

Article

Geographic Knowledge Graph (GeoKG): A Formalized Geographic Knowledge Representation

Shu Wang ^{1,2,3}, Xueying Zhang ^{1,2,3,*}, Peng Ye ^{1,2,3}, Mi Du ^{1,2,3}, Yanxu Lu ^{1,2,3} and Haonan Xue ^{1,2,3}

¹ Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing 210023, China; 141301020@stu.njnu.edu.cn (S.W.); 161301027@stu.njnu.edu.cn (P.Y.); 171301021@stu.njnu.edu.cn (M.D.); 171302083@stu.njnu.edu.cn (Y.L.); 181302106@stu.njnu.edu.cn (H.X.)

² State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023, China

³ Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

* Correspondence: zhangxueying@njnu.edu.cn; Tel.: +86-138-5149-3100

Received: 25 February 2019; Accepted: 4 April 2019; Published: 8 April 2019



Abstract: Formalized knowledge representation is the foundation of Big Data computing, mining and visualization. Current knowledge representations regard information as items linked to relevant objects or concepts by tree or graph structures. However, geographic knowledge differs from general knowledge, which is more focused on temporal, spatial, and changing knowledge. Thus, discrete knowledge items are difficult to represent geographic states, evolutions, and mechanisms, e.g., the processes of a storm “{9:30-60 mm-precipitation}-{12:00-80 mm-precipitation}-...”. The underlying problem is the constructors of the logic foundation (ALC description language) of current geographic knowledge representations, which cannot provide these descriptions. To address this issue, this study designed a formalized geographic knowledge representation called GeoKG and supplemented the constructors of the ALC description language. Then, an evolution case of administrative divisions of Nanjing was represented with the GeoKG. In order to evaluate the capabilities of our formalized model, two knowledge graphs were constructed by using the GeoKG and the YAGO by using the administrative division case. Then, a set of geographic questions were defined and translated into queries. The query results have shown that GeoKG results are more accurate and complete than the YAGO’s with the enhancing state information. Additionally, the user evaluation verified these improvements, which indicates it is a promising powerful model for geographic knowledge representation.

Keywords: geographic knowledge representation; geographic knowledge graph; formalization; GeoKG

1. Introduction

Geographic knowledge consists of the product of geographic thinking and reasoning about the world's natural and human phenomena, which plays an important role in geographic studies and applications [1]. Nearly every geographer is trying to answer the question of “how to perceive, understand and organize geographic knowledge scientifically.” [2] Generally, geographic knowledge representation is a type of human expression of the real world that is of great importance to storage and computation [3]. Especially in the era of Big Data, well-structured geographic knowledge is a benefit to all kinds of geospatial applications, because formalization is the foundation of geospatial big data computing, mining, and visualization.

At present, the most popular knowledge representation is the knowledge graph. It organizes knowledge with a set of concepts, relations, and facts, which are associated by two types {entity, relation, entity} and {entity, attribute, attribute value} [4]. There are only three basic elements in knowledge graphs: the entity, relation, and attribute. These three elements can explicitly represent general information, such as “when did the Beijing storm occur on 21 July—9:30, 21 July”. However, geographic knowledge is more complicated than general knowledge. More processes and evolutions need to be answered, e.g., “what caused the 7·21 Beijing storm”, “how did it develop”, and “what were the effects of the 7·21 Beijing storm”. Entities, relations, and attributes cannot easily and directly answer these mechanics questions. For example, the geographic knowledge graph representation of the 7·21 Beijing storm is shown in Figure 1.

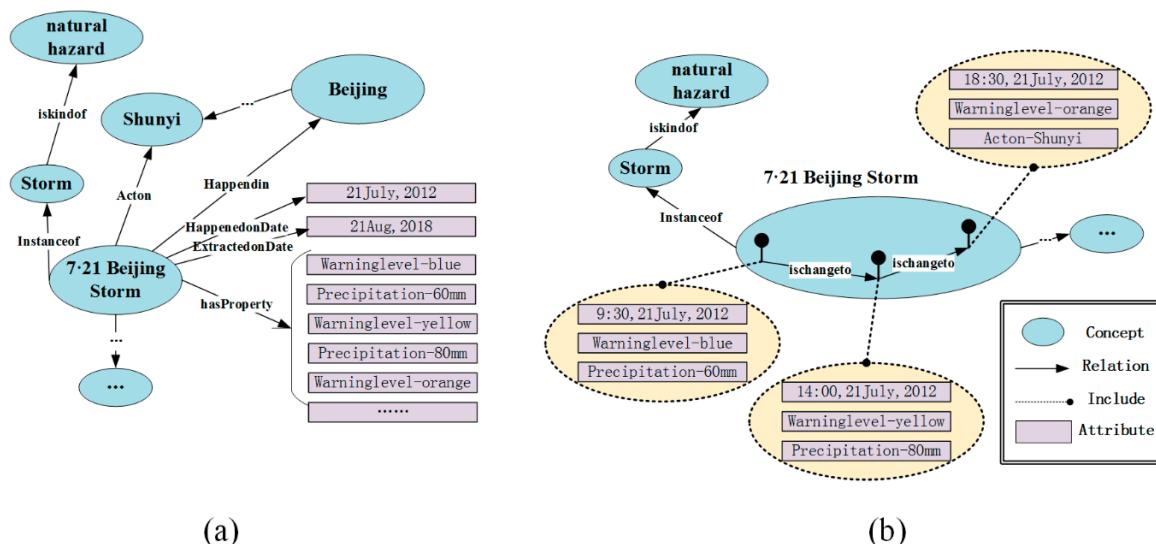


Figure 1. Different geographic knowledge representations of the 7·21 Beijing Storm. (a) Knowledge graph data structure and (b) procedural knowledge data structure.

