

# KG4NH: A Comprehensive Knowledge Graph for Question Answering in Dietary Nutrition and Human Health

Chengcheng Fu<sup>ID</sup>, Xueli Pan<sup>ID</sup>, Jieyu Wu, Junkai Cai, Zhisheng Huang, Frank van Harmelen<sup>ID</sup>, Weizhong Zhao<sup>ID</sup>, *Member, IEEE*, Xingpeng Jiang<sup>ID</sup>, *Member, IEEE*, and Tingting He<sup>ID</sup>, *Member, IEEE*

**Abstract**—It is commonly known that food nutrition is closely related to human health. The complex interactions between food nutrients and diseases, influenced by gut microbial metabolism, present challenges in systematizing and practically applying knowledge. To address this, we propose a method for extracting triples from a vast amount of literature, which is used to construct a comprehensive knowledge graph on nutrition and human health. Concurrently, we develop a query-based question answering system over our knowledge graph, proficiently addressing three types of questions. The results show that our proposed model outperforms other state-of-art methods, achieving a precision of 0.92, a recall of 0.81, and an F1 score of 0.86 in the nutrition and disease relation extraction

task. Meanwhile, our question answering system achieves an accuracy of 0.68 and an F1 score of 0.61 on our benchmark dataset, showcasing competitiveness in practical scenarios. Furthermore, we design five independent experiments to assess the quality of the data structure in the knowledge graph, ensuring results characterized by high accuracy and interpretability. In conclusion, the construction of our knowledge graph shows significant promise in facilitating diet recommendations, enhancing patient care applications, and informing decision-making in clinical research.

**Index Terms**—Knowledge graph, text mining, question answering, nutrition, human diseases.

Received 25 June 2023; revised 10 October 2023 and 15 November 2023; accepted 24 November 2023. Date of publication 1 December 2023; date of current version 7 March 2025. This work was supported in part by the National Natural Science Foundation of China under Grants 62372205 and 61932008, in part by the National Language Commission Key Research Project under Grant ZDI145-56, in part by the Fundamental Research Funds for Central Universities under Grant KJ02502022-0450, in part by the National High-end Foreign Expert Cooperation Project under Grant G2022158003L, and in part by the Natural Science Foundation of Hubei Province of China under Grant 2022CFB289. (*Corresponding authors: Xingpeng Jiang; Tingting He.*)

Chengcheng Fu and Tingting He are with the Hubei Provincial Key Laboratory of Artificial Intelligence and Smart Learning, Central China Normal University, Wuhan 430079, China, and with the National Engineering Research Center for E-Learning, Central China Normal University, Wuhan 430079, China, and with the National Language Resources Monitor Research Center for Network Media, Central China Normal University, Wuhan 430079, China, and also with the School of Computer Science, Central China Normal University, Wuhan 430079, China (e-mail: ccfu@mails.ccnu.edu.cn; tthe@mail.ccnu.edu.cn).

Xueli Pan and Frank van Harmelen are with the Department of Computer Science, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, Netherlands (e-mail: x.pan2@vu.nl; frank.van.harmelen@vu.nl).

Jieyu Wu, Junkai Cai, Weizhong Zhao, and Xingpeng Jiang are with the Hubei Provincial Key Laboratory of Artificial Intelligence and Smart Learning, Central China Normal University, Wuhan 430079, China, and with the National Language Resources Monitor Research Center for Network Media, Central China Normal University, Wuhan 430079, China, and also with the School of Computer Science, Central China Normal University, Wuhan 430079, China (e-mail: wjy221@mails.ccnu.edu.cn; tsaijk@mails.ccnu.edu.cn; wzzhao@ccnu.edu.cn; xpjiang@mail.ccnu.edu.cn).

Zhisheng Huang is with the Department of Computer Science, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, Netherlands, and with the Clinical Research Center for Mental Disorders, Shanghai Pudong New Area Mental Health Center, Tongji University School of Medicine, Shanghai 200092, China, and also with the Deep Blue Technology Group, Shanghai 200040, China (e-mail: z.huang@vu.nl).

Digital Object Identifier 10.1109/JBHI.2023.3338356

## I. INTRODUCTION

IT IS well established that maintaining proper dietary nutrition is essential for staying healthy [1]. Meanwhile, extensive evidence suggests that unhealthy diets are associated with more than two hundred diseases [2]. The interactions between food nutrients and diseases are intricate, particularly when considering the influence of gut microbial metabolism [3]. For example, a proper intake of fiber from vegetables is related to improving metabolic abnormalities and contributing to fewer symptoms of depression [4]. However, an increased intake of saturated fats from processed foods leads to an inflammatory microbial environment that triggers a series of cardiometabolic disorders [5]. Although extensive research efforts have been made to clarify the relationship between food nutrition and human health, even domain experts face challenges in systematizing and applying knowledge in this field [6]. Hence, there is an urgent need to develop a framework that integrates comprehensive knowledge and offers practical applications to support the acquisition of dietary inquiries [7].

To achieve this, a knowledge graph (KG) is utilized in this work, which is a powerful tool for representing and analyzing complex biological knowledge. Currently, KGs are extensively utilized in various related studies, owing to their numerous advantages [8]. Specifically, the employment of KGs contributes to the extraction and integration of heterogeneous data from various sources [9], and provides valuable insights into how things are linked [10]. In addition, they also facilitate the development of knowledge retrieval [11], knowledge prediction [12], and other applications [13]. However, the KGs, especially in relation

**TABLE I**  
RELATED KNOWLEDGE GRAPHS ON FOOD AND HUMAN HEALTH

Name	Ontology	Topic	Technologies
FoodKG [21]	WhatToMake ontology	Food recommendation	Semantic query
FlavorGraph [22]	-	Food recommendation	Graph node embedding
Food safety KG [23]	Food safety ontology	Food safety detection and question answering	Extraction algorithm and http service
Personal Health Ontology [24]	Personal health ontology	Dietary recommendations	Semantic query
Healthcare diet KG [25]	Food-related ontologies	Healthcare management	Text mining technologies
KG of the dietary factors associated with cardiovascular disease [26]	-	Detect potential associations between dietary factors and cardiovascular disease	Statistical analysis
Food4healthKG [27]	Food-related ontologies	Food recommendation on mental health	Semantic query

to food nutrition, still have several limitations [14]. In particular, some KGs are limited in their scale and depth, resulting in incomplete, outdated, or inconsistent information [15], leading to narrow dietary advice divorced from human health. Thus, the knowledge gaps between various interactions and ontologies in incomplete systems still persist [16], which creates barriers between local associations and domain knowledge and blocks the sustainable utilization of KGs [17]. Furthermore, there is a lack of development of effective applications of KGs [18], particularly in understanding and querying food-related knowledge for human health [19]. To address these challenges, our work focuses on developing a large-scale KG that incorporates numerous up-to-date associations and domain ontologies and can be effectively applied in a question answering system.

In this paper, we develop a comprehensive knowledge graph, which provides a valuable question answering application for dietary nutrition and human health. More specifically, our work has three contributions: (i) A comprehensive knowledge graph is developed in dietary nutrition and human health, which is constructed through automatic triples extraction and knowledge graph integration. (ii) A query-based question answering system is constructed using the knowledge graph, addressing three types of questions ranging from diet to human health, and demonstrating strong performance on a benchmark dataset. (iii) A series of experiments demonstrate that the constructed KG is well organized, contributing to a high-quality question answering system and facilitating a deeper understanding of knowledge and exploration of proper diets on human health. Our knowledge graph and relevant materials are available at: <https://github.com/ccszbd/KG4NH>.

The remainder of this paper is organised as follows. In Section II, we review the literature regarding the work about food-related KGs and outline current limitations. In Section III, we describe the workflow of constructing the KG4NH, including data collection, triples extraction and knowledge integration. Next, in Section IV we introduce the development of question answering system. In Section V, we conduct various experiments and discuss the results. Finally, in Section VI we summarise the main contributions and outline future direction of research.

## II. RELATED WORK

In this section, we provide an overview of existing methods in the literature, especially based on KGs in the context of food and human health. The relevant approaches and their summaries are shown in Table I respectively.

In recent research on nutrition and disease, various typical KGs have been proposed [20]. For example, the FoodKG is a KG with food and recipes information, which can provide food suggestions through semantic queries [21]. A similar work is FlavorGraph, which can provide food recommendations based on the relationship between recipes and flavor molecules [22].

