Received March 29, 2019, accepted May 5, 2019, date of publication May 9, 2019, date of current version May 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2915987

Review and Trend Analysis of Knowledge Graphs for Crop Pest and Diseases

LIU XIAOXUE[1], BAI XUESONG[1], WANG LONGHE[1], REN BINGYUAN[2], LU SHUHAN[3], AND LI LIN[1]

[1]College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
[2]Prevention and Control Department of National Agricultural Technology Service Center, Beijing 100125, China
[3]Computer and Information Science, The Ohio State University, Columbus, OH 43210, USA

Corresponding author: Li Lin (lilincau@126.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFD0300710.

ABSTRACT Current techniques of knowledge management have some common defects in efficiency, scalability, and applicability. Knowledge graph provides a new way for knowledge management and is a more flexible knowledge management method. Considering the specific features of crop diseases and pest data, this paper analyzed and classified the key techniques and methods of knowledge graph technology in the field of crop diseases and pest in recent years. We introduced the definition and connotation of the crop diseases and pest knowledge, and the current construction method of it was analyzed in detail from four aspects: knowledge representation, extraction, fusion, and reasoning. Furthermore, the application of crop diseases and pest knowledge graph was introduced in detail in the expert system, search engine, and knowledge question-answering system. At last, this paper summarized the challenges and important problems of diseases and pest knowledge graph and forecasted the prospect of knowledge graph according to the key points and difficulties of current knowledge graph research.

INDEX TERMS Knowledge graph, crop disease and pest, knowledge extraction, knowledge fusion, knowledge reasoning.

I. INTRODUCTION

A knowledge graph is a structured semantic knowledge base that describes concepts and their relationships in the physical world in the form of symbols. It allows knowledge representation and management to solve knowledge association problems, such as knowledge retrieval and semantic question answering. It is the basis and bridge to realize intelligent semantic retrieval [1]. Knowledge graph technology is a combination of ontology technology, knowledge extraction, fusion, query, storage, reasoning, and other technologies. It can be divided into general and vertical domain knowledge graphs. Generally, a knowledge graph is typically large-scale, in wide field, and has large amount of common sense [2]. The vertical domain knowledge graph has desirable advantages of accuracy and fine granularity, and can effectively support knowledge reasoning and knowledge retrieval applications. Crop diseases and insect pests have a vertical industry knowledge graph, which is one of the most widely used fields. Knowledge graph technology for crop pests and diseases has been widely applied for crop variety selection [3], greenhouse environment control, pest control [4], fertilization and irrigation [5], economic benefit analysis [6], and other aspects of agricultural production and application. Developing pest and disease expert systems based on knowledge graphs have promoted the conversion of achievements of agricultural technology, contributing to the development of high-yield, high-efficiency and high-quality agriculture [7].

In the field of crop diseases and insect pests, the traditional knowledge bases based on the experience of experts in the field are static and limited. However, data for agricultural production and research on diseases and insect pest are huge. Full and effective sharing and integration of pest and disease data is the basis of semantic interoperability. Expert knowledge of pests and diseases is represented by text, tables, and figures, but these need to be converted into computational resources and processed by computers; therefore, a large-scale and open-linked knowledge acquisition and management method of pests and diseases based on knowledge graph technology has emerged [8]. Knowledge graph construction includes extracting entities, relationships, and attributes from

The associate editor coordinating the review of this manuscript and approving it for publication was Feng Xia.

structured, semi-structured, and unstructured data; integrating knowledge through entity alignment and ontology matching; and evaluating the quality of the knowledge base, and then updating, completing, and retrieving knowledge [9]. The methods and technologies used in knowledge representation, extraction, fusion, and reasoning are reviewed based on the construction of knowledge graphs in the field of crop diseases and insect pests. Future developments in knowledge graph construction technology and their applications in the crop are prospected in this paper.

The main organization structure of this paper is as follows: Section 1 briefly introduces the current knowledge graph and the knowledge graph of crop diseases and insect pests. Section 2 introduces the definition and structure of knowledge graph, and gives examples of field of crop diseases and insect pests for illustration. Section 3 introduces the main technologies and methods in the knowledge graph construction of crop diseases and insect pests, which are divided into four parts: knowledge representation, knowledge extraction, knowledge fusion and knowledge reasoning. Section 4 introduces some applications of knowledge graph of crop diseases and insect pests, such as expert system, intelligent search and question answering system. Finally, the prospect of knowledge graph of crop diseases and insect pests is presented.

II. DEFINITION AND ARCHITECTURE OF KNOWLEDGE GRAPH

A. BASIC DEFINITION

A knowledge graph is a network of relational links between attributed entities. It is used to describe concepts, entities, events, and their relationships in the objective world [10]. The concept refers to the conceptual representation of objective things formed in the process of people's understanding of the world, and mainly refers to sets, categories, object types, and affair types. Event refers to activities in the objective world, such as crop disease, prevention, and control behavior. Entity is the most basic element in the knowledge map. It includes specific things in the objective world, such as crops, diseases, pathogens, symptoms, location of disease, and pesticides. Relation is an objective relationship among describing concepts, entities, and events. It exists among different entities, such as symptoms, the cause of disease, control method, selection and application methods. Properties mainly refer to the characteristics or parameters of the object, such as disease characteristics, English names of diseases and types of pesticide. Property values refer to the specific attribute values. Attribute values are generally string type. In the knowledge graph, entities are identified by a globally unique encoding, the intrinsic characteristics of entities are characterized by property–property values pairs, and the relationships between entities are described by relationships.

A knowledge graph, which is expressed by triples, is essentially a semantic web that reveals the interrelationship between entities. For example, knowledge maps are often represented as SPO (subject, predicate, object) triple form. Table 1 shows a tobacco disease description triple.

TABLE 1. Triple representation of tobacco disease description.

Subject	Predicate	Object
tobacco mosaic	symptom	flower leaf burn
tobacco mosaic	symptom	leaf curl
tobacco mosaic	route of transmission	sap transmission
tobacco mosaic	route of transmission	host plant
tobacco blight	related disease	tobacco mosaic
black rot of tobacco root	related disease	tobacco mosaic

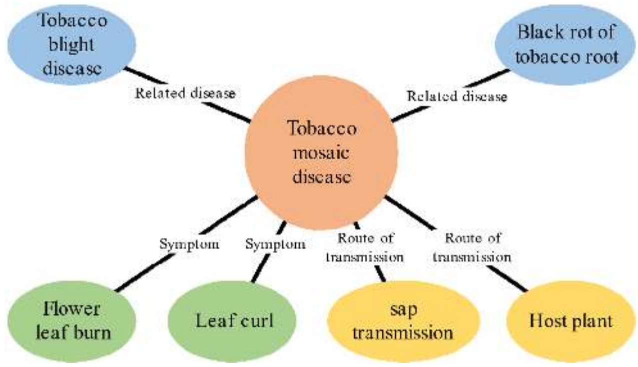

FIGURE 1. An example of knowledge graph of tobacco disease.

All the three elements are combined to form a graph. The nodes in the map represent entities, and the directed edges represent the relationships between entities. The directions of edges indicate whether the entity is the subject or the object of action. Different relationships can also be represented by different types of edges. The structure of the knowledge map consisting of all the triples in Table 1 is shown in Figure 1.

B. KNOWLEDGE GRAPH ARCHITECTURE

The structure of a knowledge graph is divided into two levels, namely, the schema and data layer [11]. The schema layer is managed on the data layer, usually through the ontology database, which stores the extracted upper knowledge. The data layer stores all the underlying knowledge of the fact class. Usually, the fact knowledge is stored in the graph database to form a relational network. Schema layer knowledge is usually extracted from data layer knowledge, which limits the data layer knowledge.

Integrity, accuracy, and data quality are important parameters to determine the usefulness of the knowledge base and are affected by building the knowledge base. The construction of knowledge map includes top-down and bottom-up approaches [12]. The top-down approach defines data schemas for knowledge maps, where in we first start with the top-level concepts, then gradually refine their subclasses to form a well-structured hierarchy. Afterward, we add each entity to the concepts. The bottom-up approach starts with entities and organizes them into bottom concepts, and then gradually abstracts them to top concepts. Figure 2 shows the structure of miniature knowledge graph of tobacco disease.

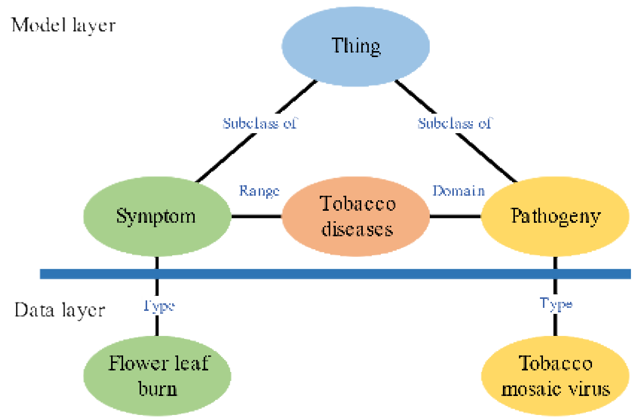

FIGURE 2. Structural chart of small tobacco diseases knowledge graph.

