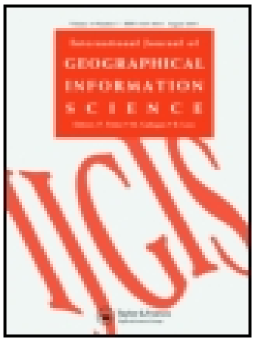

International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

A knowledge representation model based on the geographic spatiotemporal process

Kun Zheng, Ming Hui Xie, Jin Biao Zhang, Juan Xie & Shu Hao Xia

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2021.1962527

RESEARCH ARTICLE

A knowledge representation model based on the geographic spatiotemporal process

Kun Zheng[a], Ming Hui Xie[a], Jin Biao Zhang[b], Juan Xie[a] and Shu Hao Xia[a]

[a]School of Geography and Information Engineering, China University of Geosciences, Wuhan, China; [b]Guangdong Meteorological Observation Data Center, Guangdong, China

ABSTRACT

Knowledge graphs (KGs) represent entities and relations as computable networks, which is of great value for discovering hidden knowledge and patterns. Geographic KGs mainly describe static facts and have difficulty representing changes, greatly limiting their application in geographic spatiotemporal processes. By analyzing the spatiotemporal features and evolution of geographic elements, this study presents the geographic evolutionary knowledge graph (GEKG). Its representation model has five core elements: time, geographic event (geo-event), geographic entity (geo-entity), activity and property, and defines six relations: logical, semantic, evolutionary and temporal relation, participation and inclusion. It establishes a hierarchical cubical model structure and each temporal layer extends vertically and horizontally starting with the earliest geo-event. Vertical expansion refers to the connection between different kinds of element, such as the participation relation between geo-entities and geo-events. Horizontal expansion indicates the association between the same kinds of element, such as the semantic relation between geo-entities. For different layers, the spatiotemporal differences of elements produce the evolutionary relation. Finally, the comparison of GEKG with Yet Another Great Ontology (YAGO) and Geographic Knowledge Graph (GeoKG) shows that GEKG has more advantages in representing geographic evolutionary knowledge, revealing the evolution mechanism of geographic elements and the evolutionary reasons.

ARTICLE HISTORY
Received 4 December 2020
Accepted 27 July 2021

KEYWORDS
Knowledge graph; geographic spatiotemporal process; time; geo-events; geo-entities' evolution

1. Introduction

In the era of big data, the rapid increase of geographic spatiotemporal data and the lack of efficient knowledge extraction technology make it difficult to discover spatiotemporal knowledge and patterns (Lin *et al*. 2018). How to use intelligent techniques to represent and analyze spatiotemporal processes is an urgent problem, and is also key to processing of spatiotemporal data for its intelligent application (Luo *et al*. 2013, Deng *et al*. 2020).

Traditional visualization models with geographic spatiotemporal processes use spatial data as the background for visual integration. This displays the spatiotemporal changes of only a single geographic element and lacks the ability to describe the interaction and semantic relations between elements (Zheng *et al*. 2013, Shi *et al*. 2019). A geographic

CONTACT Ming Hui Xie taria@cug.edu.cn

knowledge graph (KG) establishes a semantic knowledge network for processes such as knowledge discovery, reasoning and prediction (Gong *et al*. 2014, Jiang *et al*. 2018). It provides a computable structure to analyze the association between geographic entities (geo-entities), identify hidden geographic relations, and discover geographic patterns.

Geographic KGs mainly focus on geo-entities and describe the properties and relations between geo-entities. However, for the representation of geographic spatiotemporal processes, the evolutionary knowledge is ignored, for example the development of geo-events and the formation, evolution and destruction of geo-entities. These are of great value to reveal the evolutionary patterns and mechanisms of geographic elements. For example, the spatiotemporal patterns of a typhoon can be obtained from its spatiotemporal changes. It can provide a reference for forecasting the path and impact on the weather of such a typhoon, and has important value for typhoon forecasting and early warning (Liu *et al*. 2020).

These representation models of geographic KGs organize knowledge in the form of <geo-entity, relation, geo-entity> and thus have difficulty representing changes. Their knowledge services are limited to static facts, which greatly limits the in-depth study of geographic spatiotemporal processes. To improve their expression ability regarding evolutionary knowledge, this paper presents the geographic evolutionary knowledge graph (GEKG). By representing the long-term development and changes of geo-events and geo-entities, it can provide support for revealing the formation process, the evolutionary patterns, and process-based reasoning and prediction.

2. Related work

Representing the development and change of geographic elements in a structured way lays a foundation for intelligent calculation, reasoning and prediction. It is currently the main method to represent the spatiotemporal process and can be divided into two types: those based on ontologies and those based on graphs. In addition, processes can also be represented by formulae, which reflects the effect of the process.

2.1. Representation by formulae

Using formulae to represent the process is very common. It usually takes many factors into account, reflecting the effect of the process. For example, according to the distribution of pollution concentration, the spatiotemporal model is established to describe or predict the concentration variation of air pollutants (Zou *et al*. 2016, Deng *et al*. 2018). The ensemble Kalman filter is used to obtain the best estimate of land use changes (Li *et al*. 2012). The local regression model is used to estimate the concentration of air pollutants based on spatiotemporal data (Leung *et al*. 2019). In this method, the change is represented as numerical model for prediction and calculation, which is more suitable for the change of a single geographic element.

2.2. Representations based on ontology

Geographic ontology, as a conceptual and formal norm, accurately defines the concept of knowledge and the relations between entities (Lopez-Pellicer *et al*. 2012). Ontology-based methods express the process by modeling the semantic, causal and evolutionary relations among relevant element ontologies. For example, ontologies such as sequence ontology, state ontology and process ontology are built, and the interaction between them is established to realize the modeling of complex dynamic geographic phenomena (Xue *et al*. 2019). Geographic ontologies and characterizations are used to represent geographic scenarios, and the connections among elements in different scenarios are established to simulate the evolution (Huang *et al*. 2019). By constructing a geographic process-centric ontology model, the interaction among geographic scenes, geographic processes and geographic elements are used to represent the dynamic changes in geographic phenomena (Cao *et al*. 2018). Ontology-based methods are usually based on a series of sub-processes and sub-scenarios, emphasizing their relations and the roles of the geographic elements in the process. Therefore, these methods lack the interactions between the process and the evolution of geographic elements.

2.3. Representations based on graphs

At present, the representation models of geographic KGs are organized in the form of <geo-entity, relation, geo-entity> triples, which describe the properties and relations between geo-entities (Chen *et al*. 2017, Li *et al*. 2017, Zhu *et al*. 2017, Patel *et al*. 2018, Wang *et al*. 2018, Fan *et al*. 2019). Entities have spatiotemporal features, and knowledge graphs are gradually being used to model changes. They can be divided into two categories: temporal knowledge graphs and event graphs.

