# Road Traffic Accident Data Management and Application Analysis Based on Knowledge Graph Technology

Dexin Yu*
Navigation College, Jimei University, Xiamen, 361021, China
yudx@jmu.edu.cn

Wanli Peng
Navigation College, Jimei University, Xiamen, 361021, China
wanlipeng0915@163.com

Yunjie Chen
Navigation College, Jimei University, Xiamen, 361021, China
cyunjie8080@163.com

Yu Yang
Navigation College, Jimei University, Xiamen, 361021, China
yangyu0924yy@163.com

Hongyu Chen
Navigation College, Jimei University, Xiamen, 361021, China
chen1707040621@163.com

## ABSTRACT

Amidst the digitalization wave, people realize the enormous value hidden in data. To effectively handle and scientifically manage vast amounts of road traffic accident data and achieve the integration of fragmented information, this paper proposes a method using knowledge graph technology to manage road traffic accident data. Firstly, based on the analysis of the structural characteristics of accident data, and with traffic safety management as the core, the framework of the road traffic accident knowledge graph was designed. Subsequently, entities, relationships, and attributes were extracted from road traffic accident text reports and imported into the Neo4j graph database to finalize the construction of the road traffic accident knowledge graph. Finally, the application analysis was conducted based on the constructed road traffic accident knowledge graph. The research results provide new methods and conceptual frameworks for the digitalization and refinement of traffic safety management, contributing to its more efficient and intelligent management.

## CCS CONCEPTS

• **Information systems** → Data management systems; Database design and models; Entity relationship models.

## KEYWORDS

Road traffic accident, Knowledge graph, Digitalization, Traffic safety management, Graph database

**ACM Reference Format:**

Dexin Yu*, Wanli Peng, Yunjie Chen, Yu Yang, and Hongyu Chen. 2024. Road Traffic Accident Data Management and Application Analysis Based on Knowledge Graph Technology. In *2024 International Conference on Digital Society and Artificial Intelligence (DSAI 2024), May 24–26, 2024, Qingdao, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3677892.3677940



*DSAI 2024, May 24–26, 2024, Qingdao, China*



ACM ISBN 979-8-4007-0983-8/24/05
https://doi.org/10.1145/3677892.3677940

## 1 INTRODUCTION

The rapid development of global transportation and the automobile industry, the number of vehicles worldwide has increased significantly, leading to compounded safety risks on the roads and frequent traffic accidents. Currently, an estimated 3,000 people worldwide die daily as a result of traffic accidents, with this figure continuing to rise. Road traffic accidents have become a global challenge. When road traffic accidents occur, a massive amount of accident data is inevitably generated, with diverse types and closely interrelated and intersecting elements. Furthermore, the wave of big data in transportation has arrived, with the digital application of traffic accident data has become a significant trend. Confronted with vast and semantically rich accident data, effective organization, management, understanding, and utilization can establish a data foundation for the analysis and extraction of accident data. This, in turn, provides a scientific basis and decision support for the prevention, emergency response, and policy-making related to road traffic accidents.

Under specific spatiotemporal conditions, once a road traffic accident occurs, the information related to the accident is identified, and accident data is precisely the mapping of these relevant pieces of information [1]. Road traffic accident data mainly record the information of the person involved, vehicle, road, and environmental in the accident, and their interactions collectively map the process and causes of traffic accidents. Road traffic accident data mainly consist of semi-structured and unstructured paper text information, such as accident scene records and accident investigation reports. These data are isolated and dispersedly stored in traffic management departments in various regions due to format and purpose reasons, without digital management. In the current management process of accident data, traffic accident data are stored dispersedly in different databases. Complex cross-database queries are necessary for querying, analyzing, and mining these complex and diverse traffic accident data. The dispersed storage approach impedes the direct and efficient utilization of data, resulting in a notable reduction in the utilization rate of traffic accident data. And, numerous scholars have conducted extensive research on the challenges associated with data management.

