Check for updates

Knowledge graph for integration and quality traceability of agricultural product information

Rongzhi Jing[1,2]* and Ping Li[2]

[1]College of Telecommunications and Intelligent Manufacturing, Sias University, Zhengzhou, Henan, China, [2]Henan Intelligent Manufacturing Digital Twin Engineering Research Center, Zhengzhou, Henan, China

The quality traceability of agricultural products is an important and challenging is issue due to it involves multiple activities carried out by farmers and other stakeholders who must manage multiple heterogeneous data sources collected and generated through a variety of applications, services and devices. An agricultural product information integration and quality traceability model is proposed based on knowledge graph (KG). Through modeling, extraction, fusion, storage, calculation, reasoning and application of knowledge of agricultural products, KG is constructed to integrate the knowledge of agricultural production and monitoring process. In KG, nodes represent agricultural products and their foods, and their related specific things or their property values, lines connect to the agricultural products, food, and their properties, forming network, each node and the relationship has uniqueness, the KG of agricultural products and food is stored in the Neo4j database in graph structure, which is clear and easy to reuse. The 100 batches of carrot pesticide residue test results show that the proposed method is feasible for quality traceability of agricultural products. It provides a technical approach for the quality and safety management of agricultural products "from field to Table".

KEYWORDS

quality and safety of agricultural products, quality traceability of agricultural products, knowledge graph, information fusion, multi-source and heterogeneous information

1 Introduction

The quality and safety of agricultural products have been widely concerning, as they impact people's lives and property safety. This issue involves vast amounts of diverse and heterogeneous information across production, processing, storage, transportation, and sales processes. However, agricultural products and food safety problems occur from time to time in countries around the world, such as the "poisoned eggs" in United States in 2010, "poisoned cucumber" in Germany in 2011, "horse meat" scandal in Europe in 2013, and various issues with agricultural products and food in China. These issues include pesticide overdoses, Sudan red duck eggs, poisonous cadmium rice, poisonous bean sprouts, melamine milk powder, avian influenza, illegal cooking oil, clenbuterol, bulking agent, sodium formaldehyde sulfoxylate, and plasticizers. These events and issues seriously harm the interests of consumers and the social public trust. It has affected the international reputation and even threatened life safety (Galvez et al., 2018; Corallo et al., 2020; Hou et al., 2020). It is estimated that more than 200 million foodborne diseases occur in China every year due to agricultural product safety problems. The frequent occurrence of agricultural product

safety incidents makes all sectors of society face the threat of food safety. Governments of all countries are strengthening the supervision of agricultural product quality. The traceability of agricultural product quality has become an important research in the field of artificial intelligence. Traditional quality traceability methods and systems mainly rely on the unique product identification number, barcode, two-dimensional code, and RFID tag (Aung and Chang, 2014; Zhao et al., 2019). Consumers can get some information by scanning these codes. However, these methods and systems can easily be copied and transferred to fake, and consumers generally have to passively accept the information provided by businesses rather than verify it for themselves. Based on the Internet of Things, supervised and controlled by the government and relevant departments of enterprises, all rely on the central database. However, no matter which party supervises the central database, the data security of the database depends on institutional constraints. Traceability labels for agricultural products are not unique, and tampering with the central database for private use still occurs frequently. Therefore, the existing traceability system cannot track the whole process of agricultural production and processing in real time. Merchants may completely falsify the traceability information of the entire agricultural product, and the problem of information asymmetry has not been improved. Due to the long supply chain of agricultural products, the transportation and production process is relatively scattered, and the relevant information is multi-source and heterogeneous. The complicated benefit relation between the merchants (suppliers) and consumers causes information asymmetry, which increases the difficulty of overall traceability of agricultural products and food. Because the agricultural industrial chain is long, including production, processing, storage, inspection and marketing, and other complex chains, it is difficult to fully monitor these complex chains, and consumers cannot effectively distinguish the quality of agricultural products and food. The lack of an effective way to obtain more accurate information leads to fraud (Talpur et al., 2012; Zheng et al., 2020; Nasca et al., 2021). The problems existing in the agricultural product quality traceability system are summarized as follows:

1. Traceability information contains a large amount of unstructured data, which is not standardized.
2. Supported varieties are relatively single, have poor compatibility, and are difficult to achieve dynamic expansion.
3. The identification method is single and the application cost of the enterprise is high, so it is difficult to promote.

According to different chains of transportation, production, and processing of agricultural products, quality traceability can be divided into three stages: before, during, and after processing, including pesticide and veterinary drug residues, heavy metal residues, food additives, microorganisms and nutrition, quality, and other indicators (Aik et al., 2020; Yang et al., 2020; Yuan et al., 2020), as shown in Table 1. Our country sets each detection index of agricultural products and corresponding detection technology and detection equipment; each index is a quantitative expression that can directly reflect the quality level of agricultural products. Many scholars quantified the original index data, established the corresponding grading standards, obtained the quantitative values of the index, and added various quality detection indices into the quality traceability information to improve the efficiency and accuracy of the quality traceability of agricultural products (Cocco and Mannaro, 2021; Ray et al., 2021).

As seen from Table 1, the quality traceability of agricultural products involves many problems such as multiple information and indicators, wide sources, heterogeneity and fragmentation, coexistence of information structure, semi-structure and un-structure, large scale, rapid growth, and multiple redundant indicators (Jundian et al., 2018; Mashru et al., 2020; Yang and Sun, 2020). The rapid development of agricultural information and Internet of Things technologies leads to a large increase in data, which leads to information overload and reduces the efficiency of agricultural engineers in obtaining decision-making knowledge from information. Therefore, the existing traditional methods and systems cannot quickly trace the quality problems of agricultural products. In the process of quality traceability of agricultural products, it is difficult to effectively analyze, mine, correlate, and make full use of these massive multi-source heterogeneous data and information (Carrara et al., 2020; Shen et al., 2020).