(1) Temporal knowledge graph. On the basis of typical triples, adding elements such as time and change can represent the differences of entities in different periods. As an example, using the Global Database of Events, Language, and Tone (GDELT) and Integrated Crisis Early Warning System (ICEWS) data sets, knowledge graphs are organized in the form of <entity, (relation, time), entity> (Trivedi *et al*. 2017, Liu *et al*. 2019a, Jin *et al*. 2020). Based on DBpedia, different versions of knowledge graphs are constructed, and each version is a knowledge snapshot in a specific period (Tasnim *et al*. 2019). Geographic knowledge graph (Wang *et al*. 2019) defines six basic elements to describe geo-entities, the location, time, attribute, state, change and relation. It uses entity states to represent changes. Temporal knowledge graphs take entities as research objects and can represent the differences of attributes.

(2) Event graphs. These graphs focus on events and describe their development process. For example, event logic graphs (ELGs) describe the evolutionary pattern and logical relations between events (Li *et al*. 2018, Ding *et al*. 2019). An ELG takes abstract events as nodes but is not concerned with the exact participants, location and time; it takes the sequential and causality relations as edges to represent the logical development of events. The patient event graph models the temporal relations between medical activities, and related entities exist only as actors (Liu *et al*. 2018). According to the sequence of events in the text, dialogue generation based on event graphs is performed, and multiround dialogue predictions are made by event chains (Xu *et al*. 2020). Entities generally act as participants in events, and event graphs do not direct attention to the changes of entities and the interactions between entities and events.

A geographic spatiotemporal process is the long-term evolution of geo-entities and related geo-events. The occurrence of geo-events causes the changes of geo-entities, which produce new events. In summary, there remain several problems in the representation of geographic KGs: (1) the lack of representation of complex associations between events and entities (i.e. they coevolved and are interrelated), and (2) time is not relegated to solely an attribute since it also conveys the direction of evolution.

3. The representation model of the geographic evolutionary knowledge graph

A geographic spatiotemporal process is the long-term evolution of geographic elements, containing, for example, time, geo-events and their constituent elements, geo-entities and their related properties. A geographic evolutionary knowledge graph is a completely new knowledge graph of geography, representing the evolution of geographic elements. Taking a cubic structure with a time hierarchy, its representation model establishes the complex association between geographic elements in the same layer and different layers.

3.1. The knowledge representation approach

3.1.1. Main elements in the representation model

According to the features of geographic elements in the spatiotemporal process, the model has five core elements: time, geo-entity, property, geo-event and activity. Geo-entities and geo-events form the core.

(1) Geo-entity: An individual entity that exists and can be distinguished from other entities of geography, such as a specific instance of a mountain, river or island.
(2) Time: A period of time.
(3) Property: The property specific to each geo-entity, such as the location, area and shape.
(4) Geo-event: Activity involving one or more geo-entities in a certain period (Xiang and Wang 2020).
(5) Activity: The movement or change in the geo-entities in the geo-event, such as glacier movement, river avulsion or tectonic movement of a continental plate.

3.1.2. The representation approach of geographic elements and relations

The model contains five core elements and defines six relations. The representation approach of geographic elements and their relations is shown as follows.

(1) The representation approach of geographic elements

I. The representation approach of time:

Let the time set be $T = \{ t_1, t_2, \ldots, t_n \}$, which is the whole time of the process. There is a temporal relation between any t_i and t_{i+1} in$T$. T_{tri} is the triple consisting of the times and their relation and can be expressed as:

$$T_{tri} = \langle t_i, r_{temporal}, t_{i+1} \rangle, t_i \cap t_{i+1} = \emptyset \quad (1)$$

where $r_{temporal}$ is a temporal relation.

II. The representation approach of geo-entities:

Time is the inherent feature of each geo-entity. Property is the key to distinguish geo-entities in different periods, including spatial features and other attributes. Given a geo-entity set $En = \{entity_1,\ entity_2,\ \ldots,\ entity_n\}$, a spatial features set $S_n = \left\{s_{n1},\ s_{n2}, \ldots,\ s_{nj}\right\}$ and an attribute set, each geo-entity exists for t_n. The structure of the geo-entity can be defined as follows:

$$entity_n = (t_n, S_n, A_n) = \left(t_n, \left\{s_{n1},\ s_{n2}, \ldots,\ s_{nj}\right\}, \{a_{n1},\ a_{n2}, \ldots,\ a_{nk}\}\right),\ S_n \cap A_n = \emptyset \quad (2)$$

Given the relation r_{ij} between $entity_i$ and $entity_j$, En_{tri} is the triple consisting of the geo-entities and their relations and can be expressed as:

$$En_{tri} = \ < entity_i, r_{ij}, entity_j > \ = \ < (t_i, S_i, A_i), r_{ij}, (t_j, S_j, A_j) >, t_i, t_j \in T \quad (3)$$

where r_{ij} includes two types, and $r_{evolving}$. $r_{semantic}$ is the geographic semantic relation between two geo-entities. $r_{evolving}$ is the evolutionary relation based on the change of S_n and A_n.

III. The representation approach of geo-events:

Given the geo-event set $Ev = \{Event_1, Event_2, \ldots, Event_m\}$, according to the definition of the geo-event, its structure can be expressed as:

$$Event_m = (t_m, e_{Sm}, v_m, e_{Om}) \quad (4)$$

where t_m is the time when the geo-event occurred.e_{Sm} and e_{Om} are geo-entities in a geo-event and play different roles. represents the geo-entity that produces the activity, and e_{Om} represents the geo-entity on which the activity acts. v_m indicates the movement or change of the geo-entities. They are the participators of each geo-event. In addition, can be omitted in different scenarios, e.g. the triple < Late Eocene, Indian Continent, moves northward > contains only one geo-entity, indicating that the Indian Continent moves northward in the Late Eocene.

Given the relation R_{ij} between $Event_i$ and $Event_j$, Ev_{tri} is the triple consisting of the geo-events and their relation and can be expressed as:

$$Ev_{tri} = \ < Event_i, R_{ij}, Event_j > \ = \ < (t_i, e_{Si}, v_i, e_{Oi}), R_{ij}, \left(t_j, e_{Sj}, v_j, e_{Oj}\right) > \quad (5)$$

where R_{ij} reflects the logical relation between the geo-events.

(2) The representation approach of relations

The model contains six relations: temporal, semantic, evolutionary, logical, participation and inclusion.

(I) Temporal relation is a sequential relation between two periods of time.
(II) Semantic relation is the geographic semantic relation between two geo-entities, such as the relation of adjacency and subordination.
(III) Evolutionary relation is established according to the spatiotemporal changes of each geo-entity, and contains two types: 'the same as' and 'evolves into'. The relation of 'evolves into' associates the entities before and after the change, the relation of 'the same as' connects the initial state and the state before the change.
(IV) Relation of inclusion is an inclusion relation between and its properties.
(V) Logical relation refers to a logical connection between geo-events, and mainly contains the adjoint relation and causality relation.