The management of road traffic accident data currently relies on traditional relational databases such as MySQL [2] and Oracle [3] for storage and retrieval. While this approach offers flexibility and standardization, it is not suitable for large-scale heterogeneous data with multiple semantics and has limitations in data scalability and dynamic storage. Wen Jing et al. implemented distributed management of massive wind power data storage and retrieval using the non-relational Hbase [4] database within the Hadoop framework [5]. The distributed database addresses the limitations of traditional relational databases in terms of scalability and transmission speed. Tianru Zhang et al. proposed a hierarchical storage management framework, which, based on fully considering the internal structural features of the data, utilizes reinforcement learning to autonomously and dynamically place data in different layers of the storage hierarchy, significantly reducing the processing time of complex data and achieving autonomous placement and classification of different data [6]. Dong Yuan et al. proposed a matrix-based k-means clustering strategy, accomplishing dynamic clustering of different types of data into appropriate databases, and improving the efficiency of data management work [7]. Although this method solves the problem of effective classification of similar data, it does not fully consider the strong correlation between internal data structures. Sayed Hoseini et al. conducted research on the management methods of multi-semantic data. Based on analyzing the semantic correlation features of the data, they proposed a semantic-aware data lake metadata model [8]. Yanhui Wang et al. conducted a data structure analysis of urban rail transit operation accident text data, proposed the accident semantic framework and accident risk control chain concept model, and achieved the extraction and scientific management of text accident data [9]. However, although the above data management methods consider the scalability issues of data structure and databases, the associated features and logical hierarchical levels between data have not yet been reflected in the database. In addition, in the process of data classification management and automatic placement storage, the interpretability of application data in the later stage needs to be considered.

The present study proposes a road traffic accident data management method based on knowledge graph technology as a means of effectively managing and analyzing heterogeneous accident data and further advancing the digitalization of traffic safety management. Knowledge graph is a structured representation of various entities in the objective world and their relationships [10, 11], which has significant advantages in storing and managing multi-semantic complex data. It not only possesses strong interpretability but also has the advantage of strong semantic understanding, finding wide application in fields such as railway [12] and power grid [13]. This paper designs the entities, attributes, and relationships in the road traffic accident knowledge graph separately using textual accident reports as the data foundation and completes the construction of the knowledge graph. And, application analysis is conducted utilizing the road traffic accident knowledge graph.

## 2 ANALYSIS OF CHARACTERISTICS IN ROAD TRAFFIC ACCIDENT DATA

The clarification of data characteristics and the associative structure of key elements during the process of data management facilitates the establishment of databases that better align with the structural features of the data.

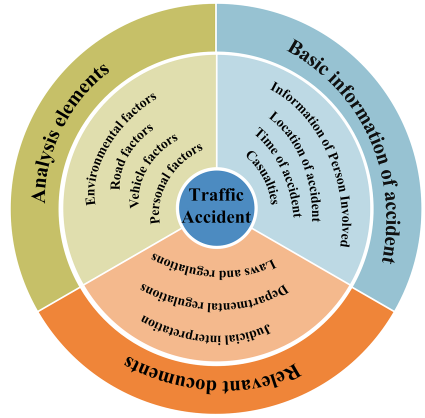


**Figure 1: Framework for the analysis of road traffic accidents.**

### 2.1 Characteristics of traffic accident data

The characteristics of road traffic accident data mainly manifest in three aspects: the form, source, and structure of the data. Firstly, road traffic accident data primarily exists in under-utilised, unstructured forms, such as accident scene photos, videos, and accident text reports. Secondly, the main sources of traffic accident data are accident records and statistical reports from transportation authorities. In addition, traffic accident data often exhibits a multi-layered structure, with each accident having multiple elements and attributes that interact with each other, resulting in complex spatiotemporal correlations in the structure of accident data.

### 2.2 Cause of traffic accident

The occurrence of road traffic accidents requires specific internal and external conditions, which encompass various aspects such as traffic conditions, environmental conditions, and road conditions. Specifically, it involves the disruption of the coordinated and stable "Human-Vehicle-Road-Environment" system under normal driving conditions, thereby fostering the conditions that lead to accidents. Therefore, in the process of analyzing the causes of traffic accidents, it is necessary to comprehensively consider factors related to humans, vehicles, environment, and roads based on the analysis of traffic-related policy documents. The detailed accident analysis framework is illustrated in Figure 1.