From the above analysis, the traceability description of agricultural product quality is described as follows. The key information in the links of agricultural production, processing, supply, and sales is first collected. Consumers, regulators, and tracing businesses can query and visit the information at any time to realize the whole process of agricultural products' traceability management, from raw material to food, the traceability of agricultural production at every link. By scanning the QR code, people can check the traceability information from raw material to the finished food of the production link. In case of food quality problems, precise recall can be carried out according to product batches so that the source can be traced, the whereabouts can be traced, and the responsibility can be investigated. Knowledge graph (KG) can integrate multiple heterogeneous data sources and store data as a graph structure, which is clear and easy to reuse and query (Xiaoxue et al., 2019). KG is an important tool in logical reasoning, knowledge and data fusion, cognition and computing fusion, and complex collaborative decision-making. It aims to effectively organize multi-source heterogeneous and fragmented data by establishing relations between data, making it easier to understand and process data by humans and machines, and providing support for search, mining, and analysis tasks (Lv et al., 2022). Big data KG is often adopted to associate a large amount of data knowledge related to crop and food quality, establish a database and knowledge set, and associate relevant data knowledge with time, place, target, and characteristics. It can be employed to provide the knowledge base representation and realize the mutual exchange of information, databases, file resources, and systems related to agricultural product traceability (Chen et al., 2020; Min et al., 2022). Sun et al. (2017) proposed an agricultural information integration framework based on KG to improve the integration and access efficiency of agricultural information. In the framework, a KG of agricultural products producing and managing was constructed, covering the basic process of planting, farming, processing, quality inspection, warehousing, transportation, and sales. Song et al. (2018) established a quality and safety traceability system of agricultural products based on Multi-Agent. In the system, artificial intelligence, intelligent control, and intelligent detection technologies were used effectively. Through a number of Agent refining divisions of labor, the quality safety early warning and supervision control for the supply chain of agricultural

TABLE 1 Traceability index of agricultural product quality.

Industrial process	Safety hazard	Significance	Judgment	Preventive measures
Raw material receiving	Organisms: pathogenic bacteria, parasites Chemistry: agricultural residues, heavy metals, nitrate Physics: leaves, sand, glass, and other impurities	Yes Yes No	Growth process by microbial contamination, parasitic invasion Improper use of pesticides or excessive heavy metals in soil, etc. Caused by growing, harvesting, or transportation	Control in the sterilization process Agricultural residues heavy Follow-up cleaning and filtering
Cleaning picking	Organisms: pathogenic bacteria, parasites Chemistry: disinfectant, cleaning agent residue Physical: no	Yes Yes No	Cleaning water can cause pollution Caused by improper disinfection, cleaning Residual disinfectant, cleaning agent	
Broken as juice	Biological: no Chemistry: disinfectant, cleaning agent residue Physics: Metal fragments	No Yes No	Cleaning and disinfection are not thorough Metal fragments introduced into the blade of a crusher	SSOP control Filtration before filling can be removed

products was carried out. Chen et al. (2019) proposed an agricultural KG, namely AgriKG, to automatically integrate the massive agricultural data from the internet. It can automatically recognize agricultural entities from unstructured text and link them to form a KG. Zhang et al. (2021) explained the basic concepts, methods, and techniques of KGs and domain KGs and introduced the status quo of research and application of KG in the field of agriculture at home and abroad. The development prospects of agricultural KGs are broad. Qin and Yao (2021) proposed a set of agricultural KG construction methods, calculation frameworks, and practical application systems. They integrated and regularized the existing storage form and representation of knowledge in the agricultural field, which makes the whole system sort out the messy agricultural knowledge and apply it better to assist learning and research. Lin (2022) summarized the development status of agricultural products traceability system at home and abroad, compared and analyzed the advantages and disadvantages of the development of agricultural products traceability system in China, designed agricultural products traceability system with network information flow as the focus, and proposes the agricultural products quality safety traceability supervision mechanism framework based on the flow of network information. Fang and Ge (2023) proposed a traceability system that utilizes blockchain technology to ensure the transparency and integrity of agricultural product supply chain information, including a data collection and storage module, a blockchain-based traceability module, and a user interface module.

Recently, with the development of information and intelligent technology, the quality and safety traceability systems of agricultural products have been significantly boosted. However, it still faces considerable challenges on how to effectively integrate large amounts of fragmented information for downstream applications. This study aims to analyze the process related to the quality and safety of agricultural products and proposes a quality traceability method for agricultural products based on KG. The contributions of this study are as follows.

1. KG is used to organically organize the multi-source heterogeneous and fragmented data related to agricultural product traceability. Through knowledge learning, the data instances of KG are enriched and the semantics of subclass relations are embodied.
2. An agricultural product quality traceability framework is constructed based on KG. This framework has a wide range of application scenarios, such as agricultural market analysis, agricultural enterprise management, and other fields requiring auxiliary decision-making. Further research is also being conducted in heterogeneous data and user search intent.
3. The constructed framework is validated on the carrots' quality traceability.

The rest of the article is organized as follows: in Section 2, the agricultural KG is introduced; Section 3 focuses on the quality traceability of agricultural products by KG in detail; The experiments and analysis are conducted in Section 4; and Section 5 concludes the research work and points out future work.