(VI) Relation of participation: Each geo-event has constituent elements, and the participators play different roles.

(3) The representation model of GEKG

Based on the representation approach of time, geo-entities and geo-events, GEKG can be represented as follows:

$$GEKG = \{ < Ev_{tri}, En_{tri}, T_{tri} > \} \tag{6}$$

where Ev_{tri} refers to the logical relation between geo-events in the same and different periods, En_{tri} represent the properties and relations of geo-entities in the same and different periods, and T_{tri}refers to the changes over time. The structure of each element also reflects the association between them.

3.2. The hierarchical cubical graph structure

In a geographic process, all elements are interrelated in the same period and there are differences at different times. According to the representation approach of geographic elements and relations, we establish a cubical graph structure with a time hierarchy to represent the evolution (Figure 1). GEKG can be divided into different time layers. Each layer reflects the association of geo-events, geo-entities and properties at the current time; between different layers, all the elements evolve in different ways.

(1) The representation of each element

In the representation model of GEKG, taking each geo-event as a whole, there is a logical relation between different geo-events (Figure 2(a)). For geo-entities, properties are the signs that distinguish them. Each geo-entity has its own unique spatial features and attributes, and there is a semantical relation between different geo-entities.

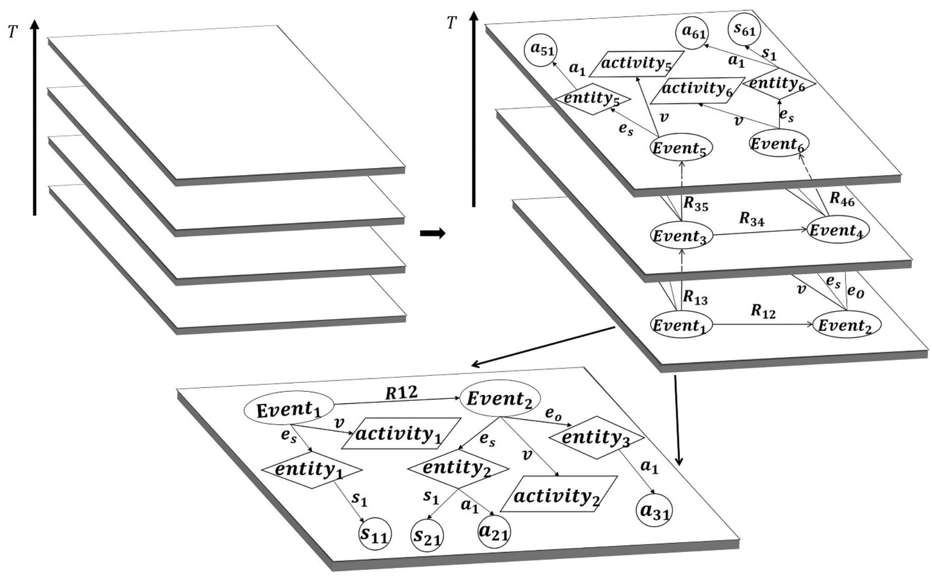

Figure 1. Graph structure of the representation model.

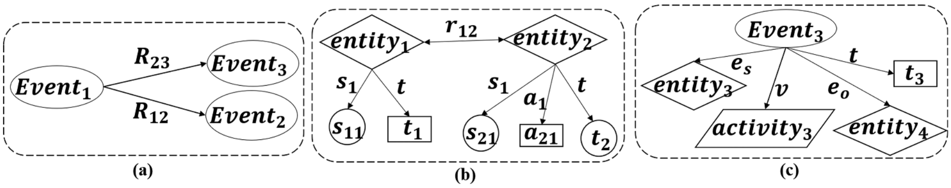

Figure 2. Example representations of each element. (a) 'geo-events and relations', (b) 'properties and relations among geo-entities', and (c) the structure of a geo-event.

Therefore, geo-entities are organized as shown in Figure 2(b). Since geo-events have a certain structure, all the elements are correlated by establishing their connections. The structure of geo-events is shown in Figure 2(c).

(2) The association in the same layer.

According to the representation in Figure 2, the association in the same period can be expressed as the structure shown in Figure 3. That is, in the same layer, all the elements belong to the current period, and the properties and relations have not changed.

(3) The association in the different layers.

For each time layer, the elements are associated in the same way. However, over time, various elements evolve in different ways (Figure 4).

The associations among geo-events in different periods are shown in Figure 4(a). In the evolutionary process, the occurrence of a geo-event causes a series of related geo-events, which reflects the logic of the development of geo-events. The properties of a geo-entity may change over time. According to the differences in different periods, evolutionary relations can be established (Figure 4(b)). From 'Middle Palaeocene' to 'Late Eocene', 'the same as' is an evolutionary relation, which indicates that 'ThesoutherntipofIndia^1' has not changed. 'Middle Palaeocene' and 'Late Eocene' represent the start-time and the end-time of the entity's state, respectively. From 'Late Eocene' to 'Late Oligocene', the value of the type has changed.

The evolutionary relation 'evolves into' associates the entities before and after the change, indicating that 'The southern tip of India^1' has evolved into 'The southern tip of India^2'. Therefore, 'Late Oligocene' is the start-time of the entity's new state. The changes of geo-entities indicate the occurrence of geo-events (Figure 4(c)). The event indicates that the property of 'East Asian monsoon region^1' has changed, and *activity* represents the change

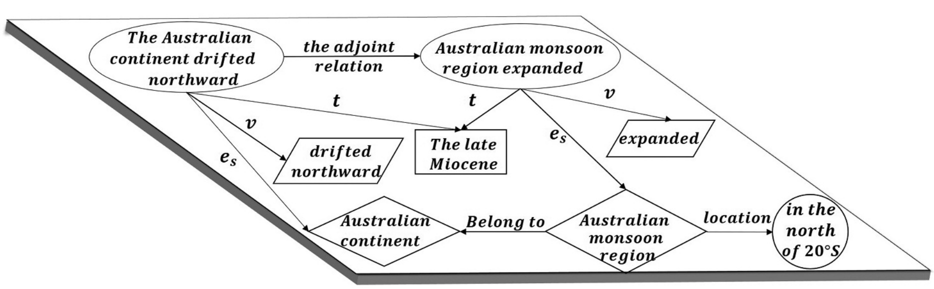

Figure 3. Example of 'the association between elements in the same period' for the development of the Australian monsoon region.

