Postprint: Research on the Construction of a Public Health Emergency Data Platform Based on Knowledge Entities

Authors: Feng Xin, Li Xue, Yan Yue, Li Jiapei, 刘梦瑶, Wu Ye

Date: 2023-10-08T00:00:00+00:00

Abstract

[Purpose / Significance] Public health emergencies pose severe threats to the lives and health of the global public. Achieving intelligent storage, querying, knowledge organization, and output of information resources for such emergencies holds significant reference value for research data integration, sharing, and domain knowledge management in the emergency response field. [Method / Process] CNKI and WOS were selected as retrieval platforms to conduct searches on public health emergencies and crawl valid data. Bibliometric software was utilized to process and fuse the extracted knowledge data, and cluster visualization analysis was performed on high-frequency terms, authors, and institutions. Knowledge entities, attributes, and relationships were stored in a Neo4j graph database to establish complex connections between knowledge entities, thereby constructing a public health emergency data platform model. [Result / Conclusion] The emergency event intelligent platform is a public-oriented knowledge data sharing resource platform based on information resources. It processes information from various domains and is constructed through the screening and fusion of internal entities. The platform enables bidirectional interaction to facilitate information resource flow, thereby effectively leveraging information resources for emergency management.

Full Text

Preamble

Study on Construction of Emergent Public Health Event Data Platform Based on Knowledge Entity

Feng Xin[1], Li Xue[2], Yan Yue[2], Li Jiapei[1], Liu Mengyao[1], Wu Ye[3],[4]

[1]Management College, Hebei GEO University, Shijiazhuang 050031
[2]Huaxin College of Hebei Geo University, Shijiazhuang 050031
[3]Computational Communication Research Center, Beijing Normal University, Zhuhai 519087
[4]School of Journalism and Communication, Beijing Normal University, Beijing 100875

Abstract

[**Purpose/Significance**] Public health emergencies seriously endanger the lives and health of the global public. Achieving intelligent storage, querying, knowledge organization, and output of information resources for such emergencies holds important reference significance for the integration and sharing of scientific research data and domain knowledge management in the emergency response field.

[**Method/Process**] This study selected CNKI and Web of Science (WOS) as retrieval platforms, searching for public health emergencies and crawling valid data. Bibliometric software was used to process and fuse the extracted knowledge data, with cluster visualization analysis performed on high-frequency terms, authors, and institutions. Knowledge entities, attributes, and relationships were stored in the Neo4j graph database to establish complex connections between knowledge entities, thereby constructing a public health emergency data platform model.

[**Result/Conclusion**] The emergency event intelligence platform is a knowledge data sharing resource platform oriented toward the public, built upon information resources that processes information from different domains through internal entity screening and fusion. The platform enables bidirectional information flow, effectively leveraging information resources for emergency management work.

Keywords: public health emergencies, emergency service, knowledge entity, COVID-19

1 Introduction

Public health emergencies refer to sudden occurrences that cause or may cause serious harm to public health, including major infectious disease outbreaks, unexplained group illnesses, major food and occupational poisoning, and other events that seriously affect public health—such as the current focus event, COVID-19. On the evening of January 30, 2020, WHO Director-General Tedros Adhanom Ghebreyesus declared at a press conference in Geneva that the novel coronavirus pneumonia outbreak constituted a public health emergency of international concern. The COVID-19 outbreak has attracted worldwide attention, making it necessary to carefully investigate the research status of public health emergencies both domestically and internationally, extract complex and abstract multidimensional data, associate and visualize professional knowledge entities in the form of entities, relationships, and attributes, and construct an emergency intelligence data platform based on comprehensive knowledge entities. By building such a platform, valuable information can be extracted from low-density data for timely sharing and intelligent dissemination, making relevant information acquisition convenient, rapid, and efficient. Ultimately, intelligent search and recommendation functions help diverse users quickly and accurately collect information to formulate emergency prevention and control measures, which holds significant practical importance for reducing losses caused by public health emergencies.

