存储数以万计的节点和关系, 且随着图谱数据量的不断增大, 关联查询的效率远高于关系型数据库, 因此利用 Neo4j 实现知识图谱表示和存储是较便捷、高效的方法.

2.3 知识图谱设计

知识图谱是一种对于事实的结构化表征, 主要由实体、关系、语义 3 部分组成. 当数据量大, 结构和来源复杂时, 用知识图谱将结构复杂、碎片化数据关联的方式来表示知识会更加清晰准确. 目前, 通用知识图谱构建主要包含数据获取与处理、知识抽取、知识融合和图谱应用 4 个阶段[30], 如图 2 所示.

数据是知识图谱的基础, 从不同结构数据源获取到的领域相关知识做预处理, 对不同来源不同类型的数据进行清洗和入库处理, 目前有很多相关工具, 如清华大学开发的 THULAC[31].

知识抽取是从预处理后的数据中自动创建实体和实体关系的技术[32], 是知识组织和信息融合的跨学科技术, 根据数据结构的不同分为结构化、非结构化和半结构化的知识抽取. 对于结构化数据, 有明确的对应关系, 可以直接构建. 而半结构化数据是指存在一定结

构但还需要加工整理的数据, 抽取时可采用构建包装器的方式. 非结构化数据处理起来较麻烦, 所使用的方法有基于模板、基于监督学习等[19].

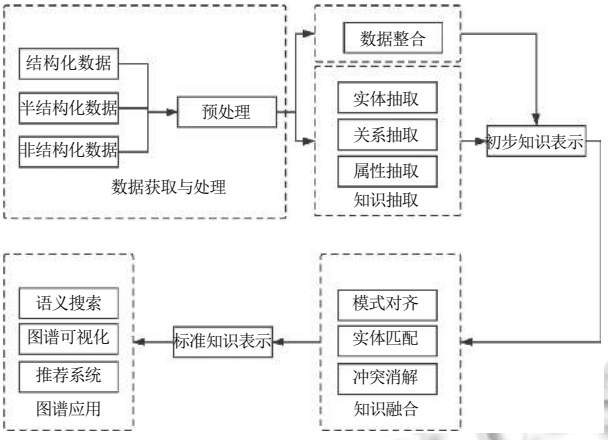

图 2 知识图谱处理流程

经过知识抽取后, 根据表 1 设计小麦的实体类型和关系模型, 从而构建“实体-关系-实体”三元组, 实体设计如表 2 所示, 关系设计如表 3 所示.

表 1 实体、关系模型

模型数据	数据结构	作用	对象范围
节点	标签	可被标记一个或多个类别标签	小麦品种、机构、互动百科数据、Wikidata数据
	属性	可存储多个相关属性	实体名称、实体ID、URL、所属类型等
关系	类型	所属关系类型	包含关系、所属关系、互斥关系、跟随关系、因果关系、组成关系、条件关系
	属性	存储关系属性	关系名称、关系ID

表 2 小麦知识图谱实体设计

实体类型	含义	示例
WinterWheat	冬性小麦	“宁麦5号”
SemiwinterWheat	半冬性小麦	“周麦18”
WeakWinterWheat	弱冬性小麦	“中麦175”
SpringWheat	春性小麦	“宁麦18”
WeakSpringWheat	弱春性小麦	“扬麦18”
OtherWheat	其它	“豫农516”
HudongItem	互动百科节点	“面粉”
NewNode	Wikidata	“绿色植物”
Organization	机构、组织	“河南农业大学”
Wheat	小麦总结点	“小麦品种”
yield	产量数据实体	“42.8”

基于实体和关系的设计, 将数据取出, 通过 Cypher 语句存入 Neo4j 数据库中, 实体和关系都能拥有特定的标签, 有利于节点和关系的分类, 也方便后期查询系统进行查询.

