
AI translation · View original & related papers at
chinaxiv.org/items/chinaxiv-202304.00543

Postprint: A Strategy for Constructing an Epidemic Knowledge Base Based on Event Ontology

Authors: Xiong Li, Chengwen Wang, Wang Kun

Date: 2023-04-01T16:02:53+00:00

Abstract

[Purpose/Significance] The fragmentation and unstructured nature of epidemic information pose challenges to emergency decision-making. To support the digitization of emergency decision-making and promote intelligent emergency management, this study combines natural language processing and event ontology to automate epidemic information management and knowledge representation.

[Method/Process] This paper proposes an automatic construction strategy for domain ontology knowledge base based on web crawler, natural language processing, and event ontology. First, web crawler and natural language processing are employed for information collection and automatic extraction of event elements, based on which an epidemic event ontology model is constructed. Subsequently, an ontology construction and update algorithm is designed to accomplish the automatic construction and expansion of event ontology.

[Results/Conclusion] Research results demonstrate that the strategy is feasible for dynamic management and automatic update of epidemic information, and that event ontology can effectively describe events while creating conditions for knowledge expansion. This study provides valuable reference for related research and practice in emergency management decision-making.

Full Text

Preamble

Volume 65, Issue 14, July 2021
ChinaXiv Cooperative Journal

Construction Strategy of Epidemic Knowledge Base Based on Event Ontology

Xiong Li¹, Wang Chengwen¹, Wang Kun^{1,2}

¹ School of Management, Shanghai University, Shanghai 200444

² Australian Artificial Intelligence Institute, University of Technology Sydney, Sydney 2007

Abstract:

[Purpose/Significance] The fragmentation and unstructured nature of epidemic information pose challenges to emergency decision-making. To support the digitalization of emergency decision-making and promote intelligent emergency management, this study combines natural language processing and event ontology to automate epidemic information management and knowledge representation. [Method/Process] An automatic construction strategy for domain ontology knowledge bases based on web crawlers, natural language processing, and event ontology is proposed. First, web crawlers and natural language processing are employed for information collection and automatic extraction of event elements, upon which an epidemic event ontology model is constructed. Then, ontology construction and update algorithms are designed to complete the automatic construction and expansion of event ontology. [Result/Conclusion] The results demonstrate that this strategy is feasible for dynamic management and automatic updating of epidemic information, and that event ontology can effectively describe events while creating conditions for knowledge expansion. This study provides a reference for related research and practice in emergency management decision-making.

Keywords: event ontology; epidemic information and knowledge; knowledge base

Classification Number: G250

DOI: 10.13266/j.issn.0252-3116.2021.14.016

Epidemic information serves as a critical basis for emergency decision-making in public health emergencies. The massive epidemic data and information generated in the big data environment provide support for effective epidemic prevention and control and enhanced emergency management capabilities. Information management solutions for emergencies based on next-generation information technologies have become a focus of attention for academia and relevant authorities. In the prevention and control of COVID-19, difficulties in epidemic information sharing and management have arisen due to fragmented information, untimely updates, information barriers, and insufficient intelligence analysis capabilities [1-2], resulting in inadequate support for epidemic prevention decision-making, ultimately affecting emergency response efficiency, constraining the effectiveness of epidemic prevention and control, and causing weak overall emergency management coordination. The large volume of unstructured data emerging at various stages of the epidemic poses challenges to traditional structured data collection and processing methods based on relational databases.

Therefore, there is an urgent need for dynamic collection and processing of epidemic information, automatic representation and extraction of epidemic knowledge, and intelligent means of information sharing to provide critical support for emergency decision-making.

1. Related Research Review

1.1 Research on Epidemic Information Management and Knowledge Representation

Information management and knowledge representation for emergencies, including epidemics, constitute important content of emergency management and can provide a decision-making basis [3]. Constructing a knowledge base is an important means to achieve effective acquisition, organization, analysis, and transmission of epidemic information [4]. Through literature review, it is found that information management and knowledge representation for emergencies are mainly based on information and intelligence system modeling, semantic ontology, and other schemes, providing references for information retrieval and decision-making in emergency response. Among them, W. B. Lee et al. proposed an unstructured information management system for emergency management [5], which organizes and represents emergency knowledge through emergency event concept relationship models and dynamic knowledge flow models. M. Dorasamy et al. solved data management, knowledge sharing, and dissemination problems through emergency management information systems [6], providing critical data, information, and knowledge for emergency management personnel and promoting the sharing of life-saving information and knowledge. Guo Hua et al. used information resource planning methods to organize emergency management intelligence flows and constructed an urban emergency management intelligence platform [7], providing dynamic information and knowledge for emergency management.