The rampant spread and further deterioration of epidemics cause severe economic shocks and social impacts, threatening lives worldwide. At a press conference, the WHO stated that under the personal command and deployment of President Xi Jinping, China has consistently adhered to the principles of openness and transparency, releasing information promptly, rapidly identifying the virus and sharing its genetic sequence, and taking decisive and forceful measures to control the spread of the epidemic. These actions demonstrate not only high responsibility for the lives and health of Chinese citizens but also strong support for global disease prevention and control. Furthermore, the extensive prevention and control measures implemented by China far exceed relevant international requirements for emergency response, setting a new benchmark for epidemic prevention in other countries. While China's epidemic has been effectively controlled, international spread poses another challenge to nations worldwide. According to WHO's daily epidemic report, as of 10:00 AM Central European Time on June 12, 2020 (4:00 PM Beijing Time), confirmed COVID-19 cases outside China increased by 136,565 from the previous day, reaching 7,325,851; deaths outside China increased by 4,925, reaching 413,649. Globally, confirmed cases increased by 136,572 to 7,410,510, with deaths increasing by 4,925 to 418,294.

Faced with the severe situation of public health emergencies, prevention-oriented and combined prevention-treatment measures must be adopted, with strengthened joint prevention and control and comprehensive response plans. In his speech on reforming and improving the major epidemic prevention and control system, General Secretary Xi Jinping mentioned that the state encourages the use of digital technologies such as big data, artificial intelligence, and cloud computing to better support epidemic monitoring and analysis, virus traceability, prevention and treatment, and resource allocation.

2 Related Research

Academic and research institutions can support and respond to public health countermeasures through their functions and characteristics, enabling information management and performance evaluation. A. L. Dunlop et al. reviewed literature and internet reports from September 11, 2001, to February 1, 2009, proposing that academic institutions, as part of social disaster response, can provide services and resources that reduce adverse consequences and lower the probability of occurrence, thereby enhancing community resilience. N. A. Vielot et al. explored whether merging public health preparedness and emergency management roles could improve emergency planning and response efficiency and effectiveness through interviews in six counties in North Carolina, discussing resulting opportunities and challenges. J. Hu et al. established permanent rural emergency management institutions with public health management functions, improved the entire emergency management mechanism process, increased investment in rural public health, and built a system with adequate emergency resource reserves. S. Zhong et al. proposed constructing an emergency rescue architecture based on edge computing models to address latency issues in traditional cloud computing for emergency services, providing a resource scheduling model that offers good scheduling time and lower costs.

Domestic scholars have also conducted in-depth analysis of resource construction for public emergencies. Regarding emergency intelligence resource guarantee for sudden incidents, Ke Danqian reviewed and analyzed the current status of emergency decision-making intelligence driven by domestic and international emergencies, laying groundwork for subsequent research. Zhang Yongling et al. introduced scenario analysis into evaluation models to comprehensively assess emergency resource guarantee capabilities, making evaluation results closer to actual needs. In achieving emergency information resource sharing and improving emergency management, Guo Lusheng et al. built catalogs to provide unified standard systems for emergency information resources, realizing the connection between information resources and emergency services and facilitating cross-departmental sharing and mutual service of emergency information resources. Qu Tengjiao et al. conducted inductive analysis of major public health emergency cases and related health emergency management literature in China from 2003-2018, expounding on the current status of health emergency management in China.

Regarding emergency decision-making and information fusion, Li Pin et al. achieved the fusion of intelligence research within think tank activities, constructing an intelligence process that supports think tank activities and assists in generating think tank products. Cao Yujie et al. compared emergency decision-making information needs with available emergency information in big data environments, constructing an overall framework for emergency information fusion services oriented toward emergency decision-making, with emergency data models as the bridge. By analyzing specific problems in emergency information fusion at the data, semantic, and service layers layer by layer, they proposed corresponding information fusion implementation paths. These research conclusions help guide the construction of big data fusion systems for emergencies and provide practical references for decision-oriented information analysis.

In summary, existing research results fully demonstrate that in the information age, emergency information resources play an indispensable role in responding to public health emergencies. However, no scholars have yet constructed knowledge entities for emergency resources. Therefore, this study will extract knowledge entities from CNKI and Web of Science, establish knowledge entity graphs based on relationships between knowledge entities, and then conduct logical reasoning at the database level to systematically present knowledge systems related to public health emergencies to users. By constructing an emergency resource platform based on knowledge entities—from popular science knowledge entity extraction to entity relationship construction, knowledge entity visualization, and finally achieving information sharing, intelligent querying, and intelligent recommendation functions—this research aims to improve the utilization rate of popular science information.