## 3 CONSTRUCTION OF ROAD TRAFFIC ACCIDENT KNOWLEDGE GRAPH BASED ON TEXT DATA

### 3.1 Data sources

The data for this study are sourced from road traffic accident investigation reports within China, which are in the form of unstructured textual data. The content of the data includes entries such as accident time, location, type, cause, weather, visibility and other items. In the process of data collection and processing, the research team collected a total of 651 road accident investigation reports in China

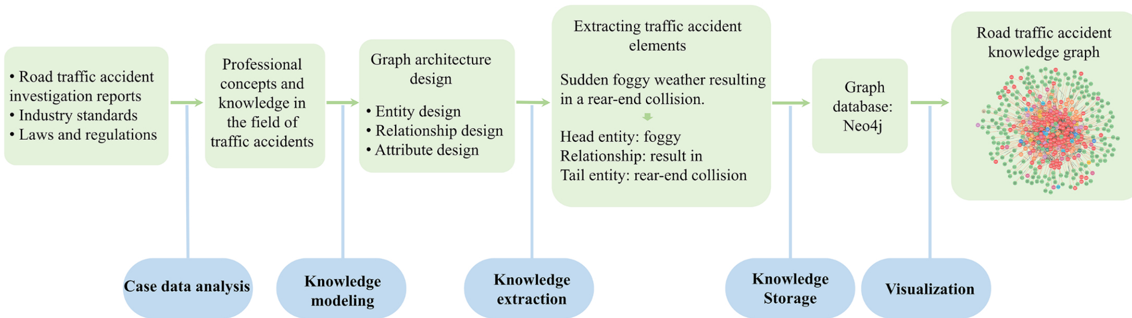


**Figure 2: Knowledge graph construction process.**

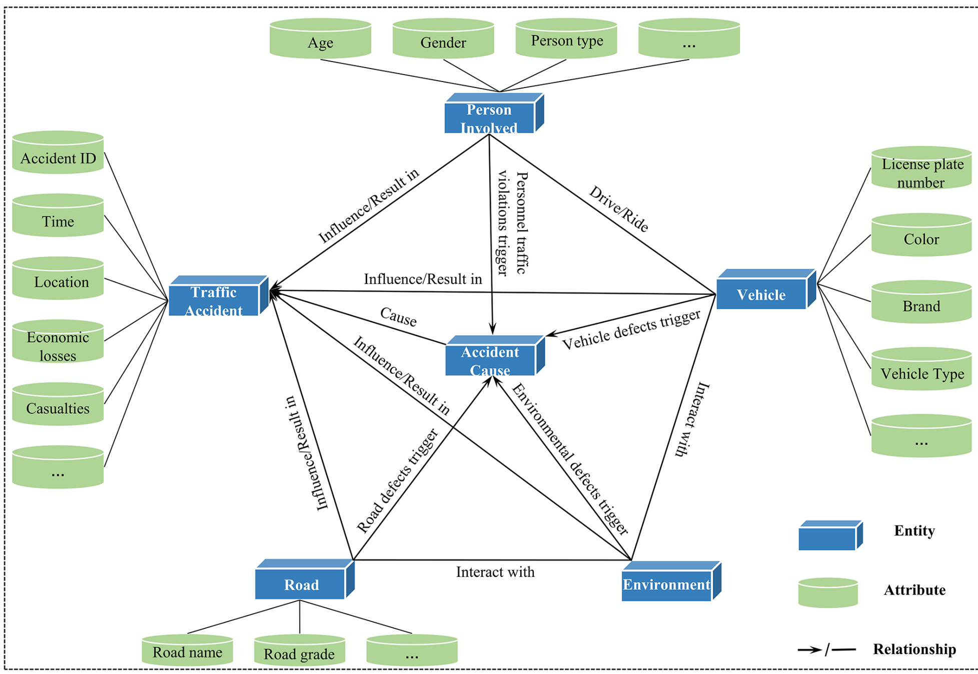


**Figure 3: The main components of the road traffic accident knowledge graph.**

over the past five years, and selected and analyzed 537 typical accident reports as the basic data for constructing the road accident knowledge graph.