2 Agricultural KG

KG describes entities, knowledge, and concepts in the real world in a structured form that can be understood by computers, and expresses multi-source heterogeneous information into a form that is closer to the cognitive world of human beings. KG provides people with a better ability to organize, manage, and understand massive information and convert the data into knowledge. KG realizes the cognition and reasoning of multi-source heterogeneous data by establishing the relationship between the multi-source heterogeneous data, mining the relationship between entities, and integrating, fusing, and visualizing multi-source heterogeneous mass data. KG is a graph-based data structure. In KG, each entity may have several different attributes according to the characteristics of the entity; the edge is used

to describe the relationship between the two entities, such as is-a relationship. The basic architecture of KG is shown in Figure 1 (Chen et al., 2020; Lv et al., 2022; Min et al., 2022). KG is constructed using the bottom-up and top-down methods. The basic steps are as follows:

1. Knowledge extraction: Through the designed domain knowledge model, entities, attributes, and entity relationships from the data sources related to the quality of multi-source heterogeneous agricultural products are extracted, and the ontological knowledge expression is formed.
2. Knowledge fusion: Carry out the ontology alignment and entity matching to eliminate the contradictions and ambiguities.
3. Knowledge processing: Evaluate the quality of the fused knowledge, including the ontology construction, knowledge reasoning, and quality evaluation.
4. Knowledge verification: Validate the knowledge quality, including automatic intelligent detection of knowledge errors and conflicts, to ensure the accuracy of KG.
5. Knowledge reasoning: Employ the engine to dig deeply into the knowledge association, connect all knowledge entities into the knowledge facts with complete meaning, and have strong knowledge reasoning ability to deduce the implied knowledge and conclusions, further enriching KG.
6. Finally, the KG knowledge is stored in the graph database.

3 Knowledge embedding and representation learning

3.1 Knowledge embedding

KG embedding aims to embed the entities and their relations in the map into a mathematical space for representation learning; that is, head, tail, or relation are projected into a low-dimensional dense vector space to preserve semantic information in the KG to obtain entity vector representation for calculation and reasoning. Different

KG embedding methods have different knowledge representation vectors due to their different model structures. TransE is the basis of the vectorization of the relation library. According to the word2vec principle, the translation invariant phenomenon of word vector is used. Relation in each triplet is translated from entity *head* to entity *tail*. By constantly adjusting *h, r,* and *t* (vectors of *head, relation,* and *tail*) so that $(h + r)$ is as equal to *t* as possible, that is, $h + r \approx t$, mathematically constrain modeling entities and relations by Eq. (1), mapping them to the same vector space.

$$d(h,t) = \| h + r - t \| \approx 0 \tag{1}$$

TransE can effectively handle simple 1:1 relationships and is suitable for sparse knowledge maps on a large scale. However, its performance is significantly reduced when dealing with complex relationships of 1: *N*, *N*: 1, and *N*: *N*.

TransD is used to embed triplet data for the complex relationship of the knowledge map CroDisKG in this study. TransD contains the entity space and the relational space related by a mapping matrix, which is reduced to the product of two vectors. By using two projection matrices, M_{rh} and M_{rt}, to project the head entity *h* and the tail entity *t* into the relation space, respectively, the modeling of entities and relations is constrained mathematically by Eq. (2).

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

Where $M_{rh} = r_p h_p + I^{m*n}$, $M_{rt} = r_p h_p + I^{m*n}$, r_p, t_p, and h_p are mapping vectors, and I^{m*n} is a unit matrix.

3.2 Knowledge representation learning

A knowledge representation learning model based on adaptive weights is proposed to solve the heterogeneity and imbalance of entities and relationships in agricultural product quality tracking.
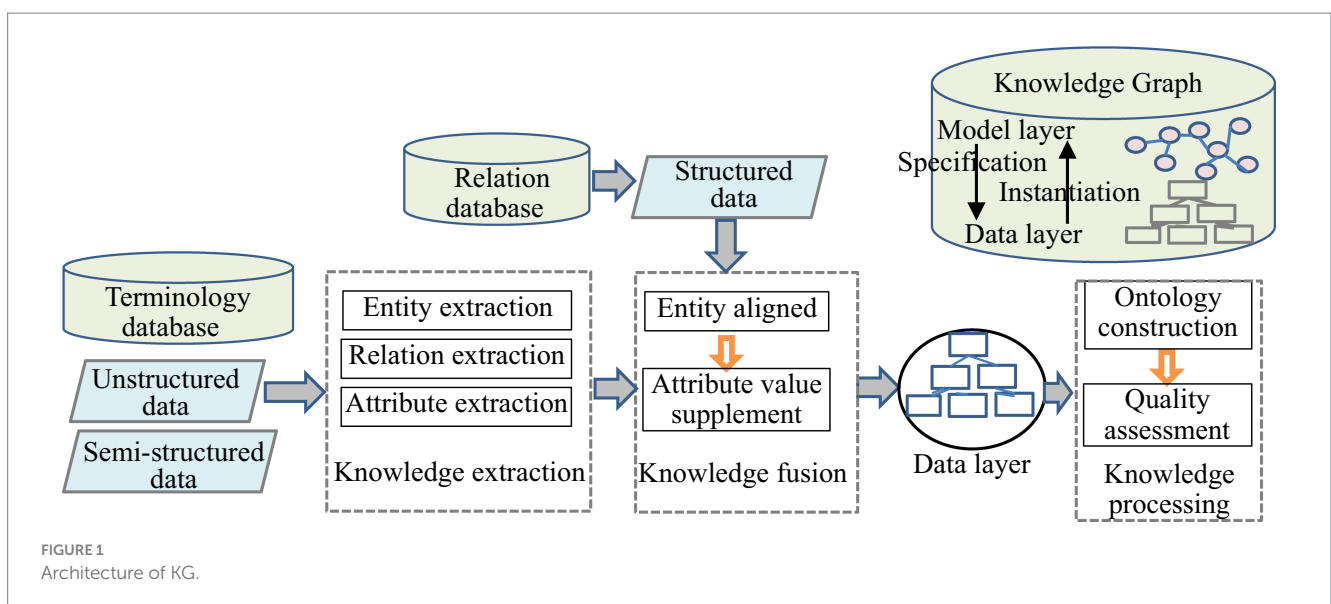

FIGURE 1
Architecture of KG.