3 Research Objects and Data Sources

This study conducted topic searches on domestic CNKI and international Web of Science databases. First, in the CNKI academic journal database, advanced search was used with "Keywords" as the search field, "public health emergencies" as the search term, and SCI/EI core journals as the limitation. The collection period was 2010-2020, and 416 valid articles were obtained after excluding irrelevant literature. Second, in the Web of Science full-text database, the Web of Science Core Collection was selected with "public health emergencies" as the topic, Article and Review as document types, SCI-EXPANDED, SSCI, and A&HCI as indexes, English as the language, and 2010-2020 as the time period, yielding 949 results after excluding irrelevant literature. After obtaining relevant data, visualization methods were used to analyze literature quantity, citation quantity, and author and institution comparisons, providing quantitative analysis of public health emergencies through scientific measurement indicators.

4 Knowledge Entity Extraction

4.1 Knowledge Entity Extraction Methods

Keywords in research papers represent the core research content of a scientific achievement and serve as an important basis for knowledge retrieval. This study used the bibliometric visualization tool Citespace to conduct statistical analysis on Chinese and English relevant topic datasets, extracting core keywords as knowledge nodes, core experts as nodes, and research institutions as nodes. Using the keyword "public health emergencies" as the main object, the main framework of knowledge entities was constructed. Document content was analyzed to automatically extract keywords as objects for establishing strongly associated knowledge entities. User association analysis was performed to mine potential knowledge associations, and universally associated knowledge entities were applied to popular science platform construction for personalized popular science knowledge recommendations to users. Starting from constructing an emergency resource platform based on knowledge entities, this research proceeds from popular science knowledge entity extraction to entity relationship construction and knowledge entity visualization, ultimately achieving information sharing, intelligent querying, intelligent recommendation, and other functions to improve the utilization of popular science information.

This study used keyword co-occurrence clustering to form networks. The Chinese dataset clustering is shown in [Figure 1: see original paper], with a total of 8 clusters: public health emergencies, student health services, infectious diseases, major public health emergencies, big data, public crisis early warning, delphi method, and chickenpox. The foreign language dataset clustering is shown in [Figure 2: see original paper], with a total of 7 clusters: resilience, ebola, synthetic cannabinoids, radiological, emergency preparedness, quarantine, and practice guidelines.

4.2 Institutional Publication Analysis

[Figure 1: see original paper] shows the clustering of Chinese public health emergency datasets, and [Figure 2: see original paper] shows the clustering of foreign language datasets.

Table 1 presents the top ten domestic institutions by publication volume on public health emergencies from 2010-2020. The top three institutions are Wuhan University School of Information Management (10 articles), Jiangsu Provincial Center for Disease Control and Prevention (7 articles), and Harbin Medical University School of Health Management Social Medicine Teaching and Research Office (5 articles), all exceeding 5 publications.

By counting publications from each institution, a list of major research institutions was obtained. In Citespace, the Institution node was selected with Top50 per time slice. Analyzing domestic and foreign institutional publication volumes separately yielded statistical tables and institutional cooperation maps.

Table 1 Top Ten Domestic Institutions by Publication Volume on Public Health Emergencies

Institution	Publications
Wuhan University School of Information Management	10
Jiangsu Provincial Center for Disease Control and Prevention	7
Harbin Medical University School of Health Management Social Medicine Teaching and Research Office	5
Chinese Center for Disease Control and Prevention Environmental Health Related Product Safety Institute	4

Institution	Publications
Hangzhou Center for Disease Control and Prevention	4
Hubei Provincial Center for Disease Control and Prevention	4
Tsinghua University Emergency Management Research Base	4
Harbin Medical University School of Health Management	4
Wuhan University Information Resource Research Center	4
Beijing Tongzhou District Center for Disease Control and Prevention	3

[Figure 3: see original paper] shows the domestic institutional cooperation network for public health emergencies, generated through Citespace visualization analysis of domestic institutional cooperation. Based on node radius, Wuhan University School of Information Management has the highest publication volume, followed by Jiangsu Provincial Center for Disease Control and Prevention, and Harbin Medical University School of Health Management Social Medicine Teaching and Research Office. However, cooperation between these institutions is minimal. In comparison, Wuhan University Information Resource Research Center cooperates more with Wuhan University School of Information Management (ranked first). Geographically, publishing institutions are distributed across the country, indicating relatively uniform geographical distribution of research institutions in China, enabling regional research according to local public health characteristics. However, China has not yet formed a pattern of in-depth and intensive cross-regional, cross-institutional research.

