Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Establishment Method of Knowledge Graphs for Public Security Cases

Mingyue Qiu (qiumy@nfpc.edu.cn)
Nanjing Forest Police College
Zhang Xueying
Nanjing Normal University

Research

Keywords: knowledge graph, text parsing, graph visualization analysis, intelligence analysis

Posted Date: August 16th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1833330/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Since the implementation of the police intensifying strategy by science and technology, how to make accurate intelligence analysis based on the modeling of massive multi-modal data has always been a real intricate problem during the applications of public security big data. Knowledge graph technique can fuse and marge multi-modal data in public security business to different intelligence entity elements including person, thing, material, time and position, and reconstruct the deep relations among data in the digital space, which can provide computable and minable data resources with logical hierarchy for intelligence analysis. This study started from the case text data in public security domain, attempted ontology modeling for knowledge mapping in public security domain and explored the establishment mode and application direction of knowledge graphs in public security domain.

1. Introduction

With continuous improvement of policing informatization level, enormous data has been accumulated in public security system, however, because of unstructured property, low data quality and diversity, most of the data cannot be effectively developed and utilized. Accordingly, police information has gradually developed towards intelligent direction. Off late, the knowledge domains have become the research hotspot in public security informatization construction leading to remarkable progress and application exploration. Overall, the public security knowledge graphs consisting of 6 types of entities (person, thing, location, object, organization and visual identity) have been preliminarily formed (Li, 2022). However, public security knowledge not only possesses the connotation and characteristics of universal knowledge but also is featured by specific temporal/spatial characteristics and the main clue of case incidents. Therefore, public security knowledge graph still poses a series of scientific problems and challenges during the establishment and applications such as: How to make structural abstraction and formal representation of public security knowledge? How to obtain rapidly and automatically multi-granularity public security knowledge from public security big data? How to mine effective information service from public security knowledge graphs for actual service application scenarios? It is undoubted that the development of public security information service towards intelligent, socialization and popularity direction can essentially be promoted, if these scientific problems can be adequately and effectively resolved.

Public security information is huge and entity relationship is complex. Therefore, how to organize the complex entities and their relations in public security information is a great challenge. The knowledge graph can intuitively describe and store the massive information such as concepts, entities, attributes and relations in the public security data, and can mine and analyze the knowledge from time, space, category and other multidimensional dimensions (Liu, 2022). The main contributions of this work are as follows:

- we attempt to conduct ontology modeling of the knowledge graph in the public security field on the basis of public security cases.

- In order to promote the in-depth mining and application of police big data resources and improve the automation level of intelligence mining, this paper explores an intelligent system that can automatically generate knowledge map based on case text, and improves the intelligence level of police intelligence.

The remainder of this paper is organized as follows. We analyze and summarize relevant research on knowledge graph in Section 2. In Section 3, we introduce the construction method of public security knowledge graph. We establish the public security knowledge graph based on the case text in Section 4. Finally, Section 5 presents the discussion and conclusion.

2. Related Research

2.1 Establishment and application of knowledge graphs

Knowledge graph is a kind of semantic network that can reveal the correlation among entities, which can be used for formal representation of things in multiple domains and the related correlations (Huang et al., 2019). Historically, knowledge graph has its origin of semantic network in the late 1950s and the early 1960s while the proposed expert system, semantic net and linked data are all the predecessors of the knowledge graph. Tim Berners Lee proposed the idea of semantic web on XML conference in 2000, with the aim of adding semantic information to web pages, supporting machine automatic processing and providing some semantic services such as information agent, search agent and information filtering. Metaweb Company, founded in 2005, is committed to develop open and sheared world knowledge library for Web semantic services. Based on some public data sets such as Wikipedia and U.S. Securities and Exchange Commission, Metaweb extracts entities (person or thing) and their correlations in the real world and stores them in the computer in the form of graph structure. In 2010, Google brought the brand of Metawebin 2010, mastered semantic search technique and proposed the concept of knowledge graph. In 2012, Google achieved the enhancement of search results using knowledge graph, which marked successful application of large-scale knowledge graph to semantic search on the Internet. Owing to its lot of characteristics such as large scale, abundant semantic information, high quality and structural friendliness, knowledge graph has gained rapid popularization in both academic and industrial circles, which also has become the most important way of knowledge representation in the era of big data (Liu, 2019).

Generally, knowledge graphs can be divided into two types–universal and industrial knowledge graphs. Universal knowledge graphs, such as encyclopedic DBpedia, zhishi.me, linguistic WordNet and Cilin, describe comprehensive common-sense knowledge and are mainly used for semantic search, which set low requirements on knowledge accuracy. Industrial knowledge graphs for specific domains, such as geological knowledge graph ‘LinkedGeoData’ and traditional Chinese medicine knowledge graph, pose high requirements on professionalism and accuracy, which can thereby take knowledge reasoning and achieve auxiliary analysis and decision support.

2.2 Research progress on public security knowledge graph

Traditional data retrieval and information understanding technique can hardly satisfy the public security business requirements under big-data environment. The unprecedented scale of public security data and the complex degree of case relation, traditional data retrieval functions impose the technical bottlenecks that can hardly be overcome in terms of semantic understanding and knowledge transformation. Knowledge graphs aim to obtain and fuse abundant semantic knowledge from massive public security data so as to form huge and inter-correlated public security knowledge network and achieve automatic mining and reasoning of hidden information. Therefore, rapid and accurate retrieval of huge accounts of public security data can be achieved through semantic understanding. Furthermore, knowledge graph can provide an effective means of data analysis, information understanding and intelligence service under public security big data environment.

