Analysis on the Construction of Knowledge Graph of Mass Events Based on Ontology

Binghua Ren, Fanliang Bu*, Zhiwen Hou, Yihui Fu and Xiaoyuan Liu
People's Public Security University of China, Beijing, China

*Corresponding author: 2019211272@stu.ppsuc.edu.cn

Abstract. Mass incidents pose a serious threat to public order and national security, building a knowledge base in the field of mass incidents, visualizing the relationship between entities from incidents, and scientifically reasoning about key information about incidents have become key links in effectively responding to mass violence crimes. In view of the lack of existing mass incident knowledge graph analysis methods and strong subjective reasoning methods, based on the BiLSTM-CRF model, this paper analyzes event elements and entity associations by extracting entities from the text data of mass events, the protege software is used to construct the mass event ontology model, and the visualization of the mass event knowledge graph is realized based on the neo4j graph database. The research results show that the knowledge graph established in this paper can better reflect the correlation between the elements in the mass incident, and is of great significance to the national security department to improve the emergency decision-making ability.

1. Introduction

Currently, with the advent of the era of big data, explosive data growth forces people to seek new data processing methods, this demand is particularly prominent in the field of mass incidents. The occurrence of mass incidents has brought painful disasters to ordinary people. In order to deal with such emergencies that may arise, not only to improve legislation, at the same time, we also need to pay attention to the utilization of data, collect and standardize information on various mass incidents scattered on the Internet. In the era of big data, mass event data presents the characteristics of multi-source heterogeneity and huge quantity. Building a large-scale knowledge base is one of the effective ways to structured storage and expression of such data. Ontology-based knowledge base rules have formulated norms for describing the logical relationship between knowledge and knowledge [1], conducive to the update and maintenance of the knowledge base and third-party integration. Knowledge graph is such an efficient mass event data processing program. Different from traditional knowledge representation methods, knowledge Graph has more standardized ontology terms and grammatical format, the graph structure model is used to describe the relationship between entities and concepts, emphasize the connection between data, And the knowledge graph has high-speed data processing capabilities, It can structure the scattered group event information. This article analyzes a large number of group event text information, study its constituent factors, Build the ontology model of mass events with the help of technologies such as knowledge extraction and semantic analysis, standardize the ontology concepts, attributes and relationships in this field. Finally realized the construction of the
knowledge graph of mass events, intuitively show the internal connections of events, people, and organizations. In order to quickly locate the data, it helps the national security department to quickly formulate emergency plans in the face of sudden public safety incidents.

2. Ontology overview
The term ontology originally belonged to the category of philosophy. Knowledge engineering scholars introduce ontology concepts in the process of developing knowledge systems. In order to obtain, describe and express knowledge in a specific field [2]. The ontology structure in the knowledge graph includes several aspects such as classes, instances, relations, axioms, and functions. Its defined terms constitute a vocabulary of related fields, and the interaction between classes is formed on the basis of the hierarchical structure of classes. Of course, the reason why the ontology model can be called the pattern layer of the knowledge graph. The most important thing is that it introduces the concept of "constraint". That is, the association rules between ontology concepts are formulated. The ontology association rules are constructed mainly through the four control factors of attributes, constraint values, disjoint descriptions and logical relationship regulations. At present, the research and application of ontology are reflected in various industries. And the system is relatively complex. The general ontology classification method is based on the core application topics of each domain ontology. Scholars such as Guarino analyzed the ontology level and domain dependency and divided ontology into top-level ontology, domain ontology, task ontology, and application ontology.

In 2012, Google officially introduced the concept of knowledge graph to its search engine. The user’s search experience and search quality are significantly improved. Knowledge graph is a data storage scheme based on the representation of knowledge based on graph model. The data storage structure is a triple, namely entity-relation-entity [3]. The knowledge graph has fast data processing capabilities and powerful semantic relationship construction capabilities. So as to quickly emerge in the knowledge base application of various industries. Knowledge graph technology can sort out the relationship between people, things, and things for all kinds of emergencies, and combined with visualization technology, it can intuitively model real scenes. The key technologies for constructing knowledge graphs include knowledge extraction, knowledge storage, knowledge fusion, etc. Among them, knowledge extraction is an important research direction in the field of natural language processing (NLP) [4, 5].

