Construction Method of National Food Safety Standard Ontology

Die Hu, Chunyi Weng, Ruoqi Wang, Xueyi Song, and Li Qin(✉)

College of Informations, Huazhong Agricultural University, Wuhan 430070, China
qinli@mail.hzau.edu.cn

Abstract. The food safety standard ontology is the schema layer of top-bottom building method for constructing the food safety knowledge graph, which can effectively ensure the professionalism and effectiveness of the knowledge graph. However, there are few relevant research in recent literature, which also restricts the application of knowledge graph in the food safety field. In order to solve this problem, this paper proposes the construction method of food safety standard ontology based on the national food safety standards. Firstly, we built the class framework of the food safety standards according to the seven steps of ontology construction, and then entities, attributes and relations between entities are extracted from the national food safety standards by employing the rule-based knowledge extraction algorithm. Finally, we import those entities, attributes and relations into the class framework to complete the ontology. According to the experimental results of our entity mapping, the food safety standard ontology can describe the mainly important concepts, terms, operation process and their relation in the standards. So it will effectively support the construction and integration of knowledge graph.

Keywords: Food safety standards · Domain ontology · Class framework · Knowledge extraction algorithm

1 Introduction

In recent years, food safety issues often attract people's attention. As food safety issues are closely related to the life of every person, frequent outbreaks of food safety accidents will affect the people's expectations for the future economic and social development, and it will affect the long-term stability of the society [1]. Food safety problems mainly include: pathogenic microorganism pollution, pesticide and veterinary drug residues, heavy metal and mycotoxin pollution, illegal and adulterated use of food additives [2]. To solve these problems, many countries formulated relevant standards to regulate the food from food raw materials, processing, packaging, transportation and sales processes. However, after years of revision and supplement, the content of food safety standards is

This document is supported by the China Fundamental Research Funds for the Central Universities (No: 2662022XXYJ001, 2662022JC004).

C. Yu et al. (Eds.): GPC 2022, LNCS 13744, pp. 50–66, 2023.
https://doi.org/10.1007/978-3-031-26118-3_4

messy and complex, if regulators and producers do not fully understand food safety standards, they will "ignorant" use some harmful technology and production environment for production, and make it into the market, causing harm to consumers [3]. For consumers, the scale of Internet data has grown exponentially and disorderly, which makes the acquisition and utilization of food safety knowledge become more difficult, all these create food safety information asymmetry between the government and consumers. Therefore, it is necessary to construct a large and standardized food safety ontology for producers and consumers to use. Moreover, it can also provide a schema basis for the top-down construction of food safety knowledge graph.

Using ontology to describe data semantically can not only excavate the essential meaning of data, but also improve the retrieval efficiency [4]. Moreover, ontology provides the pattern structure of the top layer of knowledge graph, which can greatly assist the construction process of knowledge graph [5]. We have done some research on the food safety knowledge graph [6–8], but the food safety ontology is still insufficiently. In those studies the bottom-up method is used to build the knowledge graph, that means, entities, attributes and relations in documents are extracted only through machine learning algorithm and entity merging only uses conceptual similarity. The top-down knowledge graph construction refers to the per-construction of ontology, which can improve the quality of knowledge extraction and facilitate to the integration of knowledge graph in different fields. In view of this, this paper attempts to explore ontology from national food safety standards of China. The establishment process is to mine the class framework and class relations in the standards, and then extract the instances and attributes of classes by our knowledge extraction algorithm, finally verify the validity of ontology by entity mapping.

2 Related Research

For computer science and information science, ontology is a formal, distinct and detailed description of shared concept system. According to the widely accepted definition of ontology [9], ontology is a normative and unified definition of conceptual forms and relations between concepts.

The biggest feature of ontology is sharing, and the knowledge reflected in ontology is a well-defined consensus [5], then ontology is a knowledge concept template used to describe the concept hierarchy. Ontology modeling is a very important phase in the process of knowledge graph construction, which can well sort out entities and relations among entities. In the process of building knowledge graph, the introduction of ontology will play a guiding role, which can help the machine to understand the relation between concepts and make knowledge reasoning smoothly. It is worth noting that, when studying the domain ontology model, we should follow the principles of completeness and consistency put forward by Gruber, and construct the domain ontology reasonably and normatively [10]. In addition, the research on the construction of domain ontology can be divided into two levels according to whether reusing the existing mature metadata set, the first level is reference to mature ontology, the second level is building a new domain ontology.