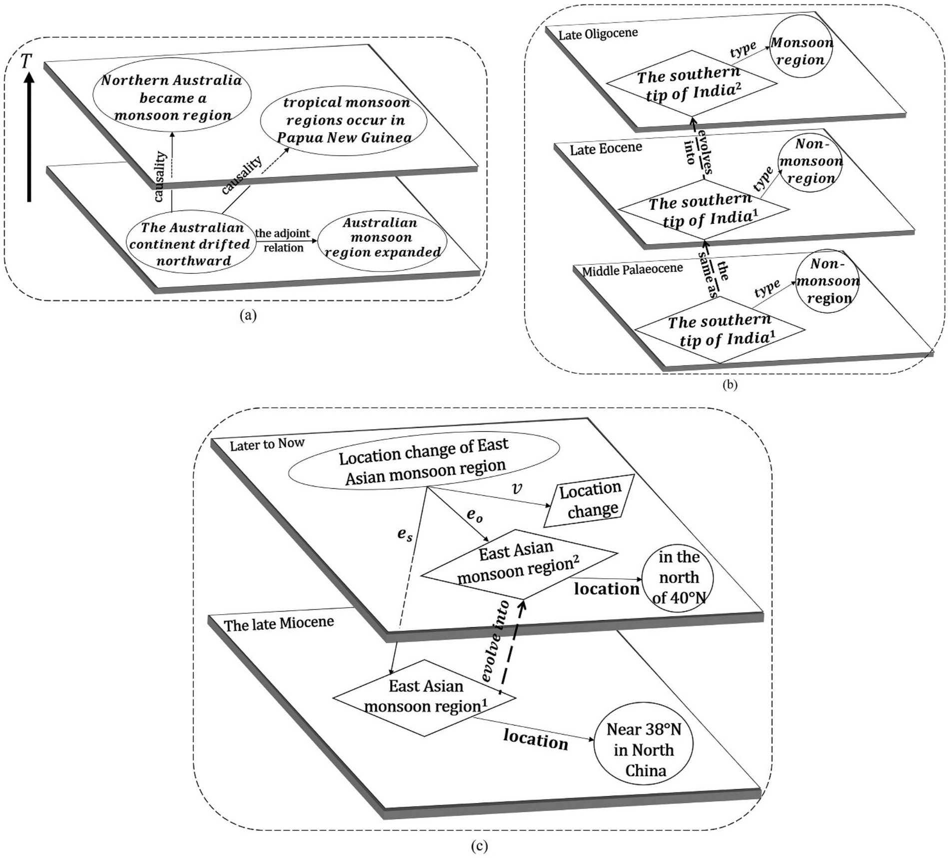

Figure 4. Associations between different layers. (a) Example of 'the association of geo-events' for the development of Australian monsoon region. (b) Example of 'the evolution of geo-entities' for the development of the southern tip of India. (c) Example of 'the association between geo-events and geo-entities' for the development of the East Asian monsoon region.

of 'location'. 'East Asian monsoon region[1]' and 'East Asian monsoon region[2]' are the entities before and after the change, respectively, and they have the evolutionary relation of 'evolves into'. Although the event involves two periods, the change actually occurred in the latter period. Therefore, it is placed in the latter time layer in Figure 4(c). In addition, the generation and extinction of entities can also be represented by Figure 4(c), which involves only one layer.

4. Experiments

4.1. Dataset description

The data in this study are from the Chinese text corpus. We crawled 56,897 Chinese texts regarding the formation and evolution of mountains, minerals, oceans and islands from online websites, such as Chinese Wikipedia, the Interactive Encyclopedia and the China National Knowledge Infrastructure (CNKI).

The Chinese text was preprocessed by the Language Technology Platform (LTP) (Che *et al.* 2010), such as segmentation, word segmentation, part-of-speech tagging and named entity recognition. On this basis, we adopted 'keyword and context awared relation extraction model' and 'event extraction model based on graph convolutional network and attentional mechanism' to extract relations and events. Finally, we built a database containing more than 60,000 events, 140,000 entities, 160,000 attributes, and 210,000 relations. According to the organizational form of GEKG, the thesis used the Neo4j graph database for graph storage and display.

4.2. The YAGO, GeoKG and GEKG

Yet Another Great Ontology (YAGO) and Geographic Knowledge Graph (GeoKG) (Wang *et al*. 2019) were compared with our GEKG. Taking 'the evolution of monsoon regions in Asia, Africa and Australia' as the example (Liu *et al*. 2019b), the relevant geographic knowledge was extracted and manually corrected. Three knowledge graphs were constructed, and then the structure and expression ability were analyzed. GeoKG is proposed to represent the changes of geo-entities. YAGO is a representative open source knowledge graph, with many items containing descriptions in temporal and spatial dimensions. Note that we compared our model with YAGO4, which is the latest version.

4.2.1. Structure comparison

In Figure 5, there are four kinds of elements: entity, attribute, time and relation. YAGO describes the attributes and relations, but it is not immutable. It can represent different properties of the entity at different times by the relations startDate and endDate. In Figure 6, GeoKG defines more variables, such as the location, time, attribute, state, change, and relation to describe entities. GeoKG clearly shows the changes in the time, attribute and location. In Figure 7, GEKG establishes the hierarchical cubical structure to represent the evolution. For each time layer, the elements are associated in the same way, reflecting the relations at the current time. The associations in different layers, include the

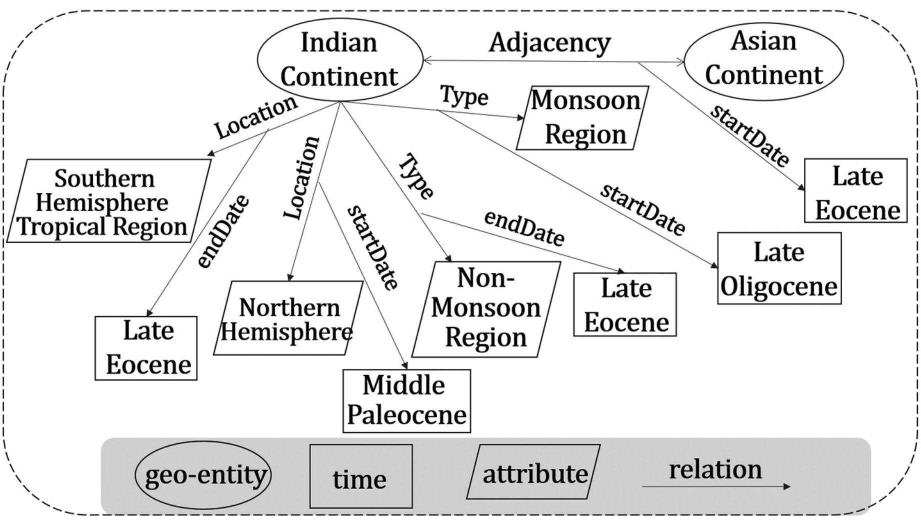

Figure 5. YAGO structure for the case of 'the evolution of the Indian Continent'.