3. Mass event ontology modeling

3.1. Data processing
Mass event information data mostly exists on the Internet in the form of unstructured (plain text), the unstructured data is mainly processed by knowledge extraction technology in the field of natural language processing (NLP) [6]. The Internet is the main source of data for mass incidents, use crawler tools to crawl information about 186 mass incidents at home and abroad from web pages, mark the text corpus with BIO, then based on the BiLSTM-CRF model to extract entities from the processed data. The entity extraction process is shown in Figure 1.

![Figure 1. Entity extraction process.](image)

In the field of natural language processing (NLP), named entity recognition is a very important data processing task. BiLSTM-CRF model is currently a more popular named entity recognition method [7], the model is divided into three layers: input layer (Look-up), hidden layer (BiLSTM), and labeling layer (CRF). When performing entity extraction, first mark the text corpus with BIO, the input layer of
the model maps the word to a vector \( x_i \) to obtain the sequence \( X(x_1, x_2, \ldots) \), in order to obtain the characteristics of the sentence, the hidden layer transforms the sequence \( X \) into a hidden state sequence \( h_i = (\tilde{h}_i; \tilde{h}_i) \), then it passes through the linear transformation layer and the output is \( p \). The CRF layer is responsible for sequence labeling [8]. This layer has a state transition matrix \( A \). The CRF layer performs label prediction scores on the sequence \( X \), that is, the score for the label sequence \( Y \) is \( \text{Score} \).

\[
\text{Score}(X, Y) = \sum_{i=0}^{n} A_{y_i, y_{i+1}} + \sum_{i=0}^{n} p_{y_i}
\]  

(1)

Use the Softmax function to get the probability \( P \) that the label sequence \( Y \) is the correct label of \( X \).

\[
P(Y | X) = \frac{e^{\text{Score}(X, Y)}}{\sum_{Y'} \exp(\text{Score}(X, Y'))}
\]  

(2)

The accuracy rate \( p = 89.54\% \) of the entity extraction experiment, the recall rate \( R = 80.23\% \)

\[
F1 = \frac{2pR}{p + R} = 84.47\%
\]  

(3)

A total of 2304 entities were extracted from the text corpus through entity extraction, including 394 personal name entities, 102 organization/institution entities, 23 event names, 804 item entities, 678 place name entities, and 303 others.

3.2. Ontology element creation
Ontology elements are also called modeling meta-language, which is the basic unit of constructing ontology [9]. According to the analysis of entity extraction results, the ontology elements of mass events are defined into seven categories: concepts, attributes, relationships, constraints, rules, axioms, and examples.

Concepts are the object types of mass incidents, that is, the classification of the ontology. The objects of mass incidents include people, events, things, organizations, regulations, etc. Attributes represent the characteristics or parameters of mass event objects and classes, such as the identity information of people, the style of control tools. Relations represent the interrelationships between classes and individuals, describing the relationships between concepts, people and people, people and things, people and organizations, etc. Constraint (restriction) is a formal statement that describes that a certain assertion can only be accepted as input when the relevant conditions are established, for example, the person involved and the person handling the case are not the same individual; The rules represent if-then (cause-consequence) statements, which can describe the logical reasoning on the temporal and spatial relationships of some group events, for example, spreading terror-related video and audio is an illegal activity; Axioms are assertions that are always correct in the ontology, including not only statements that are asserted as prior knowledge, but also theories derived from axiom statements, for example, guns are prohibited items, and the owner of the item is "human", etc. Examples are the objects at the bottom of the ontology. For example, a certain incident involved Li, who belonged to a certain organization abroad, and possessed guns.
3.3. **Ontology concept classification construction**

The constituent factors of mass events are diverse [10]. Through entity and relationship extraction of mass events and related documents, entity and relationship data is obtained, and the data is sorted and divided in multiple dimensions. Classify and construct the concept of mass incidents from seven aspects: "people", "incidents", "articles", "regulations", "emergency measures", "organizations", and "causes of incidents". And add attributes for each category. Figure 2 describes part of the ontology concepts of mass events.