Figure 1a organizes geographic knowledge of the 7·21 Beijing storm using the data structure of the current knowledge graph. This knowledge representation model can explicitly represent each fact and its relations. However, it is not able to represent evolutions or mechanisms, which are key topics in geography. Moreover, this type of knowledge representation differs greatly from procedural knowledge data structure shown in Figure 1b. In general, humans perceive objects, events, and activities through the processing of declarative knowledge, procedural knowledge, and structural knowledge [5]. And procedural knowledge gives the framework to the declarative knowledge state by state, which is benefit for underlying mechanism understanding [6,7]. The 7·21 Beijing storm includes three main stages: 9:30, 14:00, and 18:30. Each stage has a list of attributes. This procedural knowledge data structure helps people acknowledge the evolution or mechanism more explicitly. For example, people cannot directly understand that all the attributes (warning level-blue, warning level-yellow, etc.) link to “7·21 Beijing storm”, whereas people could know that the storm has different warning level on different stages.

The purpose of this paper is to improve the declarative discrete facts of knowledge graph to procedural aggregated knowledge. To address this issue, this paper presents a formalized model for geographic knowledge representation from a geography perspective, called GeoKG, and supplements the constructors of the ALC descriptive language.

The remainder of this paper is organized as follows: Section 2 reviews the related works on geographic ontology and geographic knowledge graph. Section 3 describes the methodology by stating the basic ideas from the six core geographical questions and proposes a formalized model of geographic knowledge representation called GeoKG. Section 4 gives an evolution case study of administrative

divisions of Nanjing with the formalized GeoKG model. Section 5 constructs the administrative division case by using the GeoKG model and the YAGO model, sets a series of questions, and analyzes the results. Finally, Section 6 presents the conclusions.

2. Related Works

There are two main representations of geographic knowledge for geospatial big data computing and reasoning: geographic ontology, and geographic knowledge graphs.

2.1. Geographic Ontology

Geographic ontology originates from ontology, which represents the most basic philosophical theories that represent the nature and characteristics of the real world [8]. In the 1960s, “ontology” was introduced in information science for categorization, representation, knowledge sharing, and reuse [9]. Geographic ontology is a domain concept, which explicitly and formally defines the geographic concepts and their relations within geography by hierarchical relations [10–12]. These hierarchical relations between concepts are significant to geographic knowledge representation, information integration, knowledge interoperation, and information retrieval. Thus, geographic ontology is an important geographic knowledge representation method that is widely implemented in various geographical information applications [13]. However, computer simulations not only require the standard hierarchical concept logic but also massive amounts of instance information in geographic knowledge representation. There are two types of typical geographic ontologies that are limited by the representations of geographic knowledge.

First, geographic ontology focuses on the structure of a conceptual system, which is built by strict hyponymy information [9]. These relationships are well suited for categorization, disambiguation, identification and inference but not for describing the states and phenomena of changing geographic objects [6,14]. The descriptions of these changing states and phenomena require copious information, which is lacking in geographic ontology [15]. Although hyponymy is strictly defined by hierarchical tree structures in geographic ontology, this structure cannot directly represent the relationships between multiple concepts that are important to represent evolutions and mechanisms in geography [11]. In addition, the relationships between vertices in a tree are not bi-directional, which limits the representation of the interactions between geographical objects. The cause of these problems is related to the tree structure limiting the representation of geographic knowledge [16].

Second, the logic foundation of geographic ontology is description logic (DL) of attributive concept language with complements (ALC) [17]. DL is an object-based formal knowledge representation language. It contains four components: a construction set represents concepts and roles (e.g., a river is a concept; disjunction is a role), assertion about a concept of terminology (terminology box, Tbox, e.g., each river has its own length), assertion about an individual item (assertion box, Abox, e.g., the length of the Changjiang River is 6300 km) and the reasoning mechanism of Tbox and Abox. DL can construct complicated concepts and roles with simple concepts and roles by constructors. According to the different constructors, DL can be classified as ALC, ALCN, S, SH, SHIQ, etc. ALC is the basic DL that contains intersections (\sqcap), unions (\sqcup), complements (\neg), universal restrictions (\forall), and existential restrictions (\exists). ALCN consists of basic ALC operators and number restrictions ($Q; \geq n$ and $\leq n$); ALC+ R^+ , short for S, consists of basic descriptions and enhancing relationship operators (R^+ ; role or concept transitions); the SH language, with concept inclusions and role inclusions (\sqsubseteq); and SHIQ includes inverse roles (I) and role transitions (R^+). At present, description logic SHIQ has been certified to represent changes in the field of logical theory [18]. Note that “change” is an absolutely essential element for geographic knowledge representation and means that the ALC constructors cannot represent all logical relations of geographic knowledge, especially in quantity expressions and state changes. For example, number restriction constructors are required to represent the geographic knowledge of “the Yangzi River has at least three branches” (\exists has a branch; the Yangzi River ≥ 3), and transition constructors are required to represent the geographic knowledge of “Beiping was renamed

Beijing on 27 September 1949” (Beiping ≡ Trans(Beijing)). Meanwhile, many studies theoretically demonstrated and proved the decidability, soundness and completeness of the operators of a series of DL (from ALC, S, SI, SHI to SHIQ, etc.) on the Tableau algorithm [19–22], and the complexity of SI (role or concept transitions) is PSPACE complete and the following SHI and SHIQ are EXPTIME complete [22,23].

2.2. Geographic Knowledge Graph

A knowledge graph is a graph-formed knowledge representation model with strict logic, different concepts, various relations, and massive instances [10]. It was first presented by Google in 2012, containing over 5.7 billion entities and 0.18 billion facts [24]. With this wealth of information, the real world can be explicitly described. Graph-based storage has properties of connection, direction, and multi-vertices that are suitable for representing the interactions between concepts. Thus, knowledge graphs are promising models to represent knowledge and have been widely built, e.g., YAGO [25], Freebase [26], Probbase [27], and DBpedia [28]. A geographic knowledge graph is a domain knowledge graph that is in the exploratory stage.

At present, most geographic knowledge graphs are organized as universal knowledge graphs, e.g., CSGKB [4], NCGKB [29], and CrowdGeoKG [15]. The common sense geographic knowledge base (CSGKB) uses a data structure that links the concepts of geographic features, geographic locations, spatial relationships and administrators for geographic information retrieval (GIR) instead of traditional gazetteers. Moreover, the naive Chinese geographic knowledge base (NCGKB) constructs a GIR-oriented geographic knowledge base based on Chinese Wikipedia based on given concept relations and their instances. CrowdGeoKG uses a crowdsourced geographic knowledge graph that extracts different types of geo-entities from OpenStreetMap and enriches them with human geography information from Wikidata. All of the concepts of these geographic knowledge graphs are developed based on geographic ontologies that follow the ALC descriptive language, resulting in the same problem as geographic ontology.