Using adaptive weights to balance the fusion of semi-structured, un-structured, and structured information, the total score function is defined as Eq. (3).

$$f(h,r,t) = \lambda f_D(h,r,t) + (1-\lambda) f_S(h,r,t) \qquad (3)$$

Where,

$$f_D(h,r,t) = \| h_d + r - t_d \| + \| h_d + r - M_{rt}t_s \| + \| M_{rh}h_s$$
$$+ r - t_d \|, f_S(h,r,t) = \| M_{rh}h + r - M_{rt}t \|, \lambda \epsilon(0,1)$$
is the weight, h and t are the head and tail entities, respectively, r is the relationship between h and t, h_d and t_d are the representation of the head and tail entities, respectively, based on semi-structured and un-structured information, h_s and t_s are the representation of the head and tail entities, respectively, based on structured information, and M_{rh} and M_{rt} are the projection matrix defined by the head and tail entities, respectively.

Stochastic gradient descent (SGD) is used to minimize the loss function, achieve the optimization goal, and learn the vector representation of each entity and its relationship in KG. The loss function is defined as Eq. (4).

$$L = \sum_{(h,r,t)\epsilon S(h,r,t)} \sum_{(h',r,t')\epsilon S'(h',r,t')}$$
$$\max\left(0,\left[f(h,r,t) + \alpha - f(h',r,t')\right]\right) \qquad (4)$$

Where α is the set boundary value, (h,r,t) is the positive example tuple of the triplet of KG, $f(h,r,t)$ is the score function of the positive example triplet, $S(h,r,t)$ is the set of the positive example triplet, (h',r,t') is the negative example triplet constructed for the immediately replaced head entity h and tail entity t, $f(h',r,t')$ is the score function of the negative example triplet, and $S'(h',r,t')$ is the set of the negative example triplet.

4 Agricultural product quality traceability based on KG

In agricultural product quality traceability, KG can integrate various information sources and gradually draw the outline of the entire process of agricultural product production, transportation, and processing to carry out the quality traceability. An agricultural product quality traceability method is proposed based on KG. Its schematic diagram is shown in Figure 2. The main steps are given as follows:

1. Collect and preprocess the original data.
2. Establish the entities according to the process characteristics of agricultural production, transportation, processing, and sales.
3. Combine the expert knowledge with association computing to establish the relationship between the entities.
4. Construct KG by knowledge extraction and fusion.
5. Store KG in the Neo4j database, which can provide a unified interface for intelligent traceability application of agricultural product quality. In addition, KG of agricultural product quality and safety can learn from existing knowledge and constantly generate new knowledge.

6. The quality traceability system is designed with B/S architecture and is written in Java language. It uses MySQL and Neo4j databases to store data and provides web-based retrieval interfaces. The Web front-end visualization tools, D3.Js and ECharts, are used to visualize the results.

According to KG, the quality of each chain of agricultural products can be analyzed, and information mining can be carried out through graph algorithm and deep graph traversal to detect the interaction between the specific quality problems or the related quality problems and determine whether specific agricultural products have quality problems:

✓ Screen the suspicious chains of agricultural products through specific rules.
✓ Check the key chains that have specific associations with the quality of agricultural products.
✓ Check the network characteristics and chain characteristics of subnets composed of all agricultural product chains that have specific associations with suspicious chains.
✓ Use the shortest path algorithm Dijkstra to analyze what kind of association relations are used to associate specific chains (Min et al., 2022).
✓ Analyze the multiple layers of relationships.

When quality problems occur in agricultural products, KG can be used to trace back to the supplier of processed products or raw materials, and it is also conducive to the timely correction of errors. When there is a production quality problem, the use of KG can be traced back to the specific production time, production line, and staff of the product through the batch number, which is conducive to improving the process and is more conducive to direct recall of the batch of the products with problems, without recalling all products. Through the above analysis, the workload in the investigation process is greatly reduced, and the overall efficiency of the quality traceability of agricultural products is improved.

5 Simulation experiments and analysis

In our experiment lab, a carrot traceability system is constructed in detail (Sun et al., 2017; Jing and Li, 2022). It is enlarged by KG. In this study, a carrot quality traceability system based on KG is constructed in detail. Its original big data resource is introduced as follows: The production information includes carrot seeds, planting, budding, storage, processing, carrot, packaging information, fresh storage information, sales information, five entities that need planting bases, packaging information, transport vehicles, maintenance workers, sales units, regulatory agencies, the name of agricultural products, planting address, picking date, processing type, processing unit and related management, delivery time, delivery address and delivery unit, the dealer, location and time of sale, and so on.