In food safety field, there has few research on ontology, Dooley established a comprehensive and easy-to-access domain ontology of food from farm to table (FoodOn), which accurately and consistently described the common foods in cultures around the world [11]. Li Hongwei analyzed the early warning information of food safety, and built a ontology of Hazard Analysis Critical Control Point (HACCP) [12]. The food safety standards are related to the food production process, and currently has no normalized dataset can be used. Therefore, we collected the current national food safety standards of China as data source, and analyzed its content and extracted knowledge to build ontology.

3 Model Design of Food Safety Standard Ontology

The construction process of our domain ontology mainly includes several part, that are model design of ontology, natural language process, knowledge extraction of entities and knowledge import. The specific process is shown in Fig. 1.

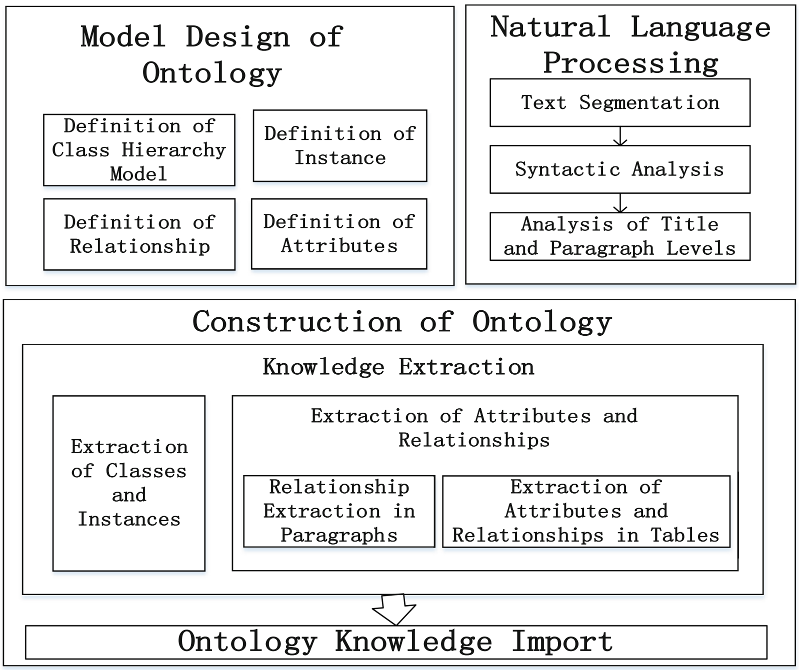

Fig. 1. Workflow of ontology construction.

3.1 National Food Safety Standards

Food safety standards are technical requirements and measures formulated after risk assessment of chemical, biological and physical substances that exist or may exist in food, food-related products and food additives. It is the most basic requirement for food to enter the market, and it is the technical regulations that should be implemented in accordance with food production and operation, inspection, import and export, supervision and management, and an important basis for food safety supervision and management [13, 14]. In our ontology study, the national food safety standard of China is the the only food safety standard enforced by the state.

We obtained 1182 documents about national food safety standards from FoodMate (http://www.foodmate.net). In our research the current national food safety standards are classified according to their content and scope of application. He Xiang [15] divides the national food safety standards into general standards, production, operation standards, inspection methods and procedures and food-related product standard. The national health administrative department has also established and formed the national food safety standard system [16] which includes four categories: general standard, product standard, product operation standard and detection method standard. Here, the standards in this paper are classified by the official classification scheme.

3.2 Model Framework of Food Safety Standard Ontology

The methods of ontology modeling are different according to different application purposes. At present, the mainly methods are the following two ways: the first is to explore the construction method from the perspective of knowledge engineering, which can be called ontology engineering; the other is the semi-automatic construction method of transforming the existing thesaurus resources into ontology. In addition, Ding Shengchun [17] proposed a comprehensive (semi-automatic) ontology construction method based on top-level ontology. The mainstream construction methods of knowledge engineering methods include Seven Steps method [18, 19], Skeleton method [20, 21], Methodology method [22], KACTUS engineering method [23], SENSUS method, IDEF-5 method. In this paper, Seven Steps method was used, Seven Step method is proposed by Stanford University School of medicine, which is mainly used to construct domain ontology.

Class Hierarchy Model of Food Safety Standard Ontology

In our research, the top-down construction mode [24] is adopted to build the ontology class model, that is, the top layer class (parent class) is abstracted first, and then their subclass are found out. If the subdivision can be continued, the third layer class will be established and refined step by step. Some class model are shown in Fig. 2.