Furthermore, there are several food-related KGs that specifically address its impact on human health. Li et al. designed a KG around food safety and built a question answering service to identify the unqualified food [23]. Seneviratne et al. constructed a KG model for personal healthcare, which can be used to generate clinically relevant diet recommendations based on semantic technologies [24]. Similarly, Huang et al. proposed a scientific KG for healthcare management [25]. They applied text mining technologies and integrated food related ontologies, so that the proposed KG model can be used for diet recommendations. Another typical work is that Milanlouei et al. constructed a KG around dietary factors and cardiovascular disease [26]. It is worth noting that the system can be utilized to detect potential avenues for further extensive experimental validation. Food4healthKG, an integrated KG for food and mental health, introduced KG techniques to food interventions for mental disorders [27]. However, previous related work paid little attention to the impact of food on gut metabolisms, which is necessary for linking food and human health. They were also limited in developing a systematic question-answering system for food recommendations. Consequently, in this work, we focus more on extracting new relationships between nutrients and human diseases from the literature and constructing comprehensive and available KGs for food-to-human health recommendations.

## III. CONSTRUCTION OF KG4NH

In this section, we detail the steps involved in constructing KG4NH (Knowledge Graph for Nutrients and Human Health), encompassing data collection, triples extraction, and knowledge integration. The initial data collection process involves data acquisition and preprocessing. Secondly, triples extraction focuses on concept recognition and relation extraction, culminating in the creation of a golden corpus that underpins a high-quality relation extraction model. Thirdly, we introduce knowledge integration to amalgamate data from diverse sources and reconcile heterogeneous knowledge, including ontology and other pertinent interactions, thereby expanding our KG. Notably, before the construction process, the utilization of valuable data sources and extension methodologies from the existing KG strengthens the viability of our work. Specifically, the foundational structure

of KG4NH and a portion of the knowledge expansion process are derived from a well-applied KG [27], representing an application of a food-related KG in human health. Overall, KG4NH stands as a comprehensive system, curated through an automated construction process and the integration of a broad spectrum of knowledge.

### A. Data Collection

There are a large number of biomedical articles published on food nutrition and human health in PubMed that describe various associations between nutrients and diseases. In line with the central theme of our research, we select three pivotal keywords: “food”, “nutrition”, and “human disease”, which are important topics in the researches on food and human health. To obtain relevant literatures for subsequent text mining, we perform a thorough search utilizing these three keywords across both abstracts and the full text sourced from PubMed. Notably, PubMed employs relevant synonyms to refine the search, ensuring the capture of contextually relevant results pertaining to food and health. Additionally, all terms are linked with the “AND” operator to ensure relevance in the literature results. Subsequently, we narrow down our selection to encompass the most recent publications released between 2012 and 2022. Given that abstracts and titles typically offer crucial information while the body of articles may contain extraneous experimental details, we opt to download titles and abstracts from a total of 230,573 articles. These will serve as the foundational data for our preprocessing phase.

According to the text from data acquisition, we use the Stanford CoreNLP tokenizer to produce more fine-grained tokens [28]. For example, the “spina bifida”, a named entity of disease, is treated as a single token rather than split into two tokens. Then, we also use Stanford CoreNLP’s sentence splitting algorithm to perform sentence segmentation. After data preprocessing, we collect a series of token streams for triples extraction automatically.

### B. Triples Extraction

The triples extraction process involves two steps: concept recognition and relation extraction. After preprocessing, the token streams serve as input for concept recognition, which are utilized to recognize named entities as well as their concept classifications. Subsequently, a state-of-the-art model is employed to extract associations between nutrients and diseases from the sentences following concept recognition. To ensure high-quality relation extraction, a golden corpus is developed specifically for training the model.

**1) Concept Recognition:** The sentences in our corpus often contain multiple heterogeneous entities and their corresponding medical background knowledge, underscoring the need for an integrated pipeline capable of recognizing both nutrition and disease named entities. To achieve this, the concept identifier (CI) tool, which is proposed in the EURECA project [29] (Enabling information reuse by linking clinical research and care), is utilized in our KG to identify unique concepts and accordingly provide their semantic knowledge. Specifically, the CI tool includes a terminology compiler that allows us to use

the SNOMED-CT [30], which is provided as terms of licensed terminologies, to identify sentences containing nutrition and disease concepts as well as the semantic links between those phrases (e.g., classification of disease entities). Fig. 1(a) depicts an example process of concept recognition. Using the CI tool, two entities within the token stream are recognized and linked to the SNOMED-CT taxonomy. For instance, type 2 diabetes is recognized as a disease entity and linked to various classifications, including diabetes mellitus and disorder of glucose metabolism. As a result, 46,807 nutrient entities and 47,749 disease entities are recognized through concept recognition. The automatic pipeline of concept recognition is well employed in identifying sentences with named entities for the next relation extraction process.

**2) Relation Extraction:** Due to the large number of associations between nutrients and disease entities, it is impractical to manually extract relations. A general relation extraction model is employed in our corpus, which exhibits unfavorable performance in our corpus. Therefore, it is crucial to obtain a model specifically for automatically extracting the relations between nutrients and human diseases. To achieve that, we develop a golden corpus for fine-tuning the model. Then, a state-of-the-art model is trained for relation extraction.

Before annotating the corpus, we first investigate the descriptions of numerous relevant articles. According to the descriptions, the sentences from the text always have unclear definitions about the relationships, such as the first example sentence in Table II, there is an association between vitamin D3 and diseases, which is no clear effect or inhibition but a general relationship. Similar examples in the articles are so numerous that we only focus on extracting the general relationships. Our relation extraction task is modeled as a binary classification task. Our task is to extract which of two relation types a given entity pair in a sentence belongs to, “related” or “unrelated”. As examples shown in Table II, the label “related” indicates that there is a description of the relations between nutrition and disease in the sentence, including but not limited to affect, cure, and undefined relations. The label “unrelated” indicates that there is a lack of description for the relation between two entities or a clear unrelated relation. It is worth noting that the second example in the “unrelated” labels, vitamin D3 and ulcerative colitis only co-occurrence in this sentence and lack a direct association. Given the prevalence of such examples, we categorize such sentences under the “unrelated” label.

Following the definition of relation extraction, we subsequently develop a golden corpus for fine-tuning the model. We filter out approximately 2,000 sentences and invite three biomedical experts to perform manual annotation for the corpus. During the annotation process, each expert works independently and follows guidelines for a shared understanding of the biomedical domain. To evaluate the quality of the annotation process, a typical Inter-Annotator Agreement (IAA) metric [31] is utilized in our work. We utilize Krippendorff’s  $\alpha$  coefficient [32], a widely accepted method for measuring the degree of disagreement between annotations, which can be represented as