Subsequently, scholars focused on criminal knowledge discovery based on the artificial intelligence. Pendharkar *et al.* proposed the multi-agent system establishment method for drug-related crime knowledge management based on the hybrid Bayesian network. As regards the component events under storage scenario, Keppens *et al.* pointed out that these component events can be recombined into useful scenarios via instantiation. The knowledge-driven crime scene establishment method is quite applicable to many cases since it allows the matching of the component events in many ways to predict the cases under investigation. Phillips *et al.* proposed a graph-based data set representation method to mine the correlation among data sets, and developed a reasoning framework for independent exploratory analysis and knowledge discovery from large-scale crime data sets. Elezaj *et al.* suggested framework based on knowledge graph method, which can utilize a hybrid agent connecting to a graphic data base, perform automatic processing of unstructured data and recognize the mode and relation hidden in the social network data.

3. Establishment Of Public Security Knowledge Graph

3.1 Components of knowledge graph

Knowledge graph mainly includes two basic semantic units namely, entity and relation. Entity refers to the things in the real world such as criminal suspect, materials and vehicles involved, while the relation is used for representing the correlation among different entities (for example, Zhangsan-sell-parrot, Zhangsan-criminal partner-Wangwu, Wangwu-higher leadership-Zhangsan). The sets of graph structures composed of triples (entity-relation-entity) act as the data layer in knowledge graph.

In addition to graph, knowledge graph still has the element of knowledge, which represents the stable mode among abstracted entities. For example, leader-member relation can be frequently found in a criminal gang, some criminal behaviors have constant crime timings and positions. The logical relation network consisting of these modes is referred to as the schema layer in knowledge graph.

3.2 Representation of knowledge graph

Knowledge representation is a data structure and data description mode in knowledge graph. Generally, data structures are mainly graphs and trees, while the schema layer is represented by tress and the data layer is described by graphs.

In JSON-LD standard, each entity can be represented by a JSON e. g, a criminal suspect can be described as:

```
{
    "name": "Zhang San",
    "telephone": "13918276387",
    "image": "http://manu.sporny.org/images/manu.png",
    "gender": "Male"
}
```

Where, ‘name’, ‘telephone’, ‘image’ and ‘gender’ are the name, the telephone, the image and the gender of the criminal suspect, respectively. However, the computer cannot understand the terms such as ‘name’ and ‘gender’. JSON-LD also supports linking each field to the semantic information in the schema layer. Accordingly, the above example can be written as:

```
{
    "http://schema.org/name": "Zhang San",
    "http://schema.org/telephone": {
        "@id": "13918276387"
    },
    "https://schema.org/image": {
        "@id": "http://manu.sporny.org/images/manu.png"
    },
    "https://schema.org/ gender": {
        "@id": "Male"
    }
}
```

By taking the data field in the first line that records the name as an example, "http://schema.org/name" points to the semantic description of ‘name’ in the schema layer. For convenience, sharing and cooperation, *schema.org* maintains a world-class semantic ontology library (the Chinese mirror website: *schema.org.cn*). For simplification, the above criminal suspect can be described as:

```
{
  "@context":"https://schema.org.cn/Person",
  "name": "Zhang San",
  "telephone": "13918276387",
  "image": "http://manu.sporny.org/images/manu.png",
  "gender": "Male"
}
```

Where, '@context' denotes the context linkage information that links the information to the semantic information of person on *schema.org.cn*. The semantic information describes the concept of person. Accordingly, JSON-LD can achieve decentralized semantic representation for all existing knowledge and flexibly establish large-scale knowledge graphs. The present knowledge graphs also adopt advanced JSON-LD standard for knowledge representation.

3.3 Knowledge modeling and representation

Knowledge modeling and representation should first establish the mode layer and then the data layer. The mode layer defines both structure and data organization mode of knowledge graph and stores abstract knowledge system. Based on the knowledge system of the mode layer, the data layer constitutes the graph by establishing the correlations among intelligence factors. The mode layer is relatively stable than the data layer. The data layer can mark the detailed knowledge system based on the model layer. Therefore, the schema layer of the knowledge graph should firstly be established so as to analyze the marks on the basis of ontological mode and form the data layer by correlating the graph data.

3.3.1 Establishment method of the mode layer of knowledge graph

The present mode layer was established by the circulation method for establishment of domain ontology. Both advantages and disadvantages of ontology construction in various domains were analyzed. It was concluded that the circulation method is applicable to constant cyclic reciprocation and spiral escalation in the construction of schema layer in public security domain (Zhang et al., 2011). Firstly, the case text data should be pre-processed and analyzed so as to extract the core concept, and the entity relationship concepts can be classified and layered in terms of intelligence factors such as person, thing, material, time and location to form a conceptual framework. Next, the public security entities and the universal entities were combined to form the ontology knowledge library. The knowledge library was then mainly updated by machine and assisted by artificial analysis.

3.3.2 Establishment method of the data layer of knowledge graph

The data sources are abundant and diverse in the establishment of data layer. Structured, unstructured and semi-structured data can be utilized in the knowledge modeling and representation. This study laid the research emphasis on knowledge annotation and relationship modeling for the case text data, and transformed the case text data into the graphs including relational triples. The case text data should first be parsed into lexical item sequences via knowledge annotation, next, the relations among the lexical items represented by entities can be analyzed via entity linking, and the nodes of an object with different names should be normalized. For example, if the crime suspects ‘Zhangsan’ and ‘Laozhang’ have the attributes of bareheaded property, tattooing and severed fingers except different names, ‘Zhangsan’ and ‘Laozhang’ can be regarded as a node. For the field that can be uniquely judged such as national identification number, the normalization of nodes can directly be taken by setting the attribute rules. Finally, the finished knowledge graph was formed after screening. The service was requested from the API of knowledge graph system, and meanwhile, the new graph pattern was fed back to the ontology knowledge library for re-updating. Accordingly, the whole life cycle of the application of knowledge application can be finished.