More importantly, three current bases organize the geographic knowledge as a set of concepts, relations, and facts, which are associated by two kinds of types {entity, relation, entity} and {entity, attribute, attribute value} [4]. Actually, there are only three basic elements in knowledge graph: entity, relation and attribute. These three elements can explicitly represent general information as “when did 7·21 Beijing storm—9:30, 21 July”. However, geographic knowledge is more complicated than general knowledge. More processes and evolutions need to be answered, e.g., “what causes the 7·21 Beijing storm”, “how did it develop”, and “what are the effects of the 7·21 Beijing storm”. Entities, relations, and attributes cannot easily and directly answer these mechanics questions.

Scholars indicated that more elements are required. PLUTO supplemented the element of time with “before” and “after” in the knowledge graph model to describe the change trajectories of geographic objects [30]. Geological knowledge graphs have been applied with the evolution element for stating changes between different geological objects [31]. YAGO also explored anchoring spatial and temporal dimensions to the knowledge base, called YAGO2 [7]. YAGO2 let time points and time intervals with standard format to describe the temporal information and set geographical coordinates associate to entities to complete their spatial information. In fact, these spatial and temporal knowledge stored in YAGO system are just regarded as general attributes by adding the predicates like “wasBornOnDate”, “occursSince”, “hasGeoCoordinates”, etc., whereas declarative discrete information cannot directly answer the proceeding questions, evolutions and mechanisms. Additionally, ten core concepts of geographic information sciences were proposed for transdisciplinary research: location, neighbourhood, field, object, network, event, granularity, accuracy, meaning, and value [23]. These concepts can cover every corner of geoscience, but they were extremely difficult to relate to one conceptualized model. More recently, six factors (geographic semantics, location, shape, evolutionary process, relationship between elements, and attribute) were proposed to describe information from geographic element, object, or phenomenon [22]. Though these factors were designed

for information representation of the geographic objects, they can also provide guidance for geographic knowledge representation. And all the above studies indicated that geographic knowledge can be represented more effectively by supplementing elements, whereas it also brings a foundation question: "how to organize geographic knowledge scientifically and cognitively?" Therefore, a conceptualized model of geographic knowledge graph from the geography perspective warrants further study.

3. Methodology

3.1. Basic Idea

3.1.1. Guiding Ideology

To address the aforementioned issues, the core question of the GeoKG model is to define the types of geographic knowledge that need to be stored. Geography (from the Greek γεωγραφία, geographia, literally "earth description") is a field of science devoted to the study of the lands, features, inhabitants, and phenomena of Earth [32]. As a type of human understanding of the geographic environment, geographic knowledge should answer questions about geography. The questions about geography have been separated into six core questions by the International Geographical Union (IGU), which is a part of the international charter on geographical education [2]. Therefore, GeoKG begins to define the basic elements and the conceptualized model using the six core questions in geography. Each question corresponds to one core issue:

- Where is it? →**space**
- What is it like? →**state**
- Why is it there? →**evolution**
- When and how did it happen? →**change**
- What impacts does it have? →**interaction**
- How should it be managed for the mutual benefit of humanity and the natural environment? →**usage**

3.1.2. Main Elements

These aforementioned questions can be used to describe six core aspects of geographic knowledge that should be represented by GeoKG. Each aspect requires some elements to describe them, and we try to find the basic elements among all aspects:

- Space →{**object**, location, time, relation, ... }
- State →{**object**, time, location, attribute ... }
- Evolution →{**object**, state, change, time, location, attribute, ... }
- Change →{**object**, time, location, attribute, relation, ... }
- Interaction →{**object**, relation, change, ... }
- Usage →{**object**, change, state, ... }

There are seven types of elements among the description of these six aspects: *object*, *location*, *time*, *attribute*, *relation*, *state*, and *change*. Three typical characteristics of these seven elements in describing the six aspects are as follows:

- **Object-centered representation.** All descriptions of the six aspects require geographic objects. Without **objects**, other elements are meaningless. Therefore, the six basic elements are formed around the object element.
- **Combined representation.** A description of a single basic element is just a statement. To represent these aspects in geography, the basic elements should be combined. Thus, all of the basic elements can be linked.

- **Stepped representation.** Note that the six aspects from the core geographical questions are not equal. Space and state focus on the static conditions of objects. Evolution and change pay more attentions to the dynamic conditions of objects. Moreover, interaction and usage rely on relationships and mechanisms between geographic objects. Thus, the basic elements cannot be treated as equals.

According to the three typical characteristics of the basic elements, we discovered that a geographic object is a type of media used to represent geographic knowledge. There are six basic elements used to describe geographic knowledge (see Figure 2). *Location*, *time*, *attribute*, *state*, *change*, and *relation* can co-efficiently represent geographic objects from different aspects. Note that these basic elements are not equivalent. *Location*, *time*, and *attribute* belong to the first level and represent a single static state of a geographic object. *State*, *change*, and *relation* describe the dynamic evolutions and relations to geographic objects.

- A geographic object is the core of geographic knowledge representation and is the minimum unit to perceive the world. The six basic elements (*location*, *time*, *attribute*, *state*, *change* and *relation*) represent geographic knowledge from different perspectives, which are linked to geographic objects.
- Static independent geographic objects can be described by elements of *location*, *time*, and *attribute*. Location shows the spatial patterns of geographic objects. Time gives the temporal dimension of geographic objects for human cognition. Attribute describes the static features of geographic objects.
- Any geographic object has an entire life cycle, including stages of generation, change, evolution and extinction. Different stages in the life cycle represent different *states*. States are represented by sets of attributes of geographic objects under a particular spatial-temporal dimension.
- Geographic objects are not always static. Any change in other elements of a geographic object will turn a state to another state or a relation to another relation. Thus, *change* is an essential part of geographic knowledge representation.
- Geographic objects are not isolated. Any scene, phenomenon, and environment consists of many geographic objects and complex relations between them. Thus, *relation* is the key descriptor of the interactions among complex geographic objects.