## 3.2 Knowledge graph construction process

The logical architecture of knowledge graphs is mainly divided into two levels, the pattern layer and the data layer. The pattern layer serves as the core of knowledge graphs, typically using domain ontology knowledge to gather and categorize data. Additionally, it acts as the initial pattern framework of the graph, offering templates for the structured data layer. The data layer is where structured, semi-structured and unstructured data is stored.

This paper constructs a traffic accident knowledge graph based on textual data using a top-down approach. Firstly, by analyzing typical accident reports, relevant industry standards, and legal regulations, the key entities and relationships leading to accidents are summarized and designed to construct the pattern architecture of the graph. Subsequently, entities and relationships are manually extracted from the accident reports based on the defined schema layer, which comprises structured samples. Finally, the extracted data is transformed into a structured form of triplets, and the knowledge fusion of accident information during data import into the Neo4j graph database, thus completing the construction of the knowledge graph. The construction process is illustrated in Figure 2.

## 3.3 Main components design of road traffic accident knowledge graph

*3.3.1 Knowledge graph architecture design.* We categorize road traffic accident entities into six major categories based on the structural

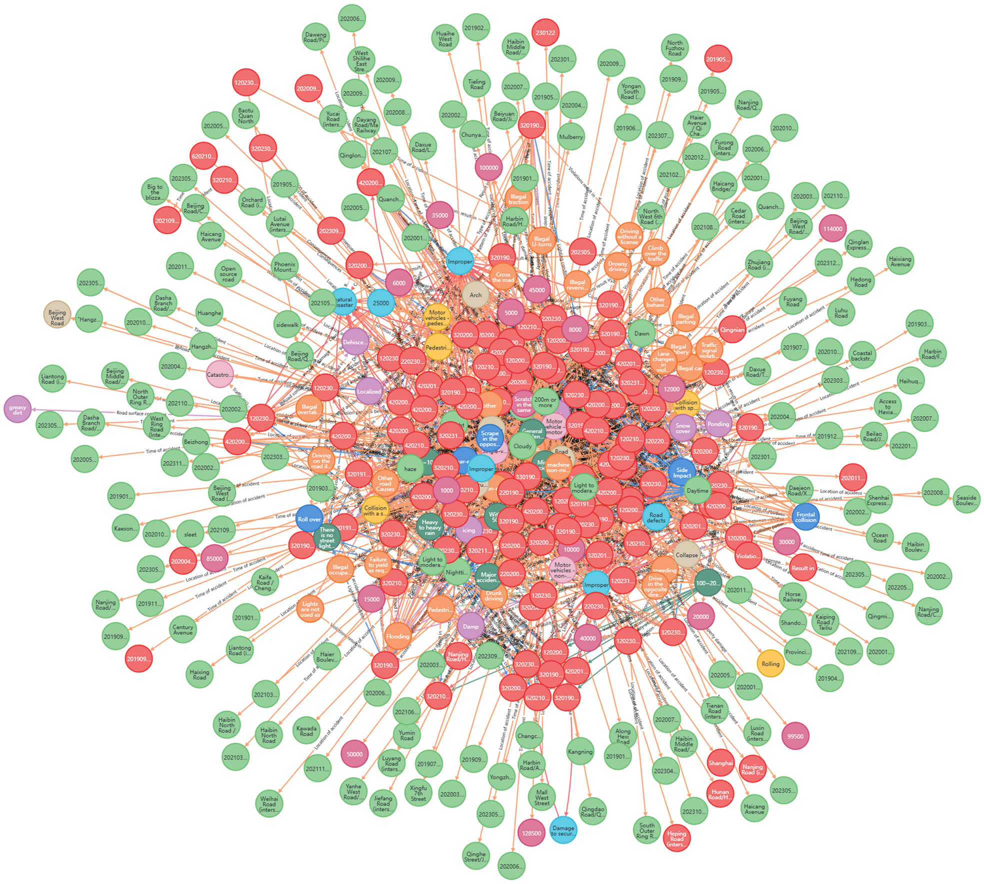

**Figure 4: Road traffic accident knowledge graph (partial).**

characteristics of accident data and the application requirements of data management. These classes include "traffic accident", "person involved", "vehicle", "road", "environment" and " accident cause". To enhance the associativity of the accident knowledge network graph in retrieval and to avoid complex entity relation logic expressions, we have summarized and designed the relationships between the six primary traffic entities based on the analysis of typical traffic accident cases. The overall graph architecture is shown in Figure 3., presenting the primary entity classification describing road traffic accidents and their accompanying attributes, as well as the types of relationships between entities.

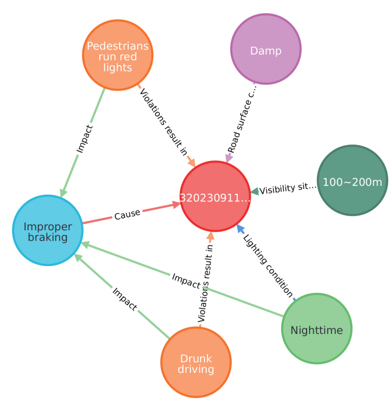


**Figure 5: Portrait of the causal process of traffic accidents.**

*3.3.2 Entity and attribute design.* A further secondary classification of entities was conducted based on the classification of the six primary entity types mentioned above, and the corresponding attributes were designed. Ultimately, the different characteristics of elements in road traffic accidents were considered, and all entities involved in road traffic accidents and their attributes were summarized as shown in Table 1.