Carrot quality traceability KG is constructed from structured, semi-structured, and unstructured data during carrot production and processing. First, the entity type is set to X = {carrots, planting, carrot food, toxic substance}, then build each entity attribute characteristics, personnel = {name, domicile of origin, unit, educational background, age}, food = {name of food, food

FIGURE 2
The schematic diagram of the proposed method.
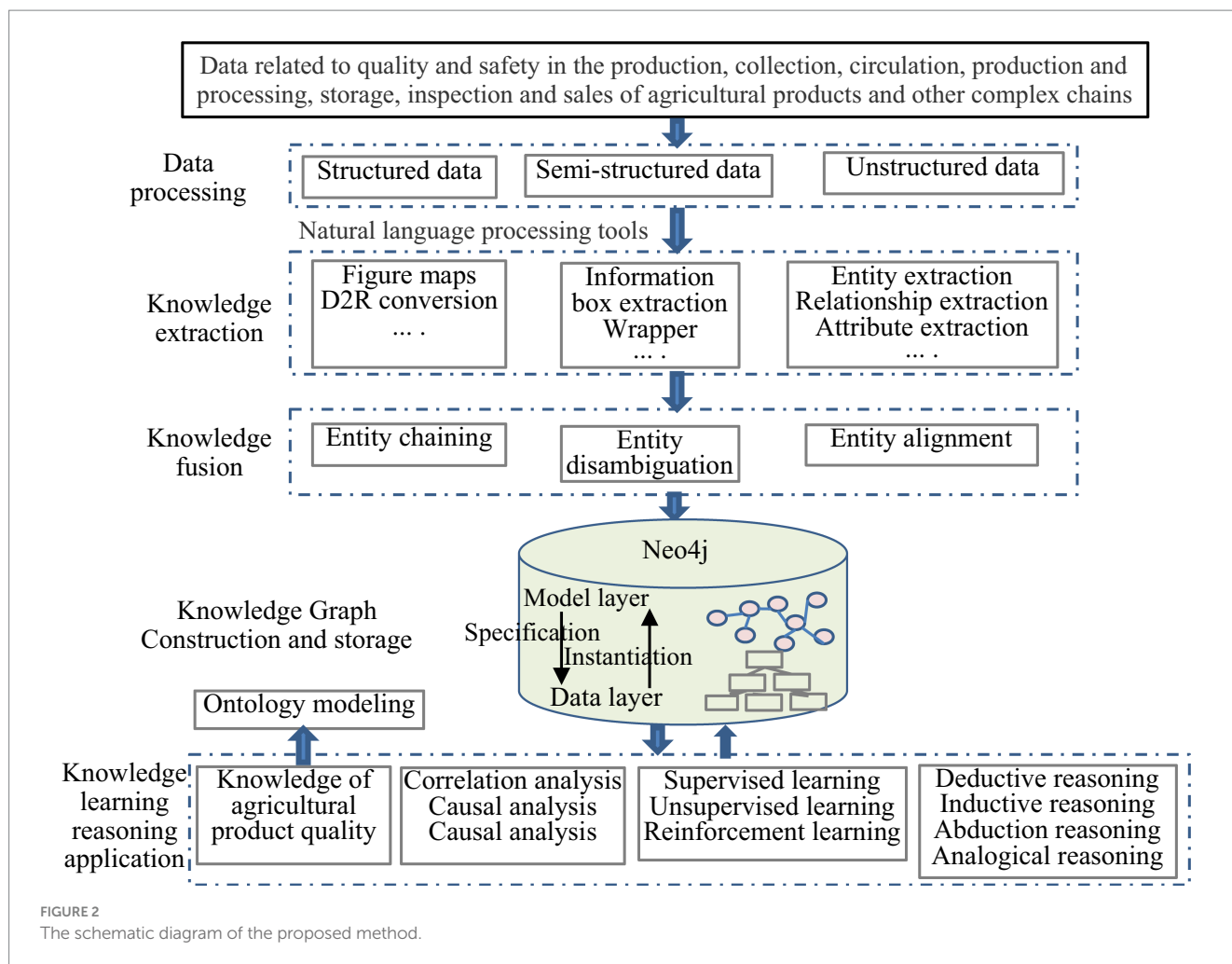
categories, food types, fine food, sampling project}, carrot = {name, food categories, food and class, type of food, subtype of food, sampling items}, and toxic substances = {name, type of toxicity, route of invasion, health hazard}. Therefore, the quality problems corresponding to the quality evaluation index of agricultural products are generated in the specific quality source type of the relationship chain belonging to a quality supervision bureau industry, considering the quality safety of agricultural products mainly exist in the following four situations: production problems, processing problems, storage problems, and sales problems. The quality problem can be set as {Production problems, Processing problems, Storage problems, and Sales problems}.

According to the characteristics of quality problems in the above four situations and the classification of geographical objects in the quality traceability ontology, the types of quality sources are divided according to the standards of production, transportation, processing, and consumption of agricultural products in the industries of different quality supervision bureaus, and the risk sources are divided into point source quality source, non-point source quality source, flow source quality source, and internal source quality source. After establishing the relationship between the production and processing of agricultural products and the quality-related industries corresponding to the agricultural product quality evaluation index, as well as the relationship between their respective categories, it is

necessary to establish the correlation between the traceability industry of agricultural product quality and the traceability category of agricultural products. Based on the established food safety entities, attributes, and relationships, Interactive Graph, an open-source project developed by Java Script, was used to construct the carrot quality traceability KG. A web-based interactive manipulation framework is provided for large-scale graph data triples. The interactive graph provides three basic features: Graph Explorer, Graph Navigator, and Rel Finder. KG data visualization and knowledge reasoning can be easily realized.

