

# 基于语义知识图谱的农业知识智能检索方法

张海瑜<sup>1</sup> 陈庆龙<sup>2</sup> 张斯静<sup>2</sup> 张子怡<sup>3</sup> 杨帆<sup>3</sup> 李鑫星<sup>3</sup>

(1. 中国农业大学烟台研究院, 烟台 264670; 2. 中国农业大学工学院, 北京 100083;

3. 中国农业大学信息与电气工程学院, 北京 100083)

**摘要:** 针对我国现存网络农业数据库同质异构、知识零散化、一物多词、歧义解析缺乏规范等问题,提出了基于语义知识图谱的农业知识智能检索方法。本文方法围绕农作物品种、农作物病虫害、农作物简介、模型方法4个要素,自顶向下构建模式层;通过本体建模形成知识图谱的概念框架,自底向上构建数据层;通过数据获取、知识抽取、融合、存储建立实体间关联关系。针对语料中歧义字段问题,本文方法在构建知识图谱中收集大量专有词汇,并对其进行分词及词性标注。为了解决在农业知识中一物多词的问题,收集了数量庞大的主要农作物别名,并作为实体赋予明确属性,采用 Bi-LSTM-CRF 进行实体识别,并通过 LSTM 将问题进行分类,利用 TF-IDF 进行关键字提取,最后将知识存储于 Neo4j 图数据库中,从而对相关农业知识数据做规范分类,解决一物多词、一义多解问题。

**关键词:** 知识图谱; 农业知识; 农业词汇; 一物多词; 智能检索

**中图分类号:** S2      **文献标识码:** A      **文章编号:** 1000-1298(2021)S0-0156-08

## Intelligent Retrieval Method of Agricultural Knowledge Based on Semantic Knowledge Graph

ZHANG Haiyu<sup>1</sup> CHEN Qinglong<sup>2</sup> ZHANG Sijing<sup>2</sup> ZHANG Ziyi<sup>3</sup> YANG Fan<sup>3</sup> LI Xinxing<sup>3</sup>

(1. Yantai Institute, China Agricultural University, Yantai 264670, China

2. College of Engineering, China Agricultural University, Beijing 100083, China

3. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

**Abstract:** Aiming at the problems of huge agricultural data, low utilization rate, complex structure and fragmented knowledge in China, a top-down and bottom-up agricultural knowledge map construction method was proposed. Focusing on the four elements of crop varieties, crop diseases and insect pests, crop introduction, and model methods, the model layer was constructed from the top down, and the conceptual framework of the knowledge graph was formed through ontology modeling, the data layer was constructed from the bottom up, through data acquisition, knowledge extraction, and fusion, storing and establishing the relationship between entities. Aiming at the problem of ambiguous fields in the corpus, this method collects large number of proprietary vocabularies in the construction of knowledge graphs to segment and mark them. In order to solve the problem of multi-word in agricultural knowledge, many main crop aliases were collected and assigned as entities. Bi-LSTM-CRF was used for named entity recognition, and LSTM was used to classify the problem, and TF-IDF was used for keyword extraction, and finally the knowledge was stored in the Neo4j graph database. The research can be used for agricultural knowledge intelligent retrieval systems, intelligent search systems and other applications to improve user experience.

**Key words:** knowledge graph; agricultural knowledge; agricultural word; multiple words with one meaning; intelligent retrieval

收稿日期: 2021-07-13 修回日期: 2021-09-08

基金项目: 国家自然科学基金青年科学基金项目(61802411)

作者简介: 张海瑜(1986—),女,讲师,主要从事知识图谱技术与应用研究,E-mail: zhanghaiyu@cau.edu.cn

通信作者: 李鑫星(1983—),男,副教授,博士,主要从事农业系统与知识工程研究,E-mail: lxxcau@cau.edu.cn

## 0 引言

随着信息与数据技术的发展和普及,网络日益成为人们获取各种知识的重要来源。目前,网络上有多种形式的农业知识库,但普遍存在着异质多元、结构散乱和知识表达不清晰、不准确等问题<sup>[1]</sup>。出现这种现象的原因十分复杂,主要可分为以下3方面:我国农业历史悠久,地域辽阔,农业生产者在长期生产活动中受地域、民族风俗与文化遗产等因素影响,对同一事物的表达具有差异性和多元化;随着现代农业技术的传播和推广,出现了一大批由国外引进的作物品种、农业技术、新型物料和装备器具等,对外来语的翻译、称谓不统一,缺少规范;知识库设计和知识获取来源方面的局限,造成知识的碎片化、零散化,缺少准确和完整性的知识表达。上述问题造成了农业知识获取和表达上的混乱,使农业生产者和管理者,特别是广大农民难以有效且准确地获取、识别与辨析其所需要的农业知识,阻碍了科学技术的普及和推广应用。