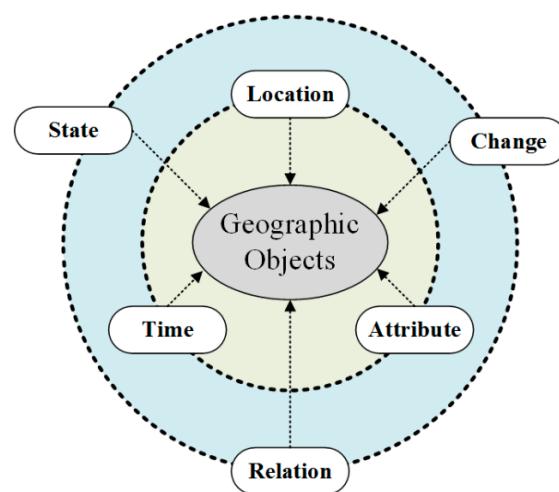


Figure 2. The six basic elements to represent a geographic object.

3.1.3. GeoKG Model

A conceptualized model of GeoKG is shown in Figure 3, which is based on the ideas mentioned above. The six core elements represent geographic objects and their information together. In this model, geographic objects consist of a series of states. Any state of a geographic object is represented

by attributes under a specific spatial-temporal condition. Any two continuous states or different states between two geographic objects could result in a change element. The change element can be categorized into time changes, location changes or attribute changes. If the essential attribute is changed, the geographic object will become another geographic object. The relation element exists between any time, location, and attribute of different states, regardless of whether they are the same geographic objects or not.

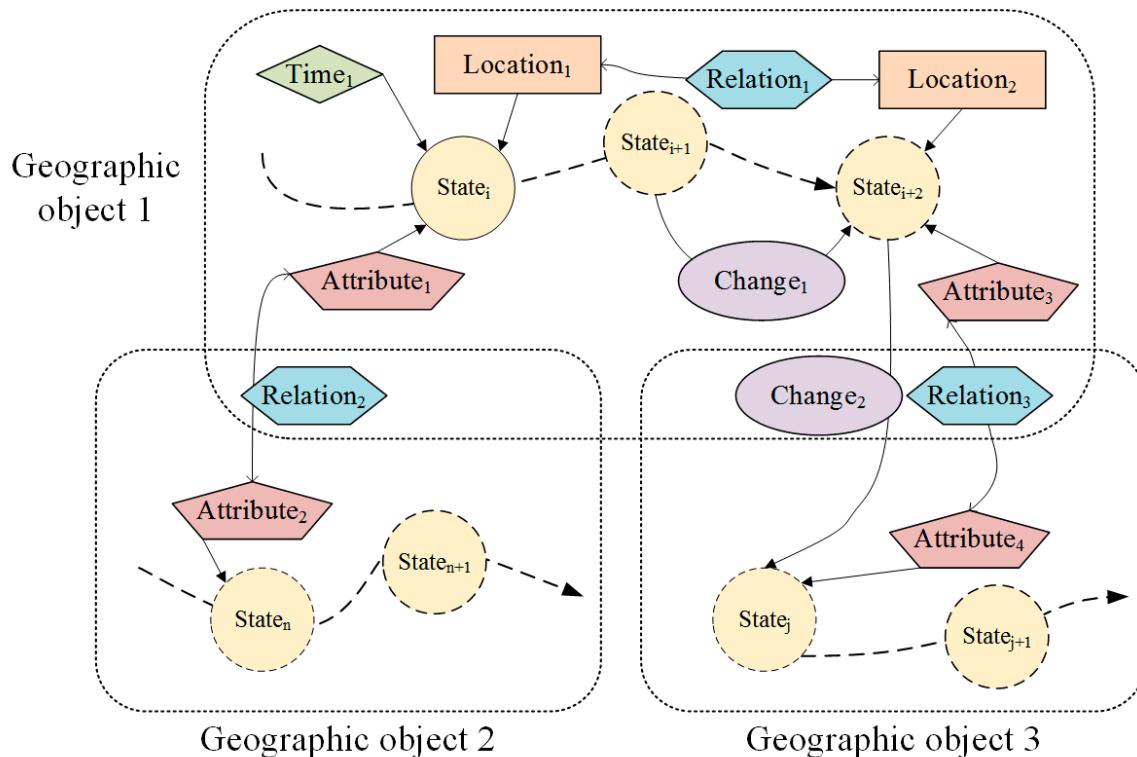


Figure 3. A conceptualized model of GeoKG based on the six basic elements.

3.2. Model Formalization

To organize geographic knowledge in consideration of the basic ideas, GeoKG must be based on a thorough and formalized model. This section provides the model semantics of GeoKG by using description logic (DL), which is not limited to only the attribute language complement (ALC) level. Using description logic, a user can create a conceptual description for the representation and computation of geographic knowledge that is clear and formal.

3.2.1. DL and Construction Operators

DL is comprised of three basic components: concepts, individuals (instances), and roles. Concepts describe the common features about individual sets, e.g., all land mass that projects well above its surroundings forms the concept of “mountain”. Individuals are the instances of concepts, e.g., a geographic entity, such as the “Rocky Mountains”. Roles can be explained as the binary relation between individuals as properties, e.g., spatial relation (conjunction, disjunction). A description logic system contains four parts. These parts include a construction set, which represent concepts and roles, an assertion about concept terminology (terminology box, Tbox), an assertion about an individual (assertion box, Abox) and the reasoning mechanism of Tbox and Abox. Tbox are sets containing the definitions of the relationship of concepts and the axiom of relationships, which contain explanations of the concepts and roles. Abox includes axiom sets describing specific situations, which contain the instance information of Tbox. Abox include two forms. One is the concept assertion, which expresses whether an object belongs to a concept. The other one is the relation assertion, which express

whether two objects satisfy a certain relation. Description logic can represent complicated concepts and relations on atomic concepts and atomic relations based on the given construction operator. The basic construction operators are and (\sqcap), or (\sqcup), not (\neg), existential quantifier (\exists), and universal quantifier (\forall), which are included in ALC DL. More operators can represent more logic, which form different types of DL.