## 3.4 Storage and visualization of Knowledge graph

In accordance with the entity and relationship descriptions designed from the knowledge graph pertaining to traffic accidents, structured accident knowledge is stored within the Neo4j graph

**Table 1: Definitions of road traffic accident entities and attributes.**

| Primary entity | Secondary entity | Tertiary entity | Attribute |
|---|---|---|---|
| Traffic Accident | — | — | Accident ID, Time, Location, Economic losses, Casualties. |
| | Accident grade | Minor accident, General accident, Major accident, Severe accident. | — |
| | Accident pattern | Front-end collision, Side-end collision, Rear-end collision, Opposite direction scrape, Same direction scrape, Pedestrian collision, Animal collision, Others. | — |
| | Accident cause | Violation, Improper braking, Improper steering, Improper throttle control, Natural disaster, Road defect, Environmental defect, Other. | — |
| Person Involved | — | — | Age, Gender, Person type. |
| | Violation | Speeding, Drunk driving, Wrong-way driving, Violation of traffic signal, Fatigue Driving, Prohibition Violation, Illegal Parking, Lane Violation, Towing Violation, Overloading, Unlicensed driving, Other. | — |
| | Transportation mode | Driving motor vehicles, Driving non-motorized vehicles, Walking, riding, Other. | — |
| Vehicle | — | — | License plate number, Color, Brand, Vehicle type, Pre-incident speed. |
| | Moving status | Straight driving, Reversing, Turning, Starting, Parking, Left turn, Right turn, Lane change, Other. | — |
| | Safety status | Normal, Brake failure, Burst tire, Steering failure, Mechanical failure, Other. | — |
| Road | — | — | Road name, Road grade, Number of lanes. |
| | Road condition | Dry, Wet, Flooding, Ice and snow, Muddy, Other. | — |
| Environment | Visibility | Within 50m, 50~100m, 100~200m, More than 200m. | — |
| | Lighting condition | Daytime, Nighttime with street lights, Nighttime without street lights. | — |
| | Weather condition | Sunny, Cloudy, Rainy, Snowy, Foggy, Windy, Dusty, Other. | — |
| Accident Cause | Violation by the person involved, Vehicle defects, Road defects, Environmental defects | — | — |

database. Utilizing the Cypher language in Neo4j software, rapid batch creation of nodes and relationships is implemented from structured CSV files, and visualized, in which a total of 7352 entities and 10,572 relationships were created, and the final. The road traffic accident knowledge graph is shown in Figure 4.