$$\alpha = 1 - (n - 1) \frac{O_{01}}{n_0 * n_1} \quad (1)$$



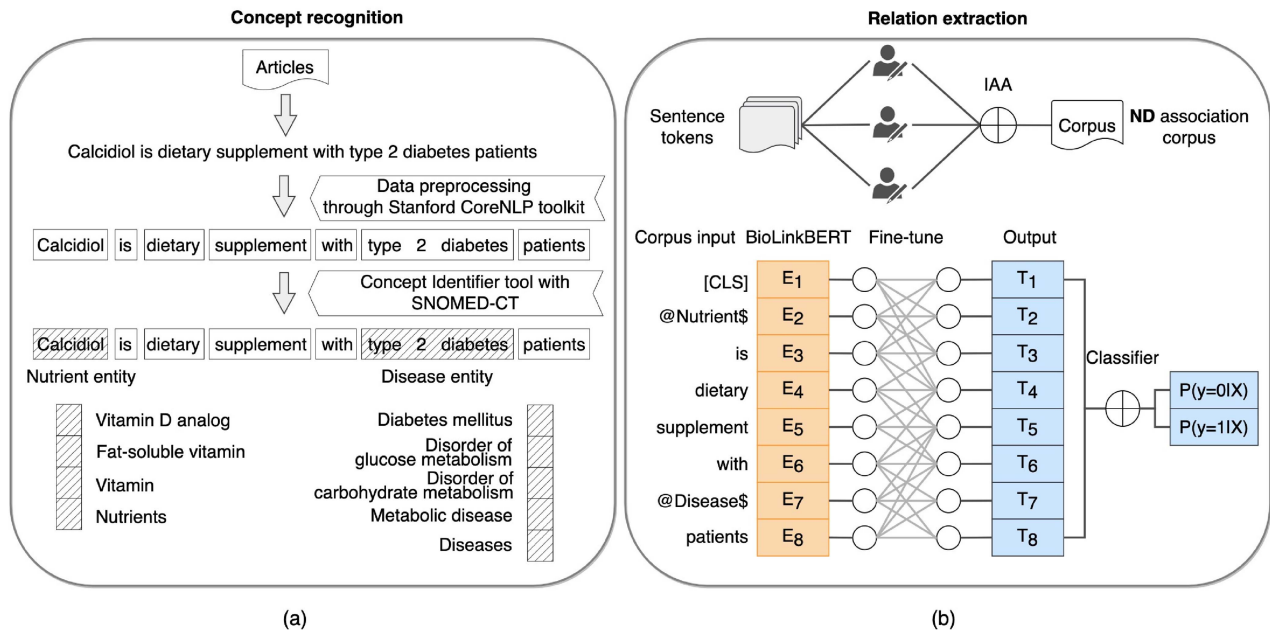


Fig. 1. Processes of triples extraction include two pipelines. (a) The pipeline of concept recognition. The input is the text from articles. (b) The pipeline of relation extraction. As the input of the model, each sentence from the corpus is converted into a series of word vector representations. The model transforms and generates an updated representation through fine-tuning, which is classified into a relation type through the final classifier.

TABLE II  
EXAMPLES OF ANNOTATED SENTENCES

Relation type	Annotated reason	Examples	PMID
Related	Nutrients have association with diseases	<b>Vitamin D3</b> and <b>beta-carotene</b> deficiency is associated with risk of <b>esophageal squamous cell carcinoma</b> .	24568502
	Nutrients reduce diseases risk	<b>Folic acid</b> (isolated or in combination with other vitamins and minerals) reduces the total risk for <b>neural tube defects</b> by 72%.	23898603
	Nutrients improve diseases risk	<b>Sucrose</b> consumption can cause <b>obesity</b> and <b>nonalcoholic fatty liver disease (NAFLD)</b> .	33612041
Unrelated	There is no explicit description of the nutritional relationship	Nutritional supplementation of <b>vitamin A</b> and health-related outcomes in patients with <b>multiple sclerosis</b> .	31232938
	There is no direct association	The aim of this study is to investigate the possible effect of <b>Vitamin D3</b> on serum ADMA levels in <b>ulcerative colitis (UC)</b> patients.	30982443
	They are not related	There was limited evidence supporting causal associations between fasting <b>glucose</b> and <b>cancer</b> .	32349989
	They are not related	Glucose and insulin showed no association with ovarian cancer.	29438162

where the  $\alpha$  ranges from 0 to 1 (indicating perfect agreement). According to common practice, an  $\alpha$  value above 0.8 is acceptable to establish the reliability of the generated data [33]. According to the annotation results of the three experts, we obtain 164 inconsistent results. Among these, 79 items are marked as related by two experts, but the third annotator disagrees, while 85 items are marked as unrelated by two experts but receive a positive annotation by a third. Finally, the coefficient  $\alpha$  value is calculated as 0.82, which is considered acceptable in common analysis. Nevertheless, we conduct a secondary annotation process to confirm complete agreement on the results. Finally, we obtain the Nutrition and Disease (ND) corpus, which consists of 234 nutrients, 278 diseases, and 1,218 interactions, including 513 related types and 705 unrelated types. Overall, our golden corpus is instrumental in fine-tuning our relation extraction model.

Several notable models have demonstrated commendable performance in relation extraction (RE), leveraging

domain-specific knowledge. For instance, models like BioBERT [34] and BlueBERT [35] have achieved competitive results. However, it's worth noting that BioBERT places greater emphasis on context features while incorporating less internal knowledge. Moreover, in the realm of nutrition-related research, the presence of external information, such as dependencies between related documents, adds an additional layer of complexity to relation extraction tasks. To address this, we employ a state-of-art model, BioLinkBERT [36], fine-tuning on our corpus. This model is well pre-trained on a wealth of PubMed abstracts and citation links, and has demonstrated impressive performance in the realm of relation extraction tasks.

The corpus is divided into training and test sets at an 8:2 ratio, which are then used to fine-tune our model. Fig. 1(b) shows the process of fine-tuning in our work. A sentence token from the corpus is represented as a series of word vectors through the pretrained language model. Then, the model generates new

vector representations through transformations and iterations. The classifier can calculate the probability of each relationship based on the result vectors and make the final relationship classification. Through the fine-tuning process, a high-quality model is obtained that is utilized to extract relations from numerous sentences automatically, which is also compared with other well-known models in further experiments. As a result, there are approximately 27,873 relations that are extracted through the RE task, including approximately 706 nutrients and 2,705 diseases. The relations can contribute to the next development of question answering applications.

### C. Knowledge Integration and Expansion

Existing knowledge in relevant studies regarding both nutrition and human disease is essential for substantiating the foundation of our knowledge graph. Initially, we leverage the foundational structure obtained from a representative KG [27]. Expanding on this, we seamlessly integrate knowledge from multiple sources into our KG4NH, filling gaps by applying structured rules to expand incomplete knowledge. Subsequently, we utilize GraphDB, a graph database tool that aligns with the resource description framework (RDF) [37], serving as the repository for our knowledge graph.

A wealth of knowledge pertaining to food, gut microbiota, and human disease exists in the current body of research. In order to facilitate knowledge integration, we invite three domain experts who specialize in the fields of nutrition, metabolism, and human health researches. Specifically, they provide suggestions for the selection of domain ontologies, which will be introduced in following description. Additionally, they contribute to defining knowledge associations after the integration of knowledge resources. In particular, they assist us in defining detailed associations between entities in our knowledge graph, as further described in the paragraph on knowledge expansion. Hence, this collaborative process ensures a comprehensive foundation for our knowledge graph.

Our integration process begins by incorporating nutrition-related associations, specifically the compositional makeup of food sourced from the FoodData Center (FDC) [38]. This encompasses the inclusive relationship between food items and their nutrient content attributes. The data from FDC, widely employed in related studies [39], plays a pivotal role in establishing connections between nutrients and food entities in our work. Then, we integrate a biological hierarchy ontology of nutrients from KEGG [40]. This resource not only provides metabolic relationships between gut microbes and nutrition but also furnishes essential information on nutrient metabolism. KEGG, known for supplying crucial metabolic associations [41], is instrumental in linking nutrients to gut metabolic processes within our system. Furthermore, our system incorporates numerous metabolic relationships between diseases and gut microbes, sourced from pertinent databases such as MENDA [42] and DMDA [43]. These additions further enrich our knowledge graph. Finally, we integrate structured ontologies, including FoodOn [44] for food ontology and SNOMED-CT for disease ontology. These widely

recognized ontologies, with their structured hierarchies and relationships, offer a nuanced and comprehensive representation of food and disease entities [45]. This expanded representation allows us to capture intricate details and interconnections that might be overlooked in a more simplified system. Overall, the integrated knowledge is encoded as the triple in our knowledge graph, comprising subjects, objects, and well-defined predicates. These predicates are determined by the real-world relationships between two entities, providing a reusable framework for knowledge expansion.

In addition to the integrated knowledge graph, there may be incomplete relations which can be augmented based on the information from neighboring nodes. To facilitate this, we implement a set of predefined rules for knowledge expansion. For instance, leveraging the transitivity property of the “subclass of” relationship allows us to automatically deduce additional connections. Specifically, if A is a subclass of B, and B is a subclass of C, then we conclude that A is also a subclass of C. This automated inference mechanism addresses situations where the original dataset lacks direct links between entities like D-glucose and monosaccharides or carbohydrates. Similarly, if there are disease-related associations with the monosaccharide class, these associations can be extended to encompass D-glucose and other instances of the class. Additionally, within the knowledge graph, a notable type of binary relation exists known as symmetrical relations. Originally, these relations are represented independently and uniquely. However, the symmetrical relation is exactly the same as its own inverse relationship. Consequently, these symmetrical properties facilitate the generation of corresponding inverse relations and triples. For example, associations like “hasChild” and “isChildOf”, “hasAssociation” and “isAssociatedWith” are symmetrical. Specifically, when A and B have an association, an equivalent triple of B being associated with A is automatically generated. This symmetrical relation supports both forward and backward reasoning in our knowledge graph. Furthermore, due to different concepts representing the same entity in our complex set of entities, instances like the nutrition “folic acid” and the metabolite “folate”, both of which refer to the same compound with biological roles. Hence, another important relation used in our system addresses lexical ambiguity. We represent the ambiguous concepts independently in the specific context and utilize the “sameAs” statement to connect instances with identical identities. This facilitates the mapping of both entities into the same compound ontologies. This approach enriches our knowledge graph with various and meaningful relationships. We effectively apply similar structured rules in the knowledge expansion process of our KG4NH, building on the framework established in prior research [27]. The demonstrated efficacy of these rules highlights their significant positive impact on the expansion process. This knowledge augmentation procedure not only bridges gaps between nodes, but also enhances the efficiency of information retrieval, thereby broadening the scope of applications based on our knowledge graph.