4. Establishment Of Public Security Knowledge Graph Based On The Case Text

4.1 Establishment of the mode layer of the public security knowledge graph based on the case text

For the establishment of the model layer of the graph, the case text data should first be pre-processed and analyzed to extract the core concepts, and the entity relationship entities were classified and layered in terms of some intelligence factors including person, thing, material, time and location, to form the conceptual framework. Next, the conceptual framework was combined with the universal ontology to form public security ontology knowledge library. The public security ontology library, as the storing carrier of intelligence knowledge, can collect, organize and share constantly iterated and updated public security intelligence knowledge and business knowledge based on logic. Accordingly, the problem that knowledge is easily lost in the public security information service can be effectively solved.

4.1.1 Case text acquisition and pre-processing

The related data were sought and downloaded from the website (https://wenshu.court.gov.cn/) of judgment document, including 36 charges and 60,746 cases in four parts. The first part includes 18,182 case texts with 10 charges of producing and selling fake and inferior commodities. The second part includes 8,802 case texts with 8 charges of infringing intellectual property rights. The third part includes 2

charges of endangering public health, with 35 case texts. The fourth part includes 33,727 case texts with 16 charges of destroying the protection of environmental resources. In this study, for each charge, 300 data was extracted from the case texts for artificial processing. The involved key industries, key sites, key parts, key personnel, main materials, main species and main crime methods were extracted from the case texts as per related standards formulated according to the characteristics of the common cases and the related national standards.

4.1.2 Establishment of public security ontology

According to the above described circulation method, the present ontology was established based on the data structured by the above case texts. The detailed procedures are described below.

1) Analyze the ontology demand and investigate the reusable ontology

In the public security domain, the ontology was established based on the case text data. After investigation of the related literature, the public domain ontology can be expanded based on the universal ontology knowledge library. The present study took the expansion on the encyclopedic knowledge tree TermTree.

2) Establish the domain core concepts

By summarizing the structured data of the above case texts and taking the statistics in terms of charges, the word frequencies of the lexical terms in each field were recorded, and the high-frequency words ranking the first quarter were screened out. Accordingly, the core concepts of the cases of the charges were obtained. For example, the key concepts of the charge of illegal logging can be obtained and the knowledge nodes in the schema layer were established on the basis of these core concepts.

3) Establish the conceptual taxonomic hierarchies and defining the knowledge nodes

These core conceptual factors were classified into five categories (person, thing, material, time and position). The object's attribute hierarchy was established by referring the semantic description of OpenSchema. For example, the semantic level of a person can be obtained by referring to https://schema.org.cn/Person. According to the hierarchy and semantic structure, the terms can be filled in so as to acquire the final ontology scheme in the public security domain.

By adding the ontology pattern in the public security domain to the universal ontology knowledge tree, the ontology in public security Termtree can be obtained. The core codes are given below.

```
 {for i in range(0, len(p)):
    thisterm = p.iloc[i]
    # Add new term
    try:
        termtree.add_term(term=thisterm["term"],
                base=thisterm["base"],
                term_type=thisterm["term_type"])
    except Exception as e:
        print(e)
}
```

Using the codes, the ontology lists in public security domain was transversed so as to extract the ontology knowledge and convert it into JSON format. By writing the ontology into the universal ontology, the public security ontology Termtree can finally be obtained.

4) Ontology evaluation and evolution

The ontology should constantly be updated and maintained in accordance with the actual requirements. New intelligence requirements should be analyzed by returning to the first step and the life cycle of ontology construction re-operated.

4.2 Establishment of the data layer of the public security knowledge graph based on the case text

Based on the knowledge system in the schema layer, the graph data layer can unearth the intelligence clues from the case texts to be processed, which can provide a creative text structured means for information mining from the case texts in public security domain. It can contribute to dig out the hidden entity relationship from huge case text library so as to form the intelligence clues in the public security work. The establishment of data layer in the knowledge graph includes two steps: knowledge labeling and entity linking. Firstly, the entity set in the texts was parsed via knowledge labeling and the entities were correlated via entity linking so as to form the graph.

To be specific, knowledge labeling was taken based on the established public security ontology and the open-source knowledge labeling tool from Baidu '*Jieyu*'., while entity linking was performed based on the

open source natural language processing tool from Harbin Institute of Technology *LTP* and the open source natural language processing tool from Baidu PaddleNLP (Zhao et al., 2020).

4.2.1 Knowledge labeling

An important problem in knowledge labeling that needs to be solved is entity disambiguation. A word can have different semantic meanings under different contexts. For example, the word *skating* in two sentences ‘Yu plays skating in the gymnasium’ and ‘Yu plays skating using his curling’ have different semantic meanings. The former refers to a kind of sports while the latter refers to taking cocaine. How to realize accurate knowledge labeling by discriminating the semantic meanings of words is the main issue? Based on the open-source knowledge labeling tool ‘Jieyu’, this study attempted to take knowledge labeling in the public security domain based on the established domain knowledge ontology.

First, the case texts were pre-processed via word separation and part-of-speed tagging so as to form a series of part-of-speech. Next, in combination with knowledge ontology tree, the noun phrases were classified via named entity recognition and the ontology mode class that the entity belonged to was obtained. Finally, the knowledge ontology nodes corresponding to the word were retrieved so as to obtain the knowledge of the entity word so as to finish knowledge labeling. By taking the sentence ‘Zhang killed Zhao and discarded the dead body into the river’ as an example, the detailed labeling process illustrated in Figure 1.