Research on epidemic information management strategies is relatively lacking, focusing mainly on the analysis and management of specific symptoms and case information of infectious diseases. Gao Shan et al. designed an infectious disease emergency case disposal ontology model [12] to model the transmission and disposal of infectious diseases. Fang An et al. developed a knowledge service platform for infectious disease symptoms and diagnosis and treatment [13], mapping concepts, objects, and their relationships into a knowledge network. W. R. Hogan et al. developed an epidemiology ontology [14] to conceptualize infectious disease information. Chen Xiaohui et al. realized a COVID-19 case activity knowledge graph [15] to support the analysis of transmission processes and case trajectories. A. Joshi et al. defined the conceptual scope of infectious diseases from dimensions including health status, transmission quantity and mode, location, and time [16], providing references for epidemic knowledge representation. These studies mainly focus on knowledge organization of epidemics from perspectives such as pathology, diagnosis and treatment, and transmission, which are helpful for epidemic monitoring and diagnosis. However, research on strate-

gies and solutions for knowledge representation of the overall dynamic evolution of epidemics from an event perspective to support epidemic prevention decision-making is scarce. Therefore, addressing the shortcomings of information utility exposed in COVID-19 prevention and control, this study focuses on the perspective of epidemic-assisted decision-making, takes events as the main thread, and explores strategies for epidemic information management and knowledge representation based on dynamic updates.

1.2 Related Research on Event Ontology

Ontology is the conceptualized representation of entities, their attributes, and relationships between entities in a domain [17]. From a basic structural perspective, ontology includes a set of basic concepts (classes) in a certain domain, a set of object properties reflecting relationships between concepts, a set of data properties defining concept features, and instances representing conceptual models in the real world [18]. Related research is mainly based on these elements for ontology development and application. An event refers to something that occurs at a specific time and location, involving multiple participants [19-20] and displaying certain action characteristics, specifically including actions, participating subjects, event objects, time, and location. An event ontology model is a cognitive representation knowledge architecture applied to occurring events [21]. Event ontology achieves formal description and sharing of event class knowledge based on this framework [22], and can perform knowledge representation, semanticization, and reasoning around event themes.

Current research on event ontology mainly focuses on two aspects: extracting and sharing domain knowledge through ontology modeling (application of event ontology), and ontology modeling and construction strategies (development of event ontology). In terms of event ontology application, it provides reference solutions for knowledge modeling of emergencies and social hotspots. Among them, a shared vocabulary-based environmental pollution event ontology model is used to extract semantic relationships in various pollution incidents [23]; a flood emergency decision support system based on multi-source data uses event ontology to model complex scenarios [24]; a safety accident knowledge event ontology model achieves dynamic representation of safety accident cases and their scenarios [25-26], providing a decision-making basis for accident analysis and prediction; research on Web service composition based on event ontology [27], knowledge representation and classification of sports events [28], and news recommendation [29] provides pathways for supply-demand information representation and interest analysis. These studies enrich the application of event ontology and provide references for domain knowledge representation and decision support.

In terms of ontology modeling and development, manual methods remain the primary approach, but this strategy is time-consuming, labor-intensive, and risks manual errors, raising the threshold for ontology research and application. Therefore, Liu Sihan et al. proposed a theoretical scheme for automatic con-

struction of event ontology combining natural language processing and neural networks [30], providing theoretical guidance for related research. At the specific development level, Zhu Wenyue et al. designed an ontology model from two levels—event category, event relationship, and instance—around event features and structures [31], providing a basic framework for event ontology modeling; J. I. Single et al. used natural language processing to extract event elements from chemical accident databases and populate event ontology [26], but did not achieve automatic construction of instance relationships; O. Gurbuz et al. determined relationships and states of organizational processes through extraction of subject-predicate-object components in sentences, and identified event objects and states based on part-of-speech [32], with effectiveness dependent on text analysis results; J. A. Reyes-Ortiz identified event elements through verbs, nouns, prepositional phrases, and affixes [33], but had poor recognition and extraction effects for proper nouns, noun phrases, and quantifiers; Wang Sili et al. automatically identified domain concepts through part-of-speech tagging [34], providing ideas for preliminary work in automatic ontology construction; Q. Mao et al. identified events and their elements through part-of-speech and syntactic analysis [35], providing formal knowledge for event evolution analysis based on semantics. These studies provide references for event element recognition and extraction, but mostly adopt manual methods in the ontology construction stage, making it difficult to meet the requirements of large-scale ontology construction.