Table 2 shows the top ten foreign institutions by publication volume on public health emergencies from 2010-2020. The Centers for Disease Control and Prevention (CDC) has the highest total publication volume at 88 articles, followed by WHO (39 articles), Harvard University (34 articles), Ministry of Health of Saudi Arabia (24 articles), and CDC (23 articles).

Table 2 Top Ten Foreign Institutions by Publication Volume on Public Health Emergencies

Institution	Publications
Centers for Disease Control and Prevention (CDC)	88
Harvard University	34
Ministry of Health of Saudi Arabia	24
University of Pittsburgh	21
Emory University	20

Institution	Publications
Johns Hopkins Bloomberg School of Public Health	19
University of Toronto	18
University of Pennsylvania	17

[Figure 4: see original paper] shows the foreign institutional cooperation network, visualized through Citespace analysis of public health emergency literature in the Web of Science Core Collection. Line thickness indicates institutional connections, while node size indicates publication frequency. The dense connections between nodes demonstrate strong cooperation among major foreign institutions.

4.3 Author Publication Analysis

In Citespace, the Author node was selected with Top50 per time slice, Thresholding set to (2,1,20), and minimum publications set to 3. Analyzing author publications from 2010-2020 with 1-year intervals yielded [Figure 5: see original paper] for domestic authors and [Figure 6: see original paper] for foreign authors.

Table 3 shows the top ten domestic authors by publication volume. Wu Qunhong leads with 10 publications over the decade, followed by An Lu (7), Liu Jing (5), and Hao Yanhua (5), indicating these researchers have made significant contributions to public health emergency research.

Table 3 Top Ten Domestic Authors by Publication Volume on Public Health Emergencies

Author	Publications
Wu Qunhong	10
An Lu	7
Liu Jing	5
Hao Yanhua	5
Li Ying	4
Wang Wei	4
Zhang Yongling	4
Chen Lu	3
Li Pin	3
Xu Ying	3

[Figure 5: see original paper] shows the domestic author cooperation network, where color indicates publication time (darker for earlier, lighter for later). The map displays expert names with 2+ publications; larger node radius and font size indicate more publications, while denser connecting lines indicate closer cooperation. Combined with Table 3, An Lu began research in this field in 2016 with substantial output, making notable contributions.

Table 4 shows the top ten foreign authors from WOS (2010-2020). Frederick M. leads with 7 publications, focusing on preventive medicine and hygiene. Elena Savoia (5 publications) specializes in preventive medicine, public health, and quantitative methods; Daniel J. Barnett (5 publications) focuses on emergency medicine; and Lainie Rutkow (5 publications) uses legal, qualitative, and quantitative methods to study public health law and policy.

Table 4 Top Ten Foreign Authors by Publication Volume on Public Health Emergencies

Author	Publications
Frederick M.	7
Elena Savoia	5
Daniel J. Barnett	5
Lainie Rutkow	5
Katherine Seib	4
Saad B. Omer	4
Niranjan Kissoon	4
James G. Hodge	4
Zhanat Carr	4
Jon S. Vernick	4

[Figure 6: see original paper] displays the distribution of WOS authors. Larger nodes indicate more publications, and color changes in node rings reflect different active periods. As shown, these authors have relatively loose connections, with the top three authors having almost no cooperative relationships, each conducting research with independent thinking.

4.4 Construction of Public Health Emergency Knowledge Entities

4.4.1 Concepts Related to Knowledge Entity Graphs Knowledge entity graphs, also known as knowledge domain visualization or knowledge domain mapping, are a series of different graphics displaying knowledge development processes and structural relationships. They use visualization technology to describe knowledge resources and their carriers, mining, analyzing, constructing, drawing, and displaying knowledge and its interconnections. Essentially, they are a form of large-scale semantic network rich in entities, concepts, and relationships—one of the methods for large-scale data knowledge representation.

Knowledge entities can store and summarize iterative knowledge, rules, metadata, and other multi-type data, or fuse structured and unstructured data, enabling analysis of complex, multi-dimensional data with intricate relationships. Knowledge entity graphs have two main storage methods: RDF-based storage and graph database-based storage. RDF stores data in triples but lacks attribute information. Graph databases generally use property graphs as the basic representation form, where entities and relationships can contain attributes, making it easier to express real business scenarios. The main difference is that RDF emphasizes easy data publishing and sharing, while graph databases focus on efficient graph querying and searching.