# 4 APPLICATION ANALYSIS OF ROAD TRAFFIC ACCIDENT KNOWLEDGE GRAPH

The road traffic accident knowledge graph encompasses all essential information about traffic accidents, effectively achieving the goal of representing all accident data through a single graph. On this basis, various applications can be further developed. This paper proposes three application scenarios from the perspective of data management.

## 4.1 Traffic accident portrait

Through the utilization of the constructed road traffic accident knowledge graph, specific accidents can be queried using Neo4j's Cypher language, and displayed as an accident portrait. As shown in Figure 5., portrait of the causal process of traffic accident is illustrated. Additionally, it is possible to structure the display of key accident elements around the involved person or vehicle. Figure 6. shows the knowledge graph centered on the vehicle involved and person involved in accident. In addition, the use of Neo4j for storage enhances the ease of data management and the timeliness of data additions, and the Cypher language allows for the dynamic updating of traffic accident data based on existing data, ensuring the accident data management system remains consistently up-to-date.

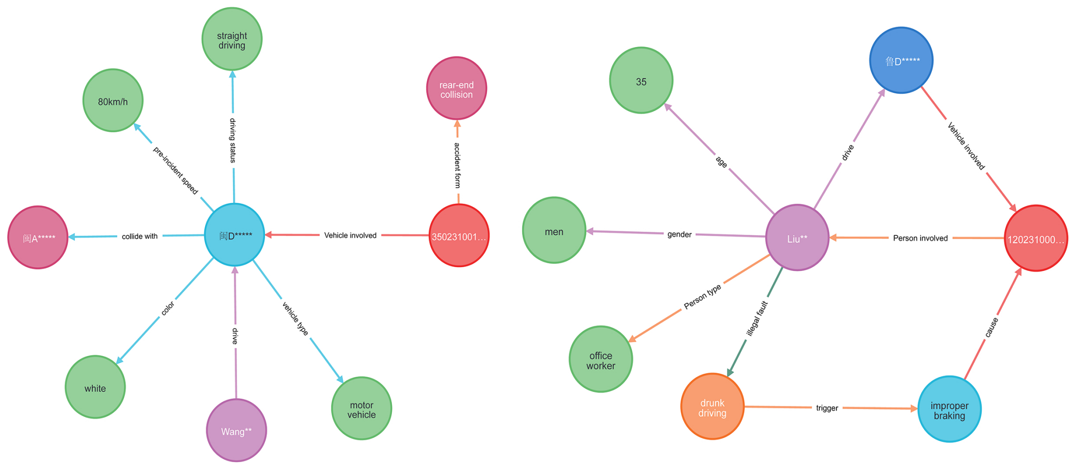


**Figure 6: Knowledge graph centred on the vehicle and person involved in accident.**

## 4.2 Intelligent traffic safety management and decision support

The road traffic accident knowledge graph provides an abundant data foundation and analytical tools for intelligent traffic safety management. Leveraging knowledge graph technology, fragmented and heterogeneous accident data are integrated into an interconnected and organized knowledge graph network. This capability empowers traffic managers to conduct knowledge retrieval, cross-regional accident analysis, and road risk assessment. For example, traffic managers can use the knowledge graph to integrate accident data from different regions, gaining a comprehensive understanding of traffic safety conditions across different areas. This facilitating the formulation of finely-tailored traffic management strategies.

## 4.3 Traffic accident data mining

A knowledge graph can essentially be viewed as a graph structure, where the road traffic accident knowledge graph integrates node and edge data related to accident information, making it a suitable subject for graph data mining. Leveraging the ability to fully exploit the value of data in a knowledge graph, this graph can be applied to traffic accident identification, accident information completion, and traffic accident prediction.