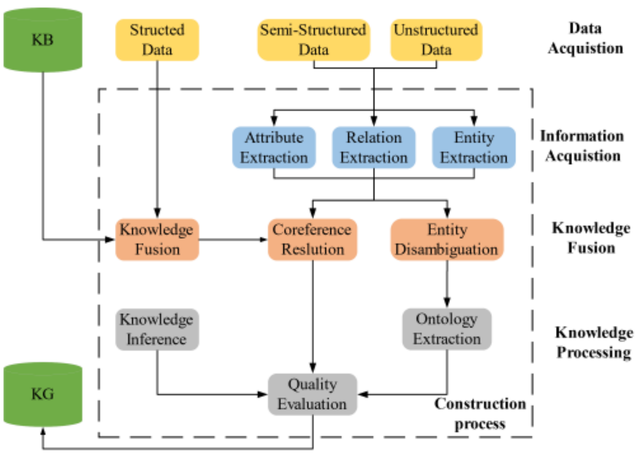

FIGURE 3. The knowledge graph construction framework.

III. CONSTRUCTION OF KNOWLEDGE GRAPH IN THE FIELD OF CROP DISEASES AND PESTS

The data structure of the crop disease and pests knowledge is complex and professional. Compared with the data characteristics in other fields, those in the field of crop pest and disease are as follows:

(1) Data have many kinds and large quantities.

(2) Distribution of resources is scattered and stability is low.

(3) Storage methods, formats, and standards are different.

Constructing knowledge graph of crop pests and diseases still faces great challenges. In this paper, the construction technology of knowledge atlas in the field of crop diseases and insect pests is summarized into four parts: knowledge representation, extraction, fusion, and reasoning. The typical structure of knowledge graph construction is shown in Figure 3. Knowledge extraction extracts the elements of knowledge graph, such as entities and relationships, from a large number of structured or unstructured data of crop pests and diseases through entity recognition, relationship extraction, and attribute extraction technology. Knowledge fusion disambiguates and links the knowledge base and entities of crop pest and disease field, enhances the logic and expressive ability of knowledge, and updates knowledge map for crop pest and disease field by aligning and fusing continuously. Lastly, with the help of ontology reasoning, assistant decision-making, and quality assessment can be automatically completed, data quality can be guaranteed, and the reliability and accuracy of knowledge map in crop pest and disease field can be improved.

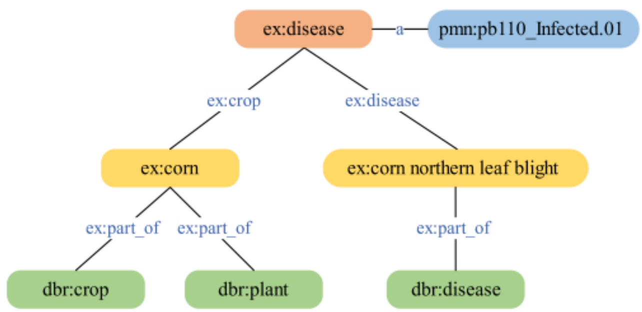

FIGURE 4. RDF diagram model example.

A. KNOWLEDGE REPRESENTATION

1) RDF AND ONTOLOGY

Knowledge representation is a set of conventions for describing the world. It is a process of knowledge symbolization, formalization, and patterns. The early knowledge representation methods used in expert knowledge base include predicate logic representation [13], production representation, frame representation [13], and semantic web representation [14]. With the growth of knowledge and the complexity of relationships, these methods are no longer the main knowledge representation methods because of their limited expressive ability and lack of flexibility. Ontology representation has become the mainstream in which knowledge is expressed in the form of networks, and two related entities are represented by SPO (subject, predicate, object) triple. Resource description framework (RDF) graph model was used to represent the binary relationship between infectious diseases (corn, corn northern leaf blight). The binary relation is represented by the node (ex: disease), whereas the relational phrase is described by the node (pmn:pb110_Infected.01). It represents the interpretation of the semantic relationship between the main actions or events and the disambiguation of entity resources in the lexical database. It also returns the correct nodes in the search. Through relationship management, we can express the executor (ex: corn) and receiver (ex: corn northern leaf blight) of actions via semantic analysis. Finally, entities belonging to the same lexical information unit are associated with the same resource through attributes (ex: part_of).

Ontology is defined as formalization, which is clear and detailed regarding the shared conceptual system. The ontology of crop pests and diseases is a model of representing and organizing knowledge of pests and diseases

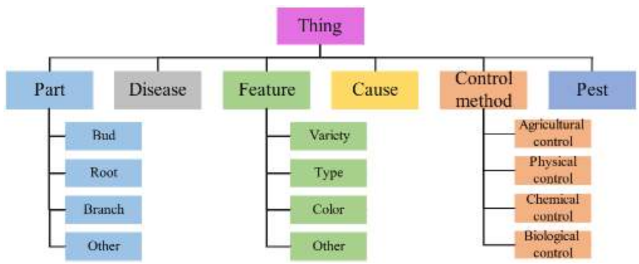

FIGURE 5. **Ontological structure of citrus pests and diseases.**

in machine-understandable formal language [15]. Ontology focuses on the intrinsic characteristics of entities. The main descriptive languages of ontology are RDF, RDF schema (RDFS), and Ontology Web language (OWL). Figure 5 shows a citrus pest ontology structure. The ontology has six data attributes and 12 object attributes to describe the basic information of disease, disease instances, and other class instances.

Ontology allows the effective organization and management of the data layer through the ontology construction of the upper layer. Experts in the field of crop diseases and insect pests have combed and integrated agricultural knowledge and constructed many ontologies of pests and diseases. For example, Wang *et al.* [16] constructed an ontology to organize and manage citrus production knowledge in the hilly areas of Chongqing, China, extracted citrus fertilizer and water ontology from documents and charts of citrus production knowledge, and developed a decision support system for citrus fertilizer and water management based on semantic ontology. Chougule and Mukhopadhyay [17] proposed a method to construct a crop pest ontology in India. Natural processing technology was used to describe the species and cases of pests and diseases, and the ontology was applied to the expert system of pests and diseases. Cañadas *et al.* [18] proposed an ontology management scheme for grape pests and developed a professional web page tool based on this ontology for quality evaluation. Lagos-Ortiz *et al.* [19] proposed an ontology-based decision support system to control pests in sugarcane, rice, soybean, and cocoa crops, to provide guidance for pest diagnosis and prevention.

2) EXPRESSING LEARNING

In recent years, with the development of machine learning, the research focus of knowledge representation has gradually shifted to representation learning. Representational learning refers to the representation of semantic information as dense low-dimensional vectors by means of machine learning, which can effectively solve the problem of data sparsity in the form of triples [20]. Representation learning is generally divided into two categories, including tensor-based decomposition method and mapping-based method. The method based on tensor decomposition is to decompose and code the whole knowledge map into a three-dimensional tensor. The corresponding value of the triple is 1 and the rest is 0. From this tensor, a core tensor and a factor matrix are decomposed.

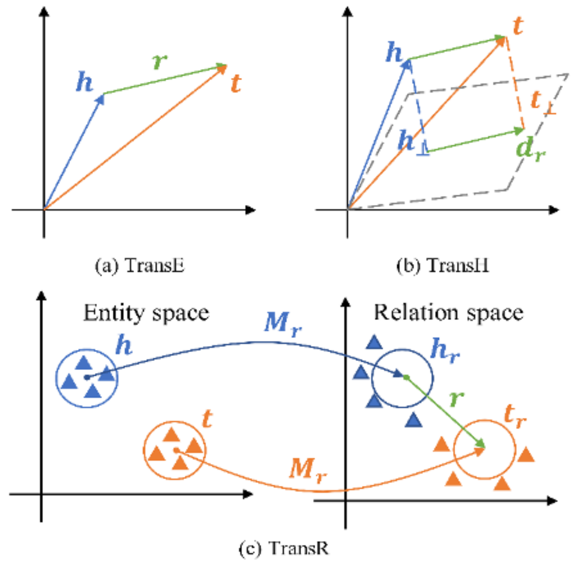

FIGURE 6. **Trans model diagram.**

The result of reduction is regarded as the probability of ternary composition. However, when there are many triple relationships in knowledge map, the dimension of tensor is too high, so this method is not effective for knowledge map with a large number of relationships and sparse knowledge.