A complete sentence was separated into lexical item series. Next, the part-of-speech information of these lexical items was obtained via part-of-speech tagging. For example, killing can be understood as an event, and its part-of-speech information is the class of the knowledge ontology that the lexical item belongs to. Killing corresponds to the class of scene event on the knowledge ontology tree. Finally, the corresponding knowledge of the word can be found from the class on the knowledge ontology tree so as to finish knowledge labeling process. The core codes are given below.

```
>>> from paddlenlp import Taskflow
>>> wordtag = Taskflow("knowledge_mining",
model="wordtag",
linking=True,
task_path="./custom_task_path/")
>>> wordtag("Zhang killed Zhao and dumped the body into the river")
```

To be specific, Taskflow executor in paddlenlp was introduced in the first line, to obtain knowledge labeling tool of ‘Jieyu,’ then self-defined catalogs that the public security knowledge ontology stored was assigned in the second line, and the knowledge labeling was performed in the third line.

```
 [
   {   "text": "Zhang killed Zhao and dumped the body into the river
",
       "items": [
          {
             "item": "Zhang",
             "offset": 0,
             "wordtag_label": "Person_entity",
             "length": 2
          },
          {
             "item": "Kill",
             "offset": 2,
             "wordtag_label": "Sceneevent",
             "length": 2,
             "termid": "Sceneevent_cb_kill"
          },
          {
             "item": "Zhao",
             "offset": 4,
             "wordtag_label": "Person_eneity",
             "length": 2
 },
          {
             "item": "and",
             "offset": 6,
             "wordtag_label": "Conjunction",
             "length": 1,
             "termid": "Conjunction_cb_and"
          },
```

```
        {
            "item": "Dump the body",
            "offset": 7,
            "wordtag_label": "Scene event",
            "length": 2,
            "termid": "Scene event_cb_dump the body"
        },
        {
            "item": "into the river",
            "offset": 9,
            "wordtag_label": "Position",
            "length": 2,
            "termid": "Position_cb_into the river"
        }
      ]
    }
]
```


Where, the *text* field represents the input text, the *items* field represents the list of parsed lexical term results (in which the field of each lexical term represents the lexical item text), *offset* represents the index at the beginning of the lexical item text, *length* represents the length of the lexical item text, wordtag_label represents the class in the ontology knowledge library after labeling, and *termid* represents the knowledge node retrieved from the class.

4.2.2 Entity linking

A. Relationship parsing based on dependency grammar analysis

Dependency grammar analysis, as a key technique in natural language processing, aims to obtain the sentence's syntactic structure by recognizing interdependence relation among vocabularies in the sentence. Based on the obtained syntactic structure after dependency grammar analysis, the relation parsing in this study was performed using the LTP-based dependency grammar analysis model Biaffine (Dozat et al., 2017). The Stanford Dependencies Chinese (Chang et al., 2009) was selected as the annotation standard. The extraction rules are described in Table 1.

Table 1. Relation extraction rules

Relation type	Label	Component explanation	Example
Subject-predicate relation	SBV	subject-verb	Yu sold cocaine to Zhao (Yu <– sell)
Verb-object relation	VOB	Direct object, verb-object	Yu sold cocaine to Zhao (sell–> cocaine)
Indirect-object relation	IOB	Indirect object, indirect-object	Yu sold cocaine to Zhao (sell–>Zhao)
Fronting object	FOB	Front object, fronting-object	Yu kills anyone (person<– kill)
Pivot	DBL	double	Yu ran off with money (with –>money)
Attributive-centered relation	ATT	attribute	First-grade goods (first-grade <– goods)
Adverbial-verb structure	ADV	adverbial	Very dangerous (very dangerous)
Verb-complement structure	CMP	complement	Finish committing a crime (commit –> finish)
Coordinating relation	COO	coordinate	Yu and Zhao (Yu –> Zhao)
Proposition-object relation	POB	preposition-object	In the trade zone (is –> in)
Left adjunction relation	LAD	left adjunct	Yu and Zhao (Yu <– Zhao)
Right adjunction relation	RAD	right adjunct	Partners (partner –> partners)
Independent structure	IS	independent structure	-
Core relationship	HED	head	-

B. Relation linking based on text matching

The present study adopted the text similarity matching model SimBERT in open-source natural language processing toolkit PaddleNLP on Baidu (Su, 2020). First, the sentences were input in the pre-training model for generating the sentence vectors, using cosine function the similarity degree between sentence vectors was calculated to form the similarity degree matrix. For example, in Fig. 2, *Yu committed a crime together with Yao* is similar to *Yu and Yao committed a crime together.* The sentence was marked by the purple box, the similar sentence was denoted as 1, and the dissimilar sentence was denoted as 0.

Accordingly, the data of Mask_1_0_0_0_0 can be obtained. The similarity matrix can be obtained by combining multiple sentences. The core codes are given below.

```
 #Generate the sentence vector
def extract_emb_feature(model,tokenizer,sentences,max_len,mask_if=False):
    mask = generate_mask(sentences,max_len)
    token_ids_list = []
    segment_ids_list = []
    for sen in sentences:
        token_ids, segment_ids = tokenizer.encode(sen,first_length=max_len)
        token_ids_list.append(token_ids)
        segment_ids_list.append(segment_ids)
    result = model.predict([np.array(token_ids_list), np.array(segment_ids_list)])
    if mask_if:
        result = result * mask
    return np.mean(result,axis=1)


#Generate the similarity matrix
def generate_mask(sen_list,max_len):
    len_list = [len(i) if len(i)<=max_len else max_len for i in sen_list]
    array_mask = np.array([np.hstack((np.ones(j),np.zeros(max_len-j))) for j in len_list])
return np.expand_dims(array_mask,axis=2)
```

By packaging above logic into the function of similarity, the similarity degree between the lexical items output after knowledge labeling and the entity items in the triple output after dependency grammar analysis was calculated and matched. For each triple relation, the similarity degree between the entity texts at the 0 and the 2nd index and the entity term text after knowledge labeling was calculated and matched. Two entities were linked so as to link the relation between entities and output the graph. The core codes are given below.

```
 for i in depout:
    a = max(similarity([[i[0], s] for s in nerout]),key=lambda x:x["similarity"])
    b = max(similarity([[i[2], s] for s in nerout]),key=lambda x:x["similarity"])
pprint(select(a)+'->'+i[1]+'->'+select(b))
Zhang–kill–Zhao
Zhang–dump the body–-into the reiver
```

C, Data format conversion and visualization

The final output triple was converted in format so as to achieve visual display. During data conversion process, the entities in the triple can be assigned with the only identifier that was randomly generated. By treating the class of ontology annotation as the entity tag, the relation can be used as the relation type. The unique identifiers of left and right entities were filed in the field of ‘source entity id’ and the field of ‘target entity id’ so as to obtain the graph data for visualization. By taking the above triple output as an example, the Excel dataset after processing is listed below, in which the Vertexes table records the entity term and the Endges table records the relation term.