In summary, current research on epidemic information management and knowledge representation is still relatively insufficient. Event ontology organizes event knowledge from perspectives such as involved subjects, time, and location around event themes, providing feasible solutions for dynamic knowledge representation and event evolution analysis, and offering ideas for epidemic knowledge representation. This article studies epidemic knowledge representation strategies from the perspective of effective organization of epidemic information and knowledge, designing an automatic representation and update scheme for network data to strengthen the supporting role of epidemic information and knowledge in emergency management decision-making.

2. Epidemic Event Ontology Model Construction

An event ontology model is the knowledge framework of events, capable of depicting relationships between events and their elements, and serves as the basic framework of event ontology. The COVID-19 epidemic has generated a large amount of dynamic epidemic information, providing information resources for epidemic analysis and knowledge extraction while also increasing the difficulty of emergency decision-making. This study takes epidemic reports crawled from the web as the basic data source, references the event class framework from literature [16,31] and domain concept and event recognition methods from literature [34-35], uses part-of-speech analysis, named entity recognition, and semantic role labeling to extract key elements of epidemic events, designs the event ontol-

ogy model around epidemic event elements, and performs ontology expansion on this basis.

2.1 Epidemic Event Element Analysis

The construction of epidemic event ontology is based on event ontology models and dynamic, objective epidemic event information. During COVID-19 prevention and control, all provinces and cities actively reported epidemic events in their jurisdictions objectively and timely, facilitating automatic and timely collection of epidemic intelligence. This study uses official epidemic reports and open knowledge graphs as data sources, employs crawler technology to regularly collect epidemic information, uses the representative Chinese natural language processing tool LTP to process epidemic information texts and standardize the epidemic intelligence corpus, and designs an event element list construction algorithm to map semantic role labeling with text blocks one-to-one. Semantic role labeling takes the predicate of the text as the core, identifies the relationship between other components and the predicate, and thereby achieves extraction of key information, which is used as the basis for determining epidemic event elements. Semantic role labeling identifies the basic elements of epidemic events and provides a basis for automatic construction of epidemic event ontology models and event ontology. Table 1 shows the process of extracting epidemic event elements from the epidemic report information of the Shanghai Municipal Health Commission (referred to as the Health Commission).

In Table 1, N_s is the set of named entities extracted through LTP named entity recognition, generally including persons, locations, etc. Since this study mainly focuses on the epidemic dynamics in specific regions, epidemic areas are treated as epidemic subjects. Therefore, the named entity set is also the epidemic subject set, representing the subjects where epidemics occur or prevention measures are implemented in various regions.

Pos is the initial event element set obtained through part-of-speech analysis of the segmented text, serving as a reference for the event element list. Initial event elements are mainly recombined based on the subject-predicate-object triple structure and verb-noun forms in part-of-speech to extract key event relationships.

Srl is the semantic role labeling set, which uses the LTP toolkit to identify relationships between components centered on text predicates. Among them, TMP represents time, $A0$ is the subject identifier in semantic roles representing the agent or trigger of the event, and $A1$ is the patient in semantic role labeling. Act is the core predicate for epidemic information, i.e., the triggering action of the epidemic event, such as “newly added” or “from,” which establishes the correspondence between related subjects and objects and provides a basis for event knowledge representation.

According to the text analysis process in Table 1, using the epidemic report information from the Shanghai Municipal Health Commission as the data source,

event elements such as occurrence time, epidemic subject, prevention object, and epidemic dynamics and status are identified. Epidemic event elements constitute the basic structure of the epidemic event ontology and lay the foundation for designing the event ontology model. To facilitate ontology expansion and reuse, this study further expands the prevention resource elements and supplements resources belonging to prevention subjects, as shown in Table 2 .

The basic structure of epidemic events identified through event element identification provides direct basis for determining the hierarchical structure of epidemic event ontology and basic relationships between elements, forming the basic prototype of the epidemic knowledge framework. Based on the key elements of epidemic events, this study defines epidemic events by combining the four-dimensional and six-tuple models of events [16,31], and describes epidemic events through a five-tuple, as shown in Equation (1):

$$E = (TIM, Sub, Act, Obj, Sta)$$

In the epidemic event five-tuple model, E represents the epidemic event, TIM is the occurrence time of the epidemic event, Sub is the event subject, mainly including the agent and patient subjects in the epidemic event, such as epidemic outbreak areas and prevention areas. Act is the event action or triggering method, such as “newly added,” “existing,” or “cumulative.” Obj is mainly the prevention object, i.e., the object of the epidemic event. Sta is the status of the epidemic event, reflecting the current development status of the epidemic. The status is extracted as its subset through different events, such as “5 newly cured and discharged cases.” Each epidemic event instance belongs to the epidemic event class.

