Food-Energy-Water Nexus
WISDOM ACADEMIC https://doi.org/10.63808/fewn.v2i3.430 Volume 2 , Issue 3

Research Article

A Study on a Food Pairing Knowledge Graph Construction Method Integrating Large Language Models and Relation Completion

Jia Yu[1,*], Tong Ji[1], Jiamin Zheng[1]
[1]School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing 102488, China
[*]**Correspondence:** Jia Yu, 3014529637@qq.com

Abstract

Aiming at scattered knowledge sources, inconsistent entity expressions and insufficient relation extraction accuracy in food pairing, this paper proposes a food pairing knowledge graph construction method integrating large language models and relation completion. A domain ontology covering dishes, ingredients, nutrients, dietary restrictions, applicable populations and health goals is designed. Candidate entities and relations are then extracted from recipe texts, nutrition descriptions and dietary-restriction materials using a large language model, and entity normalization is performed through a domain dictionary and semantic embeddings. Finally, a relation scoring function integrating LLM confidence, semantic similarity and domain rules is used to filter, review and complete candidate triples. Experiments on a food pairing corpus show that the proposed method achieves a relation extraction F1 score of 88.71%, which is 6.05 percentage points higher than direct LLM extraction, and obtains a human-verified accuracy of 87.33% for completed relations. The results indicate that the proposed method can provide a reliable structured knowledge base for healthy diet recommendation and food pairing question answering.

Keywords

Food Pairing, Knowledge Graph, Large Language Model, Entity Normalization, Relation Completion

1. Introduction

With the increasing application of artificial intelligence technologies in food science, nutrition and health, and intelligent recommendation systems, personalized food pairing services are gradually becoming an integral part of smart health management. Food pairing involves not only fundamental knowledge such as dishes, ingredients and nutrients, but also

complex constraints including allergens, dietary restrictions, disease risks, cooking methods and target populations. Traditional food pairing systems based on keywords or manual rules struggle to adequately capture these multidimensional semantic relations, resulting in recommendations that lack interpretability and traceability.

Knowledge graphs (KGs) organize domain knowledge through entities, relations and attributes, and can effectively support complex knowledge representation, path reasoning and visual interpretation. Existing studies indicate that KGs have significant value in food science and industrial applications, including food safety, nutritional analysis, food recommendation and supply chain management (Hogan et al., 2021; Min et al., 2022; Misra et al., 2020; Peng et al., 2023; Zuo et al., 2024). However, food pairing is characterized by diverse entity types, numerous synonymous expressions and implicit nutritional relations. Manual construction is costly, while rule-based extraction tends to miss cross-sentence relations and implicit semantic relations.

In recent years, large language models (LLMs) have demonstrated strong capabilities in information extraction, semantic understanding and text generation (Brown et al., 2020; Devlin et al., 2019; Vaswani et al., 2017; Wei et al., 2021), offering a new technical path for low-cost domain knowledge graph construction. Compared with traditional deep learning models, LLMs can understand domain-specific contexts through prompts and perform entity recognition, relation extraction and natural-language interpretation with limited training data. However, direct LLM-based triple extraction still faces unstable output formats, inconsistent entity aliases, non-uniform relation granularity and hallucinated relations. It is therefore necessary to combine LLM extraction with domain rules, semantic similarity and relation completion to improve the accuracy and completeness of food pairing knowledge graphs.

In light of these issues, this paper proposes a food pairing knowledge graph construction method integrating large language models and relation completion (LLM-RC-KG). The method combines the semantic understanding capability of LLMs with a knowledge-graph relation completion mechanism to construct an accurate and interpretable food pairing knowledge graph. The main contributions are as follows:

(1) A domain-specific ontology is designed for the food pairing task, covering dishes, ingredients, nutrients, dietary restrictions, applicable populations and health goals, thereby providing a structured foundation for knowledge graph construction;

(2) An LLM-assisted entity and relation extraction workflow is established, with explicit model configuration, prompt templates, confidence calculation and post-processing procedures. Entity normalization is further introduced to reduce noise caused by synonyms and aliases;

(3) A relation completion method integrating LLM confidence, semantic similarity and domain rules is proposed to improve triple quality. Comparative experiments, ablation studies, dedicated relation completion evaluation, parameter sensitivity analysis and repeated-run tests are conducted to validate the effectiveness and robustness of the proposed method.

Compared with existing food knowledge graph construction pipelines that separately use rule extraction, entity normalization, or link prediction, the main methodological contribution of LLM-RC-KG lies in a closed-loop construction strategy. Specifically, LLM extraction provides candidate triples, domain ontology and entity normalization constrain the triple space, and a confidence-fusion-and-completion mechanism jointly filters explicit triples and validates inferred triples. This design reduces noisy generative outputs while preserving interpretable domain reasoning paths.