# 5 CONCLUSION

We propose a novel approach to road traffic accident data management based on knowledge graph technology. By integrating the structural features of accident data, a road traffic accident knowledge graph is constructed, which integrates fragmented accident data, and facilitates semantic management and efficient querying of the data. Furthermore, the study analyzes the application scenarios of knowledge graphs from the perspective of data management. The road traffic accident management method proposed in this paper has significant advantages in data management applications, contributing to the scientific and refined level of road traffic safety management.

# ACKNOWLEDGMENTS

This research was funded by grant sponsor: Major Program of National Social Science Foundation of China, grant number: 23&ZD138.

# REFERENCES

[1] Fountas, G.; Sarwar, M.T.; Anastasopoulos, P.C.; Blatt, A.; Majka, K. Analysis of stationary and dynamic factors affecting highway accident occurrence: A dynamic correlated grouped random parameters binary logit approach. *Accident Analysis & Prevention* **2018**, *113*, 330-340. https://doi.org/10.1016/j.aap.2017.05.018.

[2] Tian, B.; Huang, J.; Mozafari, B.; Schoenebeck, G. Contention-aware lock scheduling for transactional databases. *Proceedings of the VLDB Endowment* **2018**, *11*, 648-662.

[3] Stephens, S.M.; Chen, J.Y.; Davidson, M.G.; Thomas, S.; Trute, B.M. Oracle database 10 g: a platform for BLAST search and regular expression pattern matching in life sciences. *Nucleic Acids Research* **2005**, *33*, D675-D679.

[4] Chang, F.; Dean, J.; Ghemawat, S.; Hsieh, W.C.; Wallach, D.A.; Burrows, M.; Chandra, T.; Fikes, A.; Gruber, R.E. Bigtable: A distributed storage system for structured data. *ACM Transactions on Computer Systems (TOCS)* **2008**, *26*, 1-26.

[5] Wen, J. Research on storage and retrieval strategy of wind turbine operation data based on Hbase. **2020**, Master's thesis, North University of China. (in Chinese).

[6] Zhang, T.; Gupta, A.; Rodríguez, M.A.F.; Spjuth, O.; Hellander, A.; Toor, S. Data management of scientific applications in a reinforcement learning-based hierarchical storage system. *Expert Systems with Applications* **2024**, *237*, 121443. https://doi.org/10.1016/j.eswa.2023.121443.

[7] Yuan, D.; Yang, Y.; Liu, X.; Chen, J. A data placement strategy in scientific cloud workflows. *Future Generation Computer Systems* **2010**, *26*, 1200-1214. https://doi.org/10.1016/j.future.2010.02.004.

[8] Hoseini, S.; Theissen-Lipp, J.; Quix, C. A survey on semantic data management as intersection of ontology-based data access, semantic modeling and data lakes. *Journal of Web Semantics* **2024**, *81*, 100819. https://doi.org/10.1016/j.websem.2024.100819.

[9] Wang, Y.; Sheng, K.; Niu, P.; Chu, C.; Li, M.; Jia, L. A comprehensive analysis method of urban rail transit operation accidents and safety management strategies based on text big data. *Safety Science* **2024**, *172*, 106400. https://doi.org/10.1016/j.ssci.2023.106400.

[10] Abu-Salih, B. Domain-specific knowledge graphs: A survey. *Journal of Network and Computer Applications* **2021**, *185*, 103076. https://doi.org/10.1016/j.jnca.2021.103076.

[11] Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Philip, S.Y. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems* **2021**, *33*, 494-514.

[12] Liu, C.; Yang, S. Using text mining to establish knowledge graph from accident/incident reports in risk assessment. *Expert Systems with Applications* **2022**, *207*, 117991. https://doi.org/10.1016/j.eswa.2022.117991.

[13] Wang, Y.; Cheng, Y.; Qi, Q.; Tao, F. IDS-KG: An industrial dataspace-based knowledge graph construction approach for smart maintenance. *Journal of Industrial Information Integration* **2024**, *38*, 100566. https://doi.org/10.1016/j.jii.2024.100566.