Let C and D be concepts; a, b, and c, individuals; and R is a role between individuals. S is a simple role, and n is a nonnegative integer. As usual, an interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consists of a non-empty set $\Delta^{\mathcal{I}}$, called the domain of \mathcal{I} , and a valuation $\cdot^{\mathcal{I}}$, which associates, with each role R, a binary relation $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. For comprehensive background reading, please refer to the referenced paper [20]. The primary operators that differ from DL are shown in Table 1.

Diagrams are supplied to illustrate the graphic meanings of the operators related to geographic objects and their relationships. A top concept indicates all concepts or objects, e.g., \top River means all the rivers. A bottom concept indicates no concepts or objects in the set, e.g., \perp River means there are no rivers in the set. An atomic concept indicates the minimize concept, e.g., Ac could be the river, ocean, city, or country. An atomic role indicates the relationships between two atomic concepts, e.g., $R \subseteq \text{river} \times \text{ocean}$ means that there exists a relationship between the river and ocean. A conjunction indicates two individuals that are joint or connected, e.g., Yangzi River \sqcap Nanjing indicates a joint part of the Yangzi River and Nanjing. A disjunction indicates the logic disjunction of two individuals, e.g., Yangzi River \sqcup Nanjing means the combination set of the Yangzi River and Nanjing. A negation indicates the set of all individuals not in the target individual, e.g., \neg Yangzi River means all individuals except the Yangzi River. An exist restriction indicates the existence of an individual or a role, e.g., \exists Yangzi River means there exists a Yangzi River and $\exists R \subseteq \text{Yangzi River} \times \text{Zhong Mountains}$ means there exists a role between the Yangzi River and Zhong Mountains. A value restriction indicates all individuals or roles, e.g., \forall River means all rivers and $\forall R \subseteq \text{Yangzi River} \times \text{Zhong Mountains}$ means all roles between the Yangzi River and Zhong Mountains. A concept inclusion indicates a concept belonging to another concept, e.g., rain \sqsubseteq precipitation means rain is a kind of precipitation. A role inclusion indicates a role belonging to a role set, e.g., $R_{\text{location_Yangzi River-Zhong Mountain}} \sqsubseteq \text{Yangzi River} \times \text{Zhong Mountains}$ indicates that the location relation between the Yangzi River and Zhong Mountains is one of the roles of the entire role set of the Yangzi River and Zhong Mountains. An inverse role indicates that a role has reversibility. A trans role indicates that a role has transmissibility. A qualifying at least/at most restriction indicates there exists at least or at most, e.g., $(\exists \geq 3 \text{ rivers}) \subseteq \text{Yangzi River}$ means the Yangzi River has at least three branches.

Table 1. Syntax and semantics of the main construction operators of the description logic.

| Category (Symbol) | Construction Operators | Syntax | Semantics | Diagrams | Category (Symbol) | Construction Operators | Syntax | Semantics | Diagrams |
|-------------------|------------------------|---------------|--|----------|-------------------|---------------------------------|-----------------------|---|----------|
| ALC | Top concept | \top | Δ^I | | H | Value restriction | $\forall R.C$ | $\{a \in C^I \forall y, (a, b) \in R^I \& b \in C^I\}$ | |
| | Bottom concept | \perp | \emptyset | | | Concept inclusion | $C_1 \sqsubseteq C_2$ | $C_1^I \subseteq C_2^I$ | |
| | Atomic concept | Ac | $Ac^I \subseteq \Delta^I$ | | | Role inclusion | $R \sqsubseteq S$ | $R^I \subseteq S^I$ | |
| | Atomic role | R | $R^I \subseteq \Delta^I \times \Delta^I$ | | | Inverse role | R^- | $\{(a, b) \in R^I (b, a) \in R^I\}$ | |
| | Conjunction | $C \sqcap D$ | $C_1^I \cap C_2^I$ | | | Trans role | Trans(R) | $\{(a, c) \in R^I \exists (a, b) \in R^I \wedge (b, c) \in R^I\}$ | |
| | Disjunction | $C \sqcup D$ | $C_1^I \cup C_2^I$ | | | Qualifying at least restriction | $R.C \geq n$ | $\{a \in C^I \#(\{b (a, b) \in R^I \& b \in C^I\}) \geq n\}$ | |
| | Negation | $\neg C$ | $\Delta^I \setminus C^I$ | | | Qualifying at most restriction | $R.C \leq n$ | $\{a \in C^I \#(\{b (a, b) \in R^I \& b \in C^I\}) \leq n\}$ | |
| | Exist restriction | $\exists R.C$ | $\{a \in C^I \exists y, (a, b) \in R^I \& b \in C^I\}$ | | | | | | |

means the number of. Note the decidability, soundness, and completeness of all these operators have been demonstrated [6,23].

3.2.2. Formalization Representation

In this section, the semantics of the GeoKG model are defined. First, we prescribe the set of geographic knowledge GK sourced from the entire world's natural and human phenomena W. GeoGK is a set of GK that can be defined as follows:

$$\text{GeoGK} = \{\langle \text{GK} \rangle | \text{GK} \in W\}$$

GK is a tuple that consists of geographic object O and its basic elements E:

$$\text{GK} = \{\langle O, E \rangle | \exists O \neq \emptyset, \exists E \neq \emptyset\}$$

The basic element set E contains six different elements: *location L*, *time T*, *attribute A*, *state St*, *change Ch* and *relation Re*. Thus, E is a six-tuple:

$$E = \{\langle L, T, A, St, Ch, Re \rangle | \exists L \sqsubseteq T \sqsubseteq A \sqsubseteq St \sqsubseteq Ch \sqsubseteq Re \neq \emptyset\}$$

Each element is identified as follows:

(1) Time

Time describes the temporal information of the state of a geographic object. Let St_i indicate a specific state of geographic object O_i ; the basic element *time T* can be defined as follows:

$$T = \{\exists T \in St_i | \forall O_i \neq \emptyset, St_i \in O_i\}$$

Time should be described by both the basic types and reference time information. The basic types are point time, interval time and reference time. Point time T_{poi} records the moment of the state of a geographic object. Interval time T_{int} indicates the time interval between two point times. Reference time T_{ref} indicates the time of other elements of a geographic object, e.g., “2018 World Cup” is an event with a unique time period that could reference the specific time accurately. Time reference knowledge $tref$ indicates the additional knowledge of time descriptions. Let tw indicate the time word. A time word indicates a point time that could contain several time descriptive parts, e.g., 12-July-2018, ten past nine and tomorrow morning. The point time T_{poi} , the interval time T_{int} and the reference time T_{ref} are defined as follows:

$$T_{poi} = \{\langle tw, tref \rangle | \forall tw \in T\}$$

$$T_{int} = \{\langle tw, tref \rangle | \forall tw \in T, \#tw \geq 2, \forall R \sqsubseteq tw\}$$

$$T_{ref} = \{\langle E, tref \rangle | \forall E \& \forall T \sqsubseteq St_i\}$$

where R is the interval relation of two time words. Time reference knowledge $tref$ is a set of reference knowledge consisting of commonality, relativity, fuzziness, continuity, and periodicity, namely, $tref = \{\langle com, rel, fuz, con, per \rangle\}$. There are some examples for each reference time word. For example, “12-July-2018” is a common time, and the Late Jurassic is a domain time description. Relativity indicates whether time is relative, e.g., “two days ago” is a relative time that refers to the absolute time “today”, “9 o'clock” is an accurate time, “around 9” is a fuzzy time, “12-July” is an instance time, and “until 12, July” is a continuous time. Periodicity can be easily understood, such as “every weekend”, “every month”, and “annually”.

(2) Location

Location describes the spatial information of the state of a geographic object. Let St_i indicate a specific state of geographic object O_i ; the basic element *location L* can be identified as follows:

$$L = \{\exists L \in St_i | \forall O_i \neq \emptyset, St_i \in O_i\}$$

According to the complexity of location descriptions, a location can be set into basic types and reference location information. The basic types include toponym, address, coordinates, and reference location. Toponym L_{top} describes a location with a common name. Address L_{add} indicates a location with orderly numbers and streets named by administrators. Coordinate L_{coo} records the location with a series of numbers organized mathematically. Reference location L_{ref} indicates the location of other elements in a geographic object. Location reference knowledge $lref$ indicates the additional knowledge of location descriptions. Let tp , ad , co indicate toponym, address, and coordinate, respectively. Toponym L_{top} , address L_{add} , coordinates L_{coo} , and reference location L_{ref} are identified as follows:

$$\begin{aligned} L_{top} &= \{\langle tp, lref \rangle \mid \forall tp \in L\} \\ L_{add} &= \{\langle ad, lref \rangle \mid \forall ad \in L\} \\ L_{coo} &= \{\langle co, lref \rangle \mid \forall co \in L\} \\ L_{ref} &= \{\langle E, lref \rangle \mid \forall E \text{ & } \forall L \sqsubseteq St_i\} \end{aligned}$$

Location reference knowledge $lref$ is a set of reference knowledge consisting of the space type, spatial reference, commonality, relativity, and fuzziness, namely, $lref = \{\langle typ, ref, com, rel, fuz \rangle\}$. Space type describes what types of space, such as reality, virtual or a specific domain location. For example, Pandora is a toponym of the virtual world of the movie Avatar. Spatial reference illustrates the system of a location description, e.g., WGS84 and Mercator projection. Commonality stores whether a location is a domain location, e.g., Beijing is a common toponym that could be coded as “-.-.-.-.-.-/-.-.-.-.-.” in a Morse code system. Relativity indicates whether a location is relative, e.g., “20 km south of Beijing” is a relative location description that refers to the absolute location “Beijing”. Fuzziness states whether a location description is accurate or not, e.g., “near Times Square” is a fuzzy location description.

(3) Attribute

An attribute describes the feature information of the state of a geographic object. Let St_i indicate a specific state of geographic object O_i ; the basic element *attribute A* can be identified as follows:

$$A = \{\exists A \in St_i \mid \forall O_i \neq \emptyset, St_i \in O_i\}$$

All the feature descriptions of a geographic object belong to an attribute, e.g., shape, color, speed, etc. To organize the attributes of a geographic object, identifying what is an attribute is key. An attribute is a single feature description of one geographic object. For example, “a typhoon is a mature tropical cyclone that develops between 180° and 100° E in the Northern Hemisphere, with peak months from August to October” describes three attributes: the typical attribute of “mature tropical cyclone”, the location attribute of “develops between 180° and 100° E in the Northern Hemisphere” and the frequency attribute of “peak months from August to October”. It is noted that attribute can be divided into two types: essential attribute A_{es} and non-essential attribute A_{ne} :

$$A_{es} = \{\exists A_{es} \in St_i \mid \forall O_i \neq \emptyset, St_i \in O_i, \#A_{es} \geq 1\}$$

$$A_{ne} = \{A_{ne} \in A' \mid \forall O_i \neq \emptyset, St_i \in O_i, A' = A_{es}^{-}\}$$

An essential attribute is a mark attribute that identifies a geographic object from others. When an essential attribute changes, a geographic object could change to another object. For example, when a mature tropical cyclone develops in the Atlantic Ocean, it cannot be a typhoon. A non-essential attribute is another feature description of a geographic object, e.g., the frequency attribute of a typhoon is “peak months from August to October”. These attributes cannot determine the nature of a geographic object.