2. Related Work

2.1. Research on Food Knowledge Graphs

Knowledge in the food sector typically originates from a variety of sources, including recipe texts, nutritional tables, dietary guidelines, food safety standards and user interaction data, and is characterised by its multi-source, heterogeneous nature, dynamic updates and complex semantic hierarchy. Min et al. (2022) summarised the applications of KG in food science and industry, noting that KG can enhance capabilities in food data organisation, knowledge reasoning and intelligent decision-making. Hogan et al. (2021) conducted a systematic review of KG from the perspectives of knowledge representation, construction and application, covering models, reasoning and applications, thereby providing a general methodological reference for the construction of domain-specific KG.

In the context of food pairing, KGs must not only represent explicit relations such as ‘dish-ingredient’ and ‘ingredient-nutrient’, but also application-oriented semantic relations such as ‘suitable for’, ‘unsuitable for’, ‘alternative ingredients’ and ‘recommended pairings’. Such relations often require judgment based on context and nutritional knowledge, and traditional rule-based extraction methods struggle to cover all scenarios. Consequently, how to improve automatic construction efficiency while ensuring accuracy is a key issue in food knowledge graph research.

2.2. Knowledge extraction aided by Large Language Models

The introduction of the Transformer architecture has driven the development of pre-trained language models (Vaswani et al., 2017). Models such as BERT and GPT have achieved remarkable results in natural language understanding and generation tasks (Brown et

al., 2020; Devlin et al., 2019; Wei et al., 2021). With the continuous improvement of instruction tuning and few-shot learning capabilities, LLMs are increasingly being applied to knowledge extraction tasks such as entity extraction, relation extraction, event extraction and question-answering generation. Compared with traditional sequence tagging models, LLMs have stronger contextual modeling and semantic understanding capabilities, enabling them to capture long-range dependencies and complex linguistic structures. Rather than relying on large-scale manually annotated data, they can rapidly transfer knowledge to specific domain tasks through prompt templates or contextual examples, thereby reducing the cost of building knowledge extraction systems.

However, the outputs of LLM are generative in nature and may suffer from issues such as inconsistent formatting, unreliable boundary detection and factual errors. For knowledge graph construction, the extracted results must be structured, verifiable and updatable. Therefore, this paper introduces entity normalization, relation validity checks and a confidence fusion mechanism following the extraction process to mitigate the uncertainty associated with generative extraction.

2.3. Relation Completion and Graph Quality Optimization

Knowledge graph completion aims to infer and fill in missing triples based on existing entities and relations, thereby enhancing the completeness and practicality of the graph. Existing methods include rule-based reasoning, link prediction based on representation learning, and relation modeling based on graph neural networks (Gao et al., 2022; Li et al., 2021; Pan et al., 2024; Vashishth et al., 2019). In the specific context of food pairing, certain relations can be reasonably inferred based on nutritional attributes, ingredient characteristics and domain-specific rules. For example, if an ingredient is rich in high-quality protein and low in fat, it may be suitable for individuals seeking to build muscle or manage fat intake; similarly, if an ingredient is a common allergen, it should be flagged as a high-risk ingredient under specific user constraints to mitigate potential health risks.

In response to the characteristics of this scenario, this paper adopts a lightweight approach to relation completion, which involves the weighted fusion of semantic similarity, rule matching, and confidence scores extracted from large language models. This method has low computational complexity, making it suitable for implementation in small- to medium-scale food knowledge graphs, while also offering a degree of interpretability.The proposed confidence fusion mechanism is motivated by complementary evidence from three sources: the LLM confidence reflects contextual extraction reliability, semantic similarity measures whether the head and tail entities are semantically compatible, and rule matching

encodes domain constraints that are difficult to learn from a small corpus. Therefore, the fusion function is not intended as a general link-prediction model; rather, it is designed as a transparent quality-control layer for small- and medium-scale food pairing knowledge graphs.

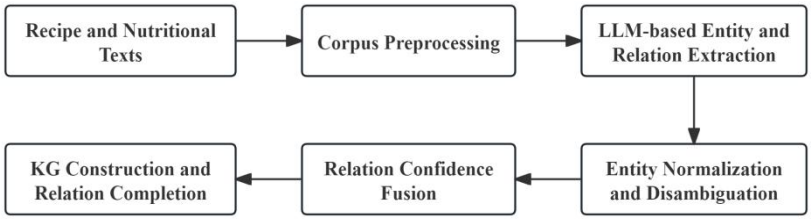
3. Methods

3.1. General Framework

The LLM-RC-KG method proposed in this paper comprises six main stages: corpus preprocessing, domain ontology design, LLM-based entity and relation extraction, entity normalization and disambiguation, relation confidence fusion, and KG construction and relation completion. Among these, domain ontology design serves as the foundational support for knowledge graph construction and is completed prior to the commencement of the process. The overall workflow of the method is illustrated in Figure 1.