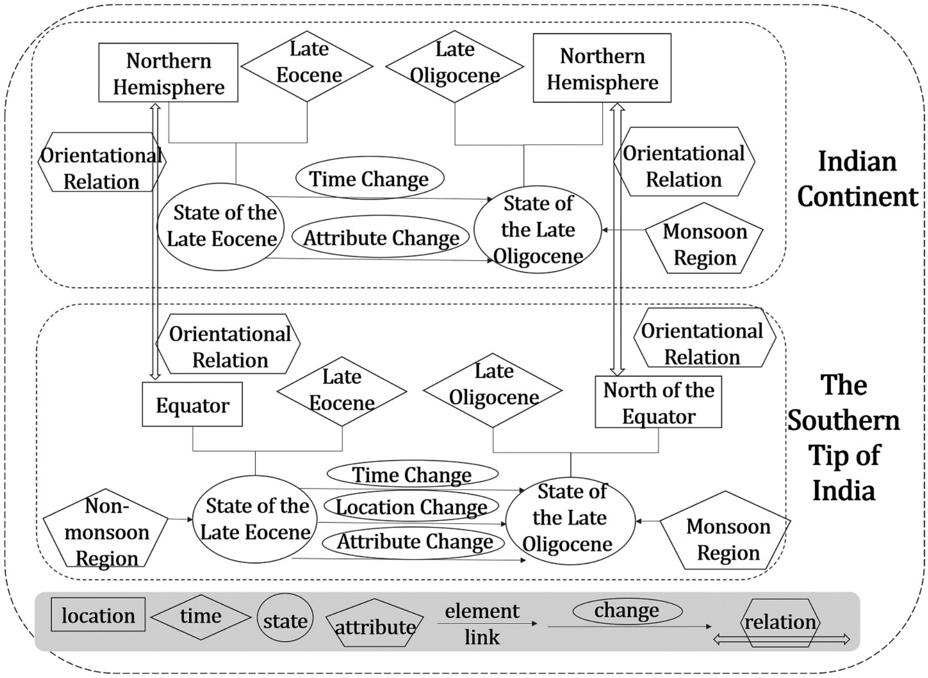

Figure 6. GeoKG structure for the case of 'the evolution of the Indian Continent and the Southern Tip of India'.

logical relation between geo-events, the evolutionary relation between geo-entities. Compared with the former, more evolutionary information is added to GEKG.

4.2.2. Comparison of knowledge expression abilities

According to the features of geographic knowledge that focuses more on the development and changes, the thesis sets up three kinds of questions namely, facts, changes and causes (Table 1). The first type of question is concerned with the detection of the representation of general knowledge, the second is concerned with the evolution, and the third with the representation of evolutionary reasons.

(1) Comparison of these questions. The Cypher language was used to carry out the graph retrieval task based on Neo4j. The questions were queried separately, and the results are shown in Table 2.

The accuracy, completeness and repetition of the three graphs is given in Table 3.

Table 1. Three types of questions for queries related to the development of monsoon regions.

Type	Questions	
Facts	Q1. Where is the South African monsoon region?	Q2. Where is the Indian Continent?
Changes	Q3. How has the type of the Indochina Peninsula changed?	Q4. How has the Indian Continent changed from Middle Paleocene to Late Oligocene?
Causes	Q5. Why has the Bay of Bengal developed into a monsoon region?	Q6. Why do tropical monsoon regions occur in Papua New Guinea?

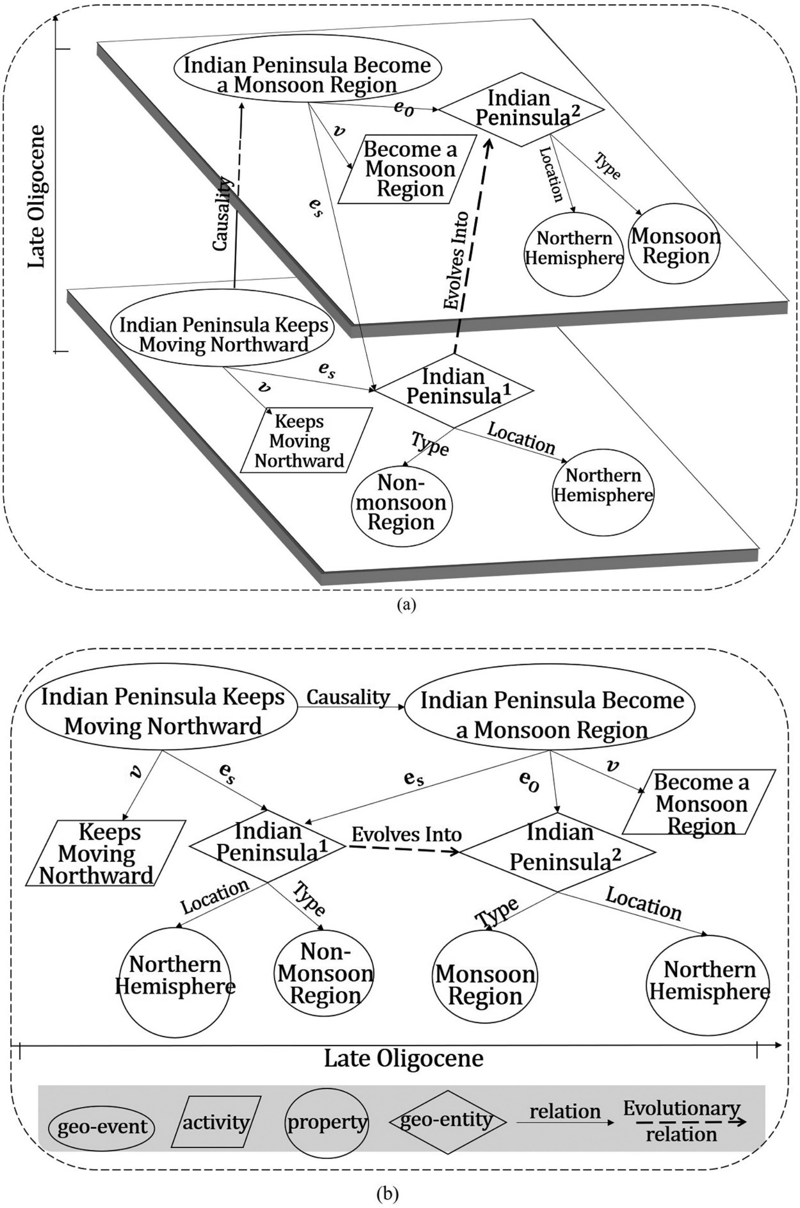

Figure 7. (a) GEKG structure for the case of 'the evolution of the Indian Peninsula'. (b) The planar representation of (a).