(4) State

The state illustrates the different stages of a geographic object. It can be seen that the above three basic elements work together to express the state. Thus, the element *state St* can be identified as follows:

$$St = \{\exists St_i \in O | \exists! L \sqsubseteq St_i, \exists! T \sqsubseteq St_i, \exists A \sqsubseteq St_i, \#A \geq 0\}$$

where $\exists!$ means the unique existence. The formulation means that the state is a part of a geographic object. As the element *state* St is represented by sets of attributes of geographic objects under a particular spatial-temporal dimension, it must depend on the element *location* L and the element *Time*. Note that the element *location* L and the element *Time* T exist uniquely, because of time and space are two dimensions to represent the stage in Euclidean space. For example, the state of a typhoon includes all features for a specific spatial-temporal reference frame, e.g., “Typhoon Maria, 23:00/10July-2018, E123.40°/N25.60°, central pressure 945 hpa, max speed 30 km/h”. The state cannot be defined without the temporal and spatial information. By contrast, the element *state* St does not depend on the element *Attribute* A . The attributes are the descriptive records that cannot affect whether the state exists. For example, “Typhoon Maria, 23:00/10July-2018, E123.40°/N25.60°” also defines a state of Typhoon Maria. Thus, the attribute element is defined different from location element and time element.

(5) Change

A change describes the changes in a geographic object from one state to another. Thus, change Ch must contain at least one difference between two states, which can be a location change, time change or attribute change. A change contains four main components:

$$Ch = \{\langle St, act, CE, type \rangle \in O | \exists St, \#St = 2, CE \in \{T, L, A\}, type \in (Ch_d, Ch_e)\}$$

where St indicates the state (including two different ones), act indicate the action of the change, CE indicate change elements and $type$ indicates the type of the change. It is noted that there are two types of changes: a developing change and an evolving change. A developing change shows the changes from one geographic object, and an evolving change describes the changes between two different geographic objects. Let Ch_d indicate a developing change and Ch_e indicate an evolving change; the formalized definitions are as follows:

$$Ch_d = \{\exists Ch_d = St_i \times St_{i+1} | \exists St_i \& St_{i+1} \in O_m, St_i \neq St_{i+1}\}$$

$$Ch_e = \{\exists Ch_e = St_{end} \times St_i | \exists St_{end} \in O_m, \exists St_i \in O_n, \exists! St_{end}.A_{es} \neq St_i.A_{es}\}$$

where O , O_m , and O_n are geographic objects, St_i and St_{i+1} indicate the continuous states of the geographic objects, St_{end} indicates the last state of the geographic objects, and A_{es} indicate the essential attribute of the geographic objects.

(6) Relation

A relation expresses the differences between the elements of geographic objects, which includes three typical types: location relation, time relation, and attribute relation. These three types describe the spatial difference, temporal difference and feature difference, respectively. A relation contains three main components: the elements of two states E , the semantic of the relation Sem , and the type of the relation $type$:

$$Re = \{\langle E, Sem, type \rangle \in O | \exists E \& \#E \geq 2, type \in (Re_l, Re_t, Re_a)\}$$

Let Re_l , Re_t , and Re_a indicate location relation, time relation, and attribute relation, respectively, L_i and L_j indicate the locations of different states, T_i and T_j indicate the times of different states, and A_i and A_j indicate the attributes of different states. The different types of relations are identified as follows:

$$Re_l = \{\exists Re_l = L_i \times L_j | \exists St_i \& St_j, St_i \neq St_{i+1}\}$$

$$Re_t = \{\exists Re_t = T_i \times T_j | \exists St_i \& St_j, St_i \neq St_{i+1}\}$$

$$Re_a = \{\exists Re_a = A_i \times A_j | \exists St_i \& St_j, St_i \neq St_{i+1}\}$$

A location relation describes the spatial relationships between different states, e.g., the location relations between the different states of a typhoon or the location relations between two different city centres under development. A time relation illustrates the temporal relationships between different states, i.e., the time span between two states, e.g., the time span of river diversion. An attribute relation describes the feature relationships between different states, i.e., the differences between two states of a typhoon, e.g., the max wind speed, central pressure, etc.

4. Case Study

In this section, a full example is shown to illustrate the geographic knowledge representation using the GeoKG model. To describe the geographic knowledge representation clearly, an evolution case of administrative divisions of Nanjing was selected. The given example includes the basic geographic objects (e.g., Yangzi River, Zhongshan Mountain), the changing area of Nanjing, and several affiliated districts in different eras.

4.1. Research Area

Nanjing, formerly romanized as Nanking and Nankin, is the capital of Jiangsu province of the People's Republic of China and the second largest city in the East China region, with an administrative area over 6000 km². The inner area of Nanjing enclosed by the city wall is Nanjing Centre District, with an area of 55 km², while the Nanjing Metropolitan Region includes surrounding cities and areas. Three representative stages were chosen to represent the revolution of Nanjing: 1368, 1949, and 2018. The sketch maps were shown in Figure 4.

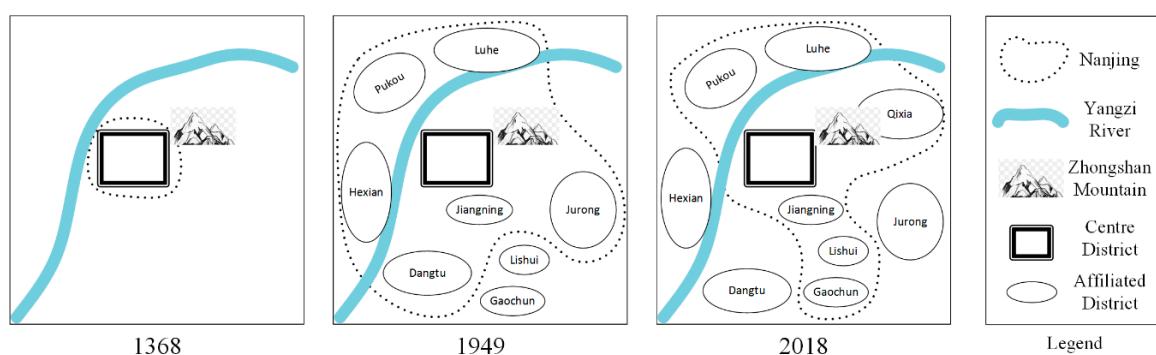


Figure 4. The sketch maps of administrative divisions evolution of Nanjing in 1368, 1949, and 2018.

The first stage is Ming dynasty, which firstly named this city in the word of "Nanjing". The first emperor of the Ming dynasty, Zhu Yuanzhang, who overthrew the Yuan dynasty, renamed the city of Nanjing, rebuilt it, and made it the dynastic capital in 1368. He constructed a 48 km long city wall around Nanjing. That is the centre district of Nanjing, which is situated in the south of the Yangzi River and to the west of the Zhongshan Mountain.