Top-Level Class Model

Food category entities, as the most important concept described in food safety, are frequently cited in all food safety general standards. For example "GB 2760-2014, Standard for The use of Food Additives" specifies the types and contents of additives allowed to be added in each food category. Based on its frequency of occurrence, we regard "food category entity" as the top-level class in food standard ontology. Besides, "name of food safety standard" as the label of each standards was also defined as another top-level class of food safety ontology. Then by analyzing the content of food safety standards, the phrases "terms and definitions" are frequently appeared in various standards. The content under "terms and definitions" is the meaning of some professional terms in food field, so we defined "standard terminology" as the third top-level class. Finally, for further show the specific content of the standard, "content of food safety" was extracted as the fourth top-level class.

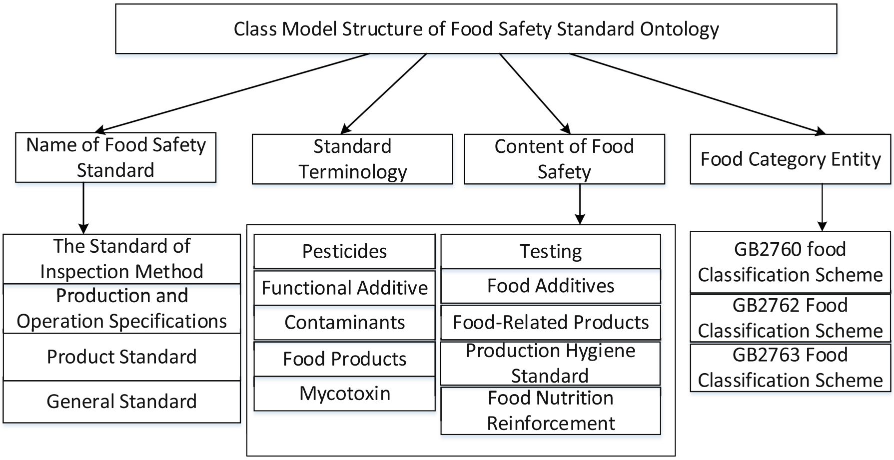

Fig. 2. Framework of top class and some subclasses

Secondary-Level Class Model

After completing the design of the top-level class, we continue analyzes its sub-class, that is, the second-level class. The model of the top-level class and its sub-class is shown in Fig. 2.

(1) Food category entities

In the existing GB/T national food safety standards, we have found three different food classification schemes, which are respectively derived from GB 2760-2014 Standards for the Use of Food Additives, GB 2763-2019 Standards for Maximum Residues of Pesticides in Foods and GB 2762-2017 Standards for Pollutants in Foods. In order to distinguish the three different food classification systems, in this paper, these three sets of food classification schemes are set up as secondary-level class, that are "GB2760 food classification scheme", "GB2763 food classification scheme" and "GB2762 food classification scheme". They are all subclass of "food category entities". In addition, all the food categories defined in the three classification systems are instances under its subclass.

(2) Name of food safety standard

At the beginning of our study, all food safety standards has been classified into different types, including general standards, production and operation standards, inspection method standards and product standards. Because each category is quite different in standard content and format, the extracted instances and relations are also quite different. So the names of these categories are also regarded as subclass of "name of food safety standard".

(3) Content of food safety

The content of food safety frequently include several common inspection items, such as food additives, pesticides, etc., and the actual extraction process proves that most standards are defined around these topics, such as the quality specifications and standards of food additives, physical and chemical testing standards. This shows that use these keywords as subclass are reasonable and feasible.

Through the process of "(1) analyzing the standard content, (2) design class name, (3) tracing back to the content, (4) determining the class name", secondary-level class of all food safety contents of the standards are extracted.

Other Subordinate Classes of Ontology
Under the secondary-level class model, you can continue to define subclass from the detailed description of standard content. For example, according to the content of production hygiene standards, as the secondary-level class, production hygiene standards include the following third-level subclass: site selection and production environment, factory buildings and workshops, requirements of raw materials and packaging materials, facilities and equipment, product traceability and recall, product hygiene standards, management system and personnel, records and documents management, etc. Factory buildings and workshops can be further subdivided, such as general requirements of factory building and workshop, special factory buildings design requirements, design and layout, workshop temperature control, etc.

Composition of Food Safety Standard Ontology
After determining the class model of the ontology, it is also necessary to determine the attributes of the class and the member instances of the class so as to realize the construction of the ontology.

Instance in Ontology
In ontology, instances represent the realization of classes, for example, propylene glycol is an instance of food additive, and silicon dioxide is also an instance of food additive. According to the analysis of the content of food safety standards, the instance definition rule are designed as follows:

(1) For the class of "food category", all names of food category in the "GB 2760" are the member instances of subclass "GB 2760, food classification scheme", such as "01.0, milk and dairy products", "01.02, fermented milk and flavor fermented milk", etc. Similarly, instances of other classification schemes are also determined according to this idea.
(2) Most instances of food safety testing items appear in general standards, usually a document hold a kind of instances. For example, "GB 2760-2014, Standard for the Use of Food Additives" lists the varieties of food additives that are allowed to be used in foods, and "GB 2761-2017, limit of mycotoxins in Foods" lists mycotoxins that may pose a greater risk to public health. Pollutants that may pose greater risks to public health are listed in "GB 2762-2017, Limits of Pollutants in Food", and the specific items listed in each standard are instances of corresponding categories.