表 3 小麦知识图谱关系设计

关系类型	具体含义	示例
品种	品种	“小麦品种”->“冬性小麦”
品种来源	品种来源	“先麦19”->“西农979”
审定编号	审定编号	“鑫麦296”->“国审麦2014011”
来源与类型	来源与类型	“扬辐麦4号”->“宁麦9号”
每穗粒数	每穗粒数(单位个)	“百农418”->“32. 55”
申请人	该品种的申请人	“涡麦11号”->“亳州市农业科学研究院”
申请单位	该品种的申请单位	“涡麦182”->“亳州市农业科学研究院”
申请者	该品种的申请者	“晨博998”->“河南省亳都种业有限公司”
研究机构	研究机构	“小麦品种”->“研究机构”
育种人	小麦的育种人	“偃亳330”->“韩红卫”
育种者	小麦的育种者	“偃亳197”->“袁灵红”
亩产量kg	单位为kg	“存麦8号”->“583.11”
亲本组合	品种的亲本组合	“西农979”->“藁优5218”
千粒重g	单位为克	“天民198”-> “37.5”
单位面积穗数万	单位为万	“天民198”-> “42.8”
belong_to	属于关系	“国麦301”->“半冬性小麦”
member_of	成员关系	“河南枣乡种业科技有限公司”->“研究机构”

在获得新知识之后, 需要对其进行整合, 以消除矛盾和歧义, 采用余弦相似度的方式表示两个实体对象的相似程度, 相似度介于−1 和 1 之间, 其中−1 表示两个对象完全不同, 1 表示完全相似. 例如, 比较两个小麦品种时, 选取小麦的重要特性 (产量、特征特性、抗性等) 作为特征值, 接着将特征向量化, 最后带入式 (1) 进行计算.

$$similarity(A,B)=\frac{A\cdot B}{\|A\|\times\|B\|}=\frac{\sum_{i=1}^{n}A_i\times B_i}{\sqrt{\sum_{i=1}^{n}A_i^2}\times\sqrt{\sum_{i=1}^{n}B_i^2}} \quad (1)$$

经过知识融合的处理, 形成较为标准知识图谱, 在知识图谱的基础上开发语义搜索、可视化管理等应用.

2.4 知识图谱数据物理存储设计

知识图谱数据类型多样化, 为了提高效率, 针对不同数据进行合理存储设计, 数据的存储架构如图 3 所示.