Table 2. Vertexes table

Entity id	Entity name	Entity tag	Attribute name_1	Attribute value_1
ID-Zhang	Zhang	Person	Identity	Criminal suspect
ID-Zhao	Zhao	Person	Identity	Criminal victim
ID-He	In the river	Position		

5. Conclusion

This study proposed a graph generation method based on the case texts. The knowledge information in the information knowledge library was mapped to individual case text and the graph describing the text semantic meaning was plotted by analyzing the entity relation in the text. Compared with the traditional text statistical analysis methods, the present method can automatically find the logical relation between semantic entities and flexibly supplement and update the entity knowledge library in accordance with the business change. In case of public security information work, the proposed method can achieve rapid and automatic cleaning of the case text data scattered in different databases and information mining, and establish the relations among different entity elements including person, thing, material, time and location thereby, effectively addressing the afore mentioned information isolated island and data iceberg problems. In the future, we will continue to maintain the ontology and system system based on the open-

source community model, update more practical functions and expand the application ecology of knowledge graph in the field of public security.

6. Declarations

6.1 Availability of supporting data

The datasets generated and/or analysed during the current study are available in the judgment document repository, https://wenshu.court.gov.cn/.

6.2 Competing interests

The authors declare that they have no competing interests.

6.3 Funding

This study is partially supported by Project on 2021 Outstanding young backbone teacher of Jiangsu Universities "Qinglan Project", No. LGZD202203 the Fundamental Research Funds for the Central Universities.

6.4 Authors' contributions

QM conducted ontology modeling of the knowledge graph in the public security field on the basis of public security cases and was a major contributor in writing the manuscript. ZX explored an intelligent system that can automatically generate knowledge map based on case text. All authors read and approved the final manuscript.

6.5 Acknowledgements

We would like to express our sincere gratitude to the anonymous reviewers and professional writing services for their valuable comments, which have greatly improved this paper.

References

1. Q.J. Li, F. Li, S.C. Li, X.Y. Li, K. Liu, Q. Liu, P.C. Dong. Improving Entity Linking by Introducing Knowledge Graph Structure Information[J]. Applied Sciences,2022,12(5).
2. S. Liu, K. Zhang, J. R. Shang. Retrospection and prospect of embeddedness theory – knowledge map analysis based on bibliometrics[J]. Journal of Data, Information and Management,2022,4(1).
3. H. Q. Huang, J. Yu, X. Liao, Review on Knowledge Graphs, Computers Sciences & Applications, 2019, 28(6): 1-12.
4. Z. Y. Liu. Research and Implementation of Knowledge Graph Mapping Technology for Public Security Intelligence Analysis, A Master thesis from University of Electronic Science and Technology of China, 2019.

5. P. C. Pendharkar, R. Bhaskar, B. Ghahramani, A hybrid Bayesian network-based multi-agent system and a distributed systems architecture for the drug crime knowledge management. International Journal of Information Technology & Decision Making, 2003, 2(04): 557-76.
6. J. Keppens, B. Schafer. Knowledge based crime scenario modeling. Expert Systems with Applications, 2006, 30(2): 203-22.
7. P. Phillips, I. Lee, Mining top-k and bottom-k correlative crime patterns through graph representations, Proceedings of the 2009 IEEE International Conference on Intelligence and Security Informatics, 2009: 25-30.
8. O. Elezaj, S. Y. Yayilgan, E. Kalemi,. Towards Designing a Knowledge Graph-Based Framework for Investigating and Preventing Crime on Online Social Networks. Proceedings of the International Conference on e-Democracy, 2019: 181-195.
9. W. X. Zhang, Q. H. Zhu, Research on Construction Methods of Domain Ontology [J], Library and Information, 2011 (01): 16-19+40.
10. M. Zhao, H.P. Qin, G.X. Zhang, Y.J. Lyu, Y. Zhu. TermTree and Knowledge Annotation Framework for Chinese Language Understanding [R], Baidu, Inc. 2020, TR: 2020-KG-TermTree.
11. T. Dozat, C. Manning, Deep Biaffine Attention for Neural Dependency Parsing[R]. Proceedings of the 5th International Conference on Learning Representations, 2017. ICLR'17.
12. P. Chang, H. Tseng, D. Jurafsky, & Manning C. Discriminative reordering with Chinese grammatical relations features[R]. Proceedings of the Third Workshop on Syntax and Structure in Statistical Translation (SSST-3) at NAACL HLT 2009, 51-59. Boulder, Colorado, June 2009. Association for Computational Linguistics, URL.
13. J. Su. SimBERT: Integrating Retrieval and Generation into BERT [R]. 2020.