知识图谱(Knowledge graph)是一种以语义网络图来描述知识和构建各种事物之间关联关系的技术方法。因其强大的知识语义处理和组织能力,知识图谱逐步在语义搜索、知识链接、智能推理、决策分析等多个领域发挥越来越重要的作用,并广泛应用于金融、商贸、医疗等方面,成为新一代认知领域的人工智能技术<sup>[2-4]</sup>。作为知识图谱的核心应用方向,智能知识检索系统通过对自然语言的语义解析,理解用户意图,获取相关知识,并通过推理计算形成自然语言表达的内容并反馈给用户<sup>[5-6]</sup>。目前农业领域的知识图谱构建已经取得了一定的成果,但针对主要农作物种植信息的知识图谱研究仍然很少<sup>[7-10]</sup>。且研究大多集中于方法、技术及应用等方面,忽略了农业各领域各种知识库关于词汇及知识表达方面的不规范现象和词义解释中一物多词与多解问题。

本文以主要粮食作物为研究对象,构建相关知识数据库;用知识图谱技术检索农作物基本信息、农作物种植技术方法等系列农业知识;通过图数据库方法表达知识与实体及实体与实体之间的关联关系,解决知识表达不清晰、不规范性 and 不准确等问题。为提高和改善用户体验,设计农业知识智能检索系统,以简明方便的界面展示给用户,便于用户理解和使用。

## 1 农业词汇知识抽取

### 1.1 知识收集

所谓知识抽取就是从不同的数据来源和现存知

识库中进行知识搜索,并提取所需信息。本文数据来源包含相关的结构化数据、半结构化数据和非结构化数据<sup>[10-12]</sup>。对此,本文做了以下工作:①整理现有部分农业数据库的知识数据。②从相关种业商务网、百度百科网站和国家农业科学数据中心网站爬取相关数据。③随机搜索收集其它网络数据与信息资源中的零散数据。知识抽取步骤如下:首先在百度百科网站找到指定农作物对应的统一资源定位系统(Uniform resource locate, URL),通过解析网页信息得到农作物的属性信息,获取农作物别名、生长环境、栽培技术、主要价值、历史文化等相关数据;然后在种业商务网站获取玉米、小麦、水稻、棉花、大豆、番薯、马铃薯7个农作物的品种名称、品种来源、育种者、特征特性、栽培技术点、产量表现等数据;最后根据国家农业科学数据中心网站中植物保护科学提供的病虫害数据,采集农作物病害中文名、病株形态特征、主要危害部位、危害症状、防治方法等数据。

采用Python语言中Scrapy框架获取网站数据。首先分析各个网页结构,根据不同的网页结构形式利用Python语言编写爬虫脚本,最后对爬取数据进行清洗并保存为CSV格式,具体采集流程如图1所示。

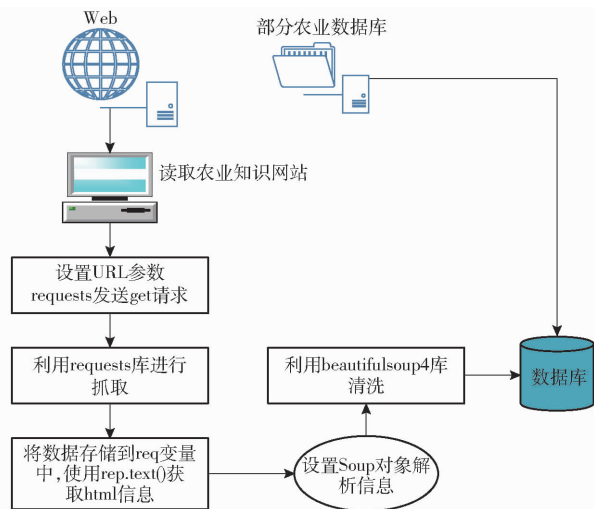


图1 农业知识数据采集流程图

Fig.1 Collection process diagram of agricultural knowledge data

### 1.2 知识处理

由于网站架构差异,按照前述方法获取的网站原始数据中存在数据无效和数据缺失问题,如“农作物别名”中包含除别名以外的农作物简介等无效数据;“农作物品种信息”中出现了二级页面的URL和未解析的html内容等。因此,需要对原始数据进行进一步清洗。通过设置停用词表,过滤诸如“农作物别名”外的无效及错误信息;对于原始数据中诸如“农作物品种信息”存在二级页面URL和页面

html 内容类问题,通过导入 Python 的 Parse 包重新解析 html 的内容,从 html 中解析出“品种来源”、“特征特性”、“栽培技术要点”等信息。为了解决网站中非结构化文本不能被直接利用的问题,使用神经语言程序学 (Neuro-linguistic programming, NLP) 中的开源工具 LTP,对文本进行分词、词性标注、依存句法分析和语义依存分析 4 个模块预处理<sup>[13]</sup>。