Figure 1

Overall workflow of the LLM-RC-KG method



Note. This figure illustrates the overall workflow of the LLM-RC-KG method (a Food Pairing Knowledge Graph Construction Method Integrating Large Language Models and Relation Completion), including corpus preprocessing, extraction, normalization, and confidence fusion.

3.2. Domain Ontology Design

For the food pairing task, the food pairing knowledge graph is formally defined as G=(E,R,A,T), where E denotes the entity set, R denotes the relation set, A denotes the attribute set, and T={(h,r,t)|h,t in E, r in R} denotes the triple set. Entity types include dishes, ingredients, nutrients, cooking methods, dietary restrictions, health goals and target populations; relation types include contains, rich in, suitable for, unsuitable for, can be substituted for, recommended pairing, belongs to a cuisine and cooking method is.

Table 1

Key Entities and Relation Types in the Food Pairing Knowledge Graph

Category	Specific Type	Example

Entity Type	Dishes, ingredients, nutrients, health goals, applicable populations	Chicken breast salad, oats, protein, people seeking fat loss
Relation Type	Contains, rich in, suitable for, not suitable for, alternatives, recommended pairings	Oats-rich in-dietary fiber; soy milk-substitute for-cow's milk
Attribute Type	Calories, protein, fat, carbohydrates, allergens, cooking methods	Calories = 320 kcal; Allergens = peanuts

Note. This table presents the key entity categories, relation types, and attribute types defined in the food pairing domain ontology.

3.3. Entity-Relation Extraction from LLM

For a given input text d, prompt templates are used to constrain the LLM to generate structured outputs. Each prompt contains a task definition, entity-type descriptions, relation-type constraints, an output schema and a small number of examples. The model output is standardized as a JSON-style candidate triple set K_c=(h_i,r_i,t_i,c_i), where h_i is the head entity, r_i is the relation, t_i is the tail entity, and c_i is the model confidence score or the confidence score derived from repeated-sampling consistency.

To improve extraction stability, a two-stage extraction strategy is adopted. In the first stage, only entities and their types are extracted; in the second stage, relations are extracted under the constraints of the recognized entity set. This strategy reduces the probability of generating invalid relations between unrecognized entities. In addition, output validation is performed to remove triples whose relation types are outside the predefined schema.

Implementation details of the LLM extraction module. The entity and relation extraction module used GPT-3.5-turbo-0125 through an API interface. The temperature was set to 0.2, top_p to 1.0, and the maximum output length to 1,024 tokens. Each text segment was sampled three times, and the final candidate triple set was obtained through majority voting and schema validation. The confidence score C_{LLM} was calculated as the proportion of consistent outputs across repeated runs and was combined with model-reported confidence when available.

The embedding model for entity normalization was text-embedding-3-small. The software stack consisted of Python 3.10, spaCy for sentence segmentation, NetworkX for graph processing, and Neo4j 5.x for graph storage. Experiments were conducted on a workstation equipped with an Intel i7 CPU, 32 GB RAM and an NVIDIA RTX 3060 GPU. Because the LLM was accessed through an API, GPU resources were mainly used for local embedding computation and baseline model training.

Prompt strategy. The prompt contains five components: task definition, allowed entity types, allowed relation types, output schema, and examples. The model is required to output a JSON list in the form {“head”:“...”, "head_type":“...”, “relation”:“...”, “tail”:“...”,

"tail_type":"...", "confidence":0-1}. Outputs that violate the ontology schema or contain unsupported relation types are discarded during post-processing.

3.4. Entity normalization and semantic disambiguation

The food domain contains many aliases and synonymous expressions, such as "tomato" and "xi hong shi" in Chinese, "potato" and "ma ling shu", and "low-fat milk" and "skimmed milk". Without entity normalization, redundant nodes appear in the knowledge graph and affect subsequent retrieval and reasoning. This paper constructs a food-domain alias dictionary and performs entity merging based on semantic vector similarity. For a candidate entity e_i and a standard entity e_j, if their types match and $sim(e_i,e_j)>=\theta_{norm}$, then e_i is mapped to e_j.