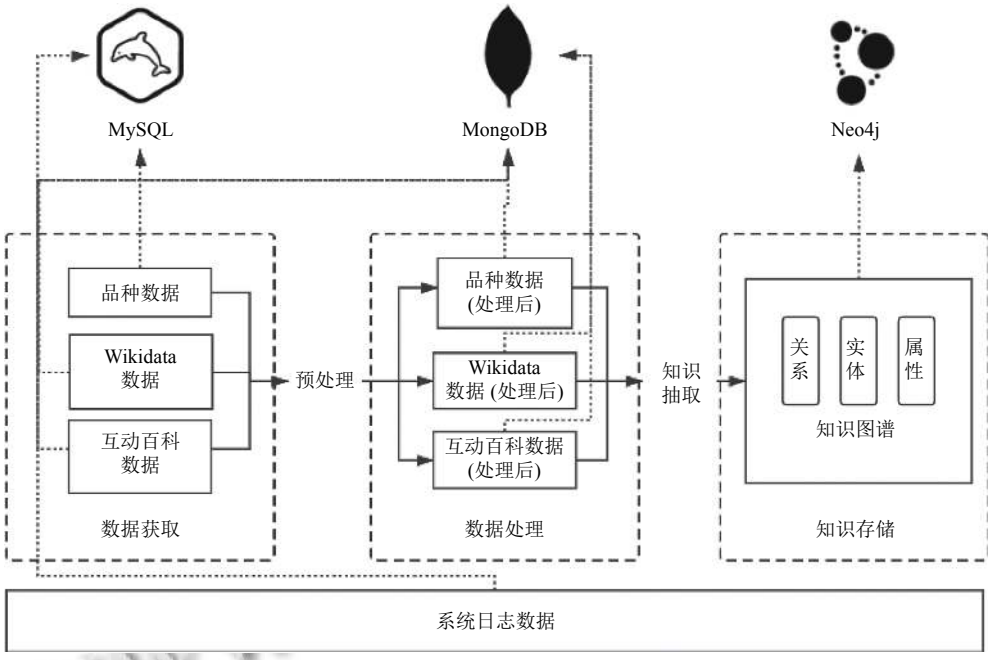

图 3 数据库物理架构

在数据获取阶段, 品种数据存储在 MySQL 结构化数据库中, 而 Wikidata 数据和百度百科词条数据存储在 MongoDB 非结构化数据库中.

在数据清洗阶段, 处理品种数据中存在的字段冗余等问题后, 品种数据含有的属性个数不一致, 选取 MongoDB 来存储处理后的数据, 以减少冗余数据, 提升空间利用率. 处理后的 Wikidata 数据和百度百科词条数据仍然存储在 MongoDB 中, 对处理后的实体、关系和属性数据存储在 Neo4j 数据库中.

3 知识图谱系统构建与应用

3.1 品种知识图谱的构建与实现

选取“种业商务网”[33] 来获取关于小麦品种的数据, 用 BJSON 的格式存储在 MongoDB 数据库. MongoDB 数据库采用, 便于保存不同的属性数据, 共获取 1852 条品种数据, 品种类型丰富, 包括冬性小麦、半冬性小麦、春性小麦、弱春性小麦、弱冬性小麦等多种. 品种的信息包括审定编号、选育单位、品种来源、特征特性、抗性鉴定、品质分析、产量结果等多个维度.

将“农业”的微百科作为种子网站, 爬取所有的微百科, 然后获取微百科中的所有词条, 共获取 735 个微百科, 词条数 102 349 个, 通过知识抽取出实体和实体与实体之间的关系, 最终构建的知识图谱共有实体 258 484 个, 关系 328 933 个, 采用图数据库 Neo4j 来存储实体和关系, 小麦知识图谱的局部结构, 如图 4 所示, 相同颜色的“圆”属于同一种实体类型, 不同“圆”代表不同的实体, “圆”之间的箭头代表实体与实体之间的关系. “圆-箭头-圆”对应 “实体-关系-实体”三元组, 例如: “徐农 029–品种来源-淮麦 20”表示“淮麦 20”是“徐农 029”的品种来源. 并且, 每种实体类型都有一个中心节点, 用来描述该类实体, 例如图中的“半冬性小麦”所指向的实体类型都是“半冬性小麦”.

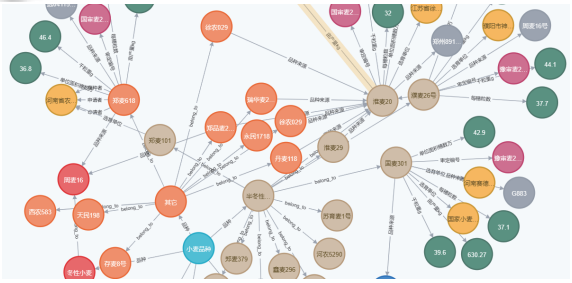

图 4 小麦品种知识图谱

3.2 品种知识精准查询与可视化

由于 Neo4j 数据库高查询性能以及查询语言可定制化, 不仅可以查询实体与实体之间的关系, 还可以实现品种的精确查询, 以返回快速、精准、结构化的知识. 品种知识的查询基于 Neo4j 图数据库的可定制化

Cypher 查询语言, 将实体和属性注入到 Cypher 查询模板中查询出相应的节点数据, 然后将数据封装利用 D3.js 可视化框架将数据可视化, 从而实现图谱中结点和有向关系的直观展示, 如图 5 所示, 可以实现品种数据的实时可视化展示分析.

