Electronics, Communications and Networks
A.J. Tallón-Ballesteros et al. (Eds.)

doi:10.3233/FAIA231242

Research on Criminal Clues Analysis Model of Illegally Added Toxic and Harmful Substances in Food

Yutian WANG[1], Qianye YANG, Haowei CHEN and Jiongkai CHEN
School of information technology, Nanjing Forest Police College, Nanjing 210023, China

Abstract. With the continuous development of the era of big data and the corresponding technical means, the technical demand for data mining and analysis research and judgment in maintaining food safety is also increasing, and how to combine the trend of The Times to strengthen the fight against food safety crimes and innovate the means of handling cases is particularly important. Therefore, in order to further improve the ability and efficiency of relevant departments in investigating food cases, this paper, based on actual case data, uses big data police analysis technology as a means to build a clue analysis model for illegal addition of toxic and harmful substances to food crimes, establishes a multi-channel data collection system, preprocesses case texts and defines standards, and innovatively uses knowledge ontology database construction technology. Then it is classified according to the business theme, and the overall framework of the analysis model of food crime clues of illegal addition of toxic and harmful substances is reused vertically, and the horizontal expansion is flexible, and the analysis, judgment and comprehensive treatment ability of hot issues of food crime of illegal addition of toxic and harmful substances in the disposal of food safety crimes the model is greatly improved.

Keywords. Food crime, model construction, proactive warning, ontology library, system application

1. Background

Food quality and safety involves thousands of households and is the most basic requirement for people's survival. If food quality and safety are not guaranteed, people's health and life safety are not guaranteed. With the continuous development of economy, food types are becoming more and more abundant, and the quantity of products are more and more abundant. While meeting the balance of food demand and supply, fake and shoddy food is frequently exposed, and mass incidents that endanger consumers' health and life safety occur from time to time, and food safety issues have become the focus of

[1] Corresponding author:Yutian WANG, School of Information Technology, Nanjing Forest Police College, Nanjing 210023, China, Email: tiantian517@foxmail.com

consumers' attention. Frequent food safety problems will not only endanger the health of consumers, but also seriously disrupt the market order, cause huge economic losses to ordinary people, and bring bad social impact. Therefore, it is necessary to strengthen the fight against toxic and harmful food crimes, so as to set up a better living environment for the people[1].

1.1. Illegal addition of toxic and harmful substances in food crime development trend

In recent years, the production and sale of toxic and harmful food crimes have shown the characteristics of great harm, wide range of influence, common crimes, forming criminal chains, and diversified criminal methods. The production and sale of toxic and harmful food crimes have formed trans-regional and multi-channel underground industrial chain, and the sales channels have also been transformed from offline to online transactions. The crime of illegally adding toxic and harmful substances has gradually shifted the main battlefield to the Internet, and the means of crime is divergent, quick, virtual and remote, the damage object is not specific, the crime place is not changed regularly. This makes the information related to these cases on the Internet more chaotic, which leads to the relevant departments often encounter problems such as difficult to obtain clues, difficult to support evidence, and difficult to eradicate crimes in the process of detection.

Therefore, it is of great practical significance to further explore the systematic management of global food crimes and conduct research on the clues analysis model of food crimes involving illegal addition of toxic and harmful substances, in order to maintain global food security and social stability.

2. **Data preprocessing**

In this paper, based on the data acquisition of relevant cases involving food with illegal addition of toxic and harmful substances, the relevant case data is desensitized, and then the desensitized data is preprocessed to create effective and usable storage knowledge. Using Python programming and other means, we conducted quantitative analysis of the cleaned and desensitized data, search for potential links and excavate the close links between key attribute fields and various crime elements in the case, analyze and summarize the extracted characteristic information, and effectively collect the data characteristic information.

2.1. Entity extraction standard definition

With reference to relevant laws and regulations, entity extraction and linkage construction are carried out for key fields such as foods involved and harmful substances involved, key personnel involved, key places involved, and amount involved, as shown in the **Table 1**.

Semantic analysis is an analytical method based on the Distibutional Hypothesis that words that often appear in similar contexts are semantically similar, and it extracts information through statistical analysis of context. On the basis of structured data, this paper further relies on semantic analysis and combines Word2Vec to model text semantic features, label entities, improve classification accuracy, and lay the foundation for the subsequent construction of knowledge ontology database[2].