![Figure 2. Partial ontology concepts of mass events.](image)

3.4. **Relationship expression**

The knowledge graph expresses knowledge in the form of triple RDF, that is, the storage type of entity-relationship-entity [11], for example: al-Qaeda-belonging to-terrorist organization. OWL is used to describe the ontology in the construction of the mass event ontology model. Based on the ontology theory, the basic relationship between the concepts, attributes, and entities of mass events is described as: The relationship between part and the whole (part-of), the inheritance relationship between concepts or classes (kind-of), the relationship between entities and concepts (instance-of), a concept is an attribute of another concept (attribute-of), constraint-on (constraint-on). The collective event ontology terminology can be divided into three categories: concept set, attribute set, and instance. The relationship structure between some ontology terms is shown in Table 1.

| Types of relationship | Concept-Concept | Concept-attributes | Concept-entity | Entity-attribute | Entity-entity |
|-----------------------|-----------------|-------------------|----------------|-----------------|---------------|
| Part-of/kind-of/attribute-of | Attribute-of/constraint-on | Instance-of | Attribute-of/constraint-on | kind-of/Part-of |

Import ontology terms and relationships into protege software to complete the visual display of the group event pattern layer, as shown in Figure 3. The model includes a collection of related concepts such as events, people, and things, and also sets the relationship between terms. Group events are a first-level concept in the ontology model, which constitutes a subordinate relationship with six second-level concepts such as person and thing. In addition, each second-level concept also includes multi-level attribute indicators. For example, the "person" category has two indicators of facial features and identity.
positioning, and facial features can be subdivided into secondary indicators such as face shape, skin color, and forehead features.

Figure 3. Visualization of mass event ontology model

4. Construction of Knowledge Graph of Mass Events

4.1. Data generation

Knowledge graph technology can model mass events in real terms, in which data is the basis for supporting the entire field graph. According to the current widely divided data sources, there are three main aspects: structured data, semi-structured data and unstructured data [12]. However, considering the particularity of mass incidents and the confidentiality of emergency plans of security departments, in order to prevent the leakage of emergency plan information, the visualization of mass incident knowledge graphs does not use real data. Here also solemnly declare that all the character information in the map does not refer to and use any real crowd information. Use excel forms to generate relevant elements of mass incidents, including personal information, social relations, and items involved in the case. Set the relationship between entities, where "person" is the subject data in the entire knowledge graph, the simulation data of "person" is mainly generated from three aspects of physical characteristics, identity characteristics and social information. The positioning of "persons" is divided into three categories, namely, personnel involved, security personnel, and key monitoring personnel, and combined with neo4j database visualization display. Figure 4 uses a person involved as an example to illustrate the composition of individual data.

Figure 4. Schematic diagram of individual simulation data of involved personnel.
4.2. Visualization of knowledge graph of mass events

This paper builds a knowledge map of mass events based on ontology, and the visual display of the map is done through the neo4j graph database. Compared with other storage solutions of the knowledge graph, the graph database has natural advantages. It is a non-relational database in the NOSQL database. The graph database discards the concepts of libraries, tables, and fields, but redefines the data and the relationships between them as "nodes" and "edges", thus forming a huge knowledge network [13]. This method of using graph theory to store data not only displays the relationship between data more intuitively, but also greatly simplifies the operation of data storage and query, and makes up for the shortcomings of relational databases in storing "relational" databases. This is due to the special data storage structure and special optimization algorithms of the graph database [14], and its operating speed will not decrease significantly with the increase of the database storage capacity.