To improve the accuracy of the quality evaluation index matching the quality safety industry and quality source type in the relationship chain of agricultural products generation, the correlation between the level 1 index of the quality security department category and quality source type is established. Through the characteristic quality evaluation index of agricultural products in various industries, the "quality safety" attribute type is added to the quality assurance industry, and the industry priority sequence of quality problems is constructed according to the safety status and processing situation of agricultural products in various industries over the years. The case graph of the correlation between industry category and quality source type of agricultural products is constructed. The experiments are carried out with carrot quality traceability. Figure 3 shows the simple KG of carrot.

FIGURE 3
Traceability of carrot KG.

The process chains of carrots and their production and processing are divided into sub-stages, and the spatial distribution structure of the upper chain is used to construct the relationship according to the connection relationship between the trunk and branch chain nodes and the trunk and branch chain in the source-sink diagram. After dividing the molecular stage, the trunk and branch nodes are represented in the form of "P+ code," the trunk and branch are represented in the form of "L+ code," and the sub-stages are represented in the form of "S+ code," according to the codes of each trunk and branch in the upper chain system. In the graph, the trunk and branch segments are still represented in the form of nodes. Therefore, the complete trunk and branch segments are represented as "P+ code" + "L+ code" + "P+ code." To complete the construction of the case graph of agricultural products, the connection relationship between the dry and branch chain nodes and the dry and branch chain segments is established, and the chain relationship between the dry and branch chain segments and the corresponding sub-stages is established, and the case graph of source-sink relationship is completed. On this basis, build a set of quality traceability systems. The system architecture and traceability data processing process are shown in Figure 4. Quality traceability of carrots requires correlation analysis among seed, water, fertilizer, and pesticide information used in carrot production and processing, carrot quality and safety information, and discrimination of the source of quality deviation. Because the KG data storage structure has the characteristics of topology, the graph calculation method is used to trace the quality.

The KG is used to trace the batch of carrot planting personnel, origin and planting place, environment, fertilizer, disease and insect control, processing place and time, inspection certificate number, preservation, inspector, supermarket, grade, salesman, and other information to realize the carrot industrial chain quality and safety

traceability of the whole process. The collection, input, traceability, and transmission of traceability information run through the whole traceability system, in which information collection is the basis, information input is the power, information traceability is the core, and information transmission is the result. In each chain of carrot planting, production, processing, transportation, and sales, the traceability information is collected comprehensively, then the collected information is selected, and the selected data are input into the system. The system obtains new alternative data through the reasoning strategy of the inference machine, and then it is input into the system.

The above steps are repeated inside the system until the inference machine has no new optional content input to the system; that is, the inference machine is finished or has been interrupted. After the above process, the system continuously searches the carrot tracing information so that the range of the information result is constantly reduced until the query result is finally given. The relevant conditions to be queried are input into the query system, clicking the query to get the desired result set. The traceability information of carrot quality can be listed in the query result. Carrot quality can be checked through KG at any time to query the function by entering the query page to view the detailed information on fresh products, browsing the carrot planting time, personnel, plant, and pesticide information used in the pest control during planting.

To verify the feasibleness of the constructed carrot quality traceability system, 100 batches of carrot pesticide residue test results are used to perform the simulation experiments. The data are collected from the carrot quality inspection department, including 40 unqualified batches. The 100 batches of sample production origin positions are input into the quality tracking system to obtain the corresponding environmental information, and the carrot quality
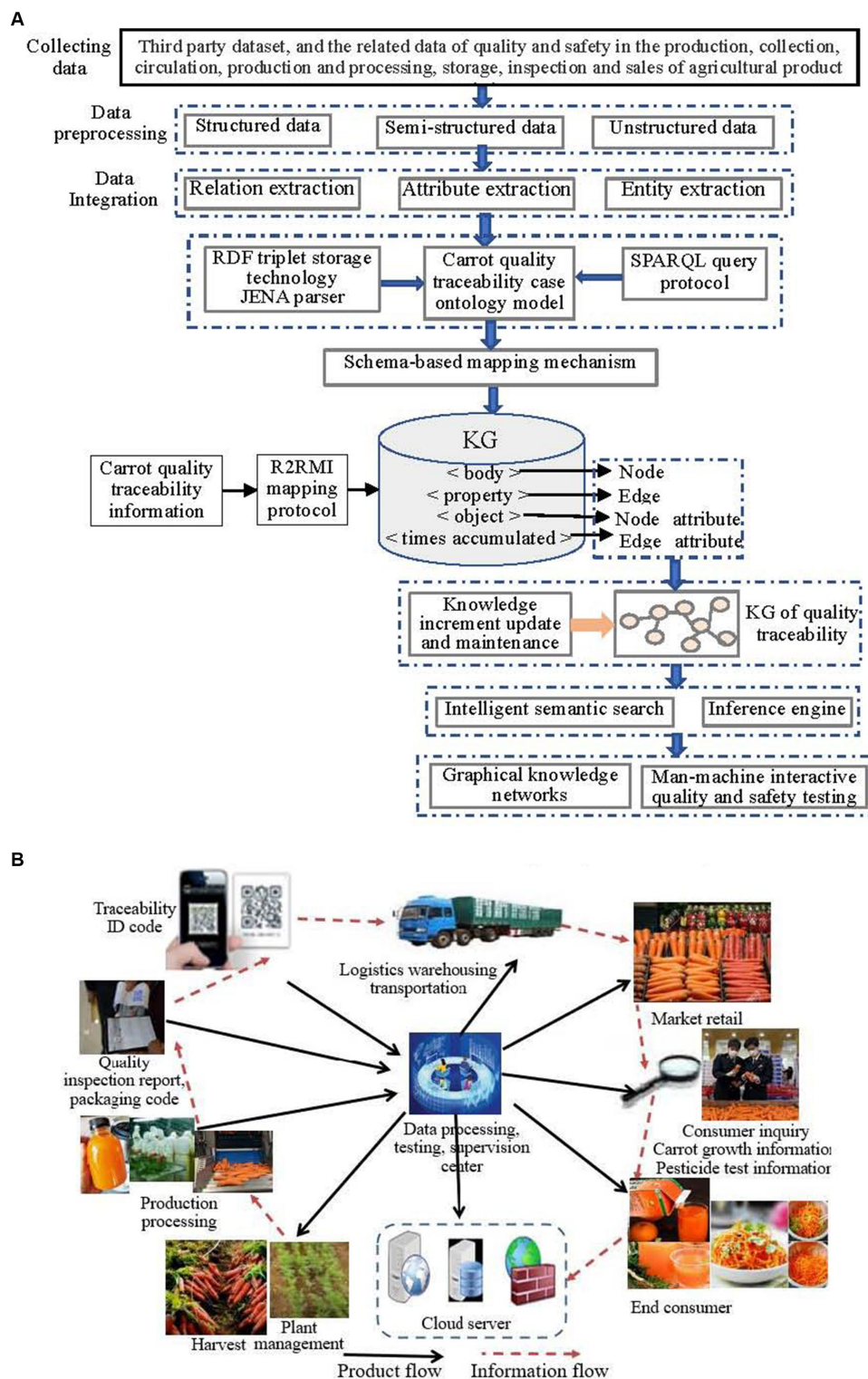
FIGURE 4
Quality traceability system of carrot. **(A)** System architecture of carrot. **(B)** Carrot traceability data processing process.