Table 1. Example of entity extraction standard

case	standard	label
The suspect Shi added about 0.1 kg of non-food raw material sodium nitrite to about 15 kg of beef	National Standard for Food Safety Food Additives	sodium nitrite
The defendant Xu purchased Sibutri, rhubarb and other substances to produce bulk weight loss capsules, and illegally sold them through Taobao and second-hand trading market websites	Notice on the Release of the List of Substances that May be Illegally Added to Health Food (the first batch) (Baohua [2012] 33 of China Food and Drug Administration)	Sibutram, ephedrine, Fenfluramine, Taobao, second-hand trading market website
Defendant Wei in Changsha City Furong District, No. 112 Weifang Second Road operating "double love health care products" shop, in addition to selling condoms and other supplies, but also sold health care products containing sildenafil ingredients.	List of Substances that May be Illegally Added to Health Food (First batch)	Sildenafil, health medicine, health medicine store

3. Construction of knowledge ontology base

3.1. The basic definition of knowledge ontology base

A knowledge ontology library is a structured collection of data used to store and organize knowledge so that computers can understand and process it. It is a formal representation for describing entities, concepts, attributes, and relationships in the real world. Knowledge ontology libraries are usually created and maintained by ontology languages and ontology editing tools[3].

The purpose of a knowledge ontology library is to establish a shared, reusable knowledge base that allows different applications and systems to use and expand upon that knowledge jointly. It can be used for knowledge graph construction, and its contents include entities (e.g., people, places, objects), concepts (e.g., animals, vehicles, food), attributes (e.g., color, size, weight), and relationships (e.g., paternity, inclusion, belonging)[4].By organizing these elements together and defining the semantic relationships between them, a model that describes the real world is constructed.

This paper will establish a knowledge ontology database related to the crime of illegally adding toxic and harmful substances to food, extract the entities, relationships, and other elements from the case, organize these elements into a structured format, and construct a clue model for the crime of illegally adding toxic and harmful substances to food with semantic analysis.

3.2. Knowledge modeling

At present, there are two principal methods of knowledge modeling: top-down data modeling method: first, ontology and inference rules are formulated for the knowledge graph, data schema is designed, and then targeted data is extracted according to the designed data schema. Bottom-up data modeling method: data extraction is carried out first, and then its characteristics are summarized according to the extracted content, and the framework is refined to form a definite data model[5].

In this paper, a top-down design and bottom-up approach are adopted to construct a crime clue analysis model for illegally added toxic and harmful substances to food. Taking the extracted characteristic information as the basic element, and referring to the relevant descriptions of toxic and harmful substances and food addition provided by authoritative agencies such as the Food and Drug Administration, and combining the relevant field regulations of relevant departments, the ontology part of the knowledge ontology database is constructed. Knowledge reasoning is adopted to explore the evolution process of cases and the knowledge expression of laws. Firstly, key field ontology and reasonable criminal reasoning rules are formulated for the knowledge ontology database of food crimes with illegal addition of toxic and harmful substances, and corresponding data mode is designed for them. Then, targeted data of food crimes with illegal addition of toxic and harmful substances are extracted according to the designed data mode. Extract key information.

After that, the ontology database is used to build A seven-step method[6] to determine the professional field and category of the food with illegal addition of toxic and harmful substances, as well as important terms such as harmful substances, key involved personnel, key places and involved amount. The top-down method is used to improve the hierarchical system and build an IS-A hierarchical structure to list and output basic classification fields. Moreover, multi-element information such as subject label, entity label and trigger word label is introduced for model fusion. Eventually, the intelligent processing of classification label data in ontology database is realized.

3.3. Knowledge representation

There are two kinds of structured representation of knowledge: symbolic representation and vector representation. This paper mainly uses the triplet form (head entity, relation entity, tail entity) based on graph data structure to symbolically represent knowledge.