Therefore, a method of independent modeling based on triples has emerged. The relationship (r) in each triple (head, relation, tail) is regarded as the translation from head (h) to tail (t). By constantly adjusting h, r and t, the ($h + r$) is as equal as possible to t, that is, $h + r \approx t$. There are two ways to express r, as a matrix or as a vector.

a: METHOD OF REPRESENTING r AS VECTOR

TransE-based translation models and a series of Trans models evolved from TransE are the main methods of expressing r as vectors, such as TransH, TransR, TransD, TransG, and TransE. The model diagram is shown in Figure 6 (a). Triple (h, r, t) needs to be satisfied:

$$d(h, r, t) = \| (h + r) - t \|_2^2 \approx 0 \quad (1)$$

The TransH model is improved on the TransE model by embedding knowledge into the hyperplane. The model is shown in Figure 6 (b). Triple (h, r, t) in TransH needs to be satisfied:

$$d(h, r, t) = \left\| \left(h - \omega_r^T \mathrm{h}\omega_r\right) + d_r - (t - \omega_r^T \mathrm{ht}\omega_r) \right\|_2^2 \approx 0 \quad (2)$$

TransR embeds entities and relationships into different spaces, and implements translation in corresponding relational spaces. For a triple (h, r, t), the mapping matrix M_r is needed to describe the relational space in which the relationship is located. The schematic model is shown in Figure 6 (c), which needs to be satisfied:

$$d\,(h, r, t) = \| h_r + r - t_r \|_2^2 = \| hM_r + r - tM_r \|_2^2 \approx 0 \quad (3)$$

TransD simplifies the mapping matrix of the relation into the product of two vectors based on TransR. $M_{rh} = r_p h_p + I^{m*n}$ and $M_{rt} = r_p t_p + I^{m*n}$ in the graphs represent the matrix mapping entities h and r to the relational space. Triple (h, r, t) needs to be satisfied:

$$d(h, r, t) = \| M_{rh}h + r - M_{rt} \|_2^2 \approx 0 \tag{4}$$

TransG Gauss mixture model uses the Bayesian non-parametric Gauss mixture model to generate multiple translation parts for a relationship and obtains the best part according to the specific semantics of the triple.

b: METHOD OF REPRESENTING r AS A MATRIX

The methods for representing r as a matrix include the use of single layer model (SLM), latent factor (LF) model, and semantic matching energy (SME). SLM defines a single-layer non-linear neural network for each triple as an energy function, takes the entity vector as the input layer, and the relation matrix as the weight parameter of the network. Numerous parameters and calculations involved are its main disadvantages, which are not suitable for sparse knowledge map. Its scoring function is as follows:

$$f(h, r, t) = u_r^T g(M_{r,1} l_h + M_{r,2} l_t) \tag{5}$$

$M_{r,1}, M_{r,2}$ are projection matrices, u_r^T are expression vectors of r, and function g is function tan h.

In SME, entities and relationships are represented by vectors, and all triples share the parameters of the model. It defines two scoring functions, and its linear form is as follows:

$$f(h, r, t) = (M_1 l_h + M_2 l_r + b_1)^T (M_3 l_h + M_4 l_r + b_2) \tag{6}$$

The bilinear form is as follows:

$$f(h, r, t) = (M_1 l_h \otimes M_2 l_r + b_1)^T (M_3 l_h \otimes M_4 l_r + b_2) \tag{7}$$

M_1, M_2, M_3, and M_4 are projection matrices, $\bigotimes$ represents multiplying by bits, b_1 and b_2 re bias vectors. The third-order tensor was also used to replace bilinear form.

LF based on bilinear transformation matrix was proposed to represent the second-order relationship between entities and relationships. The advantages of this method include simplicity and effectivity, good synergy, and low computational complexity. Its scoring function is as follows:

$$f(h, r, t) = l_h^T M_r l_t \tag{8}$$

M_r is the bilinear transformation matrix corresponding to r.

These models consider the collaboration and computational overhead between entities, embedding the relationship into the low-dimensional vector space, and then transforming the vectors or relationships representing entities into corresponding matrices. Moreover, these models introduce evaluation functions to measure the correlation between entities, which provide important reference for knowledge complementation and reasoning. Hamilton *et al.* [21] found that the low-dimensional embedding of nodes can be used for various prediction tasks, ranging from content recommendation to protein recognition. The generation of invisible data was proposed to generate node embedding by utilizing the feature information of nodes. Zuo *et al.* [22] proposed a new model for learning entity attributes and multimedia description of knowledge representation, and verified its effectiveness in knowledge map indexing and multi-model embedding. Hamilton *et al.* [23] introduced the method of embedding a single node and the whole subgraph in detail, considering the recommendation in the field of medicine as an example. Knowledge representation learning remains at the initial stage of research. The main challenges include insufficient research on the representation learning of different knowledge types (triple, tree, and network), multiple information fusion, and complex reasoning models.

B. KNOWLEDGE EXTRACTION

Knowledge acquisition in the knowledge mapping of crop diseases and insect pests extracts entities, relationships, and attributes manually or automatically from semi-structured tables and unstructured texts. Artificial knowledge extraction is usually based on the rules of plant protection experts to collect and collate information about plant diseases and insect pests. The knowledge base of crop diseases and insect pests has been constructed, including crop ontology knowledge base and diagnostic rules knowledge base. Automatic extraction method such as ontology learning technology refers to the use of machine learning technology and natural language processing technology for the knowledge extraction of pests and diseases [24]. Automatic knowledge extraction is the focus of current and future research. This section mainly introduces how to obtain knowledge of crop pests and diseases from agricultural data sources, including entities, relationships, attributes extraction, and common resources.

1) ENTITY RECOGNITION

Entity recognition refers to the automatic recognition of named entities from text, that is, how to recognize meaningful entity information, such as person and place names. Named entity recognition methods are mainly divided into three kinds: methods based on dictionary and rule, statistical machine learning based methods, and hybrid methods [25].

Early entity recognition is based on a dictionary and rules. However, it is time-consuming and laborious to build a dictionary encompassing all kinds of crop disease and pest entities. Dictionary-based methods usually require combination with other methods. Currently, pest dictionaries are mainly integrated into machine learning methods in the form of dictionary features. Additionally, as a vertical field, the method based on rules can increase accuracy. However, this method has obvious limitations, depending on pest and disease experts, writing the rules consumes a lot of manpower, and scalability is poor. Hence, adapting to the changes of data is difficult. In recent years, the use of statistical machine learning to solve the problem of named entity extraction

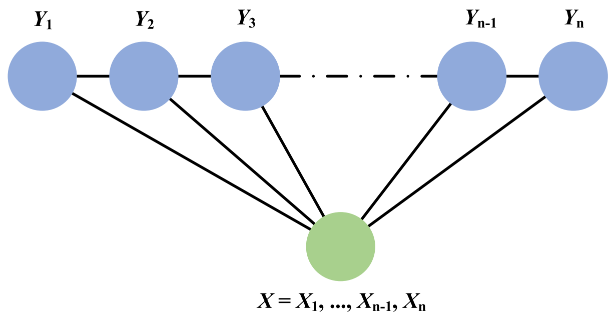

FIGURE 7. Linear Chain conditional random fields.

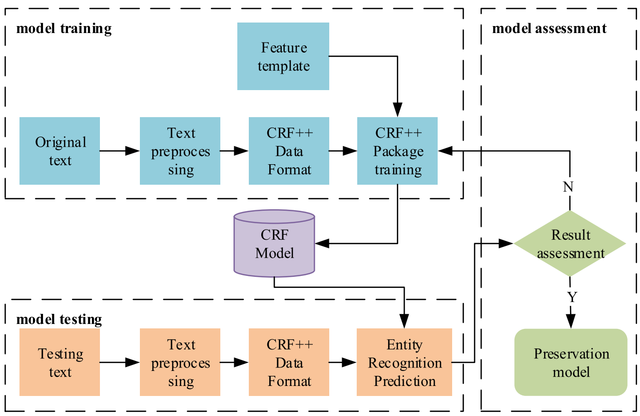

FIGURE 8. CRF model structural diagram.

has become a research hotspot. For example, the commonly used models for named entity extraction include the maximum entropy model (MEM), hidden Markov model (HMM), support vector machine (SVM) [26], and conditional random field (CRF) [27].

CRF model is an undirected graph model with conditional probability of nodes at a given node. Suppose that $X = (X_1, X_2, X_{3,} \ldots, X_n)$ and $Y = (Y_1, Y_2, Y_{3,} \ldots, Y_n)$ is a joint random variable whose structure is shown in Figure 7. If random variable Y constitutes a Markov model represented by graph $G = (V, E)$, its conditional probability distribution is called CRF.

$$P(Y_v|X, Y_w, w \neq v) = P(Y_v|X, Y_w, w \sim v) \quad (9)$$

V is the node set, E is the undirected boundary set, $w \sim v$ represents all nodes connected with node v in graph G, and $w \neq v$ represents all nodes except node v. The general steps of named entity recognition by using CRF method include corpus preprocessing, corpus annotation, feature acquisition, model training, and result evaluation. The step diagram is shown in Figure 8.

CRF model is the most widely used model in entity recognition research. At present, CRF has the best classification effect in general. CRF does not have the strict independence assumption of the HMM production model to link context content features, and feature selection is more flexible. At the same time, CRF model overcomes the shortcomings of the label bias of MEM and other non-generated digraph models. Table 2 lists the research of Chinese agricultural entity recognition based on CRF.

TABLE 2. Comparison of Chinese Agricultural named entity recognition results based on CRF.

Author	Method	Precious	Recall	F1
Li [27]	CRF + characteristics	97.72%	84%	89%
Fang [28]	simple CRF	92.7%	--	--
Wang [29]	CRF + annotation set	94%	81%	93%
Zhang [30]	entities classify	96.7%	88%	86%

The current research on named entity recognition based on CRF mainly focuses on the improvement of annotation set, feature selection, and entity classification. Other areas, such as business and biomedical entity recognition [31], have begun to use the method of combining with in-depth-learning neural networks for entity recognition. In the current research, LSTM/Bi-LSTM [32] is the most frequently combined with the CRF model. At the feature and model levels, it can avoid the problem of relying on a lot of prior knowledge, as well as the problems caused by artificial design features. The combination of CRF and neural network will become a research trend in named entity recognition of crop diseases and insect pests. In the field of crop diseases and insect pests, the difficulties of named entity recognition lie in the uneven quality of data and the high requirement of manual labeling. A special research is being conducted on how to reduce the dependence on data annotation. Its principle is to continuously improve the performance of the model by using massive unlabeled data, learn from small samples, and gradually learn new knowledge by self-exploration to form an interactive learning process.