safety of carrot production can be judged according to the carrot production technical regulations. The judgment results of this system are compared with the test results of the testing department and the subjective judgment results of the carrot quality traceability system, and the comparison results are shown in Table 2.

It can be seen from Figure 4 and Table 2 that the carrot quality traceability system based on subjective judgment cannot trace the unqualified batches. The reason is that subjective judgment can only rely on the sensory cognition or experience of the place of origin. The retrospective system in this study detected 25 unqualified

TABLE 2 Carrot quality test results by three methods.

Comparative object	Sample batch	Non-conforming batch	Judgment
Test results of agricultural products	100	38	Excessive pesticide residue
Judgment results of a vegetable traceability system	100	0	Subjective judgment
Our quality traceability system results	100	25	Pesticide residues of origin

batches, with an effective rate of 75% compared to the carrot test results.

6 Conclusion and discussion

The quality and safety of agricultural products are always a problem that governments of all countries attach great importance to. It is important to research how to find an effective supervision method for domestic agricultural products to ensure the quality and safety of agricultural products without inhibiting the healthy development of agriculture. This study proposes a knowledge graph (KG) based on agricultural product information integration and quality traceability methods. First, the collection range of product quality and safety traceability information is determined. Then, the ontology of agricultural product traceability information and knowledge map are constructed. Finally, an agricultural product quality traceability system is constructed based on KG. Taking a carrot supply chain as an example, the system is tested, and the results show that the system is effective in tracking the whole supply chain and detecting anomalies. Traceability systems can provide consumers with reliable information about the origin and quality of agricultural products and can also help producers and distributors improve their supply chain management. The next step is to study the application of the semantic model of agricultural product quality safety traceability in the actual traceability process, realize the information sharing of agricultural product quality traceability information in the agricultural product supply chain, and improve the accuracy and efficiency of the agricultural product quality traceability system.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

RJ: Data curation, Formal analysis, Methodology, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. PL: Conceptualization, Data curation, Investigation, Methodology, Software, Supervision, Writing – review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aik, J., Turner, R. M., Kirk, M. D., Heywood, A. E., and Newall, A. T. (2020). Evaluating food safety management systems in Singapore: a controlled interrupted time-series analysis of foodborne disease outbreak reports- science direct. *Food Control* 117:107324. doi: 10.1016/j.foodcont.2020.107324

Aung, M. M., and Chang, Y. S. (2014). Traceability in a food supply chain: safety and quality perspectives. *Food Control* 39, 172–184. doi: 10.1016/j.foodcont.2013.11.007

Carrara, G. R., Burle, L. M., Medeiros, D., de Albuquerque, C. V. N., and Mattos, D. M. F. (2020). Consistency, availability, and partition tolerance in blockchain: a survey on the consensus mechanism over peer-to-peer networking. *Ann. Telecommun.* 75, 163–174. doi: 10.1007/s12243-020-00751-w

Chen, Y., Kuang, J., Cheng, D., Zheng, J., Gao, M., and Zhou, A. (2019). Agri KG: an agricultural KG and its applications. *Lect. Notes Comput. Sci* 11448. doi: 10.1007/978-3-030-18590-9_81

Chen, Z., Wang, Y., Zhao, B., Cheng, J., Zhao, X., and Duan, Z. (2020). Knowledge graph completion: a review. *IEEE Access* 8, 192435–192456. doi: 10.1109/ACCESS.2020.3030076

Cocco, L., and Mannaro, K. (2021). Blockchain in Agri-food traceability systems: A model proposal for a typical Italian food product. IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER).