The accuracy of GEKG's results are better than the other two methods (Table 3). For the questions relating to facts, the results of the three graphs are completely accurate. Q3 and Q4 are about changes of geo-entities. YAGO describes only facts and cannot represent changes. In Q3 and Q4, YAGO's results are multiple property values. The results in different periods can indicate that there is a change, but they do not match the questions and are inaccurate. Due to the lack of representation of events, YAGO and GeoKG give no results for Q5 and Q6.

Table 2. The retrieval results of YAGO, GeoKG and GEKG for the questions posed in Table 1. The results are in bold and the following contents explain each answer. The contents of YAGO are triples to obtain the results. The contents of GeoKG are relevant objects and states. The explanation of GEKG is expressed in accordance with formulae (1) to (6).

Question	Results		
	YAGO	GeoKG	GEKG
Q1. Where is the South African monsoon region?	• **From the Equator to 20°S** < South African Monsoon Region, Location (startDate: Middle Paleocene), From the Equator to 20° S>	• **From the Equator to 20°S** (South African Monsoon Region, State of the Middle Paleocene) • **From the Equator to 20°S** (South African Monsoon Region, State of the Late Eocene) • **From the Equator to 20°S** (South African Monsoon Region, State of the Late Oligocene) • **From the Equator to 20°S** (South African Monsoon Region, State of the Late Miocene) • **From the Equator to 20°S** (South African Monsoon Region, State of the 'Later to Now')	• **From the Equator to 20°S** South African Monsoon Region[1] (Start-Time: Middle Paleocene, {Location: From the Equator to 20°S})
Q2. Where is the Indian Continent?	• **Northern Hemisphere** < Indian Continent, Location (startDate: Late Eocene), Northern Hemisphere> • **Southern Hemisphere Tropical Region** < Indian Continent, Location (endDate: Middle Paleocene), Southern Hemisphere Tropical Region >	• **Southern Hemisphere Tropical Region** (Indian Continent, State of the Middle Paleocene) • **Northern Hemisphere** (Indian Continent, State of the Late Eocene) • **Northern Hemisphere** (Indian Continent, State of the Late Oligocene) • **Northern Hemisphere** (Indian Continent, State of the Late Miocene) • **Northern Hemisphere** (Indian Continent, State of the 'Later to Now')	• **Southern Hemisphere Tropical Region** Indian Continent[1] (End-Time: Middle Paleocene, {Type: Nonmonsoon Region, Location: Southern Hemisphere Tropical Region}) • **Northern Hemisphere** Indian Continent[2] (Start-Time: Late Eocene, {Type: Nonmonsoon Region, Location: Northern Hemisphere}) • **Northern Hemisphere** Indian Continent[3] (Start-Time: Late Oligocene, {Type: Monsoon Region, Location: Northern Hemisphere})

(*Continued*)

Table 2. (Continued).

Question	Results		
	YAGO	GeoKG	GEKG
Q3. How has the type of the Indochina Peninsula changed?	• **Monsoon Region** < Indochina Peninsula, Type (startDate: Late Oligocene), Monsoon Region> • **Nonmonsoon Region** < Indochina Peninsula, Type (endDate: Late Eocene), Nonmonsoon Region>	• **< State of the Late Eocene, Attribute Change, State of the Late Oligocene>** (Indochina Peninsula, State of the Late Eocene, Nonmonsoon Region) &(Indochina Peninsula, State of the Late Oligocene, Monsoon Region)	• **< {Late Eocene & Late Oligocene}, Indochina Peninsula^{2}, Type of Change, Indochina Peninsula^{3}>** Indochina Peninsula^{2} (End-Time: Late Eocene, {Type: Nonmonsoon Region}) Indochina Peninsula^{3} (Start-Time: Late Oligocene, {Type: Monsoon Region})
Q4. How has the Indian Continent changed from the Middle Paleocene to the Late Oligocene?	–	• **< State of the Middle Paleocene, Location Change, State of the Late Eocene >** (Indian Continent, State of the Middle Paleocene, Southern Hemisphere Tropical Region) **&** (Indian Continent, State of the Late Eocene, Northern Hemisphere) • **< State of the Middle Paleocene, Time Change, State of the Late Eocene >** (Indian Continent, State of the Middle Paleocene) **&** (Indian Continent, State of the Late Eocene) • **< State of the Late Eocene, Attribute Change, State of the Late Oligocene>** (Indian Continent, State of the Late Eocene, Nonmonsoon Region) & (Indian Continent, State of the Late Oligocene, Monsoon Region) • **< State of the Late Eocene, Time Change, State of the Late Oligocene>** (Indian Continent, State of the Late Eocene) & (Indian Continent, State of the Late Oligocene)	• **<{Middle Paleocene & Late Eocene}, Indian Continent^{1}, Location Change, Indian Continent^{2}>** • **< Late Eocene, Indian Continent^{2}, Moves Northward>** • **< Late Eocene, Indian Continent^{2}, Contact With, Asian Continent >** • **<{Late Eocene & Late Oligocene}, Indian Continent^{2}, Type of Change, Indian Continent^{3}>** • <**Late Oligocene, Indian Continent^{3}, Keeps Moving Northward >** • <**Late Oligocene, Indian Continent^{3}, Becomes a Monsoon Region>**
Q5. Why has the Bay of Bengal developed into a monsoon region?	–	–	• **<Late Eocene, Indian Continent^{2}, Moves Northward>**
Q6. Why do tropical monsoon regions occur in Papua New Guinea?	–	–	• **< Late Miocene, Australian Mainland^{1}, Drifting North >**

Note that – means no answer. (2) Analysis of the results.

Table 3. Accuracies of YAGO, GeoKG and GEKG for the monsoon region example. It is determined according to the matching degree of the result and the correct answer. The percentage scores means how many words are included.

	Results		
Question	YAGO	GeoKG	GEKG
Q1	**100%**	**100%**	**100%**
Q2	**100%**	**100%**	**100%**
Q3	<60%	**≥ 80%**	**100%**
Q4	-	≥60%	**100%**
Q5	-	-	**100%**
Q6	-	-	**100%**

In terms of completeness, GEKG's results are better than those of YAGO and GeoKG. For the questions related to facts, the results of the three graphs are satisfactory. For the questions related to changes, YAGO's results do not match the questions. Although GeoKG can give partial results, they are not complete. For example, in Q4, the changes in the Indian Continent include not only its location and type. GEKG's results include new relations and related events, yielding a more complete representation.