清洗后的信息以 {信息类别标签:信息} 的形式进行存储,结合信息清洗后的结果和知识图谱的需求,本文设置农作物别名、生长环境、主要价值、历史文化、品种名称、品种来源、特征特性、产量表现、育种者、栽培技术要点、中文疾病名称、病株形态特征、主要危害部位、危害症状、防治方法 15 个信息标签,

最后将原始信息格式化为 CSV 文件存储到本地存储器。

## 2 农业词汇知识图谱构建

知识图谱构建是本文方法的核心,具体流程如图 2 所示。在知识图谱设计初期,首先明确该图谱所针对的领域。对于开放领域类,将海量的通用信息作为知识图谱的基础信息,通过信息补充和信息融合等方法持续丰富信息内容<sup>[14-15]</sup>。对于特定领域类,需要收集该领域的相关信息,根据信息的结构形式,采用不同的信息处理技术进行处理,并存储到信息库内,便于上层应用对信息的分析、应用和查询<sup>[16-17]</sup>。

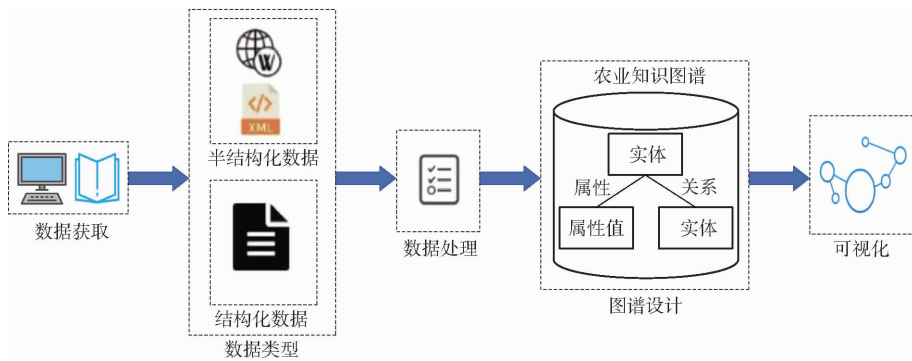


图2 知识图谱构建流程图

Fig.2 Knowledge graph construction process diagram

本文知识图谱构建为农业特定领域,在构建知识图谱过程中,需要完成对信息进行实体、关系及属性的定义。通过对所收集的农业信息进行分析,编写脚本定义信息中的实体、关系和属性<sup>[18]</sup>。具体农业知识图谱信息库的字段如表 1 所示。

表 1 数据库字段  
Tab.1 Database field

节点标签	属性
农作物	农作物名称
	生长环境
	主要价值
	历史文化
品种	品种名称
	品种来源
	特征特性
	产量表现
	育种者
	栽培技术要点
疾病	中文疾病名称
	病株形态特征
	主要危害部位
	危害症状 防治方法

铃薯、番薯 7 种农作物关联关系和别名、生长环境、主要价值、历史文化、品种名称、品种来源、产量表现、特征特性、育种者、栽培技术要点、中文疾病名称、病株形态特征、主要危害部位、危害症状、防治方法 15 类属性数据,共抽取了 8 136 个实体,28 104 个实体—关系—实体的三元组集合。

得到上述数据后,需要将数据进行保存,目前提出的关系型数据库产品种类众多,例如 Neo4j、Oracle、InfoGrid、SQLServer、OrientDB 等<sup>[19]</sup>。其中,Neo4j 提供了便捷的信息访问应用程序接口 (Application programming interface, API),通过导入 Python 语言中的 Py2neo 包,Neo4j 可以实现 CSV 格式文件批量写入图形数据库,并建立定义好的实体与实体之间关系和实体内部属性的关联。为了便于系统对上层数据的查询、处理和分析,本文选取 Neo4j 图数据库作为存储工具进行农业知识图谱构建,将收集的农业知识图谱相关数据信息导入 Neo4j 中的部分效果如图 3 所示。针对“一物多词”问题,在数据库中添加各地区对同种农作物的多种称谓,其中以绿色节点代表农作物学名或别名,属性集合,包括生长环境、主要价值、历史文化、品种名称、栽培技术;红色节点代表农作物的品种名称,属

本文共收集了玉米、小麦、棉花、水稻、大豆、马

性集合,包括品种名称、品种来源、特征特性、产量表现、育种者、栽培技术要点;黄色节点代表农作物中一种疾病名称,属性集合包括病株形态特征、主要危害部位、危害症状、防治方法。