Corallo, A., Latino, M. E., Menegoli, M., and Striani, F. (2020). The awareness assessment of the Italian Agri-food industry regarding food traceability systems. *Trends Food Sci. Technol.* 101, 28–37. doi: 10.1016/j.tifs.2020.04.022

Fang, L., and Ge, H. (2023). Research on traceability of agricultural product supply chain information. *Acad. J. Sci. Technol.* 5, 126–127. doi: 10.54097/ajst.v5i1.5470

Galvez, J. F., Mejuto, J. C., and Simal-Gandara, J. (2018). Future challenges on the use of blockchain for food traceability analysis. *TrAC Trends Anal. Chem.* 107, 222–232. doi: 10.1016/j.trac.2018.08.011

Hou, D. Y., O'Connor, D., Igalavithana, A. D., Alessi, D. S., Luo, J., Tsang, D. C. W., et al. (2020). Metal contamination and bioremediation of agricultural soils for food safety and sustainability. *Nat. Rev. Earth Environ.* 1, 366–381. doi: 10.1038/s43017-020-0061-y

Jing, R., and Li, P. (2022). Quality control system of red jujube by hybrid model: development of an efficient framework. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.888978

Jundian, S., Bingrong, D., Liwen, J., et al. (2018). Data governance collaborative method based on blockchain. *J. Computer Appl.* 38, 2500–2506. doi: 10.11772/j.issn.1001-9081.2018030594

Lin, J. (2022). Agricultural product quality and safety supervision mechanism based on information traceability system. *Front. Comput.* 747, 595–601. doi: 10.1007/978-981-16-0115-6_66

Lv, Z., Yi, K., Zhou, W., Fei, M., and Shen, Y. (2022). A review of the knowledge extraction technology in knowledge graph. 41st Chinese Control Conference (CCC). 2022: 4211–4218.

Mashru, H., Kabra, N., and Mohan, K. (2020). A systematic framework for state channel protocols identification for blockchain-based IoT networks and applications. IEEE International Conference on Communications Workshops, Dublin.

Min, W., Liu, C., Xu, L., and Jiang, S. (2022). Applications of knowledge graphs for food science and industry. *Patterns* 3:100484. doi: 10.1016/j.patter.2022.100484

Nasca, B. J., Bilimoria, K., and Yang, A. D. (2021). Quality and safety in surgery: challenges and opportunities. *Jt. Comm. J. Qual. Patient Saf.* 47, 604–607. doi: 10.1016/j.jcjq.2021.05.003

Qin, H., and Yao, Y. (2021). Agriculture KG construction and application. *J. Phys. Conf. Ser.* 1756:012010. doi: 10.1088/1742-6596/1756/1/012010

Ray, P., Duraipandian, R., and Sinha, R. (2021). A framework for identifying and managing risk impact factors for disruptions in the food supply chain. *Int. J. Eng. Appl. Sci.* 12, 1–11. doi: 10.14456/ITJEMAST.2021.84

Shen, M., Duan, J. X., Zhu, L. H., Zhang, J., du, X., and Guizani, M. (2020). Blockchain-based incentives for secure and collaborative data sharing in multiple clouds. *IEEE J Sel Areas Commun* 38, 1229–1241. doi: 10.1109/JSAC.2020.2986619

Song, Y., Lv, C., and Liu, J. (2018). Quality and safety traceability system of agricultural products based on multi-agent. *J. Intel. Fuzzy Syst.* 35, 2731–2740. doi: 10.3233/JIFS-169625

Sun, X., Wu, H., Hao, P., and Li, Q. (2017). Research on information integration method of agricultural products producing and managing based on knowledge graph. *Comput. Comput. Technol. Agric. XI* 545:22. doi: 10.1007/978-3-030-06137-1_22

Talpur, M., Shaikh, M. H., and Talpur, H. S. (2012). Relevance of internet of things in animal stocks chain Management in Pakistan's perspectives. *Int. J. Inform. Educ. Technol.* 2, 29–32. doi: 10.7763/IJIET.2012.V2.77

Xiaoxue, L., Xuesong, B., Longhe, W., Bingyuan, R., Shuhan, L., and Lin, L. (2019). Review and trend analysis of knowledge graphs for crop Pest and diseases. *IEEE Access* 7, 62251–62264. doi: 10.1109/ACCESS.2019.2915987

Yang, Y., Jiang, H., Hu, Y., and Sun, J. (2020). Development of quality and safety traceability of agricultural products based on internet plus. *Chin. J. Eng. Sci.* 22:58. doi: 10.15302/J-SSCAE-2020.04.007

Yang, C., and Sun, Z. (2020). "Data management system based on blockchain technology for agricultural supply chain," in *2020 International Conference on Data Mining Workshops (ICDMW)*. Italy: Sorrento, 907–911.

Yuan, J. J., Lu, Y. L., Cao, X. H., and Cui, H. (2020). Regulating wildlife conservation and food safety to prevent human exposure to novel virus. *Ecosyst. Health Sustain.* 6:1741325. doi: 10.1080/20964129.2020.1741325

Zhang, H., Si, H., and Ma, X. (2021). Research and application of agriculture KG. 5th International Conference on Electronic Information Technology and Computer Engineering. 680–688.

Zhao, G., Liu, S., Lopez, C., Lu, H., Elgueta, S., Chen, H., et al. (2019). Blockchain technology in Agri-food value chain management: a synthesis of applications, challenges and future research directions. *Comput. Ind.* 109, 83–99. doi: 10.1016/j.compind.2019.04.002

Zheng, Y., Cong, Y., Jiang, S., Yang, Z., and Cui, Z. (2020). Innovation of agricultural products trace-ability system in internet + eco agriculture. *IOP Conf. Series Materials Sci. Eng.* 750:012041. doi: 10.1088/1757-899X/750/1/012041