RDF(Resource Description Framework) is essentially a Data Model. In this paper, the extracted entities are RDF serialized, and multiple N-Triples are used to represent the RDF data set, as shown in the **Figure 1**:

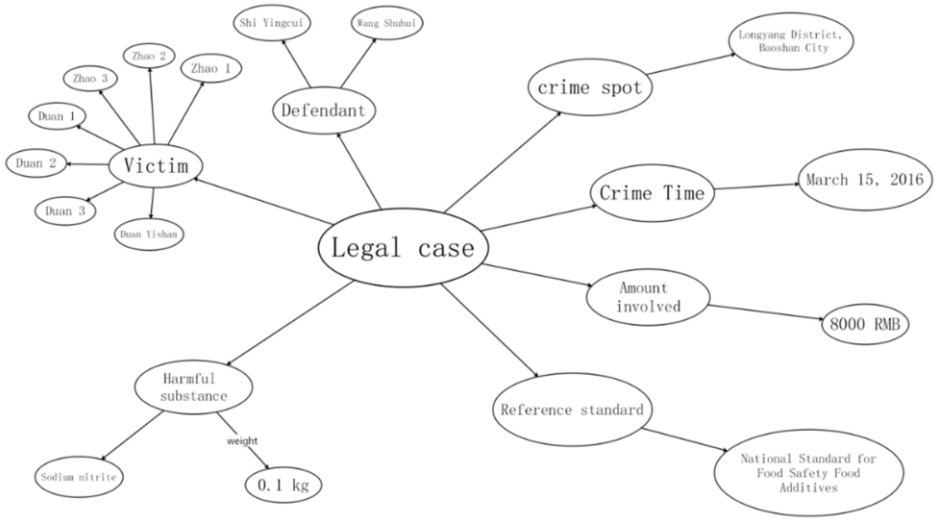

Figure 1. Resource Description Framework (RDF) data model

However, due to the limited expressive power of RDF, it cannot distinguish between categories and objects, nor can it clarify and describe the relationships/attributes of classes. For example, RDF can express the attributes of suspect Li and succinylcholine chloride and the relationship between them, but RDF has certain limitations in expressing the attributes of specific people, the attributes of places, and the relationships between people and places. Therefore, RDFS is used in this article to solve the dilemma of RDF's limited expressive power, , as shown in the **Figure 2**:

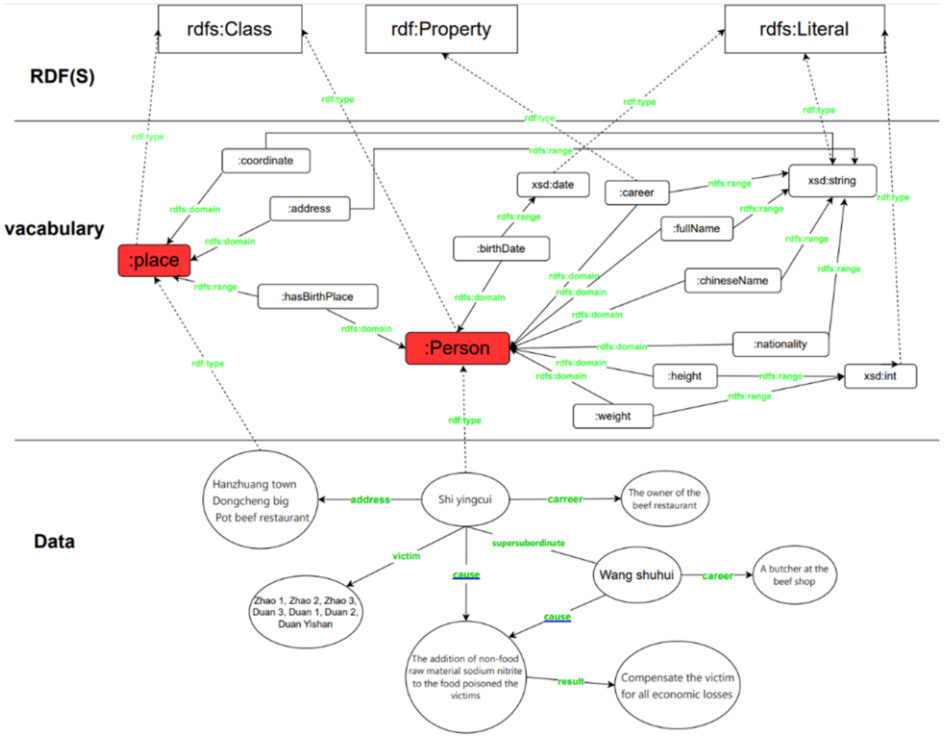

Figure 2. RDFS data representation schema

The Data layer is the specific description of the illegal addition of toxic and harmful substances with RDF, Vocabulary is some words (categories, attributes) defined by us, and RDF(S) is a predefined term. From bottom to top is a concrete to abstract process. In the figure, we use a red rounded rectangle to represent the class, a green font to represent rdf:type, rdfs:domain, rdfs:range three predefined terms, and a dashed line to represent the ownership relationship of rdf:type. In addition, in order to reduce the crossing of lines in the graph, we keep the rdf:type ownership of only one attribute, career, and omit this relationship for other attributes.