Semantic similarity is calculated using cosine similarity, as per the following formula:

$$sim(e_i,e_j)=\frac{v_i \cdot v_j}{\|v_i\|_2 \|v_j\|_2}, \quad \theta_{norm}=0.82 \tag{1}$$

Here, v_i and v_j denote the embedding vectors of the candidate entity e_i and the standard entity e_j, respectively, and theta_norm denotes the entity normalization threshold. If type e_i = type e_j and $sim(e_i,e_j)>=\theta_{norm}$, e_i is mapped to e_j; otherwise, it is retained as an independent entity pending manual review or context-based disambiguation.

3.5. Relation Confidence Fusion and Completion

To improve the quality of triples, this paper proposes a hybrid relation scoring function, expressed as follows:

$$Score(h,r,t)=\alpha C_{LLM}(h,r,t)+\beta S_{sem}(h,t)+\gamma S_{rule}(h,r,t), \quad \alpha+\beta+\gamma=1 \tag{2}$$

where C_{LLM} denotes the LLM extraction confidence, S_{sem} denotes the semantic similarity between the head and tail entities, and S_{rule} denotes the domain rule matching score. The coefficients alpha, beta and gamma satisfy $\alpha+\beta+\gamma=1$. In the default setting, $\alpha=0.50$, $\beta=0.25$ and $\gamma=0.25$, with LLM confidence serving as the primary evidence and semantic similarity and rule matching serving as auxiliary constraints. Triples with $Score>=\theta_{keep}$ are written into the graph; triples in the gray zone [θ_{review}, θ_{keep}) are submitted to manual review or secondary prompting; and triples with Score<theta_review are discarded. In this study, θ_{review} is set to 0.55 and theta_keep is set to 0.70.

Table 2

Examples of Relation Completion Rules

Rule ID	Prerequisite Path	Relational Completion	Example
R1	Food-rich in-nutrient; nutrient-contributes to-health	Food-suitable for-health goals	Tofu-rich in-protein→Tofu-suitable for-muscle gain

	goals		
R2	Food-belongs to-allergens; population-avoids-allergens	Food-unsuitable for-population	Peanuts-belong to-nut allergen→Peanuts-unsuitable for-people with nut allergies
R3	Food A-nutritionally similar to-Food B; Food A-does not contain-restricted component	Food A-substitutable for-Food B	Soya milk-substitutable for-cow's milk

Note. Examples of domain rules used for relation completion, showing the prerequisite paths and the resulting completed relations.

Relation completion primarily targets food pairing relations such as 'suitable', 'unsuitable', 'recommended pairing' and 'substitute'. This paper constructs candidate inference paths based on domain rules. For example, if ingredient x is rich in nutrient n, and nutrient n contributes to health goal g, then the candidate triple (x, suitable, g) can be generated. If ingredient x belongs to allergen category a, and user group p needs to avoid a, then the candidate triple (x, unsuitable, p) can be generated. The completed candidate triples must also be filtered by the scoring function.

To avoid unsafe health-related inference, relation completion involving diabetes, kidney disease, allergies, pregnancy, or other medical conditions is not automatically treated as a clinical recommendation. Such triples are retained only when supported by authoritative nutritional guidelines or manually verified by annotators with nutrition-related expertise. The system output is positioned as knowledge organization for food pairing and question answering, rather than medical diagnosis or treatment advice.

3.6. Algorithm Description

Table 3

Steps of the LLM-RC-KG Algorithm

Step	Operation
Input	Food text corpus D, domain ontology O, alias dictionary V, thresholds θ_{norm}, θ_{review} and θ_{keep}, and weights α, β and γ
1	Perform text cleaning, sentence segmentation, and deduplication on corpus D to obtain a set of standardized text segments
2	Invoke LLM based on prompt templates to extract candidate entity set E_c and candidate relation set T_c
3	Perform entity normalization according to entity types, named entity dictionary, and semantic similarity
4	Compute the *Score*(h,*r*,*t*) for each candidate triple, and filter low-confidence triples
5	Generate missing triples based on domain-specific rules, then recalculate confidence scores

6	Write validated entities, attributes, and relations into knowledge graph G	
Output	Food pairing knowledge graph G	

Note. The specific execution steps of the proposed LLM-RC-KG algorithm.

4. Experiments and Analysis of Results

4.1. Datasets and Experimental Setup

To validate the effectiveness of the proposed method, this paper constructs an experimental corpus for food pairing. The corpus contains 1,500 recipe texts, 500 nutrition descriptions and dietary-restriction materials, covering common scenarios such as home-cooked meals, low-fat meals, vegetarian meals, low-sugar meals and high-protein meals. After cleaning, sentence segmentation and deduplication, the corpus contains approximately 230,000 Chinese characters. The manually annotated test set contains 300 text fragments annotated with entity types, relation types and standard triples. In the experiments, alpha, beta and gamma were set to 0.50, 0.25 and 0.25, respectively; theta_norm was set to 0.82; theta_review was set to 0.55; and theta_keep was set to 0.70.