2.2 Epidemic Event Ontology Model

Based on the identified event elements, this study constructs an epidemic event ontology model to depict the basic structure of the event ontology and lay the foundation for enriching and expanding epidemic event knowledge. The epidemic event ontology model consists of concepts such as epidemic subjects and prevention objects, as well as a series of attributes that characterize relationships between concepts and instance status, forming the basic framework for event knowledge representation. Based on the characteristics of epidemic events and the needs of epidemic event ontology application, this study divides the epidemic event ontology model into three layers—concept and relationship layer, instance layer, and application layer—according to the ideas of knowledge organization, enrichment and update, and application, laying the foundation for the construction and automatic updating of epidemic event ontology. The epidemic event ontology model is shown in Figure 1 [Figure 1: see original paper], where the blue line part represents information automatically filled according to text content.

In the epidemic event ontology model, some concepts, relationships, and attributes generated during the automatic ontology filling process are derived from event triggering actions. Therefore, the concept and relationship layer includes event element classes, relationships between elements, and attributes of elements, which constitute the basic structure of the event ontology. Based on the concept and relationship layer, the instance layer includes specific epidemic events triggered by specific actions and related relationships and statuses of events, requiring filling and instantiation with epidemic data and information. The application layer provides functions such as information retrieval, dynamic maintenance of attributes and relationships, decision support, and knowledge representation for related applications based on the concept layer and instance layer. Based on the event ontology model, the basic concepts, relationships, and attributes involved in the event ontology are shown in Table 3 .

Through sorting out the basic concepts and relationships of epidemic events, this study further preliminarily builds the hierarchical relationships of the epidemic event ontology, constructs the epidemic event ontology `epi`, forms the ontology file `epieve.owl`, creates the basic architecture of the epidemic knowledge base, and prepares for automatic filling of epidemic event instances and reuse and expansion of the ontology.

3. Construction of Epidemic Knowledge Base

This study collects and extracts event elements through web crawler and text analysis technologies, constructs event element lists, designs event ontology automatic construction and filling algorithms based on epidemic event characteristics, automatically enriches event element lists into the event ontology, and continuously integrates open knowledge graphs and domain ontologies according to application requirements to form an epidemic knowledge base that provides support for decision-making. This study adopts the event ontology-driven epidemic knowledge base construction process shown in Figure 2 [Figure 2: see original paper]. The implementation of the scheme is divided into three main stages: epidemic information collection and preprocessing, epidemic event element extraction, and epidemic knowledge base construction and decision support. Taking the COVID-19 epidemic in Shanghai in 2020 as an example, the epidemic dynamics reported by the Shanghai Municipal Health Commission are used as the core data source to apply and test the proposed strategy, constructing a core epidemic event ontology library. On this basis, Chinese open knowledge graph data is introduced to continuously expand the epidemic knowledge base.

3.1 Epidemic Event Element Extraction

In Section 2.1, to introduce the epidemic event ontology model, this study briefly explained the process of using LTP for epidemic text segmentation, named entity recognition, semantic role labeling, and then constructing the event element list `eleList` (see Table 1). To clarify the basic structure of epidemic events, this section provides further details on epidemic event element extraction. The key to

epidemic event element extraction lies in testing semantic role labeling of events, supplementing missing components, and automatically mapping semantic role labels with text content one-to-one, storing them in the form of element lists. This study designs the epidemic event element list construction algorithm (see Algorithm 1) to accurately describe objective events.

Based on the LTP semantic role type conventions, A0 represents the event trigger or agent, A1 is the patient, QTY represents quantity, and TMP is the event occurrence time. In addition, the first element in the semantic role list represents the root node, i.e., the central predicate reflecting the action of the event, denoted as “Act” in this study. For cases where the event subject is missing, such as “including 1 case from Spain,” its semantic role label is (36, [(A1, 37, 37)]), with part of the event subject A0 missing. This study uses the forward traversal rule in Equation (2) to find the nearest event subject as the subject of this triple, i.e., “cured and discharged cases from Spain.”

$$\begin{aligned}
 A0'_t &= \text{segs}(s) \\
 s &= \text{Srl}_{1,n}^{i-1}(k) \\
 n &= \text{len}(\text{Srl}_1^{i=1}(k)) \\
 n &= n \quad \text{if } k \in \{A0, A1\} \\
 n &= n - 1 \quad \text{if } k \notin \{A0, A1\}
 \end{aligned}$$