Neo4j is currently the most widely used graph database, primarily describing entity properties and relationships between entities, stored in graph form. It supports 32 billion relationship nodes and 64 billion properties, has an active community, and offers high query efficiency, though its disadvantage is high cost of distributed storage and slow update speed.

This study uses major keywords, researchers, and institutional cluster names extracted from CNKI and WOS datasets as entity node names. After comprehensively integrating these knowledge nodes, public health emergency knowledge entities are divided into five categories: public health emergency level, prevention and control measures, infectious diseases, experts, and institutions. The relationships between entities, and between entities and attributes, are then defined, using CREATE statements to create nodes, properties, and relationships. Finally, all named entities and relationships are imported into the Neo4j graph database, yielding two important knowledge entity graphs: [Figure 7: see original paper] Public Health Emergency Knowledge Entity Graph and [Figure 8: see original paper] Knowledge Fusion Entity Graph of COVID-19 in Chinese and Western Medicine Research.

4.4.2 Construction of Public Health Emergency Knowledge Entities

[Figure 7: see original paper] shows the public health emergency knowledge entity graph, composed of nodes and edges, where circles represent nodes (entities in the knowledge entity graph), edges between entities represent relationships between properties and nodes, and arrows indicate direction from start to end nodes. One or more labels are added to nodes to represent entity classification, and a key-value pair collection represents properties other than entity relationship attributes. Relationships can also carry additional properties. Entity property-values are generally represented in the graph. Since properties and nodes are stored separately, properties can be quickly queried using MARCH commands and WHERE clauses. Some node properties are shown in .

The public health emergency knowledge entity graph is created based on keyword cluster analysis from [Figure 1: see original paper] and [Figure 2: see original paper], author cooperation analysis from [Figure 3: see original paper] and [Figure 5: see original paper], and institutional cooperation analysis from [Figure 4: see original paper] and [Figure 6: see original paper]. Its two primary knowledge entities are “Public Health Emergency” and “Public Health Emergencies,” with the largest circles displayed in blue to prepare for subsequent bilingual knowledge entity expansion. Based on the above keyword cluster visualization analysis, this paper divides public health emergencies into five components: public health emergency level, prevention and control measures, infectious diseases, experts, and institutions. The relationship attribute between these five entities and the "Public Health Emergency" entity node is set as "INCLUDE" (subclass relationship).

According to Article 1.3, Item 2 of the *National Public Health Emergency Response Plan*, public health emergencies can be divided into four levels based on nature, harm degree, and scope: particularly serious (Level I), serious (Level II), relatively serious (Level III), and general (Level IV). "Particularly serious" is represented in red, "serious" in orange, "relatively serious" in yellow, and "general" in green. The relationship attribute between these four entity nodes and the "Public Health Emergency Level" entity node is set as "SORT" (classification). Regarding prevention and control measures, China has actively implemented "emergency management," "health emergency," "epidemic prevention and control," and "prevention and treatment"—four entities with strong clustering and practical significance, represented in green nodes. The most severe public health emergencies are "infectious diseases," among which "COVID-19," "SARS," and "chickenpox" have caused tremendous harm to humans. This year's COVID-19 outbreak prompted first-level responses in multiple Chinese cities, highlighted in red circles. Based on the above author and institutional publication statistics, the top three authors and institutions by publication volume are included in the knowledge entity base, represented in purple and blue respectively.

4.4.3 Node Attributes of Public Health Emergencies **Table 5** shows partial node attributes of the public health emergency knowledge entity graph. Attributes describe properties of nodes or edges and can be represented using key-value pairs. In Neo4j, attribute creation is divided into Property-name and Property-value, where one property name can correspond to multiple property values.

The table header consists of node ID, three property names, and their corresponding property values. The knowledge node ID is automatically formed and unique when creating nodes. The first column contains each node's ID; the second, fourth, and sixth columns are property names set for each node (P1_{name}, P2_{name}, P3_{name}); the third, fifth, and seventh columns are their corresponding property values (P1_{value}, P2_{value}, P3_{value}). When nodes have multiple attributes, the system automatically selects the name property value to display on the graph. For example: Person:person{age:20,sex:"woman",name:"Lisa"} ultimately displays "Lisa" on the graph. The property values corresponding to the property names in the second column of Table 5 are displayed in [Figure 7: see original paper]. In addition to the displayed property values, the attribute table lists all nodes containing subclasses and some child node properties. For instance, for ID 4003, the property name "keyword" has value "public health emergency," "search range" has value 1278, and "from" has value "CNKI."