Corpus construction and provenance. The recipe texts were collected from publicly accessible recipe pages such as Xiachufang, Meishijie and Douguo, and only dish names, ingredients, cooking steps and pairing descriptions were retained. Nutrition descriptions were compiled with reference to China Food Composition Tables and open food-composition materials (Yang, 2018). Dietary-restriction rules were summarized from Dietary Guidelines for Chinese Residents (2022) and publicly available dietary guideline materials (Chinese Nutrition Society, 2022). All texts were cleaned, segmented, exactly deduplicated and filtered using SimHash-based near-duplicate detection before annotation. The corpus was split at the text-fragment level into a construction set and an independent test set to prevent leakage between graph construction and evaluation.

Manual annotation protocol. The 300 test fragments were independently annotated by two annotators with backgrounds in food science or nutrition-related data analysis. The annotation guideline defined entity boundaries, entity types, relation labels, and evidence spans. Disagreements were resolved through discussion with a third reviewer. Inter-annotator agreement was measured by Cohen’s kappa, and the obtained value was 0.86, indicating strong annotation consistency.

Although the experimental corpus covers common food pairing scenarios, it remains a medium-scale corpus. Larger and more diverse corpora will be required to further evaluate the scalability and domain robustness of the proposed framework in future work.

Table 4

Data Source, Split and Quality-Control Information

Item	Description
Recipe texts	1,500 entries from Xiachufang, Meishijie, Douguo and other public recipe pages; dish names, ingredients, cooking steps and pairing descriptions were retained.
Nutrition descriptions	500 entries based on China Food Composition Tables and open food-composition materials [22].
Dietary-restriction materials	Constraints summarized from Dietary Guidelines for Chinese Residents (2022) and public dietary guideline materials [23].
Deduplication	Exact matching and SimHash near-duplicate filtering; no overlap between construction and test sets.
Corpus split	2,000 texts for graph construction; 300 fragments for manual testing.
Annotation consistency	Two independent annotators; Cohen's kappa=0.86.

Note. Details regarding the data sources, corpus split strategy, and annotation quality control for the experimental dataset.

Table 5

Statistics on the Experimental Corpus

Item	Quantity
Recipe Texts	1,500 entries
Nutritional Knowledge Descriptions	500 entries
Manually Annotated Test Texts	300 entries
Entity Types	7 categories
Relation Types	12 categories
Constructed Triples	12,036 entries

Note. Quantitative statistics of the experimental corpus, including the number of texts, entities, relations, and constructed triples.

4.2. Evaluation Metrics

This paper evaluates the performance of entity extraction and relation extraction using Precision, Recall and the F1 score. The specific calculation formulas are as follows:

$$Precision=\frac{TP}{TP+FP} \tag{3}$$

$$Recall=\frac{TP}{TP+FN} \tag{4}$$

$$F1=\frac{2\cdot Precision\cdot Recall}{Precision+Recall} \quad (5)$$

Here, TP denotes the number of correctly extracted entity-relation triples; FP denotes the number of incorrect extractions, including instances where non-existent entities were identified or incorrect triples were generated; and FN denotes entities or relations present in the dataset but omitted by the model. For the relation completion task, given that some valid triples may not appear in the annotated dataset, this paper also calculates the human-verified accuracy of the completed triples to provide a more comprehensive reflection of the model's performance.

4.3. Comparative experiment

The proposed method is compared with seven methods: Rule, LLM-Direct, LLM+Norm, BERT-RE (Devlin et al., 2019), CasRel (Zeng et al., 2020), CompGCN-KGC (Vashishth et al., 2019) and Ours. Rule and LLM-based settings serve as incremental baselines, while BERT-RE, CasRel and CompGCN-KGC represent stronger supervised relation extraction and graph-completion baselines. The results are shown in Table 6.

Table 6

Entity Extraction and Relation Extraction Results for Different Methods (%)

Methods	Entity P	Entity R	Entity F1	Relation P	Relation R	Relation F1
Rule	84.32	70.45	76.77	78.46	61.38	68.86
LLM-Direct	88.91	84.37	86.58	86.17	79.42	82.66
LLM+Norm	91.26	86.94	89.05	89.34	82.91	86.01
BERT-RE	89.83	85.31	87.51	87.62	83.04	85.27
CasRel	90.25	86.85	88.52	88.74	84.51	86.57
CompGCN-KGC	-	-	-	87.91	84.26	86.05
Ours	92.38	89.76	91.05	90.12	87.35	88.71

Note. P = Precision, R = Recall, F1 = F1 Score. All values are presented as percentages (%).