图3 基于 Neo4j 的农业知识图谱局部结构  
Fig.3 Partial structure of agricultural knowledge map based on Neo4j

### 3 智能检索方法构建

本文搭建了基于知识图谱的智能检索系统,以更好地为不同地区农户在实际生产中出现的问題提供解决方案。

#### 3.1 分词及词性标注

针对中文语句分词的研究目前已经提出十几种方法,例如 HanLP、pynlpir、ansj、jieba、snownlp 等,并成功开发出若干个相关的系统组件<sup>[20]</sup>。这些方法总体可以分为两类,一类是基于字典分词的方法,另一类是基于知识规则分词的方法。本文选用 jieba 工具处理速度快、带有词性标注等功能,并可以自定义字典,同时该算法可直接导入数据包形式,加载进入 Python 程序,以便于实现<sup>[21]</sup>。但是 jieba 核心字典内并不包含农业领域的概念分词,因此进行分词时会产生部分专有名词拆分现象。本文将农业领域的概念词定义到用户自定义字典中,在自定义过程中,将农作物病名称、农作物品种类名词定义词性为“nr”,将农作物名称类名词定义词性为“ng”。图4包含了部分用户自定义农业字典中的专有名词。

#### 3.2 词向量

词向量(Word embedding)是自然语言处理中用于特征学习与语言建模技术的一种统称,将短语或词汇表的单词映射到实数的向量中<sup>[22]</sup>。Google 公司于 2013 年开源了一款用于词向量的工具——word2vec,通过给定的语料库优化训练模型,将词语及时有效地转换成为向量的形式<sup>[23-24]</sup>。

本文采用 word2vec 进行词向量训练,使用 Skip-grams 模型思路进行预训练词向量。该模型

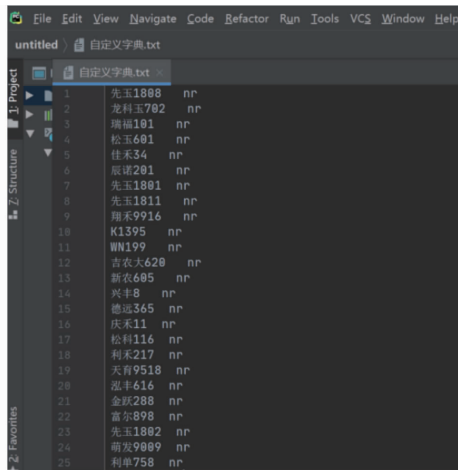


图4 农业领域部分自定义字典

Fig.4 Some custom dictionaries in agricultural field

可以根据词语的输入来预测其上下文,如图5所示,本文设置 word2vec 词向量的维度为 100,窗口尺寸设置为 5 × 5。

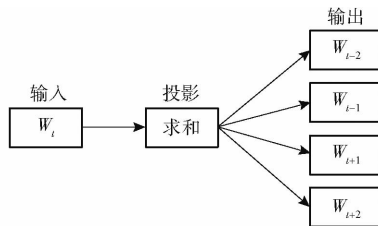


图5 Skip-grams 模型

Fig.5 Skip-grams model

#### 3.3 命名实体识别

命名实体识别(Name entity recognition,NER)是识别出给定文本中表示命名实体的成分,例如地名、人名、机构名等关键性实体<sup>[25]</sup>。该概念自 1995 年提出,被视为自然语言处理与信息检索中的一个重要领域,实体命名的技术从早期基于规则与字典的识别方法,经历传统机器学习方法再发展到深度学习的方法<sup>[26-27]</sup>。本文采用的算法模型为 Bi-LSTM-CRF,将实体标注的语料标签自动进行标注处理,即将标注的标签通过 BIOES 进行序列标注。命名实体识别任务中标记类型共有 18 种,如表 2 所示。

#### 3.4 问句分类

问句分类在检索系统中占有极其重要的作用,直接影响检索系统的检索速度。本文采用问句分类的算法,该算法可以准确解读用户意图,缩小检索结果范围,缩短响应时间,提高检索效率。如当用户输入“棒子的栽培技术是什么?”这一问句,通过问句分类模型,检索系统可以快速在玉米栽培技术范围内定位,在整体匹配过程中,可缩小答案的搜寻范围和查找时间,极大提高智能检索系统的效率。