2) RELATION EXTRACTION

Relation extraction identifies and obtains semantic relationships between entities from a large number of irregular corpus [33]. At present, more research is based on machine learning method to extract relations. Its advantages include the absence of need to construct templates manually, automatic learning of corpus, and strong portability. Relation extraction falls into the three following categories [34]:

(a) A supervised learning method. According to the training data, we designed effective features, learned various classification models, and then used the trained classifier to predict the relationship. However, this method requires manual labeling training corpus, which is usually time-consuming and laborious.

(b) Semi-supervised learning method. Bootstrapping is used to extract relations. First, several seed instances are set manually, and then relational templates and more instances are extracted iteratively from the data.

(c) Unsupervised learning methods. Assuming that entities with the same semantic relationship have similar contextual

TABLE 3. Comparison of Chinese agricultural named entity recognition results.

Author	Content	Method	Artificial demand	Field portability	Accuracy	Application size
Fader [38]	Pattern Matching Based	Pattern Matching	high	weak	high	small
Zelenko [39]	Kernel Based Function	Supervised learning	medium	medium	medium	small
Kambhatla [40]	Logical Regression					
Miller [41]	Syntactic parsing enhancement Based					
Culotta [42]	Conditional random field Based					
Min [43]	Remote Monitoring for Relation Extraction	Semi-supervised Learning	medium	medium	medium	small
Takamatsu [44]	Remote Monitoring Relation Extraction with Improved Entity Alignment Technology					
Yao [45]	Undirected graph model	Unsupervised learning	low	strong	low	large
Surdeanu [46]	Multi-instance and Multi-label					
Santos [47]	Convolutional Neural Network					
Miwa [48]	Bidirectional LSTM and Tree LSTM					
Zhang [49]	RNN-CNN binding					
Zhou [50]	Introducing Attention Mechanism to Relation Extraction Model					

information, each entity uses the corresponding contextual information to represent the semantic relationship of the entity pair and clusters all the semantic relations of the entity pairs. In addition, considering that the accuracy of relationship extraction is generally low (∼50%), relevant scholars have made further improvements, such as the combination of RNN and CNN for relationship extraction model and the introduction of attention mechanism for relationship extraction model. Table 3 lists the research status of various relationship extraction methods.

The extraction of domain relations of crop pests and diseases can be classified into two categories: hierarchical and non-hierarchical semantic relationship extraction of pests and diseases entities. The extraction of hierarchical relationships of pests and diseases entities mainly involves the is-a and is-type-of relationship. The relationship is relatively simple. Due to the unique discipline system and classification in the field of pests and diseases, such relationship extraction can be based on pest dictionaries, encyclopedia knowledge, web crawlers, regular expressions, D2R mapping. and other technologies.

The extraction of non-hierarchical semantic relationships among crop pest entities includes plant diseases-pesticides and disease symptoms-disease types. This extraction includes the text data of pests and diseases, thus, the entity type of pests and diseases is relatively limited. At present, the relationship type to be extracted is usually predefined between two entities, and the extraction task is transformed into a classification problem to be dealt with. For example, Kaushik and Chatterjee [35] defined two types of non-hierarchical relationships in the relation extraction of crop diseases and insect pests, namely, has-synonym (synonymous equivalence) and is-intercrop (intercrop crop). The extraction accuracy can reach 86.89%. The predefinition of the entity relationship generally depends on the pattern map, entity recognition, corpus, and application scenarios in constructing the knowledge map of pests and diseases.

Most non-hierarchical relationship extraction methods in crop pest and disease fields are based on pattern matching. Shuyue [36] constructed the original model of non-classification relationship according to the language model of inverse and symmetry axiom, and then extended the generated model by using lightweight ontology, which was combined with the method of co-occurrence value in statistics. Finally, the axiom of non-classification relationship in the field of crop diseases and insect pests was extracted. However, the general pattern matching method requires experts in the field of crop diseases and insect pests to exhaust the relationship template and improve it to a certain extent to increase accuracy. Ming *et al.* [37] proposed a non-classification relation extraction method for plant ontology and improved the accuracy of non-classification relation extraction by adding vocabulary to the filter and adding restrictions to patterns.

Table 3 shows that first, in the current research of relationship extraction in the field of Chinese crop diseases and insect pests, the introduction of neural network for classification remains in its infancy. Second, the neural network model based on syntactic tree has achieved good results in relation extraction, which shows that the introduction of syntactic information is helpful for relation extraction. Therefore, in the field of crop diseases and insect pests, we can study how to effectively combine multiple syntactic tree information of sentences for relation extraction. Finally, the current relationship extraction of neural networks is mainly used for pre-defined relationship types. Therefore, exploring how to use neural networks in the relation extraction of large-scale massive data and automatically discovering new relationships and their facts are greatly significant.

3) ATTRIBUTE EXTRACTION

Attribute extraction targets crop pest entities, such as pesticide attributes including production certificate number, dosage, and toxicity. The entity can be fully described by attributes. Considering that attribute and attribute value can be regarded as a relationship between entity and attribute value, attribute extraction can be transformed into relation extraction.

4) COMMON RESOURCES

The Chinese Agricultural Thesaurus (CAT) is a large and comprehensive agricultural thesaurus that contains more than 60,000 descriptions and non-descriptions including agriculture, forestry, biology, and other fields, as well as abundant semantic relations among use, pronoun, subordinate, subordinate, and participant. It has played an important role in effectively organizing and utilizing Chinese agricultural information resources. CAT, as the source of data acquisition, is authoritative, scientific, and complete. Qichen and Yawei [51] proposed an automatic generation algorithm of domain knowledge graph based on thesaurus and studied how to extract entity type and relationship type from the internal structure of thesaurus. In addition, CN-DBpedia is the largest Chinese knowledge map of open encyclopedia at present, covering tens of millions of entities and hundreds of millions of relationships. It can be used to extract upper and lower relationships of pest and link pest entities.

The AGROVOC Agricultural Data Set and the Agricultural Thesaurus of the National Agricultural Library (NAL) in the AOS (Agricultural Ontology Service) project sponsored by the Food and Agriculture Organization of the United Nations (FAO) provide ontological structure and terminological support in the field of agriculture. AGROVOC is a multilingual vocabulary maintained by the FAO. It contains more than 32,000 agricultural concepts organized by 25 top-level concepts and is described by 6,192,003 RDF triples. Similar to AGROVOC, the NALT Agricultural Thesaurus includes agricultural terms in English and Spanish. It mainly contains over 90,000 terms in agriculture, biology. and related fields and over 40,000 word-to-word relationships. It has English and Spanish versions, which can be browsed through 17 thematic categories. Table 4 lists the resources commonly used in the construction of agricultural knowledge map

TABLE 4. Mapping of agricultural knowledge construction common resource tables.

KG	KG Size	Resource
AGROVO	32,000 agricultural concepts and 6,192,003 RDF triple descriptions	Food and Agriculture Organization of the United Nations
NAL	More than 90,000 terms and more than 40,000 relations between words	English and Spanish Agricultural Thesaurus
CAT	More than 60,000 descriptions and non-descriptions, semantic relations between words	Chinese Agricultural Thesaurus
CN-Dbpedia	Tens of millions of entities, hundreds of millions of relationships	Knowledge Workshop Laboratory of Fudan University

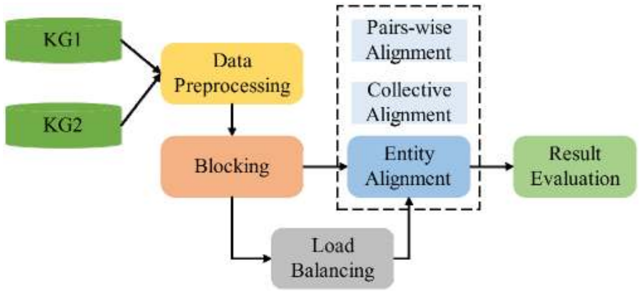

FIGURE 9. Knowledge fusion structure diagram.

5) KNOWLEDGE FUSION

In the process of building knowledge map, data should be fused from different sources, making it necessary to fuse knowledge, also known as knowledge base alignment [52]. The flow chart is shown in Figure 9. The main content of knowledge fusion is entity link. Entity link detects word sequences in a given text and identify entities in a given knowledge map. In this process, two challenges are encountered, including synonym and polysemy [25]. To solve the problem of synonyms, we need to match an entity with different names, such as abbreviations, spelling variants, and nicknames. Polysemy is solved by entity disambiguation. Two main methods are named entity disambiguation based on probabilistic topic model and ranking learning. For the field of crop diseases and insect pests, a few studies on entity disambiguation are available.

Multi-referential problem exists in crop pest entities in different data sources, that is, co-referential resolution problem, also known as entity matching, entity alignment. For example, maize leaf spot disease is also called stripe disease, coal sheath disease, leaf blight or leaf spot disease. Entity alignment algorithms are divided into paired entity alignment and collective entity alignment methods [53]. It is mainly used to eliminate inconsistencies, such as entity conflicts and unknown directions in heterogeneous data, and to form high-quality knowledge. The alignment method of paired entities mainly aims at the similarity of entity attributes but does not consider the relationship between entities. Generally, the entity alignment problem based on attribute similarity score is transformed into classification problem according to the similarity evaluation of entities [54]. Common classification methods include classification regression tree algorithm, ID3 decision tree algorithm, SVM classification method, ensemble learning framework, and hierarchical graph model [55]. The methods of calculating attribute similarity include editing distance and set similarity, such as Cosine distance, Jaccard coefficient, and TF-IDF based on vector similarity. The formula is as follows:

$$sim_{cosine}(e_1, e_2) = \frac{|A(e_1) \cap A(e_2)|}{\sqrt{A(e_1)||A(e_2)}} \tag{10}$$

$$sim_{Jaccard}(e_1, e_2) = |\frac{A(e_1) \cap A(e_2)}{A(e_1) \cup A(e_2)}| \tag{11}$$

e_1 and e_2 are given entities, and $A(e)$ represents the attribute string of entity e.

$$sim_{TF-IDF} = tf_{i,j} \times idf_i \tag{12}$$

Among them, $tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$, $idf_i = \log \frac{|D|}{1+|\{j:t_i \in d_j\}|}$.