(3) The names of all China food safety standards are instances of the corresponding categories of the standards. For example, “GB 5009.7-2016 National Food Safety Standard Determination of Reducing Sugar” stipulates the determination method of reducing sugar, which belongs to the inspection method category, that is, “GB 5009.7-2016, National Food Safety Standard Determination of Reducing Sugar” is an instance of “inspection method standard”. Also the set of terms and nouns listed below the heading of “Terms and Definitions” in each food safety standard is the set of member instances of “standard terms”.

Relations in Ontology

(1) Relations between instances

Relations definition needs to analyze the actions of various standards. First, we defined the relations in every general standards, and then the relations in other standards would be defined. The specific definition method is described as follows:

① General standards usually involve instances of food categories and food sampling inspection items, such as “benzoic acid and its sodium salt” is the instance of class “food additives” and “Blended Soy Sauce” is the instance of “GB2760 food classification scheme”. There is a relation <has_FoodAdditive> between “Blended Soy Sauce” and “benzoic acid and its sodium salt”, in which <has_FoodAdditive> is the relation name; However, “legume vegetables and potatoes” which is the instance of “GB 2762 Food Classification scheme” may contain the pollutant “lead”, this can be described as (“legume vegetables and potatoes” <has_Pollutant> “lead”), among which <has_Pollutant> is the relational name.

② Other standards, such as quality specifications of food additives and related standards, have (“instance of measurement items” <has reagents and materials> “instance of reagents and materials”), such as (“determination of total lactic acid” <has reagents and materials> “sulfuric acid”), (“determination of aluminum trioxide” <has instruments and equipment > “spectrophotometer”), the instances of the last relation are instances of “method principle” which are the subclass of “Testing” class.

③ There is also a reference relation between the standard name and the standard content specified by it. we defines a pair of reciprocal attributes of <prescribe> and <is _ prescribed _ by>. For example, “GB 10133–2014, National Food Safety Standard Aquatic Condiments” is a regulation on aquatic condiments, then the constraint (“GB10133–2014, National Food Safety Standard Aquatic Condiments” <prescribe> “Aquatic Condiments”) would be added to the ontology. If the subject is exchanged, then use the relational name <is _ predicted _ by>. However, for the standard terms, the standard terms are quoted by the standards, so the relation between them is named <has_term> and <is_term_of>, such as (“GB12694-2016, national standard for food safety, hygienic standard for livestock and poultry slaughtering and processing” <has_term> “clean area”). For non-universal standards, the more common relation name is <has_content>, which are used to represent the specific project regulations in those standards. Some reference relations in GB 8950–2016 are show in Table 1.

(2) the relation between similar instances.

There can also be a relation between instances of the same class, and it is often a pair of reciprocal relations. For the instance of class "food category", we defines each food category name in all food classification scheme as an instance, but since they are categories, there should also be a kind of relation between them, and this relation is <parent class> and <subclass>. For example, in the food classification scheme of GB2760, "01.0 milk and dairy products" is a food category class, and "01.02 fermented milk and flavor fermented milk" is a subclass of "01.0 milk and dairy products". They are all instances of "food categories", and there is an inheritance relation between them. Then add the constraints ("milk and dairy products" <subclass> "fermented milk and flavor fermented milk") and ("fermented milk and flavor fermented milk" <parent class> "milk and dairy products") in the ontology.

This kind of relation also exists between standard term instances. In order to distinguish the relation between standard term instances and food category instances, we defines reciprocal relation names as <has_term> and <is _ term _ of>, such as ("pollen" <has _ term> "bee pollen").

Table 1. Reference relations in GB 8950–2016

Standard name	Relation	Instances
GB8950-2016	prescribe	Canned food
GB8950-2016	has_range	This standard applies to the production of canned food
GB8950-2016	has_term	Commercial sterilization
GB8950-2016	has_content	The structure of equipment, tools and fixtures used in the canned food processing workshop and the installation position of fixed equipment should be convenient for thorough cleaning and disinfection

Attribute in Ontology

Because attribute values can be integers, floating-point numbers, Boolean values and strings, in order to describe an instance more accurately, attributes are often used to represent the characteristics of an instance. For example, ("benomyl" <ADI> "0.1mg/kg bw"), where ADI is an attribute name and 0.1mg/kg bw is its corresponding value.