Algorithm 1 Epidemic Event Element List Construction Algorithm

Input: epiText # epiText is epidemic text

Output: eleList # eleList is epidemic event element list

1. Segs = [ltp.seg(epiText)] # Text segmentation, Segs is the segmentation set
2. Srl = [ltp.srl(epiText)] # Semantic role labeling, Srl is the semantic role set
3. eleList = [], ek = [], ev = []
4. for srl in Srl and segs in Segs do # segs is the corresponding sentence component
 5. eve = [] # Initialize event fragment
 6. ek.append('Act'), ev.append(segs[srl[0]]) # Event triggering action
 7. for s in srl[1] do
 8. t = '', ek.append(s[0])
 9. for i in range(s[1], s[2] + 1) do
 10. ev.append(t)
 11. t += segs[i]
 12. endfor
 13. if 'A0' not in eve[ek] then
 14. eve[ek].append('A0'), eve[ev].append(A0_i ← eq.(2))
 15. endif
 16. eleList.append(eve)
5. endfor
6. return eleList

The event element extraction strategy based on natural language processing is suitable for extracting knowledge elements at the event summary description level, mainly involving part-of-speech analysis and semantic role labeling, and

has been widely applied in event element analysis research such as literature [31,35]. To enhance the effectiveness and specificity of this study's epidemic event element extraction strategy, this study introduces mechanisms such as named entity matching and part-of-speech-based initial event element comparison to correspond event elements with basic event concepts (see Figure 3 [Figure 3: see original paper]), with specific implementation logic reflected in Algorithm 2 for automatic construction of epidemic event ontology.

3.2 Automatic Construction of Epidemic Knowledge Base

Based on the extracted epidemic event elements, this study further designs Algorithm 2 to convert relevant elements into triple form and automatically create and generate triple-based graph structure semantic ontology data, forming the basic framework of the epidemic knowledge base. Based on the epidemic event ontology model, the automatic construction algorithm of the epidemic knowledge base distinguishes relevant concepts, object properties, and data properties, automatically constructs and improves the ontology framework, then divides event elements into instances, relationships, properties, and property values, and automatically fills them into the ontology architecture in triple form sequentially.

For the recognition and distinction of event subjects and objects, this study first identifies and extracts all named entities from event texts through LTP, and finds that named entities in epidemic events are all global regions, i.e., event subjects in epidemic events. The named entity matching mechanism matches event subjects labeled as A0 or A1 with elements in the identified named entity set Ns. Matched A0 elements are prevention subjects as, matched A1 elements are epidemic subjects es, and unmatched elements are prevention objects eo or others, with relevant concepts corresponding one by one.

Algorithm 2 Automatic Construction and Filling Algorithm for Epidemic Event Ontology

Input: eleList, epi, Ns, Pos # Event element list, event ontology, named entity set, part-of-speech-based element set

Output: epi

1. Class(E, ES, EO, AES) epi, AES ES
2. Property(rsub, tim, sta) epi
3. for eve[ele, seg] eleList and eve.seg segs do # seg is the corresponding sentence component
 4. as = es = eo = act = act1 = sta = time = ', num = 0
 5. ek = eve.ele, ev = eve.seg
 6. if ek == A0 or ek == A1 and ev in Ns and ev in Pos(as, eo, so) then
 7. if ek == A0 then as = ev, epi.add((as, type, AES))
 8. else es = ev, epi.add((es, type, ES))
 9. endif
 10. elif ek == A1 and ev not in Ns and ev in Pos(as, eo, so) then
 11. eo = ev, epi.add((eo, type, EO))
 12. elif ek == Act or ek == QTY or ek == TMP then
 13. if ek == Act then act = ev, epi.add((act,

```

type, objectProperty)) 14.      elif ek == QTY then num = ev
15.      else time = ev 16.      endif 17.      endif 18.      e = 'EE'
+ time, epi.add((e, type, E))    # e is the epidemic event 19.      act1 =
act + eo, epi.add((act1, type, datatypeProperty)) 20.      epi.add((act1,
subPropertyOf, sta)) 21.      epi.add((e, tim, time)) 22.      epi.add((e,
rsub, es)) 23.      epi.add((as, act, eo)) 24.      epi.add((e, act1, num))
4. endfor
5. return epi