近年来,循环神经网络(Recurrent neural network,RNN)在自然语言处理以及序列问题等领

表 2 命名实体标注标签  
Tab.2 Named entity tagging

实体标签	实体含义
O	其它
B-DRO	品种实体开头
I-DRO	品种实体中间
E-DRO	品种实体结尾
B-DIS	疾病实体开头
E-DIS	疾病实体结尾
B-SYS	症状实体开头
I-SYS	症状实体中间
E-SYS	症状实体结尾
B-DRU	药物实体开头
I-DRU	药物实体中间
E-DRU	药物实体结尾
B-V	虫害实体开头
I-V	虫害实体中间
E-V	虫害实体结尾
S-DRU	单个药物实体
S-DIS	单个疾病实体
S-SYS	单个症状实体

域取得了长足的进步,这是以一种序列数据为输入,递归地沿着序列演进的方向进行,所有节点按照链式连接,其优点在于模型具有记忆性<sup>[28]</sup>。但因其梯度爆炸或梯度消失的特点,存在不能记忆太靠前或者太靠后内容的缺点,该模型结构如图 6 所示。

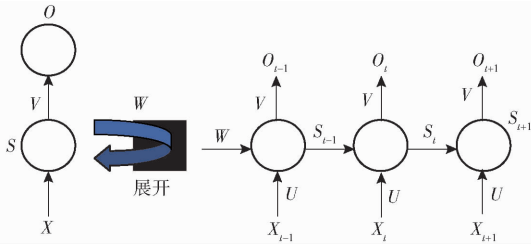


图 6 循环神经网络结构模型

Fig.6 RNN recurrent neural network model

图 6 中每个圆圈视作一个神经元,  $t-1, t, t+1$  表示时间序列,  $S$  表示样本在时间某处的记忆,  $X$  表示输入样本。  $U$  表示此刻输入样本的权重,  $W$  表示输入权重,  $V$  表示输出样本权重。在  $t=1$  时

$$h_1 = UX_1 + WS_0 \quad (1)$$

$$S_1 = f(h_1) \quad (2)$$

$$O_1 = g(VS_1) \quad (3)$$

式中  $f, g$ ——激活函数

$O_1$ ——输出

$h_1$ ——隐藏状态

随着时间的推移,可得到最终的输出值为

$$h_t = UX_t + WS_{t-1} \quad (4)$$

$$S_t = f(h_t) \quad (5)$$

$$O_t = g(VS_t) \quad (6)$$

鉴于循环神经网络存在梯度消失这一局限性,本文在研究问句分类的过程采用循环神经网络中特

殊的神经网络,长短期记忆人工神经网络(Long short-term memory, LSTM)<sup>[29]</sup>。其记忆单元增加在隐藏层的各个神经元之中,使在时间序列上的记忆信息变得可控。当各个单元在隐藏层传递时,通过可控门控制当前的信息与之前信息的记忆,使得 RNN 获得了长期记忆的功能,其优点在于增加了输入门、遗忘门、输出门,使得 RNN 梯度消失问题得到解决。LSTM 分类实例如图 7 所示。

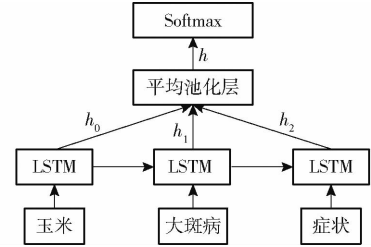


图 7 LSTM 分类实例

Fig.7 LSTM classification instances

LSTM 通过设计输入门 (Input gate) 与遗忘门 (Forget gate) 用以控制状态 C 记忆单元的信息量<sup>[30-31]</sup>。输入门决定了当前时刻输入的信息保存到单元的状态,遗忘门可以决定上一时刻有多少“信息”可以保存到当前时刻。同时, LSTM 有一个输出门 (Output gate),用以控制单元输出的状态信息。

本文采用 LSTM 来进行农业知识问句分类模型训练<sup>[32]</sup>。首先通过 Jieba 进行中文分词,然后利用 word2vec 将其转化为词向量,并将其拼接成句向量作为 LSTM 的输入。以“玉米大斑病症状”为例,通过模型训练,经过 Softmax 层得到问句属性的概率,从而得到了 3 种问句分类类别。为解决农户在病虫害防治以及品种名称混乱的问题,本文结合生产实际,将问题分为品种类、病虫害类和防治类 3 类。问句语料主要来源于百科数据和部分农业网站,通过数据爬取,得到品种类问句 1 208 条,病虫害类问句 2 908 条,防治类问句 2 631 条,其中 80% 作为训练集,20% 作为测试集,其分类算法结果如表 3 所示。其中 F1 值为准确率和召回率的调和平均值。

表 3 基于 LSTM 分类算法的问句分类结果

Tab.3 Question classification results based on LSTM classification algorithm

类别	准确率	召回率	F1 值
品种类	90.13	91.11	90.62
病虫害类	89.24	88.12	88.68
防治类	88.63	89.15	88.89