In conclusion, we obtain a large number of associations between nutrients and human diseases through triple extraction. Furthermore, the extensive external knowledge is integrated and

TABLE III  
THREE TYPES OF QUESTIONS AND EXAMPLES

Type	Description	Template <sup>1</sup>	Example
Description	You want to understand the meaning/introduction/ attribute of a concept/event.	How much {N} is in {F}? What is the biological hierarchy of {N}? What are the {DC} related to nutrition?	How much vitamin B is in apples? What is the biological hierarchy of glutamate? What are the congenital diseases related to nutrition?
Comparison	You want to compare/calculate/contrast two or more things.	Which food has higher {N} between {F <sub>1</sub> } and {F <sub>2</sub> }? How much {N} is in a dietary pattern that includes {F <sub>1</sub> } and {F <sub>2</sub> }? Are there shared nutrient associations between {DC <sub>1</sub> } and {DC <sub>2</sub> }?	Which food has higher cholesterol between beef and beans? How much vitamin B is in a dietary pattern that includes egg and milk? Are there shared nutrient associations between colon cancer and diabetes?
Reason	You want to find out the reasons of an event/effect.	How does {N} affect the gut microbiota? Why do {FC} affect human health? Why does food affect {D}?	How does vitamin A affect the gut microbiota? Why do soft drinks affect human health? Why does food affect anemia?

<sup>1</sup>: {N}:{Nutrients}, {F}:{Food}, {DC}:{Diseasecategory}, {D}:{Disease}, {FC}:{Food category}

expanded in our knowledge graph. As a result, our comprehensive knowledge graph contains approximately 255,017,496 triples, 154 semantic relationships, and 7,437,819 entities.

#### IV. DEVELOPMENT OF THE QA SYSTEM

In this section, we develop a question answering (QA) application based on our comprehensive knowledge graph. It is noted that the question answering process is template-based and provides answers through SPARQL query, which is supported by SPARQL1.1 queries. The development of the QA system is introduced in the following two aspects: the question design and the benchmark dataset. Specifically, the question design demonstrates three types of questions in our system: description, comparison, and reason questions. For further evaluation and application of our system, we also develop a benchmark dataset according to three application scenarios that can be used to demonstrate the effectiveness of our system.

##### A. Question Design

In alignment with prevalent topics in food and health research, which encompass nutrient analysis [46], nutrient metabolism [47], and the impact of food on human disease [48], we have identified a series of pertinent questions central to our study. These questions are categorized into three types based on a question answering system taxonomy [49]: description, comparison, and reason, as outlined in Table III. For each question type, distinct templates are devised with interchangeable slots. Consequently, leveraging the example questions provided by each template, we can extract answers through the SPARQL query executed on our knowledge graph. The ensuing sections elaborate on the specific characteristics of each question type.

First, the description type of questions aims to seek clarification on the meaning or nature of a particular concept or event. As illustrated in our templates, these questions are typically described as general interrogative sentences and tend to begin with “What is”. As an example in Table III, “How much vitamin B is in apples?”, which is a question for understanding the content attribute of vitamin B in apples. Second, a comparison type of question is designed to assess the similarities or differences between two or more things. The

question involves a two-step process, first assessing the relevant description of each entity that is mentioned in the question and then making a judgment or calculation to produce the final comparative result. Third, the reason type of questions is aimed at determining the reasons for an event or effect. This type of question generally begins with “Why” and needs to refer to external knowledge out of the questions. For example, “How does vitamin A affect the gut microbiota?”, in which the causes are answered by searching the complex associations between food nutrients and gut metabolism. According to the keywords provided in each template, the slots can be easily substituted with their corresponding entities to produce more diverse questions.

In summary, our question answering system exhibits proficiency in handling a diverse array of inquiries related to food and human health. These queries encompass critical areas of interest in current research, such as nutrient analysis, nutrient metabolism, and the impact of food on human diseases.

##### B. Benchmark Dataset

We realize the significance of assessing our system’s effectiveness in engaging users, considering various question types and user preferences. To achieve this, we meticulously assemble a benchmark dataset consisting of 120 questions. These questions are thoughtfully designed to cater to the distinct perspectives of three pivotal user groups: patients, medical practitioners (including doctors and nutritionists), and research professionals. Every question in our dataset undergoes a rigorous evaluation supervised by three experts who share similar backgrounds in clinical practice, nutrition science, and biomedical research. Their involvement is twofold. Firstly, they contribute to the design of questions, ensuring the questions are user-centric and appropriately categorized for specific user groups. Secondly, they participate in providing standard answers by reviewing relevant literature. The expertise of these experts enhances the credibility of our benchmark dataset. In this section, we provide a comprehensive overview of the process involved in constructing the benchmark dataset for question answering. Additionally, we present select illustrative examples, along with their corresponding SPARQL queries, providing tangible insights into the breadth of questions addressed by our system.

The first perspective of questions are designed for patients with various disorders, which aim to seek solutions for improving their health according to the answers. For example, there is the comparison question, “Which food has higher cholesterol between beef and beans?”. This question aims to compare the content of cholesterol in two different foods and return the one with higher cholesterol. To answer this question, we create a tailored SPARQL query as follows.

```
select distinct ?food?weight?nutrient where {
  ?nu rdfs:label ?nutrient;
  pq:hasNutrientWeights ?flunit.
  ?fl pq:hasNutrients ?nu;
  rdfs:label ?food;
  pq:hasNutrientWeights ?flunit.
  ?flunit rdf:value ?weight.
  filter regex(?nutrient, ‘‘cholesterol’’, ‘‘i’’)
  { filter regex(?food, ‘‘beef’’, ‘‘i’’) }
union
  { filter regex(?food, ‘‘beans’’, ‘‘i’’) }
order by desc(?weight) limit 1
```

**Listing 1.** SPARQL query for the comparison question: “Which food has higher {N} between {F<sub>1</sub>} and {F<sub>2</sub>}”, N = cholesterol, F<sub>1</sub> = beef, F<sub>2</sub> = beans

The example identifies multiple foods labeled “beef” and “beans”. The query can find the maximum cholesterol amounts of each food list and compare them to acquire the one with higher cholesterol. The question helps users obtain more focused results by comparing two objects. Furthermore, we have another similar example that, while not part of our predefined templates, is included in our benchmark dataset: “What diets are associated with obesity management”. This question seeks information about diets that are specifically linked to managing obesity. In essence, these patient-oriented questions are designed to provide more practical and suitable suggestions or solutions, enabling them to make more informed decisions regarding their health.