Figures

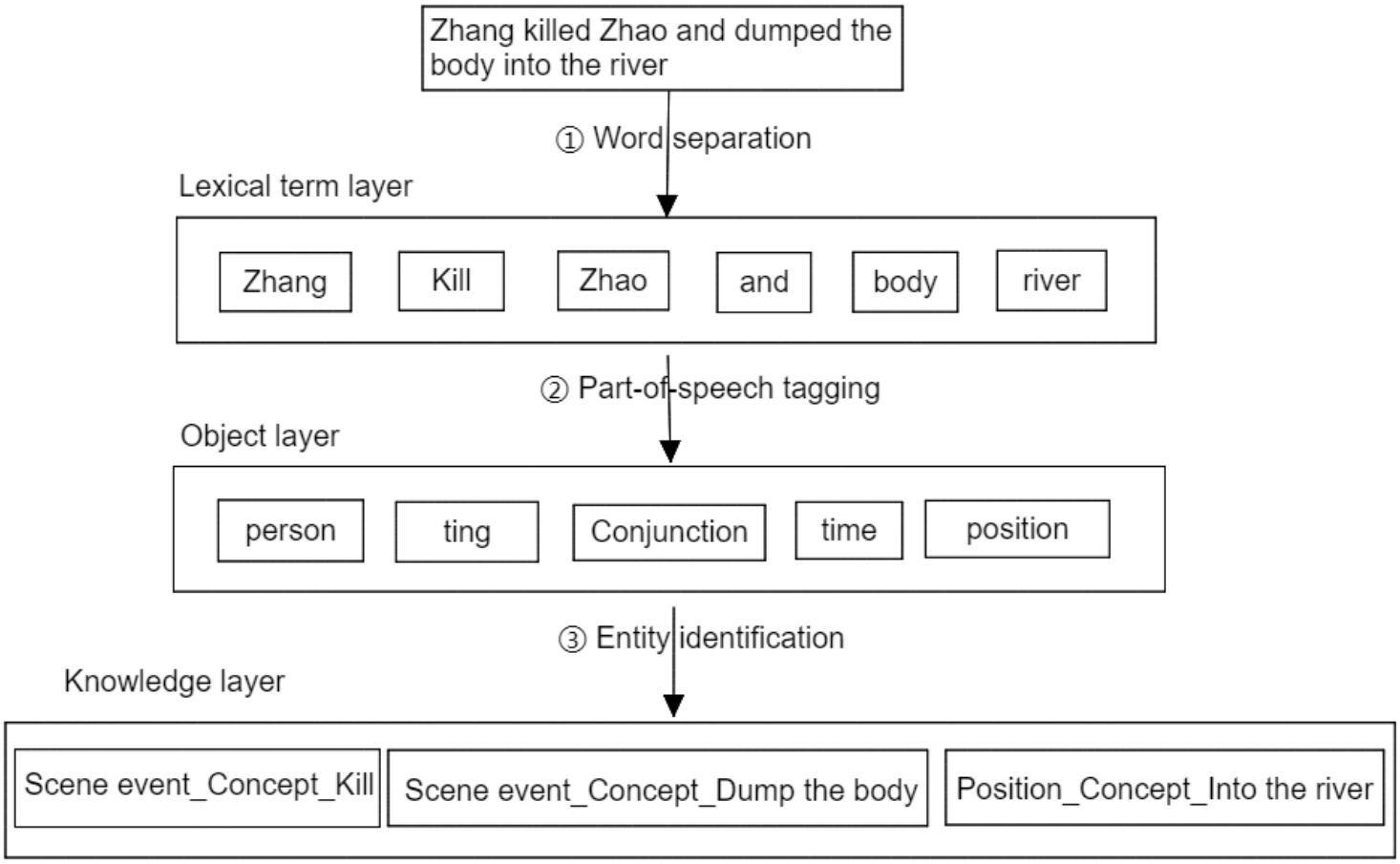

Figure 1

A case study of knowledge labeling

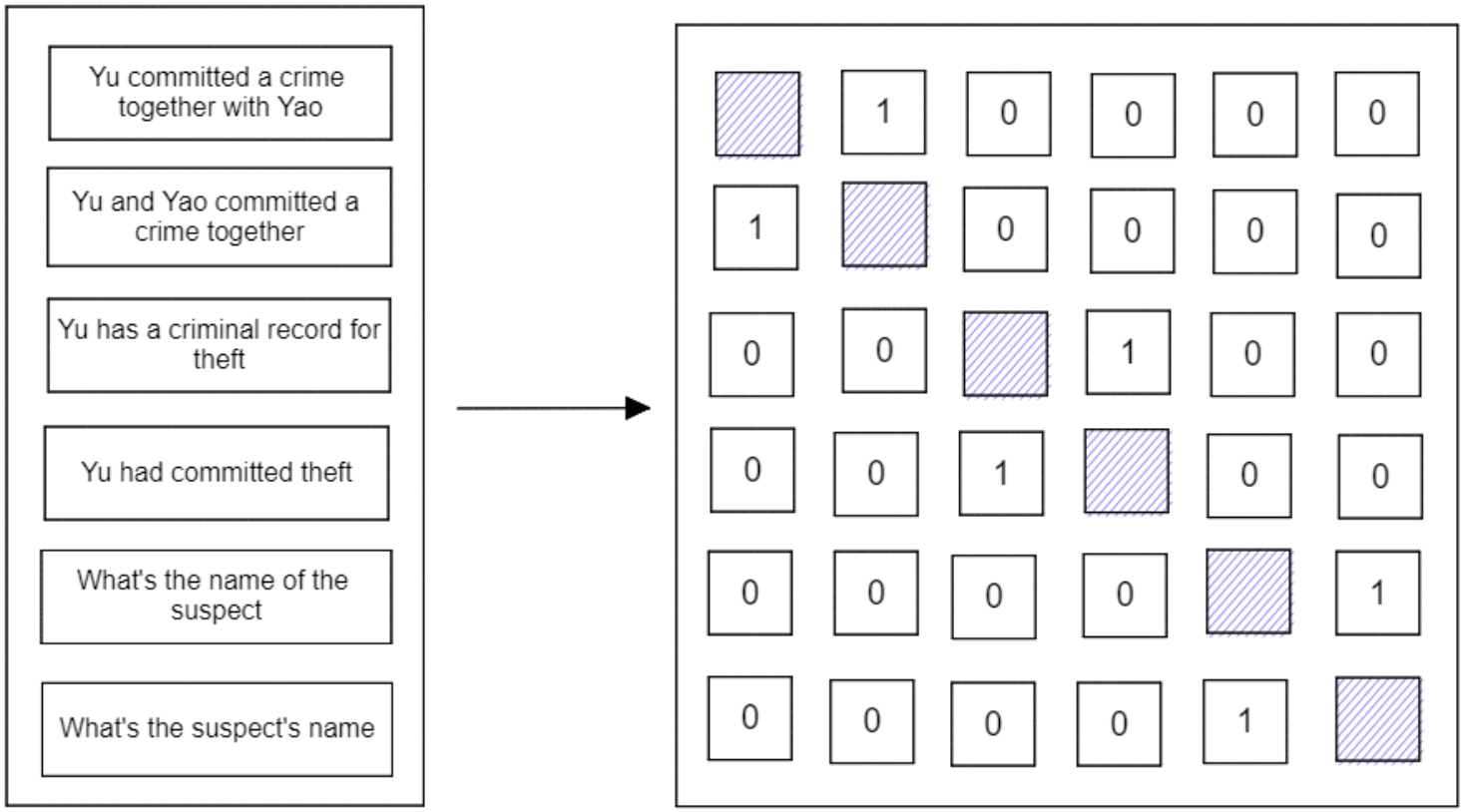

Figure 2