```

Algorithm 2 first maps epidemic event elements to event concepts, constructs relationships between concepts and instances and properties of instances in triple form, and adds classes and properties to the event ontology based on unique resource identifiers (URIs) without duplication, thereby constructing the hierarchical structure and basic relationships of the ontology. On this basis, the algorithm reads epidemic event elements through the event element list, constructs epidemic event instances, judges instance elements, and constructs URI identifiers for elements if they correspond to types such as event subject A0, time TMP, patient A1, and quantifier QTY. Then, it judges whether corresponding components exist in the ontology; if not, automatically adds relevant elements to the ontology and fills in relationships and property values between elements. Finally, the filled ontology data is returned to complete the automatic construction of the epidemic event ontology. The epidemic event instances gradually enriched through Algorithm 2 are shown in Figure 6 [Figure 6: see original paper]. Through cyclic operation of the algorithm, new epidemic information can be continuously supplemented, enabling automatic updating of the epidemic knowledge base, with epidemic knowledge gradually becoming richer.

3.3 Epidemic Knowledge Base Update Based on SPARQL

To enable the epidemic knowledge base to have basic reasoning and rule-based updating capabilities to support decision-making, this study designs a knowledge update strategy based on the query language SPARQL. On one hand, the epidemic knowledge base can automatically extract event elements and update automatically by combining Algorithms 1 and 2; on the other hand, it also needs to design personalized rules to enrich the knowledge base according to specific application requirements. This study designs SPARQL-based query rules from two perspectives—high-risk events and high-risk subjects—and combines Algorithm 3 to achieve knowledge updating. The specific query rules are shown in Table 4.

Algorithm 3 Epidemic Knowledge Base Update Algorithm Based on SPARQL

Input: epi, X, starttime, Y

Output: epi

```

1. Hre = epi.query(Q1, X, starttime)    # Execute SPARQL query
2. Hrs = epi.query(Q2, Y)
3. if Hre != [] then 4.      epi.add((HRE, subclassOf, E))    # Create "High-

```

```

    risk Event” class HRE 5.    for e in Hre do epi.add((e, type, HRE)) endfor
4. endif
5. if Hrs != [] then 8.    epi.add((HRS, subClassOf, ES))    # Create
    “High-risk Subject” class HRS 9.    epi.add((ren, type, datatypeProp-
    erty))    # Create “involved event count” 10.    for (es, en) in Hrs
    do 11.        epi.add((es, type, HRS))    # es is high-risk subject
    12.        epi.add((es, ren, en))    # en is involved event count
    13.    endfor
6. endif
7. return epi

```

Algorithm 3 extracts high-risk objects according to query rules Q1 and Q2, creates two subclasses belonging to epidemic events and epidemic subjects respectively, classifies the two types of high-risk objects separately, and performs automatic update operations for newly added information to support emergency decision assistance under risk identification. To test the algorithm, this study sets corresponding query parameters respectively, with specific results shown in Table 5 .

3.4 Results Analysis

In response to the trend of dynamic, intelligent information management and knowledge representation for public health emergencies in the big data context, this study takes the COVID-19 epidemic as an example to research epidemic knowledge representation strategies. Regarding data sources, the epidemic reports from the Shanghai Municipal Health Commission are selected as the basic data source. Web crawlers are used to dynamically crawl epidemic information since January 2020. As of December 31, a total of 320 corpora were extracted to build an epidemic corpus. In terms of methods and tools, Python is used to standardize texts, the Language Technology Platform toolkit LTP from Harbin Institute of Technology is employed for corpus segmentation, named entity recognition, part-of-speech analysis, and semantic role labeling. Algorithm 1 is used to match semantic role labels with corpora to construct the event element list eleList. Algorithm 2 is applied to fill elements from the eleList into the epidemic event ontology epi, while achieving automatic updating of epidemic instance-related information. This study uses the epidemic event ontology as the carrier of the epidemic knowledge base and employs SPARQL to implement the query rules of Algorithm 3 to complete dynamic updating of the epidemic knowledge base. Through text experiments, an epidemic domain ontology containing 6 classes, 4 object properties, 9 data properties, 410 instances, 1,121 pairs of relationships, and 918 data property values was finally obtained, preliminarily implementing an event ontology knowledge base oriented to epidemics and providing references for epidemic knowledge organization and sharing.

For knowledge expansion, this study selected prevention resources data belonging to prevention subjects from the web (<https://yyk.99.com.cn/>) (Dataset 1), the “COVID-19 Open Knowledge Graph.Events” (Dataset 2) and “COVID-19

Open Knowledge Graph.Epidemiology” (Dataset 3) from the Chinese Open Knowledge Graph platform (<http://www.openkg.cn/>), to enrich and expand the epidemic knowledge base. Prevention resources are tertiary hospitals belonging to prevention subjects (Shanghai), whose data is crawled through web crawlers and expanded according to the logic of Algorithm 2 following the hierarchy of “prevention resources → medical institutions → hospitals,” with examples shown in Figure 7 [Figure 7: see original paper].