The second view of questions is designed for doctors or nutritionists who seek a deeper understanding of the relationships between nutrition and health from a health care professional’s perspective. These questions hold significant importance, as healthcare professionals play a pivotal role in guiding treatment and offering dietary advice to their patients. For example, consider the question: “How much vitamin B is in a dietary pattern that includes egg and milk?”. Here, a dietary pattern refers to the habitual consumption behaviors and combination of foods for an individual. This question establishes a parameter around the dietary pattern and aims to ascertain the potential intake of vitamin B. This information can serve as a valuable reference for further dietary recommendations. Another example, “Why do soft drinks affect human health?”, is an instance of a reason question that focuses on the detailed impact of a specific food (soft drinks) on various human health aspects. Such questions are commonly of interest to nutritionists in clinical applications, as they provide valuable insights into the potential effects of specific foods on uncertain human diseases. In summary, these questions aim to assist users, particularly doctors and nutritionists, in making better-informed recommendations by gaining a comprehensive understanding of the underlying dynamics.

```
select distinct ?foodname?nutrientname
  ?bacterianame?diseasename where {
  ?s ztone:hasAnnotations ?o.
  ?o ztone:hasSource ?source;
  ztone:hasText ?text;
  ztone:hasAnnotation ?annotation.
  ztone:hasAnnotation ?annotation2.
  ?annotation npq:hasAssociation ?annotation2;
  ztone:hasLabel ?nutrientname;
  owl:sameAs ?compound.
  ?annotation2 ztone:hasLabel ?diseasename.
  ?food pq:hasNutrients ?nutrient;
  rdfs:label ?foodname;
  rdf:type ?foodclass.
  ?nutrient owl:sameAs ?compound.
  ?bacteria npq:hasMetabolites ?compound;
  rdfs:label ?bacterianame.
  filter regex(?diseasename, ‘‘anemia’’, ‘‘i’’)
```

**Listing 2.** SPARQL query for the reason question: “Why does food affect {D}?”, N = anemia

The third perspective of questions reflects the inquisitive nature of researchers who are interested in studying intricate relationships. There is an example, “Why does food affect anemia?”, that is an instance of a reason question for exploring the impact of food on the development or management of anemia. These questions are specifically designed to investigate potential food factors, determine the underlying mechanisms, and understand the associated events and processes. The corresponding SPARQL query is as follows.

The example allows researchers to retrieve information about the processes of food ingestion, as well as the associated nutrients and gut microbes. The other example is “Are there shared gut microbial associations between obesity and diabetes?”. This question aims to explore the presence of common gut microbial metabolisms between these two diseases. In summary, the answers to these questions contribute to users’ understanding of the intricate associations and further enable researchers to apply the implications as auxiliary suggestions for clinical experiments.

The benchmark dataset comprises a diverse range of questions beyond the predefined templates and is designed to cater to the different perspectives of three user groups. To ensure accuracy and reliability, the standard answers for these questions have been carefully examined and revised by domain experts. In the subsequent stages of our work, this benchmark dataset serves as a valuable resource to validate the effectiveness of our knowledge graph in question answering applications. Each well-structured question within the dataset can be answered by utilizing SPARQL queries on our comprehensive knowledge graph. This ensures that the performance and utility of our knowledge graph-based question answering system can be thoroughly assessed and verified using a diverse set of real-world questions.

## V. EXPERIMENTS AND DISCUSSION

In this section, we present a series of experiments to validate the efficacy of our proposed approach. Initially, we conduct contrastive experiments aimed at comparing the performance of our relation extraction model with three widely-used models.



TABLE IV  
PERFORMANCE OF CONTRASTIVE MODELS ON THE ND CORPUS

Models	Precision	Recall	F1 score
BioBERT <sub>without-finetune</sub>	0.41	0.83	0.55
BlueBERT	0.47	<b>0.89</b>	0.59
BioBERT	0.86	0.82	0.84
<b>Our method</b> <sub>BioLinkBERT<sub>base</sub></sub>	<b>0.92</b>	0.81	<b>0.86</b>
<b>Our method</b> <sub>BioLinkBERT<sub>large</sub></sub>	<b>0.92</b>	0.80	0.85

Subsequently, an interpretation study is designed to demonstrate the results and significance of triples extraction. To further illustrate the effectiveness of knowledge integration within the knowledge graph in the question answering application, we design an ablation study to qualify the performance of the application on different scales of knowledge graphs. Additionally, we engage in a comparative discussion, evaluating the capabilities of our question-answering system from four distinct perspectives and drawing comparisons with the conversational model ChatGPT. Finally, we devise a methodology for assessing the quality of the data structure within our Knowledge Graph. This approach holds the potential for extension into similar endeavors, establishing a valuable contribution to the broader research landscape.

### A. Contrastive Experiments

According to the proposed method, we proceed to fine-tune our relation extraction model, which is based on BioLinkBERT, leveraging both its base version (110 M params) and larger version (340 M params). To accomplish this, we utilize a Linux server equipped with a single Nvidia Geforce GTX 3090 (24 GB) GPU. For the fine-tuning process, we carefully set parameters tailored to our task, including selecting a learning rate from the options  $5e-6$ ,  $1e-5$ ,  $2e-5$ ,  $3e-5$ , and  $5e-5$ , choosing a batch size from 16, 32, or 64, and determining a maximum sequence length from either 128 or 256. We also experiment with varying the number of fine-tuning epochs, ranging from 3 to 10. It's worth noting that fine-tuning the base version of BioLinkBERT proves to be more computationally efficient compared to the larger version. In total, fine-tuning BioLinkBERT on our ND corpus takes less than two hours and spans 10 epochs to achieve optimal performance. Ultimately, our models demonstrate peak performance under specific hyperparameters: a learning rate of  $5e-5$ , a batch size of 32, 10 epochs, and a maximum sequence length of 128. To facilitate a fair comparison with other popular models, we use these same parameters for fine-tuning the models on our ND corpus. In terms of evaluation, we employ precision, recall, and F1 score as the metrics for our contrastive experiments [50].

As depicted in Table IV, our approach outperforms other relevant models in the realm of relation extraction. Across all versions of BioLinkBERT, our method attains the highest precision score, registering at 0.92. However, BlueBERT demonstrates exceptional recall, a strength likely stemming from its pre-training on PubMed abstracts and clinical notes, heavily relying on annotations. This suggests its proficiency in capturing a larger portion of true positives, although this may come at the expense of some false positives. For our specific objective of performance

in relation extraction, striking a balance between precision and recall is paramount. In this regard, our model leveraging the base version of BioLinkBERT achieves the most harmonized performance, yielding an F1 score of 0.86. It's noteworthy that while both the base and large versions of BioLinkBERT exhibit similar performance, the latter necessitates three times the duration for fine-tuning compared to the former. Overall, these outcomes collectively affirm the precision and efficacy of our relation extraction method in accurately delineating associations between nutrients and diseases.

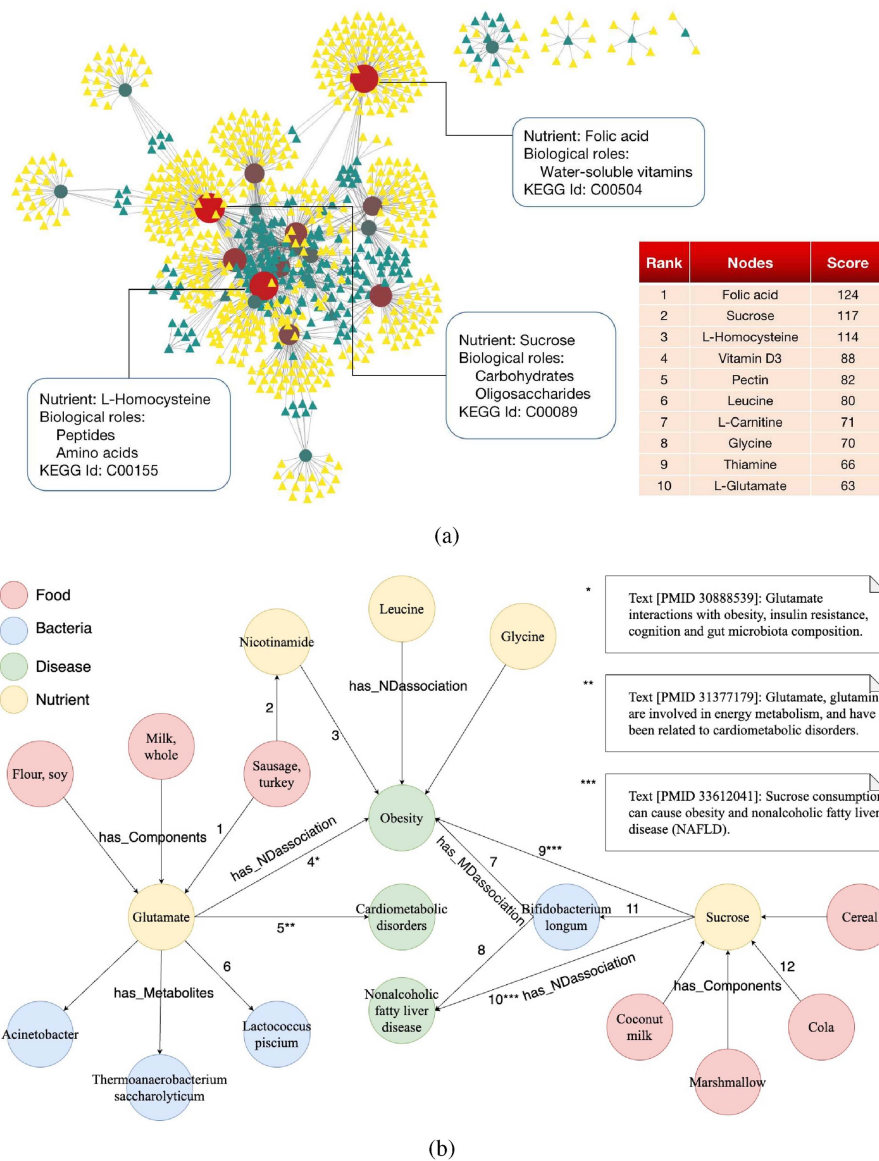
### B. Interpretation Study

In this section, we calculate the importance of nodes in a heterogeneous graph composed of relation extraction results, where nodes represent nutrients or diseases, and edges represent relations between them. Furthermore, to demonstrate the significance of the results from relation extraction in the question answering application, we discuss a multipath interpretation scheme for answering the same question.