### 3.5 问句关键词提取

用户提出的问题被正确分类后,就可以通过关

关键词提取模块获得问句中的语义关键词。用户提出的问题被转化为问题的模板形式后,计算两者之间的相似度。若相似度大于已设定的阈值且相似度最高的模板作为用户输入问题,并执行相对应的 Cypher 查询语句,返回用户输入问题答案。

本文基于 TF-IDF 关键词提取算法,对用户输入的问句进行处理,并根据提取到的信息构造农业领域知识图谱查询语句,进行信息检索<sup>[33]</sup>。本文构建的农业领域的知识图谱,首先对问题和模板进行分词,通过统计每个词条出现的次数,计算出每个词条的逆文档词谱。当出现  $n$  条词语时,将每个模板的问句转换成向量表示 ( $N$  维),每个维度的数值为本词在模板或问句里出现的次数乘以本词对应的 IDF 值。计算两者之间的余弦值,余弦值越大,表示两者之间相似度越高。通过试验设定一个阈值  $m$ ,当余弦值大于阈值  $m$  时,由相似度最大的问题模板作为用户问题。问题解答模块对于大部分问题能够得出一个初始的答案,但也会存在无效的情况,此刻运用答案生产模块对用户提出的问题进行分析,然后采用相似度计算方法选出答案库中相似度最高的问题,检测相似度是否大于已设定的阈值,若是,则输出答案,若不是,则在检索页面中显示“无法检索出答案”。

本文采用 Neo4j 作为知识存储,Neo4j 提供了高性能的信息查询语言 Cypher,基于 Cypher 构造了 Cypher 查询模板。当问句通过基于 LSTM 来进行主要粮食知识问句分类和基于 TF-IDF 关键词提取处理后,系统会根据用户输入问句的类型,获取相应的 Cypher 模板,将输入问句的关键词替换成模板中的变量,最后传递给 Neo4j,查询获取问题的答案,图 8 为 Cypher 执行语句生成过程。

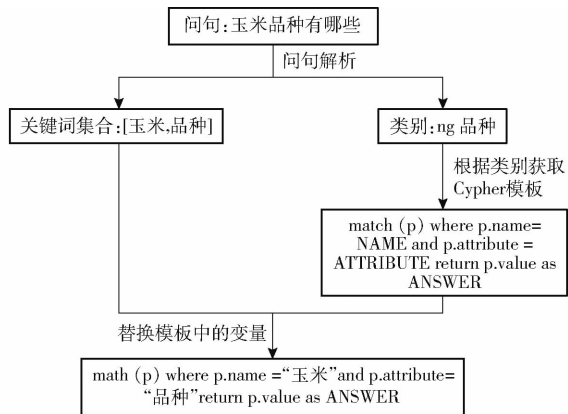


图 8 Cypher 执行语句生成过程

Fig. 8 Cypher execution statement generation process

### 3.6 农业知识智能检索系统设计与实现

#### 3.6.1 智能检索系统设计

根据实际需求设计农业知识智能检索系统,

图 9 为检索系统流程图,主要包括 3 方面:根据用户输入的内容进行解析,将所述问题进行分类;将问句向量化,转换为计算机能够识别的内容;在答案库中检索,根据相似度匹配相应的答案。

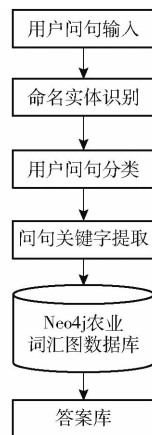


图 9 检索系统流程图

Fig. 9 Retrieval system flow chart

#### 3.6.2 智能检索系统实现

本文搭建检索系统采用 Java 语言,系统采用 MVC 模式、B/S 架构,前端界面使用 Bootstrap 框架进行构建,同时还采用了 Html5 和 Css3 等技术。检索系统通过用户检索解析,结合 Neo4j 图形数据库自带 Cypher 查询语言,在问答知识库中匹配答案,返回前端页面。图 10 为用户登录界面。图 11 为检索系统主界面。图 12 为用户搜索返回的结果界面。搭建此检索系统旨在更好地为本领域农户服务,使农户在使用此系统时更加准确、及时地解决问题,使农户在实际生产过程中提高效率。