For additives such as food additives and pesticides, they also have their own characteristic attributes. For example, food additives have CNS number, INS number, sensory requirements and other attributes, while pesticides also have numeric type attributes such as pesticide residues and ADI value of daily allowable intake. The specific contents are shown in Table 2.

Table 2. Attributes of additives

Subject class	Subject	Attribute	Object
Food additives	Calcium hydrogen phosphate	CNS number	06.006
Food additives	Calcium hydrogen phosphate	INS number	341ii
Food additives	Calcium hydrogen phosphate	Color requirement	White
Pesticide	Benomyl	ADI	0.1 mg/kg bw

4 Data Import of Food Safety Standard Ontology

4.1 Data Extraction

The creation and maintenance of ontology often takes a lot of time. Youn Jason et al. proposed a semi-supervised framework of automatic ontology population from the existing ontology support by using the method of word embedding [25]. In our research, a rule-based optimal semantic matching algorithm is adopted to realize semi-automatic knowledge extraction.

The knowledge extraction rules we defined rely on keyword analysis. The key words are recurring in the food safety standards, such as maximum limit, instruments and equipment, reagents and materials, etc. we obtained keywords through statistical word frequency and semantic analysis. In general, the document format of the same type standard is approximately the same, and the content needs to be extracted also has similar core words. Therefore, in the process of automatic extraction, the same type standards will be extracted in batches, and obtain all the class, attribute and instances. The following two methods are used to mine the instance relation:

(1) For unstructured text, in order to identify the document structure, context relation should be kept as much as possible, and the relation between instances should be found through the context relation, those relation usually is <include> or <has_content>.
(2) For the structured data in the table, the relation between instances is discovered through the header of the table, and the relation name is usually the field name in the header.

Furthermore, the document layout is mainly judged by the subtitles and title numbers of all levels, and the title numbers can be considered as the relations between classes, such as the inclusion relation and the instance relation.

Extraction of Classes and Instances

In food safety standards, all levels of subtitles often contain core words. we puts these words into the class name pool as candidate class names. The text under the all subtitles is taken as an instance of title class. However, when the all levels of subtitles and the text behind the subtitles are actually extracted, there will be a problem that the layers of title number in different paragraph is not same. Because, there are probably have two-levels titles, three-level titles or more in the text, and sometimes there are no text behind the subtitle. Therefore, we must design algorithms to automatically identify the layouts of the document, so as to extract the title classes and content instances more accurately. The relation name between a title class and its subtitles class is <include>.

The algorithms of document layout identification use digital label to mark the all levels of subtitles and text behind them. For example, the marking rules of document layout of "GB 8955-2016, National Food Safety Standard, Hygienic Specification for the Production of Edible Vegetable Oils and Their Products" is listed as follows:

(1) If the title number is subtitle, for example, there is the subtitle "4.1 General Requirements" under the title "4 Workshop and Workshop", and there is a <include> relation between them.
(2) If there is no subtitle after a title, for example, if the next paragraph behind subtitle "4.1 General Requirements" dose not have a subtitle number, then the text in the next paragraph is a description of the subtitle, and we considered them as an instance of the subtitle "4.1 General Requirements".
(3) Marking for each class and instance. For example, the title candidate class "4 factories and workshops" is the first-level title, marked as 1; The candidate title subclass "4.1 General Requirements" is a second-level title, marked with 2; The text behind "4.1 General Requirements" is its instance, and there is a <has_content> relation between "4.1 General Requirements" and its instance, so it is marked as 3.
(4) The title candidate subclass "4.2 Design and Layout" is the same level with the title candidate subclass "4.1 General Requirements", so it is marked as 2. In this way, each title class would be marked.
(5) If there is no subtitle and no text behind a title, for example, subtitle 5.1 and subtitle 5.2 has title numbers, but there is no text and no subtitle behind them, so they can be regarded as instances of the title candidate class "5 Facilities and Equipment".

After the above marking process, all classes and instances need to be extracted by marking values. In order to save the paths of all classes and instances, we try to use the stack as an intermediate medium to extract all data. The specific method are as follows:

(1) According to the marking values, the marked data in the document are stored in the stack in turn. Each data in the stack has an inclusive relation from the bottom of the stack to the top of the stack, so their marking values are also increasing in sequence. When the marking values are not increasing, the next data may temporarily suspend store, and the label of the top data of the stack would be recorded, which is an instance of the previous data. Then store all the data in the stack in an ordered list, which saves a path from vertex to leaf in the tree, this path also reflects the < include> relation from class to its subclass and its instance.
(2) When a new data re-enters the stack, it is necessary to check whether the marking value of new data is greater than that of the top data in the stack. If not, the data in the stack need to be popped out in turn until the marking value of the top data is less than the newly added data. Then the new data is pushed into the stack.
(3) return to step (1).