As shown in Table 6, the rule-based method achieves acceptable precision but has a low recall rate, indicating that fixed templates struggle to capture diverse expressions in food-related texts. Direct LLM extraction improves both entity and relation extraction performance, but inconsistent entity aliases and relation hallucinations remain. After entity normalization, the entity F1 score increases to 89.05% and the relation F1 score rises to 86.01%. Compared with BERT-RE, CasRel and CompGCN-KGC, the proposed method

improves the relation F1 score by 3.44, 2.14 and 2.66 percentage points, respectively, demonstrating that it improves recall while maintaining precision.

For fair comparison, all baselines use the same ontology, corpus split and evaluation protocol. Rule and LLM-Direct are implemented as lightweight baselines, whereas BERT-RE, CasRel and CompGCN-KGC represent supervised relation extraction and graph-completion settings. As shown in Table 6, the proposed method achieves the best relation F1 score among all compared methods.

4.4. Ablation Experiment

To analyze the impact of each module on overall performance, ablation experiments are conducted by sequentially removing entity normalization, relation completion, rule-based scoring and semantic similarity. Their effects on relation extraction performance are reported in Table 7.

Table 7

Ablation Experiment Results (%)

Methods	Relation P	Relation R	Relation F1
Ours (complete method)	90.12	87.35	88.71
Remove Entity Normalization	86.73	81.96	84.28
Remove Relation Completion	89.34	82.91	86.01
Remove Rule-based Scoring	88.21	85.86	87.02
Remove Semantic Similarity	87.90	85.61	86.74

Note. Results of the ablation study analyzing the contribution of each module to relation extraction performance. Values are percentages (%).

Ablation experiments indicate that entity normalization has the greatest impact on overall performance; removing this module resulted in a 4.43 percentage point decrease in the relation F1 score, primarily because entity aliases led to the same relation being stored in separate locations. Removing relation completion caused a significant drop in recall, suggesting that the completion strategy is effective at identifying latent relations in the text that are not explicitly stated but are supported by domain rules. Both rule scores and semantic similarity contribute to suppressing erroneous triples and improving the quality of the knowledge graph.

4.5. Dedicated Evaluation of Relation Completion

To separate extraction performance from completion performance, relation completion is evaluated independently. Completed triples are not judged only against text-extracted triples, because a valid completion may be absent from the original sentence. Therefore, the number of generated candidate triples, retained triples, manually verified triples and human-verified accuracy are reported separately in Table 8.

Table 8

Dedicated Evaluation of Relation Completion

Metric	Value
Generated completion candidates	1,436
Retained after confidence filtering	682 (47.49%)
Manual verification sample size	300
Human-verified accuracy	87.33% (262/300)
Main error type	Unsupported health inference: 39.47%; insufficient numerical evidence: 34.21%; ambiguous entity: 26.32%

Note. Independent evaluation metrics for the relation completion module, including the human-verified accuracy of the generated candidates.

Representative cases are used to show why a completed relation is accepted or rejected. For example, a relation such as "oats-rich in-dietary fibre → oats-suitable for-high-fibre diet" is accepted when supported by nutritional evidence, whereas a relation involving a disease-specific dietary recommendation is rejected or marked for expert review if no authoritative source is available.

4.6. Parameter Sensitivity and Repeated-Run Analysis

To evaluate robustness, this paper conducts sensitivity analysis for the confidence-fusion weights, the entity normalization threshold θ_{norm}, and the relation retention threshold θ_{keep}. In addition, repeated-run statistics are reported because LLM outputs may vary across sampling runs.

Table 9

Parameter Sensitivity Analysis

Parameter setting	Relation P	Relation R	Relation F1	Note
α/β/γ=0.50/0.25/0.25	90.12	87.35	88.71	Default
α/β/γ=0.40/0.30/0.30	89.34	87.12	88.22	Higher semantic/rule weight
θ_{norm}=0.78	88.76	87.92	88.34	Looser normalization

θ_{norm}=0.86	90.51	85.98	88.19	Stricter normalization
θ_{keep}=0.65	88.43	88.66	88.54	Looser retention
θ_{keep}=0.75	91.06	85.41	88.15	Stricter retention

Note. Parameter sensitivity analysis for confidence-fusion weights (α, β, γ), entity normalization threshold (θ_{norm}), and relation retention threshold (θ_{keep}).