For foreign knowledge entity construction, this paper mainly divides the "Public Health Emergencies" event into three parts based on clustering results: the first part is "Emergency," including ebola, radiological, novel coronavirus, and quarantine; the second and third parts are "Experts" and "Institutions," respectively. Based on foreign author publication statistics, authors such as Frederick M., Elena Savoia, and Daniel J. Barnett, along with foreign institutions, are added to the knowledge entities.

4.4.4 Knowledge Entity Query The Neo4j graph database implements graph query operations through the Cypher language. In Neo4j, graph traversal query speed is a constant independent of the amount of data traversed—only when a node or relationship needs to be accessed does the graph database traverse and return it. When using the Cypher statement MATCH with regular expressions, Cypher automatically invokes built-in traversal search algorithms without requiring manual specification of traversal methods, enabling traversal of the entire graph.

[Figure 8: see original paper] was obtained by applying Cypher query statements to [Figure 7: see original paper]. The query statement is:

```
MATCH(Sara:sara)
MATCH(Sara:sara)-[r1:IN]->(n1)-[r2:IN]->(n2)-[r3:IN]->(n3)
RETURN Sara,n1,r1,n2,r2,n3,r3
```

MATCH and RETURN are Cypher keywords; Sara is a variable to save nodes; sara is a label; r1, r2, r3 represent relationships between nodes; n1, n2, n3 refer to all nodes with "IN" relationships to the previous node. When executing the first statement, a red node named "COVID-19" is queried. When reaching "n1," the database displays two nodes (in purple): "Traditional Chinese Medicine" and "Western Medicine." When reaching "n2," all subclasses of these two nodes are queried—diagnosis, treatment, and prevention (blue nodes). Only after executing the entire statement are all nodes in [Figure 8: see original paper] queried, and only the RETURN statement displays these nodes on the graph.

This knowledge entity summarizes COVID-19 research from Chinese and Western medicine perspectives, summarizing relevant conclusions. It not only promotes deeper understanding of COVID-19 but also helps maximize the role of medical systems in diagnosis and treatment, facilitating better formulation of prevention and control measures and rescue plans. Due to the widespread impact of this epidemic, researchers globally have studied various aspects of COVID-19, including virus origin, transmission routes, epidemic prediction, diagnosis, and treatment. Chen Ying et al.'s visualization research demonstrates domestic and international attention to COVID-19, and with researchers' efforts, breakthroughs are expected soon.

5 Construction of Emergency Public Health Event Data Platform Based on Knowledge Entities

As mentioned earlier, knowledge entity graphs are symbolic representations of physical world networks, describing concepts existing in real events and their relationships. The emergency resource platform is built upon knowledge entity graphs using a bottom-up operational model. By constructing an emergency resource platform based on knowledge entities, using public health emergencies as the foundation, mining diversified relationships within and between current COVID-19 events, and visualizing them for users, information acquisition becomes convenient, rapid, and efficient. Through intelligent search and recommendation, users can quickly and accurately collect information to better utilize information resources for emergency prevention and control measures.

As shown in [Figure 9: see original paper], the emergency resource platform comprises four layers: data layer, knowledge base construction, data storage, and data application.

The data layer lies at the platform's bottom, primarily including data collection, cleaning, conversion, and standardization. Processed unstructured and semi-structured data serve as the data source for building the knowledge entity base. This layer masters domestic and international research status on public health emergencies, conducts logical reasoning, analyzes patterns of public health emergencies from big data, making development traceable and forming the platform's underlying architecture.

Original data can be divided into structured, semi-structured, and unstructured data based on structural degree. Structured data (row data) is stored in databases and can be logically expressed through two-dimensional table structures, primarily managed through relational databases. Semi-structured data, a form of structured data, doesn't conform to relational database models but contains relevant markers to separate semantic elements and hierarchically structure records and fields. Unstructured data has irregular or incomplete structure without predefined data models, making it inconvenient to express through database two-dimensional logic tables. This includes all formats of office documents, texts, images, HTML, various reports, images, audio, and video information about certain emergency incidents.