For example, title “4 workshops and workshops”, subtitle “4.1 general requirements” and text after “should meet ……” are stored in the stack in sequence, with marking values 1, 2 and 3 respectively. At this time, the next data “4.2 design and layout” that needs to be pushed into the stack which has a marking value 2. When the data were judged needs to be popped out of the stack, the sequence of the data in the stack will be saved in a order list. At the same time, the relation between the the standard name class and its instance need to be retained, so the standard name class needs to be inserted into the order list in the head. Then the final sequence is [“GB 8955-2016, National Food Safety Standard, Hygienic Code for the Production of Edible Vegetable Oil and Its Products”, “4 factories and workshops”, “4.1 general requirements” and “should meet ……”]. After the path being saved, the data with marking values 2 and 3 would be popped out from stack and new data “4. 2 Design and Layout” would be pushed into the stack, and then repeat the previous operation with the marking value. The specific algorithm description is shown in Algorithm 1.

Algorithm1: Paths extraction method of classes and instances

```
Input: Paragraphs in document C = {c1, c2, ..., cn}, n is the total of paragraphs.
Output: Instances paths from upper classes.
1. Text layout marking:
  Set the marking values as L = {l1, l2, ... ln}
  i = 1
  For i<n {
 Read ci,
    If ci contains the P-level title number {
       li = P, P ∈ (1,2,3..m), where m is the maximum number of all levels.
    } else if ci-1 contain a P-level title number {
                li = li-1 + 1
          }else{   li = li-1 }
   i = i+1
}
2. Stack initialization: Create a stack and empty it, and set the pointer cur to point
to the bottom of the stack;
3.  Save the path for all classes and their instances:
 i = 1
  For i<n {
 Reading ci and li
 If cur points to the bottom of the stack {
    push ci and li into the stack
    cur points to ci and li (current top of stack)
 } else if  li > lcur {
         push ci and li into the stack
         cur points to ci and li (current top of stack)
    } else{
        read the standard name of the document, and joint with the data from the
bottom of the stack to the top of the stack in turn, and store path in a sequential list.
         While li <= lcur{ pop out the top element}
          push ci and li into the stack
          cur points to ci and li(current top of stack)
        }
   }
 i = i +1
 save the path.
}
4. Get the instances: Search each path list, and the last element of the order list is
the instance.
```

Extraction of Attribute and Relations

Relations Extraction from Documents

For we have defined <include> relations between classes, also between classes and instances, and algorithm 1 has saved the instance paths we need, so the <include> relations between them can be obtained by reading the paths. In order to distinguish whether the relations are between classes or between classes and instances, algorithm 2 stores the <include> relation in two tables, named as "class_ relations" and "classes_ instances" respectively.

Algorithm 2: Save < include > relation between classes and instances.

```
Input: Path for all classes and their instances
Output: "class_relations" between classes, and "class_instances"  between classes
and instances.
1: initialization: "class _ relations", "class _ instances"  are all empty lists.
2. Get the relation between classes:
For each path {
   i = 1
   For i < length(path)-1 {
       relation = (Path[i-1],subclass,Path[i])
   save the relation in the class_relations list
   i = i+1
 }
}
3. Get the relation between class and instance:
For each path {
     relation = (Path[-2],instance,Path[-1])
 save the relation in the class_instances list
}
```

Attribute Extracting from Tables

Tables appearing in standards usually have a table header, which defines the attribute meaning of the table. we matched the table header to the class name and instance name and judged whether the data in table have the relation we need. If the data in table is necessary, we extracted attribute from tables. In the table, the header is the semantics of each attribute, it is also the relation name of class and its instance. So if the relation triplet in table is (? S,? P,? O), predicate (? P) is the header name and the object (? O) is the attribute value, at last the content of subject (? S) needs to be analyzed through table data.

For example, we extracted relational triples in Table 3. The relational triples are ("calcium hydrogen phosphate "<color requirement> "white"), ("calcium hydrogen phosphate" <state requirement> "powder"), ("calcium hydrogen phosphate" <color inspection method> "take appropriate samples"), ("Calcium hydrogen phosphate" <State inspection method> "Take appropriate sample"). It can be found that the relation name in this table is actually the attribute value (such as color and state) joint with the table header (such as requirements). While the subject of the relation triple is the instance (such as calcium hydrogen phosphate) in the food safety standard. Therefore, the extraction method of relational triples in tables is as follows:

(1) First, read the instance in the food safety standard. we did it by segmenting the name of food safety standard, then matched keywords in the instance name set;
(2) Then jointed the relation names;
(3) Finally, found the corresponding relational attribute values. See algorithm 3 for specific implementation.