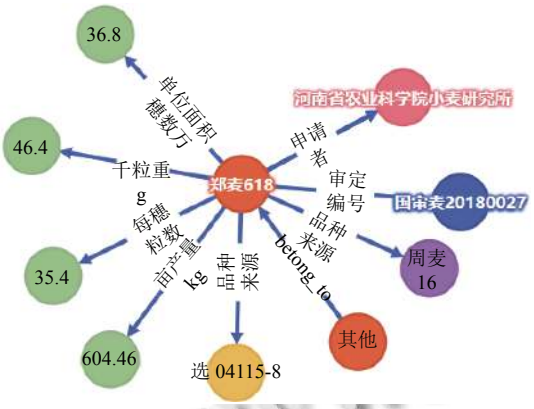

图 5 知识图谱检索

4 结论与展望

本研究基于爬虫技术, 利用 Neo4j、NLP 以及图谱构建技术, 经过数据收集与整理, 知识获取, 知识融合, 知识存储等步骤, 解决现存的知识重复、知识间的关联不够明确等问题. 建立了标准的小麦品种知识图谱体系, 在此基础上, 使用 Neo4j 图数据库存储小麦知识图谱, 建立了小麦品种知识图谱查询系统, 提供品种知识的关系查询、实体查询、品种推荐等功能, 实现了品种知识的精准查询与可视化分析.

基于 Neo4j 图数据库的定制化 Cypher 查询, 利用 D3.j 进行数据可视化, 为农业知识的精确查询和可视化提供了新的途径, 同时也为知识图谱技术在农业生产的应用与落地提供了技术参考. 在未来的研究工作中, 要不断的充实建立的知识图谱体系与系统, 实现知识的及时更新与充实. 此外, 利用 NLP 技术, 结合知识问答系统, 实现农业知识的智能问答推荐也是一个很有价值的应用方向.

参考文献

1 梅方权. 农业信息化带动农业现代化的战略分析. 中国农村经济, 2001, (12): 22–26.

2 张兰廷. 大数据的社会价值与战略选择 [博士学位论文]. 北京: 中共中央党校, 2014.

3 李涛, 王次臣, 李华康. 知识图谱的发展与构建. 南京理工大学学报, 2017, 41(1): 22–34.

4 Steiner T, Verborgh R, Troncy R, *et al*. Adding realtime coverage to the Google knowledge graph. Proceedings of the 2012 International Conference on Posters & Demonstrations Track. Boston, MA, USA. 2012. 65–68.

5 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述. 电子科技大学学报, 2016, 45(4): 589–606. [doi: 10.3969/j.issn.1001-0548.2016.04.012]

6 蒋秉川, 游雄, 李科, 等. 利用地理知识图谱的 COVID-19 疫情态势交互式可视分析. 武汉大学学报 (信息科学版), 2020, 45(6): 836–845.

7 车金立, 唐力伟, 邓士杰, 等. 基于百科知识的军事装备知识图谱构建与应用. 兵器装备工程学报, 2019, 40(1): 148–153. [doi: 10.11809/bqzbgcxb2019.01.031]

8 Liu XX, Bai XS, Wang LH, *et al*. Review and trend analysis of knowledge graphs for crop pest and diseases. IEEE Access, 2019, 7: 62251–62264. [doi: 10.1109/ACCESS.2019.2915987]

9 张善文, 王振, 王祖良. 结合知识图谱与双向长短时记忆网络的小麦条锈病预测. 农业工程学报, 2020, 36(12): 172–178. [doi: 10.11975/j.issn.1002-6819.2020.12.021]

10 Project webpage. https://github.com/qq547276542/Agriculture_KnowledgeGraph.

11 叶帅. 基于 Neo4j 的煤矿领域知识图谱构建及查询方法研究 [硕士学位论文]. 徐州: 中国矿业大学, 2019.

12 张青岭, 李显正, 李航宇, 等. 知识图谱在农业中的应用. 电子技术与软件工程, 2019, (7): 245–247.