Knowledge entity construction, located above the data layer, includes knowledge extraction, entity alignment, and knowledge design, serving as the platform's crucial component with a connecting role. A knowledge graph typically consists of entity sets (number of entities), relationship sets (number of relationships), and corresponding triples. The ontology construction module mainly expresses public health emergency concepts, using abstract concepts to represent real existence through an iterative process. Using ontology constraints to regulate relationships between entities, relationships, and entity attributes can effectively solve the "information silo" problem, forming ontological knowledge expression.

This study extracted entities and relationships of public health emergencies, expressing facts as triples of entity-attribute-attribute value. First, two primary knowledge entities were built: "Public Health Emergency" and "Public Health Emergencies." Then, based on keyword, author, and institution clustering data from CNKI and WOS literature, lower-level knowledge entities were divided, completing knowledge entity construction. Node attributes of public health emergencies were analyzed to characterize intrinsic properties between entities using relationships to connect two entities and deeply mine internal connections.

During this process, key technologies for knowledge extraction include entity extraction, relationship extraction, and attribute extraction. This paper extracted key knowledge entities such as keywords, authors, and institutions from unstructured and semi-structured data—specifically, literature on public health emergencies retrieved from domestic CNKI and foreign WOS. The top ten authors and institutions were compiled and standardized to complete the ontology module model construction, with Citespace visualization software providing preliminary knowledge display. After obtaining entity, relationship, and attribute information, these fragmented pieces were fused through knowledge entity disambiguation and co-reference resolution. This process includes entity linking (applicable to semi-structured and unstructured data extracted through information extraction) and knowledge merging (primarily processing structured data such as external knowledge bases and relational databases).

The data storage layer is located above the knowledge entity base and is built upon entity data from the knowledge entity base. When knowledge entities involve large amounts of complex data, traditional relational data storage methods appear inefficient and time-consuming. Therefore, using graph databases significantly improves work efficiency in querying, making graph queries and searches easier. At this level, COVID-19 knowledge entities are extracted, divided from Traditional Chinese Medicine and Western medicine perspectives, and relevant conclusions are summarized. The underlying data is transformed into information with decision-making value, and valuable information is precipitated and fused with human knowledge systems to form a complete knowledge system, stored in graph form. Neo4j graph database is selected as the storage medium for describing entity attributes and relationships. Establishing nodes, relationships, and properties in Neo4j forms a knowledge graph ready for querying.

Storing data on networks with full transaction management functionality well supports application requirements for dynamic data characteristics. Based on concepts corresponding to the ontology module, relationships between entity data are established to achieve knowledge entity visualization of concepts, relationships, and attributes. Theoretically, applying knowledge entity graph theoretical methods to think tank construction enriches the theoretical methods for building intelligent data platforms and effectively solves issues of platform vitality and sustainability. Practically, relying on authoritative literature database platforms to obtain knowledge entities makes information acquisition convenient, rapid, and efficient. Extracting entities from public health emergencies over the years allows querying 借鉴 able prevention, treatment, and response methods for similar events, providing strong support for public health emergency decision-making.

The data application layer is the platform's top layer and final form, primarily serving general public, government staff, medical workers, researchers, and educators. It is an intelligent search engine based on knowledge that can combine real-life situations to find solutions through general patterns, achieving transformation from understanding problems to knowing how to solve them. Implementing intelligent querying of emergency prevention and control and intelligent searching of emergency events helps users obtain needed knowledge more accurately and effectively, providing practical response strategies for emergency rescue decision-making. With big data support, knowledge entity graphs visualize and systematize search results, making knowledge systems form network nodes and display more precise information. The emergency event big data platform is not simple data input and output but a shared resource platform oriented toward the public that processes information from different domains through internal entity screening and fusion.

Currently, knowledge entity construction in China is still in its early development stage, with many technologies and knowledge acquisition algorithms needing improvement. However, it should be particularly noted that constructing a public health emergency data platform is a complex process involving different environmental backgrounds, subjects, cultures, and operational mechanisms. The emergency event big data platform based on knowledge entities constructed in this paper is only a model. Existing research data has certain limitations, and relevant conclusions need further verification and improvement in the face of variable and difficult-to-grasp environmental factors, with equal emphasis on theory and practice and integration of humanities and technology. In future planning, more datasets will be added to construct emergency event big data platforms based on knowledge entities, applied to different fields to effectively respond to different types of emergencies.