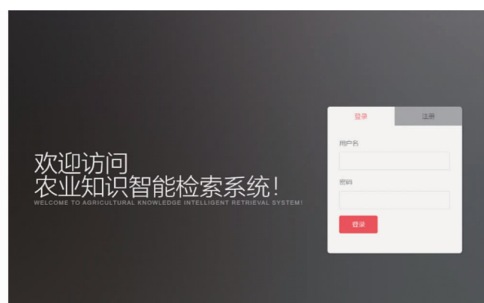


图 10 用户登录界面

Fig. 10 User login interface

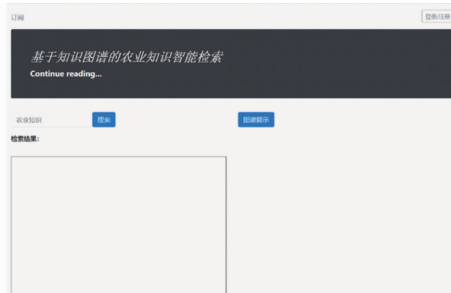


图 11 检索系统主界面

Fig. 11 Main interface of retrieval system



图 12 用户搜索返回的结果界面

Fig. 12 Results returned by user search

## 4 结束语

构建了农业语义领域知识图谱和智能检索系统,扩展了知识图谱技术在农业语义领域的应用范围。本文方法解决了中国各地区对于农作物的称呼不同而造成网络资源搜索时信息检索不准确以及不匹配等问题,并在知识图谱中增加了农作物病虫害以及栽培技术等相关内容。搭建的智能检索系统使农户更加容易地了解农作物的相关知识,可以自行诊断农作物的病虫害。

## 参 考 文 献

- [1] 马超,胡婉聆,张爽. 我国农业大数据技术的应用探析[J]. 南方农机, 2017, 48(3):42.  
MA Chao, HU Wanling, ZHANG Shuang. Analysis on the application of agricultural big data technology in my country[J]. Southern Agricultural Machinery, 2017, 48(3): 42. (in Chinese)
- [2] 王昊奋. 大规模知识图谱技术[EB/OL]. (2014-06-12) [http://www.China-cloud.com/zhongyunxy/20140612\\_38070.html](http://www.China-cloud.com/zhongyunxy/20140612_38070.html).
- [3] SOWA J F. Principles of semantic networks; explorations in the representation of knowledge[M]. Morgan Kaufmann, 2014.
- [4] 袁培森,李润隆,王翀,等. 基于BERT的水稻表型知识图谱实体关系抽取研究[J]. 农业机械学报, 2021, 52(5): 151-158.  
YUAN Peisen, LI Runlong, WANG Chong, et al. Entity relationship extraction from rice phenotype knowledge graph based on BERT[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(5): 151-158. (in Chinese)
- [5] ZANG F, YUAN N J, LIAN D, et al. Collaborative knowledge base embedding for recommender systems[C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016: 353-362.
- [6] 徐增林,盛泳潘,贺丽荣,等. 知识图谱技术综述[J]. 电子科技大学学报, 2016, 45(4): 589-606.  
XU Zenglin, SHENG Yongpan, HE Lirong, et al. Overview of knowledge graph technology[J]. Journal of University of Electronic Science and Technology of China, 2016, 45(4): 589-606. (in Chinese)
- [7] 刘齐胜,周锦,程蜜,等. 农业叙词管理系统的设计与实现[J]. 电脑与信息技术, 2017, 25(4): 14-17.  
LIU Qisheng, ZHOU Jin, CHENG Mi, et al. Design and implementation of agricultural narrative management system [J]. Computer and Information Technology, 2017, 25(4): 14-17. (in Chinese)
- [8] 吴赛赛. 基于知识图谱的作物病虫害智能问答系统设计与实现[D]. 北京:中国农业科学院, 2021.  
WU Saisai. Design and implementation of intelligent question answering system for crop diseases and insect pests based on knowledge graph [D]. Beijing: Chinese Academy of Agricultural Sciences, 2021. (in Chinese)
- [9] 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.  
DING Haochen, WANG Zhongming. Construction and application of ontology-based Chinese knowledge map of *Camellia oleifera* [J]. World Forestry Research, 2020, 33(4): 50-55. (in Chinese)
- [10] 赵明,董翠琴,董乔雪,等. 基于BIGRU的番茄病虫害问答系统问句分类研究[J]. 农业机械学报, 2018, 49(5): 271-276.  
ZHAO Ming, DONG Cuicui, DONG Qiaoxue, et al. Question classification of tomato pests and diseases question answering system based on BIGRU[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(5): 271-276. (in Chinese)
- [11] 刘峤,李杨,段宏,等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2006, 53(3): 582-600.  
LIU Qiao, LI Yang, DUAN Hong, et al. Summary of knowledge graph construction technology[J]. Computer Research and Development, 2006, 53(3): 582-600. (in Chinese)
- [12] DODDS K. Popular geopolitics and audience dispositions: James Bond and the Internet Movie Database (IMDb) [J]. Transactions of the Institute of British Geographers, 2006, 31(2): 116-130.
- [13] 吴赛赛,周爱莲,谢能付,等. 基于深度学习的作物病虫害可视化知识图谱构建[J]. 农业工程学报, 2020, 36(24): 177-185.  
WU Saisai, ZHOU Ailian, XIE Nengfu, et al. Construction of visual knowledge graph of crop diseases and pests based on deep learning[J]. Transactions of the CSAE, 2020, 36(24): 177-185. (in Chinese)
- [14] CHE W, LI Z, LIU T. LTP: a Chinese language technology platform[C]//Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations, 2010: 13-16.
- [15] MIWA M, BANSAL M. End-to-end relation extraction using LSTMs on sequences and tree structures[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016.