We initiate a relation extraction task to extract a substantial volume of connections between nutrients and diseases. Following the elimination of duplicate objects, we utilize the cytoHubba interface to compute the Matthews Correlation Coefficient (MCC) score for each node. These results are subsequently visualized using the Cytoscape tool [51], as depicted in Fig. 2(a). In the left network, the color and size of the nodes signify their significance within the graph. Several nutrients, including sucrose, L-homocysteine, and vitamin D, emerge as pivotal entities, a fact further substantiated by their biological implications in gut metabolism as outlined in the KEGG knowledge base. For instance, folic acid stands out as the most critical node in our findings. Extensive literature establishes folic acid as a heavily researched food supplement in relation to human health. Its presence can be attributed to rice, pasta, bread and it could have effects on pregnancy, a baby's brain (anencephaly), and a baby's spine (spina bifida) [52]. Additionally, the right table of the figure lists the ten most important nodes along with their MCC scores. In this table, sucrose also holds importance in the results due to its multiple relations with human diseases. As literature reports, sucrose is a common nutrient found in a range of foods, such as molasses, sweet potato, orange juice, etc., and is also a natural nutrient with implications for mood swings, weight gain, insulin sensitivity, etc [53]. In general, the interpretation experiment showcases the effectiveness of our method for relation extraction, wherein the significance of nutrients also underscores the indispensable role of food supplements in human health.

Furthermore, we discuss a multipath interpretation scheme by employing the results of relation extraction to the question answering application. Take a reason question as an example, "Why does food affect obesity?". Based on the answers to the question, a portion of the subgraph centered around obesity is visualized in Fig. 2(b), which illustrates the associations among obesity, food nutrients, and gut microbes. According to the figure, multiple pathways can be obtained between food and obesity. For example, the impact of Turkey sausage on obesity





**Fig. 2.** (a) The visualized results of relation extraction. In the left network, the triangle signs represent human disease entities, while circle signs represent nutrients. The size and color of each node, ranging from yellow to red, indicate their importance in the network. The right table shows the 10 most important nodes in the graph based on MCC scores. (b) The visualized results of a question, “Why does food affect obesity?”. There are four entity types in the figure: food (pink circles), bacteria (blue circles), disease (green circles), and nutrients (yellow circles). Their relations are specifically defined based on the type of the head and the tail entities.

can be attributed to different nutrients, as shown by the connections from “1” to “4” and from “2” to “3”. Additionally, there is a complex pathway involving gut microbes, with the route from “12” through “11” to “7”. This pathway is inferred from the KG, which needs to be further verified. In general, the results show that our system can provide multipath interpretations based on a simple question by going from a one-to-one relation to a many-to-many network.

### C. Ablation Study

To ascertain the effectiveness of the knowledge integration process on our knowledge graph, we conduct an ablation study. We generate answers for the questions in the benchmark dataset

by employing SPARQL queries on our comprehensive knowledge graph. These generated answers are then compared with the standard answers using a set of evaluation metrics, including accuracy, mean precision, recall, and F1 score. Here, accuracy signifies the average similarity between generated answers and standard answers, measured using cosine similarity.

We mainly integrate five key components during knowledge integration process: ND associations, Food Ontology (FO), Disease Ontology (DO), Microbes and Diseases (MD) associations, and other integrated knowledge sources. The set of ND associations underlines the importance of our work in generating triples. Additionally, FO and DO serve as representative ontologies in the context of food and human health, which find extensive application in our question-answering system. The MD

**TABLE V**  
ABLATION STUDY OF QUESTION ANSWERING RESULTS ON BENCHMARK DATASET

Knowledge Graph <sup>†</sup>	Accuracy	Precision	Recall	F1 score
KG4NH <sub>complete</sub>	<b>0.68</b>	<b>0.65</b>	<b>0.58</b>	<b>0.61</b>
" -ND	0.47	0.52	0.36	0.43
-FO	0.37	0.41	0.29	0.34
-DO	0.57	0.52	0.48	0.50
-MD	0.62	0.59	0.52	0.55
-EK	0.21	0.18	0.09	0.12

<sup>†</sup> "-": Removing the latter content; ND: Nutrients and Diseases associations; FO: Food Ontology; DO: Disease Ontology; MD: Microbes and Diseases associations; EK: other Existing integrated Knowledge (more than FO, DO, MD, etc).

associations play a crucial role in illustrating the impact of gut microbiota, hence their inclusion in our ablation study. Finally, the broader integrated knowledge encompasses more than FO, DO, and MD associations, encompassing nutrient classifications and associations between nutrients and microbes. These are harmoniously represented through existing knowledge (EK).

The results of the ablation study are demonstrated in Table V. As shown, the removal of each part causes different degrees of loss of the effects, of which the EK part causes the biggest loss after removal from our knowledge graph. The possible cause of this effect is that our questions in the benchmark dataset are commonly required to retrieve the relevant relationships on our knowledge graph, which is also the largest component of our knowledge graph. The effects of the knowledge graph without ND associations also suffer a larger loss, which is mainly shown in the questions related to the associations between nutrients and human health. It also shows that the triple generation results contribute to the efficient implementation of question answering applications. The other FO part also promotes the effectiveness of our question-answering task because the benchmark questions are biased toward food-related relationships. In another aspect, the effect of the question-answering task suffers mild loss without the DO and MD associations, which is also due to the bias of the question design. In other words, compared with the other incomplete knowledge graphs, our well-constructed knowledge graph can obtain a better effect on the question answering task.

#### D. Comparative Discussion

To bridge the gap in tailored knowledge graphs for food and human health-related question answering, we engage in a comparative analysis between our KG4NH and the renowned conversational language model, ChatGPT. While ChatGPT's adaptability spans to applications in food and human health question answering [54], we recognize that making direct comparisons can be intricate due to the divergence in domains. Therefore, our discussion centers on delineating the unique capabilities and attributes of each system across four pivotal criteria [55].

**Correctness:** We evaluate the correctness of the answers by comparing them to the ground truth in the benchmark dataset. As shown in Table V, our comprehensive knowledge graph consistently provides precise answers. In contrast, ChatGPT may require additional human prompting for desired outcomes. Moreover, its responses may be broad descriptions that are challenging to directly benchmark against ground truth.

**Determinism:** To assess the determinism of the answers, we conduct a comparative analysis by running each question three times on both systems and examining the differences in the answers. Our knowledge graph, utilizing identical SPARQL queries, maintains uniformity. Conversely, ChatGPT's responses may vary due to differences in training data.

**Robustness:** We assess the robustness by introducing variations in the questions, including adding grammar errors and word substitutions. ChatGPT exhibits remarkable adaptability, even in the face of notable grammar mistakes. In comparison, our KG4NH faces challenges in consistently generating desired outcomes, particularly when core elements like predicates or spelling are affected. Our SPARQL queries, based on templates, may limit their tolerance for minor query errors.

**Interpretability:** Given the medical context, result interpretability is a crucial metric. We evaluate it through manual comparison with the benchmark dataset and expert assessments. While our system's structured knowledge from SPARQL queries yields more specific biological insights through knowledge inference, it may lack the natural language descriptions provided by ChatGPT. Nevertheless, the traceability of our results is supported by the extensive knowledge graph and its interconnected data, ensuring credibility. In contrast, ChatGPT offers more readable and general responses based on its expansive language model, but the interpretability may be restricted.