Figure 5. Visualization of knowledge graph

As shown in Figure 5, the neo4j database displays the map information of a few group events, including the name of the event, the type of the event, the information of the person involved, and the social relationship of the person. And define the attributes of event elements, formally express the relationship between event concepts and attributes. Due to the powerful correlation capabilities of the Knowledge Graph, we can easily perform correlation analysis on this type of event, which is convenient for the security department to quickly prepare emergency plans. Cypher is the query language of neo4j database [15], it is a declarative language, it allows expressive and efficient query of the graph storage without writing the traversal code of the graph structure, and can accurately locate a certain character or event node.

5. Conclusions

This article mainly explores the construction of a knowledge graph of mass incidents based on ontology, standardize some ontology terms in this field, the formal expression of the relationship between ontology concepts and attributes, finally, the simulation data is generated by combining the examples, visual display of knowledge graph through neo4j database. Mass incidents are mostly emergencies, therefore, the relevant massive data needs an efficient storage solution, in order to quickly retrieve and locate and generate emergency plans in time. The knowledge graph has a graph model to design storage solutions, it has high-speed data processing capabilities and powerful semantic association capabilities, not only can the node relationship be displayed intuitively, but also new information can be inferred. It is an effective solution for rapid analysis of mass incidents. In this article,
there is room for improvement in some steps and methods in the process of constructing the knowledge map of mass events. I will use this as the direction of my in-depth research in the future study.

References

[1] Wang S, Wang W, Zhuang Y, et al. An ontology evolution method based on folksonomy[J]. Journal of Applied Research and Technology, 2015.

[2] Westerinen A, Tauber R. Ontology development by domain experts (without using the "O" word)[J]. Applied ontology, 2017, 12(3-4):1-13.

[3] Kristiadi A, Khan MA, Lukovnikov D, et al. Incorporating Literals into Knowledge Graph Embeddings[J]. 2018.

[4] Shixiang L, Yanling C. Research on Text and Sentiment Classification of Particle Swarm Optimization in Natural Language Processing[J]. Machine tools and hydraulics, 2018, 46(24):156-161[Chinese].

[5] Fei W, Li C, Mianzhu Y, et al. Advances in natural language processing driven by new technologies[J]. Journal of Wuhan University (Engineering Edition), 2018, 051(008):669-678[Chinese].

[6] Tom Y, Devamanyu H, Soujanya P, et al. Recent Trends in Deep Learning Based Natural Language Processing [Review Article][J]. IEEE Computational Intelligence Magazine, 2018, 13(3):55-75.

[7] Ling L, Zhihao Y, Pei Y, et al. An Attention-based BiLSTM-CRF Approach to Document-level Chemical Named Entity Recognition[J]. Bioinformatics(8):8.

[8] Sheping Z, Hongyu D. BiLSTM-CRF-based knowledge graph entity extraction method[J]. Computer applications and software, 2019, 36(05):269-274[Chinese].

[9] Feiliang R, Jikun S, et al. A Survey of Domain Ontology Construction from Text[J]. Journal of Computer, 2019, 42(03):654-676[Chinese].

[10] Tong J, Zuo L. Weibo communication and government legitimacy in China: a computer-assisted analysis of Weibo messages on two 'mass incidents'[J]. Information Communication & Society, 2014, 17(1):66-85.

[11] Zhu H, Tian F, Wu K, et al. A multi-constraint learning path recommendation algorithm based on knowledge map[J]. Knowledge-Based Systems, 2018, 143(MAR.1):102-114.

[12] Xiao X, Skitmore M, Li H, et al. Mapping Knowledge in the Economic Areas of Green Building Using Scientometric Analysis[J]. Energies, 2019, 12(15):3011.

[13] Drakopoulos G, Kanavos A, Mylonas P, et al. Defining and evaluating Twitter influence metrics: a higher-order approach in Neo4j[J]. Social Network Analysis & Mining, 2017, 7(1):52.

[14] Drakopoulos G, Kanavos A, Makris C, et al. On converting community detection algorithms for fuzzy graphs in Neo4j[J]. 2016.

[15] Faming, Gong, Yuhui, et al. Neo4j graph database realizes efficient storage performance of oilfield ontology.[J]. PloS one, 2018.