The attribute and structural similarity of entity pairs are considered in the calculation of entity similarity by collective entity alignment. The similarity functions of two entities e_1 and e_2 are defined as follows:

$$sim(e_1, e_2) = (1-\alpha) sim_{Attr}(e_1, e_2) + \alpha sim_{NB}(e_1, e_2) \tag{13}$$

$sim_{Attr}(e_1, e_2)$ is the attribute similarity function corresponding to the entity pair, $sim_{NB}(e_1, e_2)$ is the structural similarity function corresponding to the entity pair, and $0 \leq \alpha \leq 1$ is their adjustment parameter.

Collective entity alignment is divided into local and global collective entity alignment [56]. The typical algorithms of local collective entity alignment are vector space model and cosine similarity calculation, which have low accuracy but good recall rate and running speed. Global collective entity alignment adjusts the similarity between entities through the interaction of different matching decisions. It is divided into two methods: collective entity alignment based on similarity propagation and probability model. The method based on similarity propagation generates new matches iteratively by bootstrapping initial matches. The probabilistic model-based approach builds complex probabilistic models for entity matching and decision-making, including relational Bayesian network model, Latent Dirichlet Allocation assignment model, and Markov logic network model. It can improve the matching effect, but the efficiency needs to be improved. With the expansion of the scale of knowledge base and the increase of the number of entities, increasing attention has been paid to the entity alignment in knowledge base. How to align entities accurately and efficiently is one of the key points of future knowledge fusion research.

6) KNOWLEDGE REASONING

Knowledge reasoning extracts hidden information from existing knowledge. Knowledge reasoning focuses on the selection and application of knowledge and methods, minimizes manual participation, deduces missing facts, and completes problem-solving. Moreover, knowledge reasoning methods can be divided into three categories: knowledge-based, logic-based, and in-depth learning methods [57]. The method of relational reasoning based on knowledge representation is similar to that in knowledge representation learning, that is, knowledge is represented as a low-dimensional vector or matrix, and then a series of matrix changes are employed

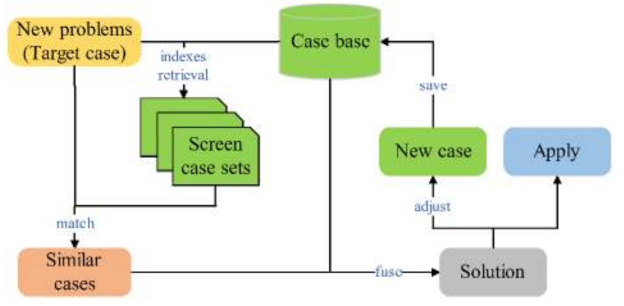

FIGURE 10. Case-based reasoning process diagram.

for relational reasoning. The methods based on logical rules include description logic (DL), rule-based reasoning (RBR) and case-based reasoning (CBR).

For complex entity relations, descriptive logic can be used for reasoning. Knowledge reasoning based on descriptive logic generally includes TBox and ABox, where TBox is a set of axioms used to describe the relationship between concepts and relationships, and ABox is a set of axioms used to describe specific facts, known as assertions. Rule-based reasoning, such as reasoning based on production rules, is generally expressed as IF *condition* THEN*action*. *condition* are sets of conditions, and *action* are sequences of actions. When conditions are satisfied, corresponding rules are triggered. Kamsu-Foguem *et al.* [58] proposed a decision-making method combining concept maps with semantic web, which uses rule-embedding to carry out automatic knowledge reasoning. Lecue and Wu [59] used rule-based reasoning on knowledge graph to predict large-scale abnormal costs and extracted all prediction rules through the association mining of semantic description. Finally, the scalability, accuracy, and consistency of prediction were achieved. Case-based reasoning refers to reasoning from matching in case base. For example, in the field of crop diseases and insect pests, Juan [60] constructed a case-based reasoning model of tobacco mosaic disease based on knowledge map and proposed a set of applicable prevention and a control scheme of tobacco mosaic disease. The main variables of environmental indicators are: average temperature, minimum temperature, average relative humidity, average rainfall, maximum rainfall, average sunshine hours, maximum soil moisture of 10 cm, average soil moisture of 20 cm, maximum soil temperature. The process of CBR is shown in Figure 10.

Traditional reasoning based on logical rules promotes the automation process of reasoning for pest diagnosis and control measures to a certain extent, but it experiences obvious defects, such as insufficient learning ability, low data utilization rate, and inaccurate rate to be improved, which do not meet the requirements of practical application. Considering the increasing data on diseases and insect pests, problems are inevitable, such as missing information and prolonged time. Therefore, a relational reasoning method based on deep learning is employed. For example, the combination of KNN network and CBR can improve the efficiency and accuracy of relational reasoning in reasoning failure situations [61], [62].

In the knowledge graph of crop diseases and insect pests, knowledge reasoning is applied to the assistant Crop Planting Expert system and the diagnosis and control of crop diseases and insect pests, such as the construction of diagnostic model and intelligent expert system of Junzi orchid diseases and insect pests [63], wheat water saving expert system [64], and maize disease diagnosis system based on rule-based reasoning [65]. At present, even for the same disease, plant protection experts make different diagnoses according to the actual disease situation, that is, the construction of knowledge inference engine of pests and diseases must deal with many repeated and contradictory information, and this step increases the complexity of the knowledge inference model of pests and diseases. Knowledge reasoning based on in-depth learning method should be investigated in future research.

IV. APPLICATION OF THE KNOWLEDGE GRAPH OF CROP DISEASES AND INSECT PESTS

Knowledge graph provides a more effective way for the expression, organization, management and utilization of massive, heterogeneous, and dynamic large data information, which improves the level of intelligence and is closer to human cognitive thinking. At present, knowledge graph technology is mainly used in crop pest expert system, intelligent semantic search, crop pest knowledge question and answer.

A. EXPERT SYSTEM

Many expert systems have been early-developed for the diagnosis and control of crop diseases and insect pests, but only few have been popularized and applied, because the amount of knowledge in knowledge base is insufficient. Most of them can only realize the query function of knowledge. Essentially, they are only an information publishing system, and the intelligent range and decision-making ability of experts are difficult to surpass [66]. Traditional expert systems for crop diseases and pests have weak decision-making and reasoning abilities. For example, most expert systems for crop cultivation are knowledge- and rule-based, which cannot be closely integrated with crop growth models. The reasoning process of human experts can hardly be imitated. They only use the shallow knowledge of the system but lack dynamic prediction function and reasoning interpretation [67]. The expert system based on knowledge atlas technology solves the fundamental problem of the lack of correlation between crop pest knowledge and plays a great role in crop pest knowledge retrieval and visualization.

The application of knowledge map in crop pest expert system has promoted the development of precision agriculture. For example, Rajendra *et al.* [68] proposed a framework- and rule-based knowledge representation method to determine whether peanut plants are infected by certain diseases. Nascimento *et al.* [69] developed a hand-held diagnostic tool for pests and diseases by using expert knowledge to assist the diagnosis of important pests in commercial teak. Mendes *et al.* [70] constructed a variable-speed irrigation control expert system for the rational fertilization and irrigation of crops. Rahim *et al.* [71] used uncertain rule reasoning to construct plant diseases and insect pests expert system, provided crop management strategy, formulated suitable fertilizer/pesticide/herbicide program, and obtained reliable and accurate pesticide application suggestions. Kalita *et al.* [72] developed a rice expert system in which users or farmers input problems or diseases that occur throughout the life cycle of rice plants and then identified possible diseases by comparing them with regular knowledge bases. Ballot *et al.* [73] used case matching to construct an expert system and proposed a model to simulate continuous crop yield. Considering winter wheat and broad bean, the farm soil, weather, economic, and social data were combined to provide the most valuable agricultural guidance for farmers, and to evaluate the long-term sustainable operation of the system. Damos [74] used ontology and semantic knowledge representation technology to classify pests and developed a pest expert system to simulate the ability of pests and diseases to affect crops. Babalola *et al.* [75] reviewed the development of combining pest and disease models with crop models and proposed a modular method for pest and disease diagnosis modeling. Lasso and Corrales [76] used machine learning and knowledge mapping technology based on the analysis of crop conditions by using imprecise graph matching to establish pest and disease prediction model for the early warning of pest diseases.

The knowledge map of crop diseases and insect pests not only provides data, model, and knowledge products, but also uses its own advantage of knowledge relevance to meet the needs of decision support. The expert system based on the knowledge map of crop diseases and insect pests is highly comprehensive, accurate, and scalable. In the future, more crop pest expert systems based on machine learning technology will provide a highly accurate decision support for crop pest prediction and diagnosis.