In conclusion, our question answering system excels in correctness and determinism compared to the ChatGPT. However, improving the system's robustness remains a priority for future work. Integrating the conversational large language model with our system could enhance the interpretability and readability of question answering results.

#### E. Quality Assessment

To assess the quality of our KGs' data structure, we design three constraint schemata using the Shapes Constraint Language (SHACL) framework [56]. The rules focus on basic food properties, numerical type records, and class hierarchy. The three constraint schema are summarized as follows:

- 1) Each nutrient in food has a unique weight.
- 2) The sum of the nutrient amount in each food is no more than 100 g.
- 3) All classes and instances match their definition.

```
prefix sh: <http://www.w3.org/ns/shacl#>
prefix ex: <http://example.com/ns#>

ex: TargetObjectsOfHasNutrientWeightShape
a sh:NodeShape ;
sh:targetObjectsOf npq:hasNutrientWeights ;
sh:property [
sh:path rdf:value ;
sh:minCount 1 ;
sh:maxCount 1 ;
sh:datatype xsd:double ; ] .
```

**Listing 3.** Each nutrient in food sample should have a unique weight.

The first rule is designed to restrict the number of values for each nutrient based on the basic food properties. Each food consists of multiple nutrients, each with a unique weight value.

The second rule is designed specifically for our KGs: the amount of each food sample is 100 grams from the original dataset (defined by FoodData Center) [38]. When we transform them into KGs, there are some missing compositions that can hardly be randomly filled in, so we calculate the sum of all nutrients in each food and set a maximum of 100 grams. This rule is designed to verify whether there are abnormal values in the food and nutrients, which have a close relation with the answers to description and comparison questions in the case studies. The third rule is about the definition of the concept, which is classified into classes or instances. For example, “Metabolic disease” and “Beverage” are subclasses of disease and food and will be subclasses of “OWL: Classes”. Items such as “Major depression disorder” and “cola” are instances of these classes. The constraint is designed in case of confusion among concepts to ensure more accurate results.

Examining the results of the quality assessment through the SHACL test, some errors are indeed discovered through the tests. For example, the first test yields an outstanding report, which refers to multiple weight values of a nutrient in one food. Upon further investigation during the second verification, it was identified that these errors arose during the data import process. Except for this, the third rule helps identify missing values about the classes definition in our KGs. We apply the third test to food entity assessment, and the results report a large number of food instances that have no type in the KGs. This means that there are many concepts that have relations with nutrients and have other similar food properties, which have not yet been recognized as food entities in our KGs. Relying on the quality assessment, various abnormal results have been found in our system, which could have adverse effects on the applications. In this case, we revised the abnormal records in our KGs and supplemented the incomplete triples according to the quality assessment. In general, the quality assessment is applied throughout our whole work and can also be extended in future work.”

## VI. CONCLUSION AND FUTURE WORK

In conclusion, we herein develop a comprehensive, constantly updated knowledge graph for dietary nutrition and human health, in which the automatic pipelines of triples extraction and knowledge integration indeed contribute to the construction of KGs. Based on the knowledge graph, a query-based question answering system is built to produce credible answers to three types of questions. Subsequently, we provide a benchmark dataset based on three different scenarios for the evaluation of our system. Finally, five well-designed experiments are conducted to verify the effectiveness of our work. Overall, our work demonstrates a comprehensive and systematic knowledge graph on dietary nutrition and human health. This resource aims to empower researchers, clinicians, and patients with insights into the intricate relationships between food and human health. The systematic approach results in positive implications for exploring diets in maintaining and improving human health.

In consideration of our proposed method, it is crucial to acknowledge its limitations. One notable constraint is the dependency of the relation extraction model’s accuracy on our

manually annotated corpus. While we’ve taken measures to ensure the quality of the corpus, there is still room for improvement in future research. We’re actively exploring the integration of large language models and unsupervised learning techniques, which hold potential for further enhancing relation extraction accuracy. Furthermore, we recognize that our question answering system, while effective, may not cover all possible questions related to food topics with the current three-question type classification. In future work, we plan to refine this classification to encompass a wider range of inquiries in the realm of food and health. This refinement will lead to a broader coverage of food-related questions, ultimately enhancing the system’s usefulness. Additionally, we’re aware of the limitations of our template-based question answering system, particularly in its capacity for natural language conversations. We’re dedicated to incorporating state-of-the-art natural language understanding technologies to enhance the system’s adaptability and responsiveness.

## ACKNOWLEDGMENT

All authors contribute to the writing and review of the final manuscript.

## REFERENCES

- [1] T. G. Dinan et al., “Feeding melancholic microbes: Mynewgut recommendations on diet and mood,” *Clin. Nutr.*, vol. 38, no. 5, pp. 1995–2001, 2019.
- [2] J. C. Clemente, L. K. Ursell, L. W. Parfrey, and R. Knight, “The impact of the gut microbiota on human health: An integrative view,” *Cell*, vol. 148, no. 6, pp. 1258–1270, 2012.
- [3] J. M. Posma et al., “Nutriome–metabolome relationships provide insights into dietary intake and metabolism,” *Nature Food*, vol. 1, no. 7, pp. 426–436, 2020.
- [4] M. A. Martínez-González and A. Sánchez-Villegas, “Food patterns and the prevention of depression,” *Proc. Nutr. Soc.*, vol. 75, no. 2, pp. 139–146, 2016.
- [5] L. A. Bolte et al., “Long-term dietary patterns are associated with pro-inflammatory and anti-inflammatory features of the gut microbiome,” *Gut*, vol. 70, no. 7, pp. 1287–1298, 2021.
- [6] F. F. Anhe et al., “Type 2 diabetes influences bacterial tissue compartmentalisation in human obesity,” *Nature Metab.*, vol. 2, no. 3, pp. 233–242, 2020.
- [7] D. Mozaffarian et al., “Food compass is a nutrient profiling system using expanded characteristics for assessing healthfulness of foods,” *Nature Food*, vol. 2, no. 10, pp. 809–818, 2021.
- [8] S. Zheng et al., “PharmKG: A dedicated knowledge graph benchmark for biomedical data mining,” *Brief. Bioinf.*, vol. 22, no. 4, 2021, Art. no. bbaa344.
- [9] V. D. Badal et al., “Challenges in the construction of knowledge bases for human microbiome-disease associations,” *Microbiome*, vol. 7, no. 1, pp. 1–15, 2019.
- [10] J. Hu, R. Lepore, R. J. Dobson, A. Al-Chalabi, D. M. Bean, and A. Iacoangeli, “Dglinker: Flexible knowledge-graph prediction of disease–gene associations,” *Nucleic Acids Res.*, vol. 49, no. W1, pp. W153–W161, 2021.
- [11] Q. Liang, W. Zhang, H. Wu, and B. Liu, “LncRNA-disease association identification using graph auto-encoder and learning to rank,” *Brief. Bioinf.*, vol. 24, no. 1, 2023, Art. no. bbac539.
- [12] S. Wang et al., “KG4SL: Knowledge graph neural network for synthetic lethality prediction in human cancers,” *Bioinformatics*, vol. 37, no. Supplement\_1, pp. i418–i425, 2021.
- [13] C. Zhu, Z. Yang, X. Xia, N. Li, F. Zhong, and L. Liu, “Multimodal reasoning based on knowledge graph embedding for specific diseases,” *Bioinformatics*, vol. 38, no. 8, pp. 2235–2245, 2022.
- [14] D. N. Nicholson and C. S. Greene, “Constructing knowledge graphs and their biomedical applications,” *Comput. Struct. Biotechnol. J.*, vol. 18, pp. 1414–1428, 2020.