Table 3. Sensory requirements in "GB 1886.3-2016, National Food Safety Standard, Food Additive Calcium Hydrogen Phosphate"

Project	Ask	Test Method
Colour and lustre	White	Take a proper amount of sample and put it in a clean and dry white porcelain plate, and observe its color and state in natural light
Condition	Powder	Take a proper amount of sample and put it in a clean and dry white porcelain plate, and observe its color and state in natural light

Algorithm 3 Save the attributes relations in table

Input: Tables in standards
Output: Attribute and relation
1. Initialization: attribute relations is an empty list.
2. Get the instance *obj*: get the *obj* name from the instance name by word segmentation and string matching.
3. Get the attribute relation: extracting the values of attributes, row[*i*, *j*] means the value of row *i* and column *j*, and when *i*=0, it point to the header row .

```
i =1
For i < = R {//R is the maximum number of records in the table.
    relation = (obj,row[i,S]+row[0,V],row[i,V])
      //S is the column number where the prefix name of the relation is located, and V is the column number where the relation value is located.
save relations
}
```

4.2 Data Import

Finally, the entities and relations are imported into the class model framework. When all entities and relations are imported into the ontology model, ontology visualization can be shown. Figure 3 show an diagram of part instances and relation. In the Fig. 3, the green inner ring is an instance of class "food safety standard name". Actually, they belong to different classes, but for the size limit of the diagram, the class is not shown here. And the gray outer circle instance is instance of the "range", in which the dotted line points from the instance of "standard name" to the instance of "range". Those dotted line indicated that there is a relation between them, and the name of the relation can be viewed by clicking the dotted line.

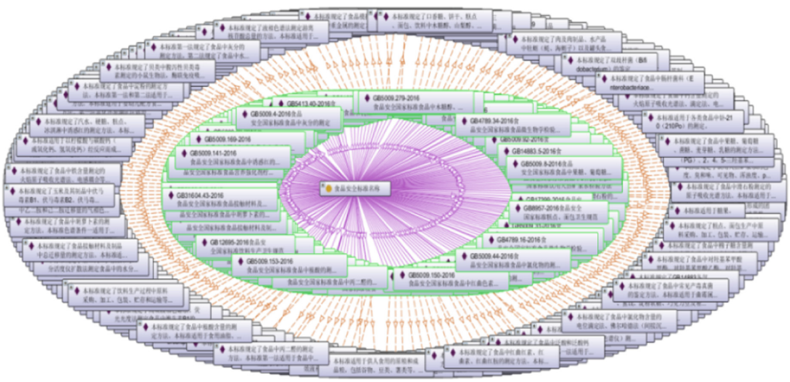

Fig. 3. Entity and relation diagram of food safety standard ontology (part)

4.3 Validity Analysis

In order to verify the correctness and effectiveness of the ontology we have built, we mainly use two methods: one is to invite research experts in food safety to manually modify the data and structures in the ontology; the other is to use the entities mapping method to do concept mapping with some food name corpora we have collected. Here, we use the NER technology to achieve this, and then we need obtained the relationship between the food name entities and the food category class or subclass in the ontology, Here we adopt C-norm [26] (a new share neural method to resolve the few shot learning entity linking problem) as our classifier to achieve this goal. At last we manually identify the mapping results to determine the precision and the recall of the entity mapping. My experimental results show that the precision and the recall is 0.85 and 0.72.

5 Summary

For a long time, food safety knowledge lack of standardization which will lead to ambiguity in understanding, though the study of food safety ontology is an effective way to solve this problem, but few researches were done on it. We collected 1182 national food safety standards from web, and build class framework for them. Then a new rule-based knowledge extraction algorithm was proposed to extract the instances and relations in all standards. The whole ontology includes 236 classes, 48 relation names, 823 attribute names, 8812 instances and 131,406 constraints (relations).

Our ontology describes various concepts and semantic relations related to food safety stipulated in national food safety standards, including food classification system, additive limits of various foods, pesticide and veterinary drug residues, pathogenic microorganism pollution, heavy metal and mycotoxin pollution, and food inspection, detection and physicochemical analysis methods. The significance of ontology is to provide schema layer for the top-down construction of food safety knowledge graph.

In order to verify the effectiveness of ontology, we use manual verification and entities mapping experiments to prove the effectiveness of our ontology. The current accuracy and recall can meet our basic needs, we will adopt BERT model to further improve them in future study.