B. INTELLIGENT SEMANTIC SEARCH

Faced with massive data, intelligent semantic search in the field of crop diseases and insect pests establishes a large-scale knowledge base of diseases and insect pests, retrieves entities from the knowledge maps of diseases and insect pests, and returns entity pairs and attribute values associated with entities, thus expanding the query results and improving the search methods and accuracy [77]. At present, most search engines at home and abroad have been improved to semantic search based on knowledge map. Semantic search based on knowledge graph technology enables computers to truly understand user's needs and provide accurate answers rather than related link sequences [78]. When a user enters a query question, the system first processes the sentence, including entity recognition, syntactic analysis, and semantic analysis. At the same time constructs the corresponding grammar analysis tree and then submits the processing results to the knowledge search and distribution module, finds the search template matching the processing results, and clarifies the user's search intention. After identifying the search intention,

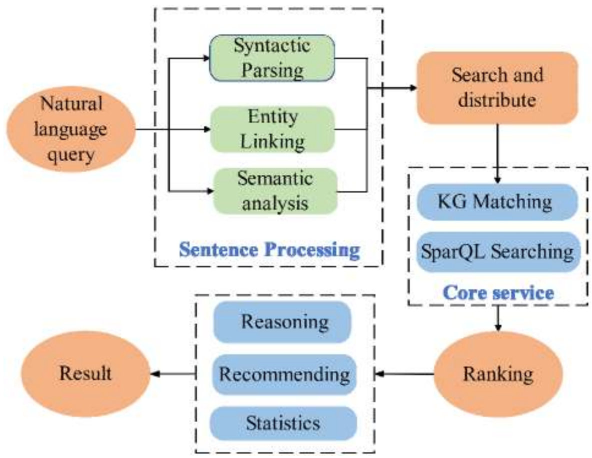

FIGURE 11. Search engine framework based on knowledge graph.

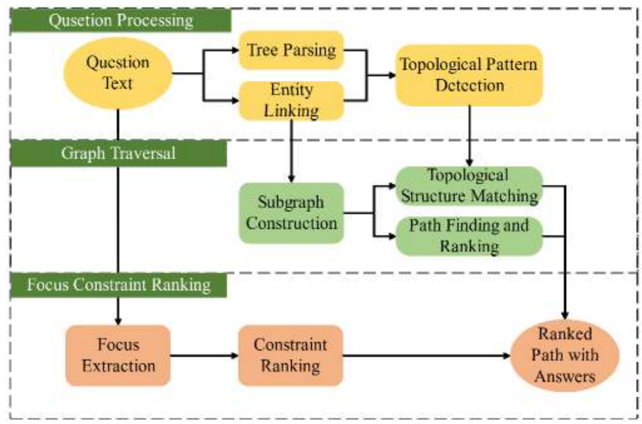

FIGURE 12. Question and answering system architecture based on knowledge graph.

it is forwarded according to the intention. Finally, the processing results of requests are integrated, sorted, and recommended. The search engine framework based on knowledge map is shown in Figure 11. Blondet *et al.* [66] proposed an advanced search engine based on knowledge mapping technology to search and list different types of biomedical entities, including genes, diseases, drugs, targets, and transcription factors associated with user queries with fast response.

Google took the lead in incorporating knowledge graph into search engines. The main domestic agricultural search engines are ''Nongsou Search'', ''Huanong Online'', and ''Sounong Search''. Nongsou Search has 6 million agricultural cooperative websites, and achieves an intelligent search engine of full-text and semantic retrieval. Huanong Online uses natural language semantics analysis technology to realize information-processing application and realizes vertical search application in agricultural industry. The search engine for Sounong Search provides information retrieval of supply and demand, price, market dynamics, agricultural technology, video, news, diseases, insects, and weeds.

C. QUESTIONS AND ANSWERING SYSTEM

The knowledge map technology of crop diseases and insect pests can be applied to a new generation of agricultural system model and knowledge question-and-answer products, and can help accelerate agricultural innovation and meet the demand for the automatic acquisition of agricultural knowledge. At present, the methods used in question answering system mainly include the following:

(a) Information-retrieval-based approach. It is a shallow semantic analysis method based on keyword matching and information extraction, which uses question information and knowledge base resources to obtain candidate answers.

(b) Semantic parsing-based approach. Natural language questions are parsed into a logical form of expression through which the answers can be found from the knowledge base.

(c) Modeling method based on vector space. Vector space is used to describe natural language questions and entities and relationships in knowledge maps. Question answering models are generated by machine and in-depth learning. Question answering system based on knowledge map maps the expression of natural language parsing into the vocabulary of knowledge map elements, which enhances the performance and expansibility of question answering system [79]. Figure 12 shows the flow chart of QA system based on knowledge map.

Hu *et al.* [80] joint disambiguation method is added to the question answer system based on a knowledge map, which improves query accuracy and query performance. Zhang *et al.* [81] proposed the architecture of deep learning and the application of end-to-end variational learning in knowledge-map-based question answering system, which can deal with multi-hop reasoning problems. Abdi *et al.* [82] integrated natural language processing, ontology, and information retrieval technology to develop knowledge question answering systems in physical fields. Yih *et al.* [83] applied entity link technology and deep convolution neural network model matching question and predicate sequence in question answering system, and greatly improved the system performance.

Questions and answers in the field of crop diseases and insect pests are no longer limited to knowledge matching, but also involve determining accurate answers after understanding deeper questions [84]. Lacasta *et al.* [85] combined the strategy of pesticide application, knowledge map, and intelligent question-and-answer to accurately recommend pest and disease application program. In other fields, Chen *et al.* [86] proposed the design and development of an intelligent question answering system for agrotechnical knowledge based on a knowledge atlas, which showed the interrelation of agrotechnical knowledge related to questions through knowledge cards, atlases, and related links; additionally, it improved the knowledge acquisition and enhanced the interactivity. Chen *et al.* [86] constructed the knowledge atlas of TCM, including disease, syndrome, symptom, Chinese herbal medicine, and prescription databases, and conducted questions and answers and the assistant diagnosis and treatment of TCM based on the knowledge atlas. In the future, knowledge question-answering based on representation learning will become the main research direction of knowledge -map-based question-answering system [87]. It transforms entities, relationships, and questions into numerical vectors in low-dimensional semantic space through representation learning technology. By using numerical computation, it directly matches the most similar answers to users' question semantics. Subsequently, in-depth question answering system oriented to knowledge base will become a research hotspot.

V. CONCLUSION

With the development of information technology, data in the field of crop diseases and insect pests have been accumulated. Constructing knowledge graph can improve the management, the extent of sharing and application of crop pest and diseases knowledge. The development of intelligent agriculture and precision agriculture is greatly significant and is also a future research hotspot. In this paper, the background, technology, and application of knowledge map of crop diseases and insect pests were summarized from the perspective of the application of knowledge map in crop diseases and insect pests. The main challenges faced by the knowledge map of crop diseases and insect pests were summarized, and future research directions were prospected. The combination of knowledge atlas and pest knowledge will promote the automation and intellectualization of crop pest and disease related systems. Many attempts have been made to study the knowledge map of crop diseases and insect pests, but they were not perfect and in-depth, and require further research. In the future, the knowledge map of crop diseases and insect pests can be deeply studied from the following aspects:

(a) implement the overall solution to improve the quality of knowledge maps;

(b) explore how the knowledge map can be extended and the general method of knowledge map construction;

(c) implement and establish the evaluation criteria of the knowledge atlas in the field of pests and diseases; and

(d) achieve fast and accurate construction of large-scale domain knowledge map.

With the emergence of large-scale networks, the expansion and automation of knowledge map will be the trend of future developments in this field.

REFERENCES

[1] H. Arnaout and S. Elbassuoni, "Effective searching of RDF knowledge graphs," *J. Web Semantics*, vol. 48, pp. 66–84, Jan. 2018.

[2] C. B. Wang, X. Ma, J. Chen, and J. Chen, "Information extraction and knowledge graph construction from geoscience literature," *Comput. Geosci.*, vol. 112, pp. 112–120, Mar. 2018.

[3] P. K. Thornton *et al.*, "Site selection to test an integrated approach to agricultural research for development: Combining expert knowledge and participatory geographic information system methods," *Int. J. Agricult. Sustainability*, vol. 4, no. 1, pp. 39–60, 2006.

[4] M. M. Yusof, N. F. Rosli, M. Othman, R. Mohamed, and M. H. A. Abdullah, "M-DCocoa: M-agriculture expert system for diagnosing cocoa plant diseases," in *Recent Advances on Soft Computing and Data Mining* (Advances in Intelligent Systems and Computing), vol. 700, R. Ghazali, M. Deris, N. Nawi, and J. Abawajy, Eds. Cham, Switzerland: Springer, 2018.

[5] Y. Yao and W. Xu, "Construction of ontology knowledge base for water and soil conservation measures suitability evaluation," in *Journal of Fuzhou University* (Natural Science Edition), 2016, pp. 188–195.

[6] W. D. van Eeden, J. P. de Villiers, R. J. Berndt, W. A. J. Nel, and E. Blasch, "Micro-Doppler radar classification of humans and animals in an operational environment," *Expert Syst. Appl.*, vol. 102, pp. 1–11, Jul. 2018.

[7] B. Stewart-Koster *et al.*, "Expert based model building to quantify risk factors in a combined aquaculture-agriculture system," *Agricult. Syst.*, vol. 157, pp. 230–240, Oct. 2017.