3.4. Knowledge storage

The original data types of knowledge graph generally fall into three categories: structural data, such as relational database; Unstructured data: such as pictures, audio, video; Semi-structured data: such as XML, JSON, encyclopedia. Generally, graph databases are used for knowledge graph storage[7], such as Neo4j, ArangoDB, NebulaGraph, etc. Graph databases[8] can simultaneously support queries and perform real-time updates on big

data, and have a flexible online schema environment. Therefore, this paper uses Neo4j to store knowledge in triad databases, , as shown in the **Figure 3**:

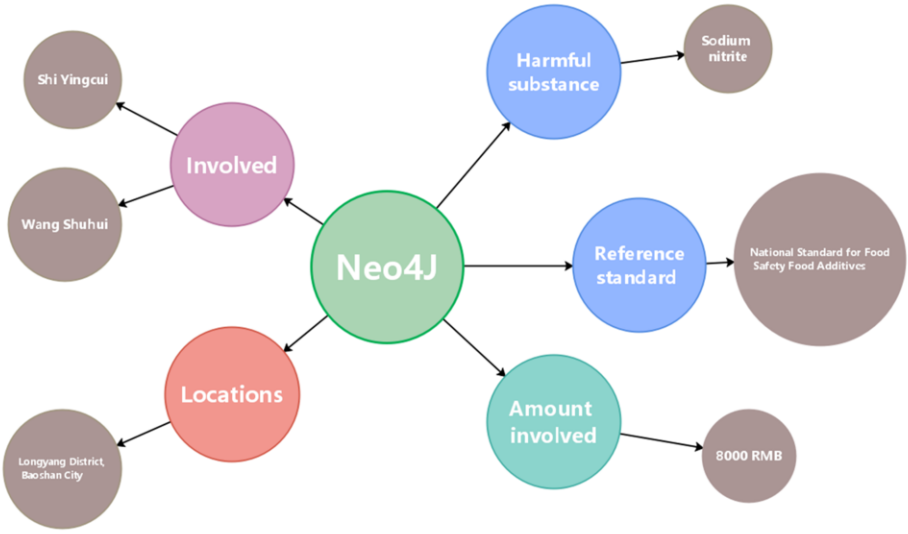

Figure 3. Neo4j graph database storage schema

Based on the constructed RDFS triplet data model, this paper uses nodes to represent an entity record, and connects to other nodes through relationships, adding related properties in the process, and creating multiple composite indexes. Build a more complete relational data platform to realize real-time early warning.

3.5 Overall architecture

In this paper, a five-layer design pattern of data source - storage layer - interface layer - application layer - decision layer is used to construct the model. The data sources are relatively extensive, mainly from domestic and foreign open source intelligence information about the crime of illegally adding toxic and harmful substances to food and information provided by relevant departments. Through the use of big data research and classification to separate all kinds of data information, the obtained structured data and unstructured data were pre-processed to extract information such as harmful names, personnel involved and key positions. RDFS triplet technology [9]was used to sort out the data, and then Neo4j was used to store the data knowledge. In the interface layer, we carry on the data materialization and relational processing, and then import the data into the application layer for the next operation. At the application level, we use graph analysis and data visualization to make the model more intuitive[10]. Through visualization operations such as graph exploration analysis and relationship modeling[11], we dig deeply into the intersection area of each case data, find out the neglected data contact points, and make the case context clear. Finally, the real-time early warning of the crime of illegally adding toxic and harmful substances is realized at the decision-making level, and the front-end technology in the artificial intelligence domain is used in the actual combat to improve the decision-making utilization rate, as shown in the **Figure 4**:

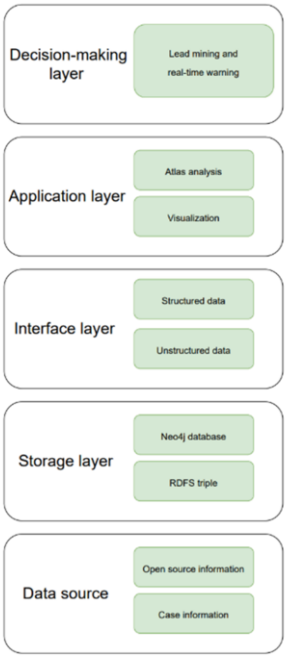

Figure 4. Overall structure drawing

4. Innovative highlights and practical results

4.1. Build multiple scenario application sub-models to meet the needs of multiple categories and levels of cases

By using the data related to the production and sale of fake and shoddy food crime on the Internet, combined with the relevant algorithms, the structure of each data is analyzed to find out the needs such as: characteristic information is extracted from key attribute fields such as main items, key personnel, key places, and involved amount to establish a set of analysis and judgment model that can be directly used by relevant institutions, so as to provide convenience for combating food safety crimes. The case information of different scenarios is integrated and processed, and multiple scenario application sub-models are built to enhance the universality of the models, so as to meet the needs of multiple categories and levels of cases.

4.2. Diversified data analysis tools build models to realize automated and intelligent analysis and judgment process

Through the use of information extraction, data cleaning, knowledge graph construction and other related technologies and multidimensional data analysis, not only to provide theoretical support for the future production and sale of shoddy food investigation and combat behavior, but also for the future related department to respond to the production and sale of shoddy food detection and early warning basis. Using a variety of data analysis means and tools, the model is built to realize the automation and intelligence of

the analysis, research and judgment process and the trend of smart policing in the era of big data.

5. Summary and prospect

The construction of crime clue model is based on information extraction, data cleaning, knowledge graph construction and multi-dimensional data analysis. This article will optimize the following aspects:

- To study the criminal characteristics and trends of illegal addition of toxic and harmful substances in food at home and abroad.
- Contacting the actual work of combating food crimes to understand the difficulties and pain points encountered in the actual detection of such cases, which is one of the key goals of the model built in this paper, so as to promote high-quality work of combating food crimes.
- Establish and design the experimental framework and test the model system according to the practical difficulties encountered in collecting data and real cases.
- The knowledge meta-indexing method is tried to further construct the knowledge graph containing food crime monitoring and early warning, and provide new ideas for combating food crime.

Acknowledgement

This research is supported by the 2023 College Students Innovation and Entrepreneurship Training Program (Approval number: 202312213036Z).

References

[1] Lin Haiying, Liang Yongheng, Ren Yanqi et al. Discussion on current situation of food safety and Food quality management [J]. Food Industry,2023,44(06):341-343. (in Chinese)

[2] Wang Yajie, Yang Bing, DAI Jiao et al. Risk analysis algorithm and application of big Data for food toxic and hazardous substances detection based on cloud computing [J]. Anhui agricultural science, 2017, (21) : 216-220 + 233. DOI: 10.13989 / j.carol carroll nki. 0517-6611.2017.21.146.

[3] Yang Haobo. Research and implementation of innovative knowledge expression and retrieval methods based on functional ontology [D]. Hebei University of Technology,2011.

[4] Tong Yuting.Study on problems related to production and sale of toxic and harmful food crimes [J]. Food Industry,2022,No.112(11):117-119.

[5] Ou Jingwen. On the characteristics and preventive measures of the crime of producing and selling poisonous and harmful food [J]. Law and Society, 2012(30):238-239. (in Chinese) DOI:10.19387/j.Cnki. The 1009-0592.2012.30.122.

[6] Dong Jiangang, Zhang Feng. Based on circulation in the evolution of IETM domain ontology construction research [J]. Journal of electronic design engineering, 2012, 20 (14) : 1-4 + 7. DOI: 10.14022 / j.carol carroll nki DZSJGC. 2012.14.039.

[7] Dong Jiangang, Zhang Feng. Based on circulation in the evolution of IETM domain ontology construction research [J]. Journal of electronic design engineering, 2012, 20 (14) : 1-4 + 7. DOI: 10.14022 / j.carol carroll nki DZSJGC. 2012.14.039.

[8] Li Junyi, Wang Zhuo, Ma Pengwei. Research on the development trend of graph database technology [J]. Information and Communication Technology and Policy, 2019,47(05):67-72. (in Chinese)

[9] Gao Xiangbin. Entity Relationship model construction based on RDF(S) resource query [D]. Jilin University,2017.

[10] Gao Xiangbin. Entity Relationship model construction based on RDF(S) resource query [D]. Jilin University,2017.
[11] Xia Wei, Wang Shanlei, Yin Zidu et al. Entity association modeling and completion of Knowledge Graph based on Mutual Information [J]. Journal of Computer Science and Exploration,2018,12(07):1064-1074. (in Chinese)