Repetition is generally for questions related to facts. The results of GeoKG have more duplicates than the results of YAGO and GEKG. In Q1, the position of the South African monsoon region has not changed, so the repetition rates of YAGO and GEKG are both zero. Although GeoKG's results are in five different periods, they are the same, and the repeatability is much higher than for the other two. In Q2, the position of the Indian Continent changed from the Middle Paleocene to the Late Eocene. YAGO's results are still not redundant. GeoKG has stored four identical locations. The reason GEKG has duplicates is that it treats all the properties of the geo-entity as a whole state. The Indian Continent has the same position in the Late Eocene and Late Oligocene, but the former is a nonmonsoon region, and the latter is a monsoon region. The overall state of the Indian Continent in the two periods is different. Therefore, when the attribute changes, the current entity is different from the previous period.

Given a combination of the accuracy, completeness and repetition (Table 4), YAGO can answer only factual questions accurately, but it has low repeatability. GeoKG has high

Table 4. Comparison of YAGO, GeoKG and GEKG in terms of comprehensive ability. 'Strong' refers to accurate and complete answers with no duplication. 'Relatively strong' means the answer is correct and complete, but the expression is not clear enough or there are repetitions. 'Moderate' indicates that the result is incomplete or some components are inaccurate. 'Weak' means the result does not match the question.

	Results		
Question	YAGO	GeoKG	GEKG
Q1	Strong	Relatively strong	Strong
Q2	Strong	Relatively strong	Relatively strong
Q3	Weak	Relatively strong	Strong
Q4	-	Moderate	Strong
Q5	-	-	Strong
Q6	-	-	Strong

Table 5. The application and limitations of YAGO, GeoKG and GEKG.

Knowledge graph	Application scenarios	Limitations
YAGO	• Describe attributes and relations	Cannot represent changes
GeoKG	• Describe attributes and relations • Describe the evolution of geo-entity	High repetition and lower retrieval efficiency
GEKG	• Describe attributes and relations • Describe the evolution of geo-entity • Describe the evolution of geo-event • Describe the evolutionary reason	Simple reasoning is needed to obtain the attributes of the entity in the intermediate period

accuracy but also high repeatability. For the questions related to changes, GeoKG's results are not complete, so the comprehensive ability is moderate. GEKG gives accurate answers to all the questions and has special advantages for the representation of process. GEKG aims to represent evolutionary information, so it has low redundancy.

In summary, in terms of structure, GEKG highlights the evolution of geo-entities and geo-events over time, which contain more procedural knowledge than GeoKG and YAGO. GEKG has a better expression ability. For the questions related to facts, the results of the three graphs are accurate, but the results from GEKG are substantially better than those from GeoKG and YAGO for process-related questions. GEKG has an absolute advantage in representing evolutionary knowledge and is more complete than the other two graphs.

4.2.3. Discussion

According to the characteristics of the three graphs, their limitations and application scenarios are compared as follows (Table 5):

Take the question 'Q1: Where is the South African monsoon region?' as an example. The result of YAGO and GEKG is 'From the Equator to 20°S'. They can both represent static facts succinctly and efficiently. GeoKG stores a large number of duplicate items, which results lower retrieval efficiency. For the questions about changes (for example 'Q3: How has the type of the Indochina Peninsula changed?'), the results of YAGO are multiple values of the type. GeoKG and GEKG can directly represent how the type changes. For the questions about reasons, only GEKG has clear advantages. It represents evolutionary reasons by geo-events, and the other two have no results. In general, YAGO is more suitable for describing static facts. GeoKG stores different states of entities in different periods, which is more suitable for representing frequent changes of entities in a short time. For GEKG, it has an advantage in representing evolutionary processes and reasons. Sometimes, the retrieval of attributes requires simple reasoning, which has relatively little impact on performance.

5. Conclusion

This study presented a hierarchical cubic knowledge representation model to represent the geographic spatiotemporal process. The main innovations can be summarized as follows. First, based on the spatiotemporal features and evolutionary relations, the model presented a knowledge representation approach of geo-entities, which established the relations between geo-entities in different spatiotemporal states and could clearly represent their evolution. Second, the model provided a relation representation mechanism between geo-events and geo-entities, showing the complex associations and interactions between them. Third, the cubic model structure with a time hierarchy was established, which is helpful to represent the evolution of geographic elements over time.

Geographic KG was expanded to describe geographic evolutionary knowledge, and then the geographic evolutionary knowledge graph (GEKG) was generated. For GIScience, the GEKG provides a structured and computable knowledge representation model, which could help to discover hidden geographic knowledge and spatiotemporal patterns and reveal the evolutionary mechanisms of geographic elements and the components of geographic phenomena.

The experimental results show that GEKG has a strong expression ability to not only describe relations and properties but also represent the formation, development and change of monsoon regions. Overall, the representation model of GEKG is a structured knowledge organization approach based on geographic spatiotemporal processes. Compared with geographic KG, it provides important support for the representation of evolutionary knowledge, spatiotemporal analysis and reasoning based on geographic spatiotemporal processes.

Future work should still consider the following issues: the current knowledge extraction methods are not systematic and mature. For the GEKG model, it is necessary to improve the existing methods to form a relatively complete framework of automatic knowledge extraction based on process.

Data and codes availability statement

The data that support the findings of this study are openly available in [figshare] at [https://doi.org/10.6084/m9.figshare.13078061].

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Natural Science Foundation of China [No. 61501412,No. 61672474]; the National Key R&D Program of China [No. 2018YFB1004600]; Science and Technology Planning Project of Guangdong Province, China [No. 2018B020207012]; Science and National Science and Technology Major Project of China [No. 2017ZX05036-001-010].

Notes on contributors

Kun Zheng received his Ph.D. from China University of Geosciences in 2006. He is currently an associate professor with the School of Geography and Information Engineering, China University of Geosciences, Wuhan, China, and his current research interests involve spatio-temporal big data

analytics, visualization and high-performance computation. His contribution was study design and paper writing.

Ming Hui Xie received the B.S. degree from Wuhan University of Science and Technology in 2018, and the M. S. degree from China University of Geosciences (Wuhan) in 2021. Her main research direction is geographic knowledge graph. Her contribution was paper writing and data processing.

Juan Xie received the B.S. degree from Shandong Agricultural University in 2019. She is currently pursuing the M.S. degree in surveying and mapping engineering from China University of Geosciences (Wuhan), and her main research direction is spatio-temporal big data analysis. Her contribution to the paper was data acquisition and analysis.

Jin Biao Zhang received the B.S. degree from Nanjing University of Information Science and Technology in 1991. He is currently a professor-level senior engineer of Guangdong Meteorological Observation Data Center, and his research interests involve meteorological information technology and application. His contribution to the paper is sample analysis and data testing.

Shu Hao Xia received the B.S. degree in geographic information science from China University of Geosciences (Wuhan) in 2020, where he is currently pursuing the M.S. degree in cartography and geographic information engineering. His research interest is natural language processing. His contribution was data processing.