[8] Z. Xu, Y. P. Sheng, L. R. He, Y. F. Wang, "Review on knowledge graph techniques," *J. Univ. Electron. Sci. Technol. China*, vol. 45, no. 4, pp. 589–606, 2016.

[9] L. Qiao, L. Yang, D. Hong, L. Yao, and Q. Zhiguang, "Knowledge graph construction techniques," *J. Comput. Res. Develop.*, vol. 53, no. 3, pp. 582–600, 2016.

[10] T. Yu *et al.*, "Knowledge graph for TCM health preservation: Design, construction, and applications," *Artif. Intell. Med.*, vol. 77, pp. 48–52, Mar. 2017.

[11] J. L. Martinez-Rodriguez, I. Lopez-Arevalo, and A. B. Rios-Alvarado, "OpenIE-based approach for knowledge graph construction from text," *Expert Syst. Appl.*, vol. 113, pp. 339–355, Dec. 2018.

[12] X. Zhang, X. Liu, X. Li, and D. Pan, "MMKG: An approach to generate metallic materials knowledge graph based on DBpedia and Wikipedia," *Comput. Phys. Commun.*, vol. 211, pp. 98–112, Feb. 2017.

[13] C. Dong, "Design and implementation of forest management knowledge service system based on frame representation," *J. Zhejiang A F Univ.*, vol. 34, no. 3, pp. 491–500, 2017.

[14] Z. Shaohua, "Object-oriented representation for owl ontology knowledge base," *Softw. Eng. Appl.*, vol. 07, pp. 132–141, 2018.

[15] J. Dou, J. Qin, Z. Jin, and Z. Li, "Knowledge graph based on domain ontology and natural language processing technology for Chinese intangible cultural heritage," *J. Vis. Lang. Comput.*, vol. 48, pp. 19–28, Oct. 2018.

[16] Y. Wang, Y. Wang, J. Wang, Y. Yuan, and Z. Zhang, "An ontology-based approach to integration of hilly citrus production knowledge," *Comput. Electron. Agricult.*, vol. 113, pp. 24–43, Apr. 2015.

[17] A. Chougule, V. K. Jha, and D. Mukhopadhyay, "Adaptive ontology construction method for crop pest management," in *Proc. Int. Conf. Data Eng. Commun. Technol.*, 2016, pp. 665–674.

[18] J. Cañadas, I. M. del Águila, and J. Palma, "Development of a web tool for action threshold evaluation in table grape pest management," *Precis. Agricult.*, vol. 18, pp. 974–996, Dec. 2017.

[19] K. Lagos-Ortiz *et al.*, "An ontology-based decision support system for insect pest control in crops," in *Proc. Int. Conf. Technol. Innov.* Cham, Switzerland: Springer, 2018.

[20] N. Guan, D. Song, and L. Liao, "Knowledge graph embedding with concepts," *Knowl.-Based Syst.*, vol. 164, pp. 38–44, Jan. 2019.

[21] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," Tech. Rep., 2017.

[22] Y. Zuo, Q. Fang, S. Qian, X. Zhang, and C. Xu, "Representation learning of knowledge graphs with entity attributes and multimedia descriptions," in *Proc. IEEE 4th Int. Conf. Multimedia Big Data (BigMM)*, Sep. 2018, pp. 1–5.

[23] W. L. Hamilton, R. Ying, and J. Leskovec, "Representation learning on graphs: Methods and applications," *IEEE Data Eng. Bull.*, to be published.

[24] W. J. Vlietstra *et al.*, "Automated extraction of potential migraine biomarkers using a semantic graph," *J. Biomed. Informat.*, vol. 71, pp. 178–189, Jul. 2017.

[25] G. Zhu and C. A. Iglesias, "Exploiting semantic similarity for named entity disambiguation in knowledge graphs," *Expert Syst. Appl.*, vol. 101, pp. 8–24, Jul. 2018.

[26] Z. Zhou, J. Yang, Y. Qi, and Y. Cai, "Support vector machine and back propagation neutral network approaches for trip mode prediction using mobile phone data," *IET Intell. Transp. Syst.*, vol. 12, no. 10, pp. 1220–1226, 2018.

[27] L. Xiang, W. Xiaohong, J. Lu, C. Xin, L. Lei, and Z. Yan, "Recognition of crops,diseases and pesticides named entitiesin chinese based on conditional random fields," *Trans. Chin. Soc. Agricult.*, vol. 48, no. S1, pp. 178–185, 2017.

[28] F. Ying, "Research on application of C-CRF model on named entity recognition of the agricultural product," *Guangdong Agricult. Sci.*, vol. 38, pp. 197–199, 2011.

[29] C. Wang and F. Wang, "Study on recognition of Chinese agricultural named entity with confitional random fields," *J. Agricult. Univ. Hebei*, vol. 37, no. 1, pp. 132–135, 2014.

[30] Z. Jian *et al.*, "Chinese agricultural named entity recognition based on conditional random fields," *Jisuanji Yu Xiandaihua*, no. 1, pp. 123–126, 2018.

[31] T. H. Dang, H.-Q. Le, T. M. Nguyen, and S. T. Vu, "D3NER: Biomedical named entity recognition using CRF-biLSTM improved with fine-tuned embeddings of various linguistic information," *Bioinformatics*, vol. 34, no. 20, pp. 3539–3546, 2018.

[32] R. Kadari, Y. Zhang, W. Zhang, and T. Liu, "CCG supertagging via bidirectional LSTM-CRF neural architecture," *Neurocomputing*, vol. 283, pp. 31–37, Mar. 2018.

[33] V. Devisree and P. C. R. Raj, "A hybrid approach to relationship extraction from stories," *Procedia Technol.*, vol. 24, pp. 1499–1506, 2016.

[34] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, "A review of relational machine learning for knowledge graphs," *Proc. IEEE*, vol. 104, no. 1, pp. 11–33, Jan. 2016.

[35] N. Kaushik and N. Chatterjee, "Automatic relationship extraction from agricultural text for ontology construction," *Inf. Process. Agricult.*, vol. 5, no. 1, pp. 60–73, 2018.

[36] M. Shuyue, "Axioms extraction method of the non-taxonomic relationship in crop diseases and insect pests," *Comput. Digit. Eng.*, vol. 43, no. 10, pp. 1746–1750, 2015.

[37] Z. Ming, D. Yaru, D. Huifang, Z. Jiajun, W. Hongshuo, and C. Ying, "Research on ontology non-taxonomic relations extraction in plant domain knowledge graph construction," *Trans. Chin. Soc. Agricult. Machinery*, vol. 47, no. 9, pp. 9–38, 2016.

[38] A. Fader, S. Soderland, and O. Etzioni, "Identifying relations for open information extraction," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2011, pp. 1535–1545.

[39] D. Zelenko, C. Aone, and A. Richardella, "Kernel methods for relation extraction," *J. Mach. Learn. Res.*, vol. 3, pp. 1083–1106, Feb. 2003.

[40] N. Kambhatla, "Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations," in *Proc. ACL*, 2004, pp. 1–4.

[41] S. Miller, H. Fox, L. Ramshaw, and R. Weischedel, "A novel use of statistical parsing to extract information from text," in *Proc. HLT-NAACL*, 2006, pp. 1–8.

[42] A. Culotta, A. McCallum, and J. Betz, "Integrating probabilistic extraction models and data mining to discover relations and patterns in text," in *Proc. Hum. Lang. Technol. Conf. North Amer. Chapter ACL*, 2006, pp. 296–303.

[43] B. Min, R. Grishman, L. Wan, C. Wang, and D. Gondek, "Distant supervision for relation extraction with an incomplete knowledge base," in *Proc. NAACL-HLT*, 2013, pp. 1–6.

[44] S. Takamatsu, "Reducing wrong labels in distant supervision for relation extraction," in *Proc. ACL*, 2012, pp. 721–729.

[45] L. Yao, S. Riedel, and A. McCallum, "Collective cross-document relation extraction without labelled data," in *Proc. EMNLP*, 2010, pp. 1013–1023.

[46] M. Surdeanu, J. Tibshirani, R. Nallapati, and C. D. Manning, "Multi-instance multi-label learning for relation extraction," in *Proc. ACL-HLT*, 2011, pp. 455–465.

[47] C. N. D. Santos, B. Xiang, and B. Zhou, "Classifying relations by ranking with convolutional neural networks," *Comput. Sci.*, vol. 86, no. 86, pp. 132–137, 2015.

[48] M. Miwa and M. Bansal, "End-to-end relation extraction using LSTMs on sequences and tree structures," Tech. Rep., 2016.

[49] X. Zhang, F. Chen, and R. Huang, "A combination of RNN and CNN for attention-based relation classification," *Procedia Comput. Sci.*, vol. 131, pp. 911–917, 2018.

[50] P. Zhou, J. Xu, Z. Qi, H. Bao, Z. Chen, and B. Xu, "Distant supervision for relation extraction with hierarchical selective attention," *Neural Netw.*, vol. 108, pp. 240–247, Dec. 2018.

[51] H. Qichen and Z. Yawei, "Initial seed set automatic generation algorithm for domain knowledge graph based on thesaurus," *J. Chin. Inf. Process.*, vol. 32, no. 8, pp. 1–8, 2018.

[52] W. Liu *et al.*, "Representation learning over multiple knowledge graphs for knowledge graphs alignment," *Neurocomputing*, vol. 320, pp. 12–24, Dec. 2018.