Illustration of similarity degree in text matching

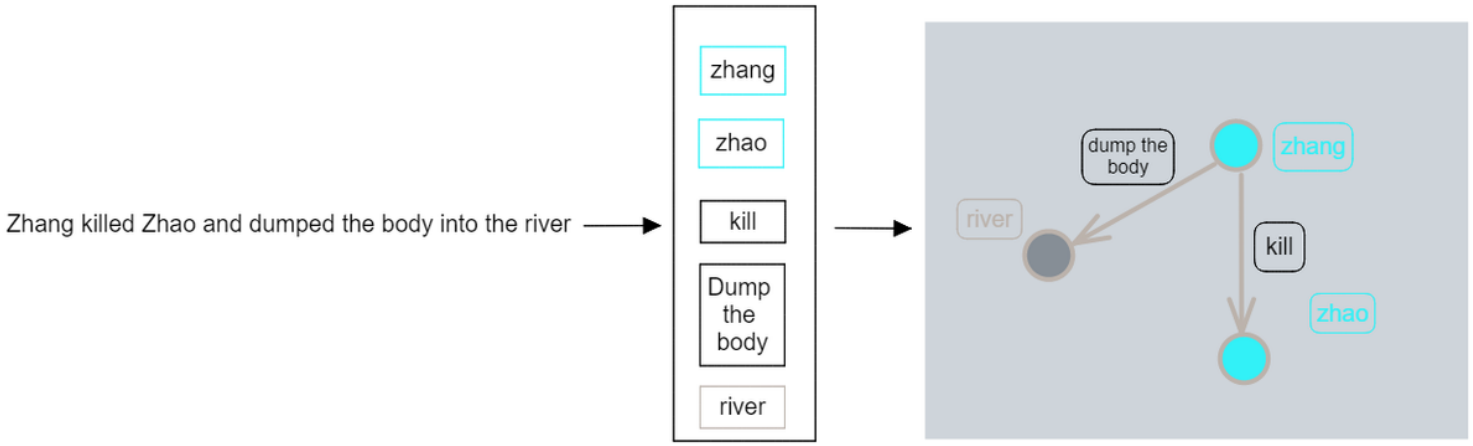

Figure 3

Graph visualization result

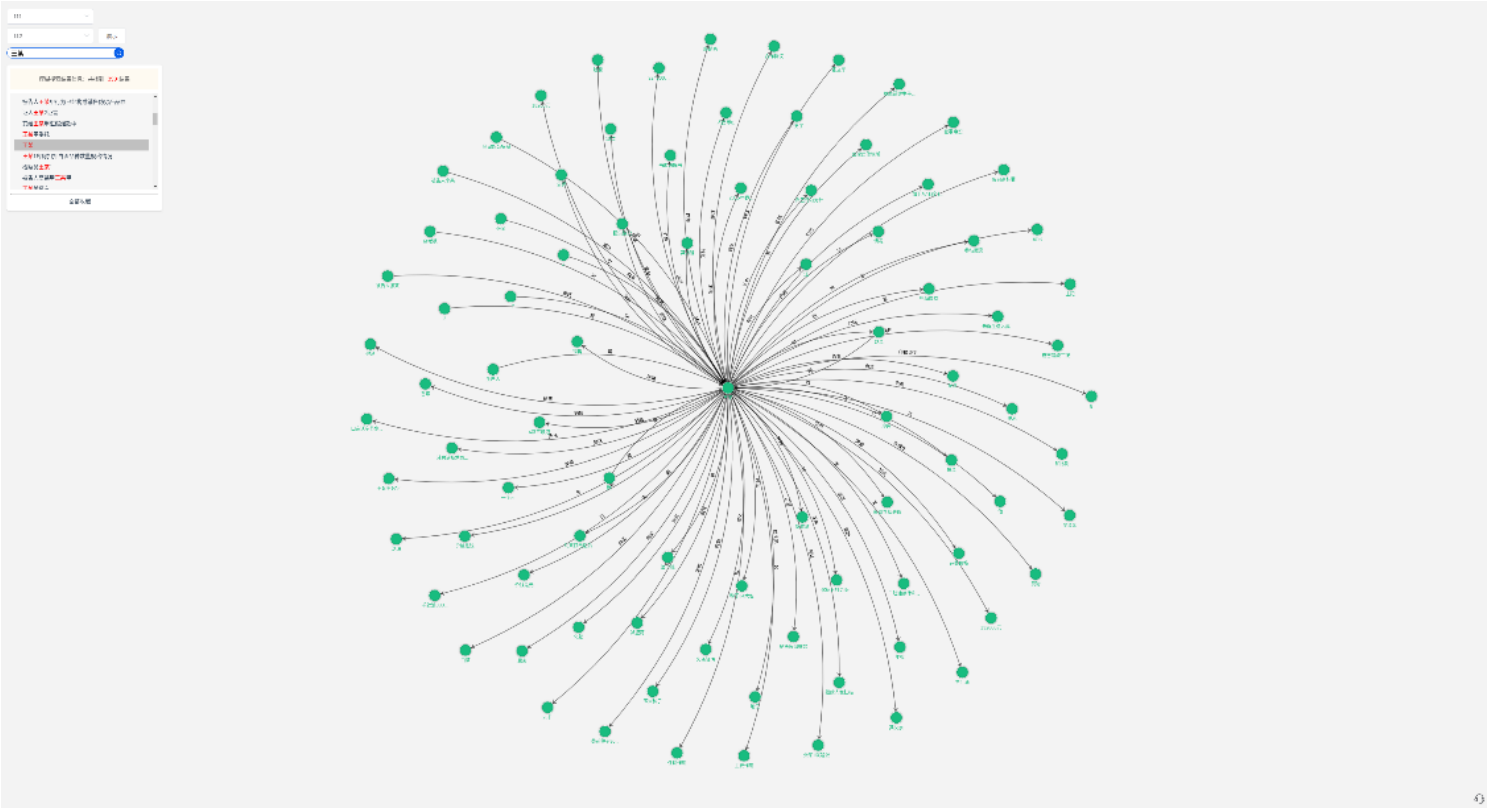

Figure 4