Table 10

Repeated-Run Results and Significance Testing

Method	Relation F1 mean±std	Significance vs. Ours
Rule	68.79±0.42	p<0.001
LLM-Direct	82.54±0.63	p<0.001
LLM+Norm	85.93±0.51	p=0.003
BERT-RE	85.18±0.56	p=0.002
CasRel	86.41±0.48	p=0.006
CompGCN-KGC	85.97±0.52	p=0.004
Ours	88.64±0.37	-

Note. Mean and standard deviation of the relation F1 score across repeated sampling runs, along with statistical significance compared to the proposed method.

As shown in Tables 9 and 10, the default parameter setting achieves the highest relation F1 score, and the standard deviation across repeated runs remains small. This indicates that the proposed method is relatively stable under moderate parameter changes and repeated LLM sampling.

4.7. Results of KG Construction and Case Studies

The final food pairing knowledge graph comprises 5,329 entities, 12,036 triples and 12 types of relations. Among these, food ingredient entities account for the largest proportion, at approximately 42.6%; dish entities account for approximately 28.4%; while entities related to nutrients, health objectives and target populations collectively form key semantic nodes for food pairing inference. Taking ‘tofu’ as an example, the graph includes relations such as ‘tofu-rich-in-plant-protein’, ‘tofu-suitable-for-high-protein-diet’ and ‘tofu-can-replace-certain-meat-ingredients’, which provide a structured basis for subsequent vegetarian, high-protein and low-fat food pairing tasks.

Error analysis reveals that the method still has three limitations. First, some nutritional relations require precise numerical support, and reliance solely on textual descriptions may lead to judgment bias. Second, relations involving diabetes, kidney disease, allergies or other health conditions are medically sensitive and must be constrained by authoritative nutritional guidelines and expert validation. Third, LLMs may still miss cross-sentence relations when processing complex long sentences. Future work will incorporate authoritative nutritional databases and expert review mechanisms to further improve knowledge quality.

5. Conclusion

This paper proposes an LLM-RC-KG framework for constructing a food pairing knowledge graph under the challenges of scattered knowledge sources, inconsistent entity expressions and insufficient relation extraction accuracy. The method constrains LLM outputs through a domain ontology, improves triple quality through entity normalization and confidence fusion, and enhances graph completeness through lightweight relation completion rules. Experiments show that the proposed method improves relation extraction performance compared with both incremental and stronger baselines. The dedicated completion evaluation, parameter sensitivity analysis and repeated-run results further support the effectiveness and robustness of the framework. Future work will enlarge the corpus, integrate authoritative nutritional databases and expert validation, and explore graph neural network-based completion methods to enhance reliability and reasoning capability.

Appendix A. Representative Prompt Template

Entity extraction prompt: You are a food-domain information extraction assistant. Given the following text, identify entities belonging only to the predefined types: Dish, Ingredient, Nutrient, CookingMethod, DietaryRestriction, HealthGoal, and Population. Output a JSON list with fields entity, type, and evidence. Do not output entities outside the schema.

Relation extraction prompt: Based on the recognized entities, extract relations only from the predefined relation set: contains, rich_in, suitable_for, unsuitable_for, substitute_for, recommended_pairing, belongs_to_cuisine, and cooking_method_is. Output JSON triples with fields head, head_type, relation, tail, tail_type, evidence, and confidence. If evidence is insufficient, return an empty list.

Post-processing: JSON format validation, ontology label checking, entity normalization, duplicate removal, confidence fusion, and manual/expert review are applied sequentially. Any health-related relation without authoritative evidence is marked as review-required instead of being automatically accepted.

Acknowledgments

The authors have no additional acknowledgments to declare.

Author contributions

Conceptualization, Jia Yu and Tong Ji; methodology, Jia Yu; software, Jia Yu and Tong Ji; validation, Jia Yu, Tong Ji and Jiamin Zheng; formal analysis, Jia Yu; investigation, Jia Yu and Tong Ji; resources, Jiamin Zheng; data curation, Jia Yu and Tong Ji; writing—original draft preparation, Jia Yu; writing—review and editing, Jia Yu, Tong Ji and Jiamin Zheng; visualization, Jia Yu and Tong Ji; supervision, Jia Yu; project administration, Jia Yu; funding acquisition, Jia Yu. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the 2026 Undergraduate Innovation and Entrepreneurship Training Program of Beijing Technology and Business University, grant number S202610011005.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Conflict of interest

The authors declare that there is no conflict of interest.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

All authors have given their consent.