13 王儒敬. 我国农业信息化发展的瓶颈与应对策略思考. 中国科学院院刊, 2013, 28(3): 337–343. [doi: 10.3969/j.issn.1000-3045.2013.03.007]

14 齐金山, 梁循, 李志宇, 等. 大规模复杂信息网络表示学习: 概念、方法与挑战. 计算机学报, 2018, 41(10): 2394–2420. [doi: 10.11897/SP.J.1016.2018.02394]

15 胡芳槐. 基于多种数据源的中文知识图谱构建方法研究 [博士学位论文]. 上海: 华东理工大学, 2015.

16 Bollacker K, Evans C, Paritosh P, *et al*. Freebase: A collaboratively created graph database for structuring human knowledge. Proceedings of 2008 ACM SIGMOD International Conference on Management of Data. Vancouver, BC, Canada. 2008. 1247–1250. [doi: 10.1145/1376616.1376746]

17 Auer S, Bizer C, Kobilarov G, *et al*. Dbpedia: A nucleus for a web of open data. Proceedings of the 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference. Busan, Republic of Korea. 2007. 722–735.

18 王丹丹. 宁夏水稻知识图谱构建方法研究与应用 [硕士学位论文]. 银川: 北方民族大学, 2020.
19 刘峤, 钟云, 李杨, 等. 基于图的中文集成实体链接算法. 计算机研究与发展, 2016, 53(2): 270–283. [doi: 10.7544/issn1000-1239.2016.20150832]
20 互动百科, 基于中文维基技术 (维客, wiki 百科) 的网络百科全书. https://www.baike.com/.
21 Suchanek FM, Kasneci G, Weikum G. Yago: A core of semantic knowledge. Proceedings of the 16th International Conference on World Wide Web. Banff, AB, Canada. 2007. 697–706.
22 Xu B, Xu Y, Liang JQ, *et al*. CN-DBpedia: A never-ending Chinese knowledge extraction system. Proceedings of the 30th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. Arras, France. 2017. 428–438.
23 刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述. 计算机研究与发展, 2016, 53(3): 582–600. [doi: 10.7544/issn1000-1239.2016.20148228]
24 贾君枝, 薛秋红. Wikidata 的特点、数据获取与应用. 图书情报工作, 2016, 60(17): 136–141, 148.
25 李肖俊, 邵必林. 多源异构数据情境中学术知识图谱模型构建研究. 现代情报, 2020, 40(6): 88–97.
26 Beyan O, Decker S. An RDF based semantic approach to model temporal relations in health records. Proceedings of the 9th International Conference Semantic Web Applications and Tools for Life Sciences. Amsterdam, the Netherlands. 2016.
27 Wang M, Zhang JH, Liu J, *et al*. PDD graph: Bridging electronic medical records and biomedical knowledge graphs via entity linking. Proceedings of the 16th International Semantic Web Conference. Vienna, Austria. 2017. 219–227.
28 Webber J. A programmatic introduction to Neo4j. Proceedings of the 3rd Annual Conference on Systems, Programming, and Applications: Software for Humanity. Tucson, AZ, USA. 2012. 217–218.
29 Fatima R, Ahmed A. Role of graph databases in social networking sites: A performance comparison between graph database neo4j and relational database MySQL in social networking sites. Journal of Independent Studies and Research, 2012, 10(2): 22–25.
30 胡楠. 基于开放领域知识库的自动问答研究 [硕士学位论文]. 武汉: 华中科技大学, 2019.
31 Sun MS, Chen XX, Zhang KX, *et al*. Thulac: An efficient lexical analyzer for Chinese. http://thulac.thunlp.org/. (2017-01-17)[2019-04-02].
32 徐健, 张智雄, 吴振新. 实体关系抽取的技术方法综述. 现代图书情报技术, 2008, (8): 18–23.
33 种业商务网. https: //www.chinaseed114. com/seed/xiaomai. [2020-07-12].