Public security graph obtained by searching *Wang*

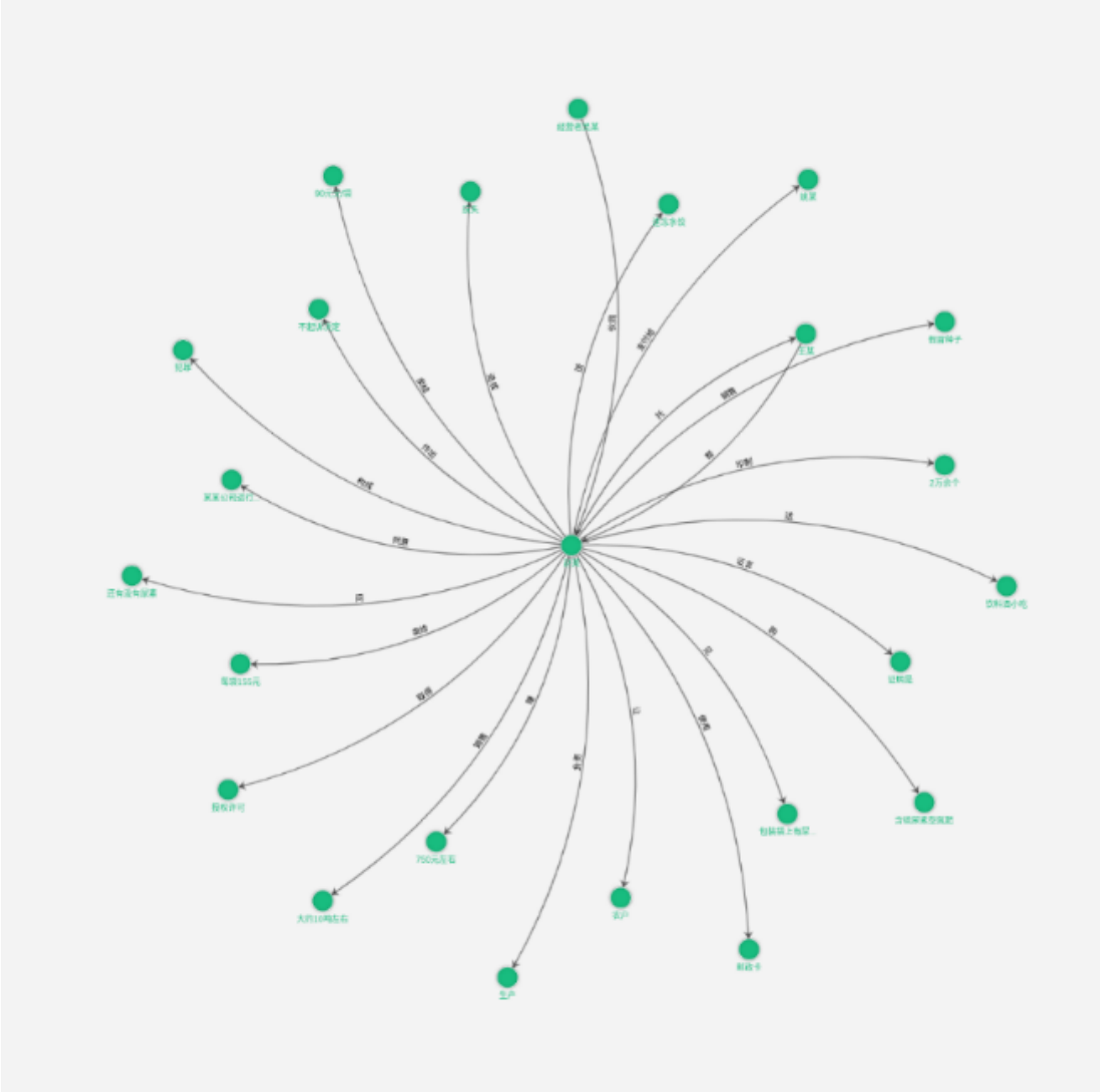

Figure 5

Public security knowledge graph obtained by clicking *Zhao*