Datasets 2 and 3 correspond to event schema and epidemiology data in `xlsx` and `json` formats (see Figure 8 [Figure 8: see original paper]). This study uses Python to parse them, converts key fields such as `id`, `type`, `label`, `domain`, `range`, `subClassOf`, and `subPropertyOf` into concepts, instances, and property types, builds auxiliary ontologies `event` and `covid` based on the idea of Algorithm 2, and merges these two ontologies into the epidemic event ontology using the protégé ontology editing tool. The final basic structure of the epidemic knowledge base is shown in Figure 9 [Figure 9: see original paper].

Through expansion with three datasets, epidemic knowledge has been extended, demonstrating the feasibility of the event ontology knowledge expansion strategy based on multi-source data. This scheme provides references for integrating epidemic events with knowledge in fields such as resources, geography, populations, diseases, and policies, offers a referable solution for emergency decision-making based on multi-source data, and lays the foundation for further epidemic knowledge semantic reasoning and reuse. The results of automatic construction of the epidemic knowledge base based on event ontology show that the ontology automatic construction strategy combining natural language processing technologies such as named entity recognition, part-of-speech analysis, and semantic role labeling can achieve epidemic information management and knowledge representation from data collection, data preprocessing, event element extraction, event ontology automatic construction and filling, to knowledge base automatic updating and expansion, improving knowledge base construction efficiency and reducing manual costs. However, the automatic construction strategy for epidemic event ontology largely depends on text preprocessing and natural language processing strategies, requiring development of corresponding algorithms to drive ontology automation construction.

References

- [1] Cao Zhenxiang, Chu Jiewang, Guo Chunxia. Construction of a theoretical framework for emergency intelligence support system for major epidemic prevention and control—Taking the 2019 novel coronavirus pneumonia epidemic prevention and control as an example[J]. *Library and Information Service*, 2020, 64(15): 72-81.
- [2] Li Jinhua. Theoretical framework and application of major epidemic information management in the era of big data[J]. *Modern Information*, 2020, 40(7): 25-33, 51.

- [3] Ni ZJ, Rong LL, Wang N, et al. Knowledge model for emergency response based on contingency planning system of China[J]. *International journal of information management*, 2019, 46: 10-.
- [4] Ding Botao. Epidemic information management in the era of artificial intelligence: Challenges and changes[J]. *Library and Information Knowledge*, 2020(6): 109-116.
- [5] Lee WB, Wang Y, Wang WM, et al. An unstructured information management system (UIMS) for emergency management[J]. *Expert systems with applications*, 2012, 39(17): 12743-12758.
- [6] Dorasamy M, Raman M, Kaliannan M. Integrated community emergency management and awareness system: a knowledge management system for disaster support[J]. *Technological forecasting and social change*, 2017, 121: 139-167.
- [7] Guo Hua, Qu Fang, Zhan Peizhi. Research on construction of urban emergency management intelligence platform[J]. *Library and Information Service*, 2018, 62(6): 93-104.
- [8] Xu Beibei, Yang Zijiang, Zhu Shiwei, et al. Research on construction of public opinion knowledge base for emergencies based on ontology[J]. *Journal of Intelligence*, 2019, 38(4): 132-137.
- [9] Zhao Youlin, Pang Shuo, Wu Zongda. Research on construction of spatio-temporal semantic model for emergency management driven by social perception data[J]. *Information Science*, 2021, 39(2): 44-.
- [10] Wang Fang, Yang Jing, Xu Lulu. Research on ontology construction for fire emergency management[J]. *Journal of the China Society for Scientific and Technical Information*, 2020, 39(9): 914-925.
- [11] Wu HT, Zhong BT, Medjdoub. An ontological metro accident case retrieval using CBR and NLP[J]. *Applied sciences*, 2020, 10(15): 5298.
- [12] Gao Shan, Wang Wenjun, Du Lei, et al. Research on infectious disease emergency case sharing ontology model[J]. *Computer Applications*, 2010, 30(11): 2924-2927.
- [13] Fang An, Hong Na, Gao Dongping, et al. Infectious disease ontology construction and its application in knowledge service platform[J]. *New Technology of Library and Information Service*, 2012(1): 7-12.
- [14] Hogan WR, Wagner MM, Brochhausen M, et al. The Apollo structured vocabulary: an OWL2 ontology of phenomena in infectious disease epidemiology and population biology for use in epidemic simulation[J]. *Journal of biomedical semantics*, 2016, 7: 50.
- [15] Chen Xiaohui, Liu Junnan, Xu Li, et al. Construction of COVID-19 case activity knowledge graph—Taking Zhengzhou city as an example[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(6): 816-825.