References

Cao, Y., *et al*., 2018. Geographic process modeling based on geographic ontology. *Open Geosciences*, 10 (1), 782–796. doi:10.1515/geo-2018-0061

Che, W., Li, Z., and Liu, T., 2010. LTP: a Chinese language technology platform. *COLING 2010, 23rd International Conference on Computational Linguistics*, 8, 23–27.

Chen, J., Deng, S., and Chen, H., 2017. CrowdGeoKG: crowdsourced geo-knowledge graph. *In*: *China conference on knowledge graph and semantic computing*. Singapore: Springer.

Deng, M., *et al*., 2018. Heterogeneous space–time artificial neural networks for space–time series prediction. *Transactions in GIS*, 22 (1), 183–201. doi:10.1111/tgis.12302

Deng, M., *et al*., 2020. Spatio-temporal analysis methods for multi-modal geographic big data. *Journal of Geo-information Science*, 22 (1), 41–56.

Ding, X., *et al*., 2019. ELG: an event logic graph. *arXiv Preprint*, arXiv:1907.08015.

Fan, R., *et al*., 2019. Deep learning-based named entity recognition and knowledge graph construction for geological hazards. *International Journal of Geo-Information*, 9 (1), 15. doi:10.3390/ijgi9010015

Gong, J., Geng, J., and Wu, H., 2014. Geospatial knowledge service: a review. *Geomatics Information Science of Wuhan University*, 39 (8), 883–890.

Huang, Y., *et al*., 2019. Using geographic ontologies and geo-characterization to represent geographic scenarios. *ISPRS International Journal of Geo-information*, 8 (12), 566. doi:10.3390/ijgi8120566

Jiang, B., *et al*., 2018. Geographic knowledge graph building extracted from multi-sourced heterogeneous data. *Acta Geodaetica et Cartographica Sinica*, 47 (8), 1051–1061.

Jin, W., *et al*., 2020. Recurrent event network for reasoning over temporal knowledge graphs. *In*: *Proceedings of the 2020 conference on empirical methods in natural language processing*, 6669–6683. online.

Leung, Y., *et al*., 2019. Integration of air pollution data collected by mobile sensors and ground-based stations to derive a spatiotemporal air pollution profile of a city. *International Journal of Geographical Information Science*, 33 (11), 1–23. doi:10.1080/13658816.2019.1633468

Li, J., Liu, R., and Xiong, R., 2017. A Chinese geographic knowledge base for GIR. *IEEE International Conference on Computational Science & Engineering & IEEE International Conference on Embedded & Ubiquitous Computing*, 1, 361–368.

Li, X., *et al*., 2012. Assimilating process context information of cellular automata into change detection for monitoring land use changes. *International Journal of Geographical Information Science*, 26 (9), 1667–1687. doi:10.1080/13658816.2011.643803

Li, Z., Ding, X., and Liu, T., 2018. Constructing narrative event evolutionary graph for script event prediction. *The International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, 4201–4207.

Lin, H., *et al*., 2018. Prospect of geo-knowledge engineering in the era of spatio-temporal big data. *Geomatics and Information Science of Wuhan University*, 43 (12), 452–458.

Liu, J., *et al*., 2019a. Evolving knowledge graphs. *In*: *IEEE conference on computer communications*, Paris, France, 2260–2268.

Liu, N., *et al*., 2020. Diagnostic analysis of heavy rainstorm induced by two typhoons with similar paths. *Meteorological Disaster Prevention*, 27 (4), 19–23.

Liu, X., *et al*., 2018. PatientEG dataset: bringing event graph model with temporal relations to electronic medical records. *arXiv Preprint*, arXiv:1812.09905.

Liu, X., *et al*., 2019b. Continental drift, plateau uplift, and the evolutions of monsoon and arid regions in Asia, Africa, and Australia during the Cenozoic. *Science China Earth Sciences*, 62 (7), 1053–1075. doi:10.1007/s11430-018-9337-8

Lopez-Pellicer, F.J., *et al*., 2012. An ontology for the representation of spatiotemporal jurisdictional domains in information retrieval systems. *International Journal of Geographical Information Science*, 26 (3–4), 579–597. doi:10.1080/13658816.2011.599811

Luo, W., *et al*., 2013. Principal tensor-based spatial-temporal structure linkage visualization. *Journal of Basic Science and Engineering*, 21 (2), 276–286.

Patel, H., Paraskevopoulos, P., and Renz, M., 2018. GeoTeGra: a system for the creation of knowledge graph based on social network data with geographical and temporal information. *In*: *International conference on advances in social networks analysis and mining (ASONAM)*, Barcelona, Spain, 617–620.

Shi, J.S., *et al*., 2019. Simulation and expression of radioactive pollutant dispersion process based on 3D grid. *Geomatics World*, 26 (2), 52–59.

Tasnim, M., *et al*., 2019. Summarizing entity temporal evolution in knowledge graphs. *In*: *Companion proceedings of the 2019 World Wide Web conference*, San Francisco, USA, 961–965.

Trivedi, R., *et al*., 2017. Know-evolve: deep temporal reasoning for dynamic knowledge graphs. *Proceedings of the 34th International Conference on Machine Learning*, 70, 3462–3471.

Wang, C., *et al*., 2018. Information extraction and knowledge graph construction from geoscience literature. *Computers&Geosciences*, 112, 112–120.

Wang, S., *et al*., 2019. Geographic knowledge graph (GeoKG): a formalized geographic knowledge representation. *ISPRS International Journal of Geo Information*, 8 (4), 184. doi:10.3390/ijgi8040184

Xiang, W. and Wang, B., 2020. A survey of Chinese event extraction research. *Computer Technology and Development*, 30 (2), 1–6.

Xu, J., *et al*., 2020. Enhancing dialog coherence with event graph grounded content planning. *In*: *Twenty-Ninth international joint conference on artificial intelligence and Seventeenth Pacific Rim international conference on artificial intelligence {IJCAI-PRICAI-20,* Yokohama, Japan, 3941–3947.

Xue, C., *et al*., 2019. A novel process-oriented graph storage for dynamic geographic phenomena. *International Journal of Geo-Information*, 8 (2), 100. doi:10.3390/ijgi8020100

Zheng, M., Jin, M., and Guo, F., 2013. Modeling and simulation of toxic gas dispersion in urban streets supported by GIS. *Geomatics and Information Science of Wuhan University*, 38 (8), 935–939.

Zhu, Y., *et al*., 2017. Intelligent learning for knowledge graph towards geological data. *Scientific Programming*, 2017, 13. doi:10.1155/2017/5072427

Zou, B., *et al*., 2016. An optimized spatial proximity model for fine particulate matter air pollution exposure assessment in areas of sparse monitoring. *International Journal of Geographical Information Science*, 30 (4), 727–747. doi:10.1080/13658816.2015.1095921