- [16] KATIYAR A, CARDIE C. Going out on a limb: joint extraction of entity mentions and relations without dependency trees[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017.
- [17] 王达山. 商业银行大数据建设探讨[J]. 中国金融电脑, 2016(3): 83-87.  
WANG Dashan. Discussion on the construction of big data in commercial banks[J]. China Financial Computer, 2016(3): 83-87. (in Chinese)
- [18] KAUR S, KAUR K. Visualizing class diagram using orientdb NOSQL data-store[J]. International Journal of Computer Applications, 2016, 145(10): 11-15.
- [19] 赵月斋. 一种基于词频统计的中文分词方法[J]. 科技展望, 2016, 26(10): 283  
ZHAO Yuezhai. A Chinese word segmentation method based on word[J]. Technology Outlook, 2016, 26(10): 283. (in Chinese)
- [20] 程童凌, 李娟子. 基于维基类百科知识资源的实体关系发现和语标注[J]. 电子技术与软件工程, 2015(18): 170-173.  
CHENG Tongling, LI Juanzi. Entity relationship discovery and language tagging based on Wikipedia knowledge resources [J]. Electronic Technology and Software Engineering, 2015(18): 170-173. (in Chinese)
- [21] HAN D J, LV Q J, YANG J. Research and analysis of browsing habits of new media audiences on the Chinese campus[C]// Proceedings of 2017 2nd International Conference on Wireless Communication and Network Engineering(WCNE 2017), 2017.
- [22] 朱小燕, 李晶, 郝宇, 等. 人工智能知识图谱前言技术[M]. 北京: 电子工业出版社, 2020.
- [23] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]// Advance in Neural Information Processing Systems, 2013: 3111-3119.
- [24] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. arXiv: 1301.3781.
- [25] 郭喜跃, 何婷婷. 信息抽取研究综述[J]. 计算机科学, 2015, 42(2): 14-17, 38.  
GUO Xiyue, HE Tingting. Summary of information extraction research[J]. Computer Science, 2015, 42(2): 14-17, 38. (in Chinese)
- [26] 林广和, 张绍武, 林鸿飞. 基于细粒度词表示的命名实体识别研究[J]. 中文信息学报, 2018, 32(11): 62-71.  
LIN Guanghe, ZHANG Shaowu, LIN Hongfei. Research on named entity recognition based on fine-grained word representation [J]. Journal of Chinese Information Processing, 2018, 32(11): 62-71. (in Chinese)
- [27] 丁雅琴. 基于知识图谱的医疗问答系统研究与开发[D]. 武汉: 华中师范大学, 2020.  
DING Yaqin. Research and development of medical question answering system based on knowledge graph[D]. Wuhan: Huazhong Normal University, 2020. (in Chinese)
- [28] 陈梅梅. 基于知识图谱的医疗问答系统设计与实现[D]. 厦门: 厦门大学, 2019.  
CHEN Meimei. Design and implementation of medical question answering system based on knowledge graph[D]. Xiamen: Xiamen University, 2019. (in Chinese)
- [29] SUNDERMEYER M, SCHLUTER R, NEY H, et al. LSTM neural networks for language modeling[C]// Conference of the International Speech Communication Association, 2012.
- [30] GERS F A, SCHMIDHUBER J, CUMMINS F. Learning to forget: continual prediction with LSTM[J]. Neural Computation, 2000, 12(10): 2451-2471.
- [31] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase. representations using RNN encoder-decoder for statistical machine translation [J]. arXiv: 1406.1078.
- [32] MIKOLOV T, KARAFIÁT M, BURGET L, et al. Recurrent neural network based language model[C]// Conference of the International Speech Communication Association, 2015.
- [33] 赵胜辉, 李吉月, 徐碧路, 等. 基于 TFIDF 的社区问答系统问句相似度改进算法[J]. 北京理工大学学报, 2017, 37(9): 982-985.  
ZHAO Shenghui, LI Jiyue, XU Bilu, et al. Improved algorithm of question similarity in community question answering system based on TFIDF[J]. Journal of Beijing Institute of Technology, 2017, 37(9): 982-985. (in Chinese)