[53] A. Annane, Z. Bellahsene, F. Azouaou, and C. Jonquet, "Building an effective and efficient background knowledge resource to enhance ontology matching," *J. Web Semantics*, vol. 51, pp. 51–68, Aug. 2018.

[54] G. Zhu and C. A. Iglesias, "Sematch: Semantic similarity framework for knowledge graphs," *Knowl.-Based Syst.*, vol. 130, pp. 30–32, Aug. 2017.

[55] X. Xue and X. Yao, "Interactive ontology matching based on partial reference alignment," *Appl. Soft Comput.*, vol. 72, pp. 355–370, Nov. 2018.

[56] M. Chen, Y. Tian, K. W. Chang, S. Skiena, and C. Zaniolo, "Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment," Tech. Rep., 2018.

[57] F. Yang, Z. Yang, and W. W. Cohen, "Differentiable learning of logical rules for knowledge base reasoning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 2319–2328.

[58] B. Kamsu-Foguem, F. H. Abanda, M. B. Doumbouya, and J. F. Tchouanguem, "Graph-based ontology reasoning for formal verification of BREEAM rules," *Cognit. Syst. Res.*, vol. 55, pp. 14–33, Jun. 2019.

[59] F. Lecue and J. Wu, "Explaining and predicting abnormal expenses at large scale using knowledge graph based reasoning," *J. Web Semantics*, vol. 44, pp. 89–103, May 2017.

[60] W. Juan, "Research of tobacco disease prevention and control model based on case study reasoning and knowledge map," M.S. thesis, College Comput. Sci., Anhui Agricult. Univ., Hefei, China, 2016.

[61] D. López-Sánchez, J. M. Corchado, and A. G. Arrieta, "A CBR system for image-based webpage classification: Case representation with convolutional neural networks," in *Proc. 13th Int. Florida Artif. Intell. Res. Soc. Conf.*, 2017, pp. 1–6.

[62] S. Y. Sharma, D. P. Roy, and B. Chakraborty, "Health services support system (HESS): An ontology supported CBR system for cognitive diagnosis applied on HIV/AIDS detection," in *Proc. 2nd Int. Conf. Comput. Sustain. Global Develop.*, Mar. 2015, pp. 284–290.

[63] S. Hao, "Clivia plant diseases and pests' diagnosis model building and the realization of intelligent expert system," M.S. thesis, Tianjin Univ., Tianjin, China, 2015.

[64] C. Liping, Y. Maishun, and L. Xiaodong, "Reasoning mechanism of wheat irrigating expert system for rational use of water," *Comput. Eng. Appl.*, vol. 40, pp. 216–219, 2004.

[65] M. Bo, L. Junchen, and W. Na, "Design and realization of the maize disease diagnosis system based on image rules," *J. Maize Sci.*, vol. 17, no. 6, pp. 132–135, 2009.

[66] G. Blondet, J. Le Duigou, and N. Boudaoud, "A knowledge-based system for numerical design of experiments processes in mechanical engineering," *Expert Syst. Appl.*, vol. 122, pp. 289–302, May 2019.

[67] J. Gu, H. Wui, and X. Sun, *Realization of the Cross-Platform Agriculture Expert System With Java and JESS*, C. Zhao, Ed. ISIITA, 2007, pp. 323–327.

[68] L. R. Haidar, I. M. I. Subroto, and A. Marwanto, "Expert system on soybean disease using knowledge representation method," *J. Telematics Inform.*, vol. 5, no. 1, pp. 36–46, 2017.

[69] D. A. Nascimento, R. M. Anunciação, A. Arnhold, A. C. F. Filho, A. D. Santos, and J. C. Zanuncio, "Expert system for identification of economically important insect pests in commercial teak plantations," *Comput. Electron. Agricult.*, vol. 121, pp. 368–373, Feb. 2016.

[70] W. R. Mendes, F. M. U. Araújo, R. Dutta, and D. M. Heeren, "Fuzzy control system for variable rate irrigation using remote sensing," *Expert Syst. Appl.*, vol. 124, pp. 13–24, Jun. 2019.

[71] R. Rahim, N. N. D. S, and M. Zamroni, "Pest plant disease prevention and identification system using certainty factor method," *Int. J. Eng. Technol.*, vol. 7, pp. 763–765, 2018.

[72] H. Kalita, S. K. Sarma, and R. D. Choudhury, "Expert system for diagnosis of diseases of rice plants: Prototype design and implementation," in *Proc. Int. Conf. Autom. Control Dyn. Optim. Techn.*, Sep. 2016, pp. 723–730.

[73] R. Ballot *et al.*, "First cropping system model based on expert-knowledge parameterization," *Agronomy Sustain. Develop.*, vol. 38, p. 33, Jun. 2018.

[74] P. Damos, "Modular structure of web-based decision support systems for integrated pest management. A review," *Agronomy Sustain. Develop.*, vol. 35, pp. 1347–1372, Oct. 2015.

[75] A. Babalola, O. Abiodun, F. Aderanti, and A. John, "Development of a Web based expert system for managing pests and diseases of moringa oleifera," in *Engineering and Information Technology*. CSEIT, 2018.

[76] E. Lasso and J. C. Corrales, "Towards an alert system for coffee diseases and pests in a smart farming approach based on semi-supervised learning and graph similarity," in *Proc. Int. Conf. ICT Adapting Agricult. Climate Change*. Cham, Switzerland: Springer, 2017.

[77] M. Arenas, B. C. Grau, E. Kharlamov, Š. Marciuška, and D. Zheleznyakov, "Faceted search over RDF-based knowledge graphs," *J. Web Semantics*, vols. 37–38, pp. 55–74, Mar. 2016.

[78] J. Johansson, M. Contero, P. Company, and F. Elgh, "Supporting connectivism in knowledge based engineering with graph theory, filtering techniques and model quality assurance," *Adv. Eng. Informat.*, vol. 38, pp. 252–263, Oct. 2018.

[79] D. Diefenbach, V. Lopez, K. Singh, and P. Maret, "Core techniques of question answering systems over knowledge bases: A survey," *Knowl. Inf. yst.*, vol. 55, no. 3, pp. 529–569, 2017.
[80] S. Hu, L. Zou, J. X. Yu, H. Wang, and D. Zhao, "Answering natural language questions by subgraph matching over knowledge graphs," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 5, pp. 824–837, May 2018.
[81] Y. Zhang, H. Dai, Z. Kozareva, A. J. Smola, and L. Song, "Variational reasoning for question answering with knowledge graph," in *Proc. 22nd AAAI Conf. Artif. Intell.*, 2017, pp. 1–8.
[82] A. Abdi, N. Idris, and Z. Ahmad, "QAPD: An ontology-based question answering system in the physics domain," *Soft Comput.*, vol. 22, pp. 213–230, Jan. 2018.
[83] W. Yih, M. Chang, X. He, and J. Gao, "Semantic parsing via staged query graph generation: Question answering with knowledge base," in *Proc. Joint Conf. 53rd Annu. Meeting ACL 7th Int. Joint Conf. Natural Lang. Process. AFNLP*, 2015, pp. 1–12.
[84] M. Dubey, D. Banerjee, D. Chaudhuri, and J. Lehmann, "EARL: Joint entity and relation linking for question answering over knowledge graphs," in *Proc. Int. Semantic Web Conf.*, 2018, pp. 108–126.
[85] J. Lacasta, F. J. Lopez-Pellicer, B. Espejo-García, J. Nogueras-Iso, and F. J. Zarazaga-Soria, "Agricultural recommendation system for crop protection," *Comput. Electron. Agricult.*, vol. 152, pp. 82–89, Sep. 2018.
[86] L. Chen, J. Gao, Y. Yuan, and L. Wan, *Agricultural Question Classification Based on CNN of Cascade Word Vectors*. Cham, Switzerland, 2018, pp. 110–121.
[87] K. Liu, Y.-Z. Zhang, G.-L. Ji, S.-W. Lai, and J. Zhao, "Representation learning for question answering over knowledge base: An overview," *Acta Automatica Sinica*, vol. 42, no. 6, pp. 807–818, 2016.

LIU XIAOXUE received the B.S. degree in computer science and technology from Shandong Agricultural University, Shandong, China, in 2015. She is currently pursuing the M.S. degree in computer science and technology with the College of Information and Electrical Engineering, China Agricultural University, Beijing, China. Her research interests include natural language processing and knowledge engineering.

BAI XUESONG received the B.S. degree in software engineering from Tianjin Polytechnic University, Tianjin, China, in 2016. He is currently pursuing the M.S. degree in computer technology with the College of Information and Electrical Engineering, China Agricultural University, Beijing, China. His research interests include computer vision and machine learning.

WANG LONGHE received the M.S. degree in computer science from China Agricultural University, in 2009. He is currently with China Agricultural University, mainly responsible for informatization construction of national research facility for phenotypic and genotypic analysis of model animals informatization construction.

REN BINGYUAN received the M.S. degree in plant protection from Nanjing Agricultural University, in 2014. He is currently with the National Agricultural Technology Extension and Service Center, mainly responsible for crop diseases and insect pest control.

LU SHUHAN is currently pursuing the B.S. degree in computer and information science from The Ohio State University, Columbus, OH, USA. Her research interests include computer vision and machine learning.

LI LIN is currently a Professor and a Doctoral Supervisor with China Agricultural University. Her main research interests include knowledge engineering and machine learning.