Additional information

Received: 25 May 2026

Accepted: 29 June 2026

Published online: 16 August 2026

Publisher's Note

References

[1] Zuo, M., Wang, F., Song, S., Yan, W., & Dai, X. (2024). “Intelligence+ food regulation”: Development process, application status, and future direction. *Journal of Food Science and Technology, 42*(3), 1–10. https://doi.org/10.12301/spxb202400331

[2] Misra, N. N., Dixit, Y., Al-Mallahi, A., Bhullar, M. S., Upadhyay, R., & Martynenko, A. (2022). IoT, big data, and artificial intelligence in agriculture and food industry. *IEEE Internet of Things Journal, 9*(9), 6305–6324. https://doi.org/10.1109/JIOT.2020.2998584

[3] Min, W. Q., Liu, C. L., Xu, L. Y., & Jiang, S. Q. (2022). Applications of knowledge graphs for food science and industry. *Patterns, 3*(5), Article 100484. https://doi.org/10.1016/j.patter.2022.100484

[4] Peng, C. Y., Xia, F., Naseriparsa, M., & Zhu, X. (2023). Knowledge graphs: Opportunities and challenges. *Artificial Intelligence Review, 56*(11), 13071–13102. https://doi.org/10.1007/s10462-023-10465-9

[5] Hogan, A., Blomqvist, E., Cochez, M., d’Amato, C., de Melo, G., Gutierrez, C., Kirrane, S., Labra Gayo, J. E., Navigli, R., Neumaier, S., Ngonga Ngomo, A.-C., Polleres, A., Roomp, K., Sack, H., Sequeda, J., & Simperl, E. (2021). Knowledge graphs. *ACM Computing Surveys, 54*(4), 1–37. https://doi.org/10.1145/3447772

[6] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). DBpedia: A nucleus for a web of open data. In *Proceedings of the 6th International Semantic Web Conference (ISWC)* (pp. 722–735). Springer. https://doi.org/10.1007/978-3-642-04930-9_2

[7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 5998–6008). https://doi.org/10.48550/arXiv.1706.03762

[8] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

[9] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A.,

Ziegler, D., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems* (Vol. 33, pp. 1877–1901). https://doi.org/10.5555/3495724.3495883

[10] Wei, J., Bosma, M., Zhao, V. Y., Guu, K., Yu, A. W., Lester, B., Du, N., Dai, A. M., & Le, Q. V. (2022). Finetuned language models are zero-shot learners. In *Proceedings of the 10th International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.2109.01652

[11] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2022). LoRA: Low-rank adaptation of large language models. In *Proceedings of the 10th International Conference on Learning Representations*. https://openreview.net/forum?id=nZeVKeeFYf9

[12] Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023). QLoRA: Efficient finetuning of quantized large language models [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2305.14314

[13] Vashishth, S., Sanyal, S., Nitin, V., & Talukdar, P. (2020). Composition-based multi-relational graph convolutional networks. In *Proceedings of the 8th International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.1911.03082

[14] Gao, J., Luo, X., & Wang, H. (2022). Chinese causal event extraction using causality-associated graph neural network. *Concurrency and Computation: Practice and Experience, 34*(3), Article e6572. https://doi.org/10.1002/cpe.6572

[15] Li, Z., Sun, Y., Zhu, J., Tang, S., Zhang, C., & Ma, H. (2021). Improve relation extraction with dual attention-guided graph convolutional networks. *Neural Computing and Applications, 33*(6), 1773–1784. https://doi.org/10.1007/s00521-021-06410-2

[16] Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering, 36*(7), 3580–3599. https://doi.org/10.1109/TKDE.2024.3352100

[17] Ibrahim, N., Aboulela, S., Ibrahim, A., & Kashef, R. (2024). A survey on augmenting knowledge graphs (KGs) with large language models (LLMs): Models, evaluation metrics, benchmarks, and challenges. *Discover Artificial Intelligence, 4*(1), 76. https://doi.org/10.1007/s44163-024-00137-0

[18] Ma, P., Tsai, S., He, Y., Jia, X., Zhen, D., Yu, N., Wang, Q., Ahuja, J. K. C., & Wei, C.-I. (2024). Large language models in food science: Innovations, applications, and future. *Trends in Food Science & Technology, 148*, Article 104488. https://doi.org/10.1016/j.tifs.2024.104488

[19] Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems* (Vol. 26, pp. 2787–2795). https://doi.org/10.5555/2999611.2999715

[20] Wei, Z., Su, J., Wang, Y., Tian, Y., & Chang, Y. (2020). A novel cascade binary tagging framework for relational triple extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1476–1488). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.136

[21] Yao, L., Mao, C., & Luo, Y. (2019). KG-BERT: BERT for knowledge graph completion [Preprint]. arXiv. https://doi.org/10.48550/arXiv.1909.03193

[22] Yang, Y. X. (2018). *China food composition tables: Standard edition*. Peking University Medical Press.

[23] Chinese Nutrition Society. (2022). *Dietary guidelines for Chinese residents (2022)*. People's Medical Publishing House.