- [15] R. T. Sousa, S. Silva, and C. Pesquita, "Evolving knowledge graph similarity for supervised learning in complex biomedical domains," *BMC Bioinf.*, vol. 21, no. 1, pp. 1–19, 2020.
- [16] G. Cenikj, B. K. Seljak, and T. Eftimov, "Foodchem: A food-chemical relation extraction model," in *Proc. IEEE Symp. Ser. Comput. Intell.*, 2021, pp. 1–8.
- [17] J. T. Morton et al., "Learning representations of microbe-metabolite interactions," *Nature Methods*, vol. 16, no. 12, pp. 1306–1314, 2019.
- [18] Y. Chen, A. Subburathinam, C.-H. Chen, and M. J. Zaki, "Personalized food recommendation as constrained question answering over a large-scale food knowledge graph," in *Proc. 14th ACM Int. Conf. Web Search Data Mining*, 2021, pp. 544–552.
- [19] H. Yu, H. Li, D. Mao, and Q. Cai, "A relationship extraction method for domain knowledge graph construction," *World Wide Web*, vol. 23, pp. 735–753, 2020.
- [20] W. Min, C. Liu, L. Xu, and S. Jiang, "Applications of knowledge graphs for food science and industry," *Patterns*, vol. 3, no. 5, pp. 1–21, 2022.
- [21] S. Haussmann et al., "Foodkg: A semantics-driven knowledge graph for food recommendation," in *Proc. Int. Semantic Web Conf.*, 2019, pp. 146–162.
- [22] D. Park, K. Kim, S. Kim, M. Spranger, and J. Kang, "Flavorgraph: A large-scale food-chemical graph for generating food representations and recommending food pairings," *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, 2021.
- [23] L. Qin, Z. Hao, and L. Zhao, "Food safety knowledge graph and question answering system," in *Proc. 7th Int. Conf. Inf. Technol.*, 2019, pp. 559–564.
- [24] O. Seneviratne, J. Harris, C.-H. Chen, and D. L. McGuinness, "Personal health knowledge graph for clinically relevant diet recommendations," in *Proc. 3rd Annu. Automat. Knowl. Base Construction Conf.*, 2021.
- [25] L. Huang, C. Yu, Y. Chi, X. Qi, and H. Xu, "Towards smart healthcare management based on knowledge graph technology," in *Proc. 8th Int. Conf. Softw. Comput. Appl.*, 2019, pp. 330–337.
- [26] S. Milanlouei, G. Menichetti, Y. Li, J. Loscalzo, W. C. Willett, and A.-L. Barabási, "A systematic comprehensive longitudinal evaluation of dietary factors associated with acute myocardial infarction and fatal coronary heart disease," *Nature Commun.*, vol. 11, no. 1, 2020, Art. no. 6074.
- [27] C. Fu, Z. Huang, F. van Harmelen, T. He, and X. Jiang, "Food recommendation for mental health by using knowledge graph approach," in *Proc. Health Inf. Sci. 11th Int. Conf.*, 2022, pp. 231–242.
- [28] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky, "The stanford corenlp natural language processing toolkit," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics: Syst. Demonstrations*, 2014, pp. 55–60.
- [29] H. Kondylakis et al., "Patient empowerment for cancer patients through a novel ICT infrastructure," *J. Biomed. Inform.*, vol. 101, 2020, Art. no. 103342.
- [30] K. Donnelly et al., "Snomed-CT: The advanced terminology and coding system for ehealth," *Stud. Health Technol. Inform.*, vol. 121, pp. 279–290, 2006.
- [31] R. Artstein, "Inter-annotator agreement," in *Handbook of Linguistic Annotation*. Berlin, Germany: Springer, 2017, pp. 297–313.
- [32] K. Krippendorff, "Computing Krippendorff's alpha-reliability," Annenberg Sch. Commun., Philadelphia, PA, USA, Departmental Papers, 2011.
- [33] J. Carletta, "Assessing agreement on classification tasks: The Kappa statistic," *Comput. Linguist.*, vol. 22, pp. 249–254, 1996.
- [34] J. Lee et al., "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, 2020.
- [35] Y. Peng, S. Yan, and Z. Lu, "Transfer learning in biomedical natural language processing: An evaluation of bert and ELMO on ten benchmarking datasets," in *Proc. 18th BioNLP Workshop Shared Task*, 2019, pp. 1–8.
- [36] M. Yasunaga, J. Leskovec, and P. Liang, "LinkBERT: Pretraining language models with document links," in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics*, 2022, pp. 8003–8016.
- [37] R. H. Güting, "Graphdb: Modeling and querying graphs in databases," in *Proc. 20th Int. Conf. Very Large Data Bases*, 1994, pp. 297–308.
- [38] K. McKillop, J. Harnly, P. Pehrsson, N. Fukagawa, and J. Finley, "Fooddata central, USDA's updated approach to food composition data systems," *Curr. Develop. Nutr.*, vol. 5, no. Supplement\_2, pp. 596–596, 2021.
- [39] N. K. Fukagawa, K. McKillop, P. R. Pehrsson, A. Moshfegh, J. Harnly, and J. Finley, "USDA's fooddata central: What is it and why is it needed today?," *Amer. J. Clin. Nutr.*, vol. 115, no. 3, pp. 619–624, 2022.
- [40] M. Kanehisa, M. Furumichi, Y. Sato, M. Ishiguro-Watanabe, and M. Tanabe, "KEGG: Integrating viruses and cellular organisms," *Nucleic Acids Res.*, vol. 49, no. D1, pp. D545–D551, 2021.
- [41] P. Smith-Brown, M. Morrison, L. Krause, and P. Davies, "Dairy and plant based food intakes are associated with altered faecal microbiota in 2 to 3 year old australian children," *Sci. Rep.*, vol. 6, no. 1, 2016, Art. no. 32385.
- [42] J. Pu et al., "Menda: A comprehensive curated resource of metabolic characterization in depression," *Brief. Bioinf.*, vol. 21, no. 4, pp. 1455–1464, 2020.
- [43] C. Fu, R. Zhong, X. Jiang, T. He, and X. Jiang, "An integrated knowledge graph for microbe-disease associations," in *Proc. Int. Conf. Health Inf. Sci.*, 2020, pp. 79–90.
- [44] D. M. Dooley et al., "Foodon: A harmonized food ontology to increase global food traceability, quality control and data integration," *npj Sci. Food*, vol. 2, no. 1, pp. 1–10, 2018.
- [45] G. Popovski, B. K. Seljak, and T. Eftimov, "A survey of named-entity recognition methods for food information extraction," *IEEE Access*, vol. 8, pp. 31586–31594, 2020.
- [46] D. Osborne et al., *The Analysis of Nutrients in Foods*. Cambridge, MA, USA: Academic Press, 1978.
- [47] Y. Chen, M. Michalak, and L. B. Agellon, "Focus: Nutrition and food science: Importance of nutrients and nutrient metabolism on human health," *Yale J. Biol. Med.*, vol. 91, no. 2, pp. 95–103, 2018.
- [48] D. Mafra, N. A. Borges, B. Lindholm, P. G. Shiels, P. Evenepoel, and P. Stenvinkel, "Food as medicine: Targeting the uraemic phenotype in chronic kidney disease," *Nature Rev. Nephrol.*, vol. 17, pp. 153–171, 2021.
- [49] V. Bolotova, V. Blinov, F. Scholer, W. B. Croft, and M. Sanderson, "A non-factoid question-answering taxonomy," in *Proc. 45th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2022, pp. 1196–1207.
- [50] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *Proc. Eur. Conf. Inf. Retrieval*, 2005, pp. 345–359.
- [51] C.-H. Chin, S.-H. Chen, H.-H. Wu, C.-W. Ho, M.-T. Ko, and C.-Y. Lin, "cytoHubba: Identifying hub objects and sub-networks from complex interactome," *BMC Syst. Biol.*, vol. 8, no. 4, pp. 1–7, 2014.
- [52] K. S. Crider, L. B. Bailey, and R. J. Berry, "Folic acid food fortification—its history, effect, concerns, and future directions," *Nutrients*, vol. 3, no. 3, pp. 370–384, 2011.
- [53] M. Daly, "Sugars, insulin sensitivity, and the postprandial state," *Amer. J. Clin. Nutr.*, vol. 78, no. 4, pp. 865S–872S, 2003.
- [54] P. Niszczota and I. Rybicka, "The credibility of dietary advice formulated by chatGPT: Robo-diets for people with food allergies," *Nutrition*, vol. 112, 2023, Art. no. 112076.
- [55] R. Omar, O. Mangukiyi, P. Kalnis, and E. Mansour, "ChatGPT versus traditional question answering for knowledge graphs: Current status and future directions towards knowledge graph chatbots," 2023, *arXiv:2302.06466*.
- [56] H. Knublauch and D. Kontokostas, *Shapes Constraint Language (SHACL)*, W3C Candidate Rec., 2017.