- [16] Joshi A, Karimi S, Sparks R, et al. Survey of text-based epidemic intelligence: a computational linguistics perspective[J]. *ACM computing surveys*, 2019, 52(6): 119.
- [17] Tudorache T. Ontology engineering: current state, challenges, and future directions[J]. *Semantic web*, 2020, 11(1): 125-.
- [18] Lv ZM, Peng R. A novel meta-matching approach for ontology alignment using grasshopper optimization[J]. *Knowledge-based systems*, 2020, 201: 106050.
- [19] Zhong Z, Liu Z, Liu W, et al. Event ontology and its evaluation[J]. *Journal of information and computational science*, 2010, 7(1): 95-101.
- [20] Liu W, Jiang L, Wu YS, et al. Topic detection and tracking based on event ontology[J]. *IEEE access*, 2020, 8: 98044-98056.
- [21] Zhong ZM, Liu ZT, Li CH, et al. Event ontology reasoning based on event class influence factors[J]. *International journal of machine learning and cybernetics*, 2012, 3(2): 133-139.
- [22] Goy A, Magro D, Rovera AM. On the role of thematic roles in a historical event ontology[J]. *Applied ontology*, 2018, 13(1): 19-39.
- [23] Liu Wei, Ding Ning, Yang Junhui, et al. Event ontology schema for environmental pollution emergency domain[J]. *Computer Science and Exploration*, 2016, 10(4): 466-480.
- [24] Fertiera A, Barthe-Delanoë AM, Montarnal A, et al. A new emergency decision support system: the automatic interpretation and contextualisation of events to model a crisis situation in real-time[J]. *Decision support systems*, 2020, 133: 113260.
- [25] Zhu Yuqian, Liao Huimin, Luo Xiaojuan, et al. Research on event ontology model of safety information cognition accidents[J]. *Journal of Intelligence*, 2020, 39(7): 151-158.
- [26] Single JI, Schmidt J, Deneckere J. Knowledge acquisition from chemical accident databases using an ontology-based method and natural language processing[J]. *Safety science*, 2020, 129: 104747.
- [27] Tang Yingying, Liu Wei, Liu Feijing, et al. Dynamic composition of Web services based on event ontology[J]. *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, 2013, 41(S2): 22-.
- [28] Kuptabut S, Netisopakul P. Event extraction using ontology directed semantic grammar[J]. *Journal of information science and engineering*, 2016, 32(1): 79-96.
- [29] Zhu Wenyue, Liu Wei, Liu Zongtian. News personalized recommendation based on event ontology[J]. *Computer Engineering*, 2019, 45(6): 267-272, 279.
- [30] Liu Sihan, Liu Xuhong, Liu Xiulei. A survey of event ontology representation models and construction[J]. *Journal of Beijing Information Science &*

Technology University, 2018, 33(2): 35-40, 7.

[31] Zhu Wenyue, Liu Zongtian. Domain knowledge modeling for emergencies based on event ontology[J]. Computer Engineering and Applications, 2018, 54(21): 148-155.

[32] Gurbuz O, Rabhi F, Demirors O. Process ontology development using natural language processing: a multiple case study[J]. Business process management journal, 2019, 25(6): 1208-1227.

[33] Reyes-Ortiz JA. Criminal event ontology population and enrichment using pattern recognition from text[J]. International journal of pattern recognition and artificial intelligence, 2019, 33(11): 1940014.

[34] Wang Sili, Zhu Zhongming, Liu Wei, et al. Research on automatic domain ontology concept acquisition method based on deep learning[J]. Information Studies: Theory & Application, 2020, 43(3): 145-152, 160.

[35] Mao Q, Li X, Peng H, et al. Event prediction based on evolutionary event ontology knowledge[J]. Future generation computer systems, 2021, 115: 76-89.

Author Contributions:

Xiong Li: Determined the research topic, proposed the research ideas, and guided paper revision;

Wang Chengwen: Designed the research scheme, conducted experiments, and wrote the paper;

Wang Kun: Validated algorithms and revised the paper.

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

Source: ChinaXiv — Machine translation. Verify with original.