References

[1] XI Jinping. Comprehensively improve law-based prevention and control capabilities and improve the national public health emergency management system[J]. Contemporary Guangxi, 2020(5): 4-5.

[2] DUNLOP A L, LOGUE K M, BELTRAN G, et al. Role of academic institutions in community disaster response since September 11, 2001[J]. Disaster medicine and public health preparedness, 2011, 5(3): 218-226.

[3] VIELOT N A, HORNEY JENNIFER A. Can merging the roles of public health preparedness and emergency management increase the efficiency and effectiveness of emergency planning and response?[J]. International journal of environmental research and public health, 2014, 11(3): 2911-2921.

[4] HU JIAXIANG, CHEN CHAO, KUAI TINGTING. Improvement of emergency management mechanism of public health crisis in rural China: a review article[J]. Iranian journal of public health, 2018, 47(2): 156-165.

[5] ZHONG S, HE T, LI M, et al. An emergency resource scheduling model based on edge computing[M]// Artificial Intelligence for Communications and Networks. Cham: Springer, 2019.

[6] KE Danqian. Research status and prospects of emergency decision-making intelligence driven by domestic and foreign emergencies[J]. Modern Intelligence, 2015, 35(8): 12-16.

[7] ZHANG Yongling, CHEN Lu. Research on evaluation model of emergency resource guarantee capability based on scenario analysis[J]. China Safety Science Journal, 2014, 24(12): 151-158.

[8] GUO Lusheng, LIU Chunnian, LI Ruinan. Research on construction of emergency information resource catalog system for public service[J]. Library Science Research, 2016(7): 41-49, 23.

[9] QU Tengjiao, GU Shiyan, LI Mengzhu, et al. Development status and challenges of health emergency management in China[J]. Chinese Journal of Public Health Management, 2019(4): 433-437.

[10] LI Pin, XU LinYu, YANG Jianlin. Intelligence research for think tank services[J]. Journal of the China Society for Scientific and Technical Information, 2020, 39(2): 135-147.

[11] CAO Yujie, LI Gang, MAO Jin, et al. Research on emergency information fusion for decision-making processes in big data environments[J]. Library and Information Knowledge, 2018(5): 41-49.

[12] XIANG Linghui, GU Jinguang, WU Gang. Distributed storage of RDF data based on graph databases[J]. Computer Applications and Software, 2014, 31(11): 35-39.

[13] LIN Qisheng, WANG Lei, ZHOU Xi, et al. Optimization and implementation of literature retrieval methods based on graph databases[J]. Microelectronics & Computer, 2017, 34(10): 35-39.

[14] Animal Husbandry and Fisheries. National public health emergency response plan[J]. Chinese Journal of Food Hygiene, 2006, 18(4): 366-373.

[15] LIN Feng. Design and implementation of distributed key-value storage system[D]. Guangzhou: South China University of Technology, 2017.

[16] ZHANG Fengjun. Research and application of social network data based on Neo4j graph database[D]. Changsha: Hunan University, 2016.

[17] JIA Ming, WANG Hong, SONG Chunli, et al. Literature review and research overview of COVID-19[J]. Shaanxi Medical Journal, 2020, 49(3): 259-263, 267.

[18] ZHAO Gang, CAO Yingying, YU Wenhui, et al. Discussion on prevention and treatment of COVID-19 based on classical Chinese medicine theory[J]. Jiangsu Journal of Traditional Chinese Medicine, 2020, 52(4): 10-13.

[19] CHEN Ying, GUO Yibo, GUO Ran, et al. Visual analysis of COVID-19 research based on bibliometrics[J]. China Journal of Chinese Materia Medica, 2020, 45(10): 2239-2248.

[20] LIU Qiao, LI Yang, DUAN Hong, et al. Survey on knowledge graph construction technology[J]. Journal of Computer Research and Development, 2016, 53(3): 582-600.

Author Contributions

FENG Xin: Data collection and analysis, experimental verification
LI Xue: Theoretical construction, paper writing
YAN Yue: Data collection and analysis, experimental verification, paper writing
LI Jiapei: Conclusion analysis, paper guidance
LIU Mengyao: Language polishing, paper revision and final version revision
WU Ye: Conclusion analysis, paper guidance

Note: Figure translations are in progress. See original paper for figures.