References

1. Xu, X.: Food safety in China: problems, causes and countermeasures. Issues in Agric. Econ. **23**(10), 45–48 (2002)
2. Xu, R., Pang, G.: Study on the current situation, problems and countermeasures of food safety in China. China Science Publishing & media Ltd. (CSPM), BeiJin (2015)
3. Zhou, D., Yang, H.: Information asymmetry and government supervision mechanism in food quality and safety management. Chin. Rural Econ. **6**, 29–35, 52 (2002)
4. Gao, X.: Research on the construction and analysis method of domain Ontology of film and television works. Zhejiang Normal University of China (2014)
5. Liu, Q., Li, Y., Duan, H., Liu, Y., Qin, Z.: Knowledge graph construction techniques. J.Comput. Res. and Dev. **53**(3), 582–600 (2016)
6. Qin, L., Hao, Z., Li, G.: Construction and correlation analysis of national food safety standard graph. J. Comput. Appl. **41**(4), 1005–1011 (2021)
7. Li, Q., Zhigang, H., Liang, Z.: Food safety knowledge graph and question answering system. In: ICIT 2019: Proceedings of the 2019 7th International Conference on Information Technology: IoT and Smart City, pp.559–564. ACM, New York(2019)
8. Qin, L., Hao, Z.: National food safety standard graph and its correlation research. In: Liang, Q., Wang, W., Mu, J., Liu, X., Na, Z., Cai, X. (eds.) Artificial Intelligence in China. LNEE, vol. 653, pp. 405–411. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-8599-9_47
9. Gruber, T.R.: A translation approach to portable ontology specifications. Knowl. Acquis. **5**(2), 199–220 (1993)
10. Gruber, T.R.: Towards principles for the design of ontologies used for knowledge sharing. Int. J. Hum.-Comput. Stud. **43**(5–6), 907–928 (1995)
11. Dooley, D.M., Griffiths, E.J., Gosal, G.S., et al.: FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. NPJ Sci. Food **2**(1), 1–10 (2018)
12. Li, H., Huang, W., Hong, X.: Research on ontology building in food security pre-warning. Comput. Technol. Dev. **23**(9), 238–240, 244 (2013)
13. Chen, J., Li, B.: China's food safety standard system: problems and solutions. Food Sci. **09**, 334–338 (2014)
14. Wang, H.-X., Jian, F., Zhang, L., Huang, Q.: Analysis and suggestion on the current situation of standards in food inspection in China. J. Food Saf. Qual. **011**(006), 2001–2006 (2020)
15. He, X.: Research on national food safety standard system construction. Central South University (2013)
16. Letter of the general office of the national health and Family Planning Commission on notifying the conclusion of the cleaning and integration of the catalogue of national food safety standards and food related standards. Chinese Journal of Food Hygiene (2017). CNKI:SUN:ZSPZ.0.2017-04-034
17. Ding, S., Li, Y., Gan, L.: Research on integrated domain ontology construction method based on top-level ontology. Inf. Stud. Theory Appl. **30**(2), 236–240 (2007)
18. Swartout, B., Patil, R., Knight, K., et al.: Toward distributed use of large-scale ontologies. In: Proceedings of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems, vol. 138. no. 148, p. 25 (1996)
19. Noyn, F., Mc Guinness, D.L.: Ontology development 101: a guide to creating your first ontology (2001). http://protege.stanford.edu/publications
20. Uschold, M., Gruninger, M.: Ontologies: principles, methods and applications. Knowl. Eng. Rev. **11**(2), 14–17 (1996)

21. Pinto, H.S., Martins, J.P.: Ontologies: how can they be built? Knowl. Inf. Syst. **6**, 441–464 (2004)
22. Fernández-López, M., Gómez-Pérez, A., Juristo, N.: Methontology: from ontological art towards ontological engineering, no. 5, pp. 33–40 (1997)
23. The KACTUS booklet version 1. 0. esprit project 814. (2015). http://www.swi.psy.uva.nl/prjects/NewKACTUS/Reports.html
24. Chen, P., Lu, Y., Zheng, V.W., et al.: KnowEdu: a system to construct knowledge graph for education. IEEE Access **6**, 31553–31563 (2018)
25. Youn, J., Naravane, T., Tagkopoulos, I.: Using word embeddings to learn a better food ontology. Front. Artif. Intell. **3**, 1–8 (2020)
26. Ferré, A., Deléger, L., Bossy, R., Claire, N.: C-norm: a neural approach to few-shot entity normalization. BMC Bioinform. **21**, 579 (2020)