The second stage is the founding of the People's Republic of China. The government set Nanjing as a province unit, which directly controlled by the government. At that stage, Nanjing administrated the centre district and several affiliated districts. The centre district included districts 1–10 and affiliated districts involved Jiangning, Jurong, Dangtu, Hexian, Pukou, and Luhe. In 1949, Nanjing had been expanded through Yangzi River and Zhongshan Mountain.

The third stage is 2018, which refers to the current administrative boundaries of Nanjing. After a series of administrative division adjustments, Gaochun and Lishui was supplemented into Nanjing and Jurong, Dangtu, and Hexian was removed from the boundaries.

During over 600 years development of Nanjing, numerous elements were changed including the boundaries, affiliated districts, the relations between Nanjing and other geographic objects (e.g., Yangzi River and Zhongshan Mountain). Different relations happened in different stages among

these geographic objects. Thus, the GeoKG model was used to represent these changing geographic knowledge. The formalization is introduced in the next section.

4.2. Formalization

In this example, administrative division evolution was organized by using the GeoKG model. A geographic object is the key to represent geographic knowledge. First, this case identifies six relevant geographic objects: Nanjing O_{nj} , Yangzi River O_{yr} , Zhongshan Mountain O_{zm} , Centre District O_{cd} , Jiangning O_{jn} , and Gaochun O_{gc} . Jiangning and Gaochun are representative affiliated districts which were selected in this case. Jiangning is always been part of Nanjing in 1949 and 2018 and Gaochun has an administrative division adjustment. Each geographic object consists of a series of states, changes and relations. For example, Nanjing O_{nj} contains three states $S_{nj} = \{S_{nj1}, S_{nj2}, S_{nj3}\}$, six changes $C_{nj} = \{C_{nj11}, C_{nj12}, C_{nj13}, C_{nj21}, C_{nj22}, C_{nj23}\}$, and 12 relations $R_{nj} = \{R_{nj11}, R_{nj12}, R_{nj13}, R_{nj21}, R_{nj22}, R_{nj23}, R_{nj31}, R_{nj32}, R_{nj33}, R_{nj34}, R_{nj35}\}$. Thus, Nanjing O_{nj} can be defined as follow and the corresponding diagram is shown in Figure 5.

$$O_{nj} = \left\{ \begin{array}{l} S_{nj} \sqsubseteq O_{nj}, C_{nj} \sqsubseteq O_{nj}, R_{nj} \sqsubseteq O_{nj} \\ S_{nj} = \{S_{nj1}, S_{nj2}, S_{nj3} | S_{nj}.number \leq 3, S_{nj}.number \geq 3\}, \\ C_{nj} = \{C_{nj11}, C_{nj12}, C_{nj13}, C_{nj21}, C_{nj22}, C_{nj23} | C_{nj}.number \leq 6, C_{nj}.number \geq 6\} \\ R_{nj} = \{R_{nj11}, R_{nj12}, R_{nj13}, R_{nj21}, R_{nj22}, R_{nj23}, R_{nj31}, R_{nj32}, R_{nj33}, R_{nj34}, R_{nj35} | R_{nj}.number \leq 12, R_{nj}.number \geq 12\} \end{array} \right\}$$

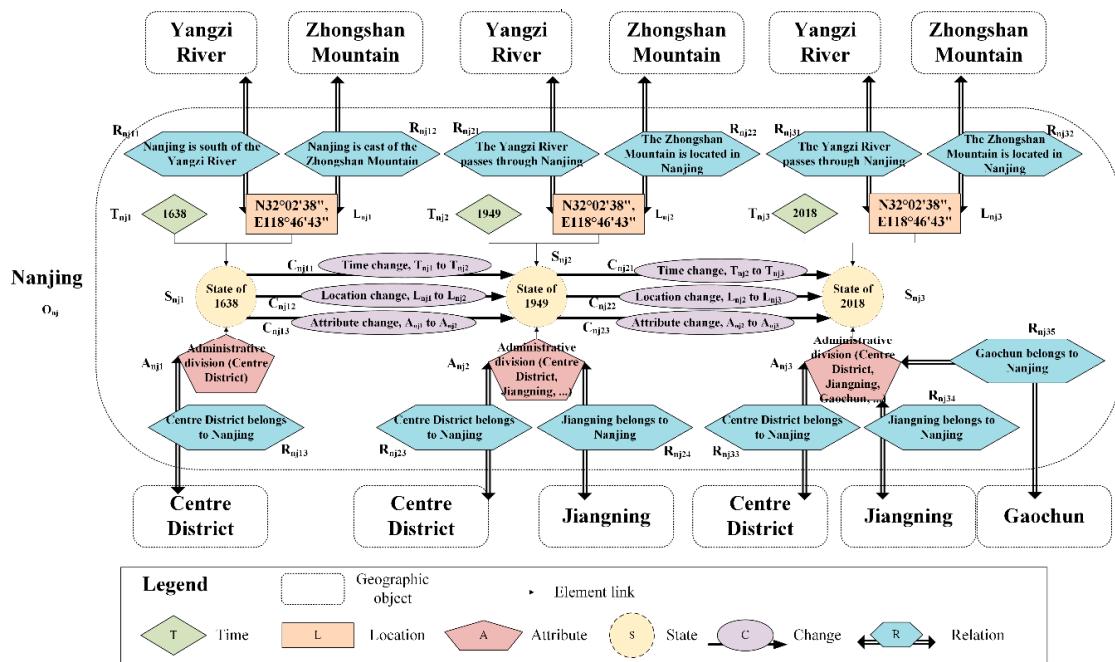


Figure 5. The diagram of different elements of Nanjing by using the GeoKG model.

Actually, different states of Nanjing $\{S_{nj1}, S_{nj2}, S_{nj3}\}$ indicate three different stages of 1368, 1949, and 2018. Each state contains different time, location, and attribute elements. For example, the state S_{nj1} of Nanjing contains time element T_{nj} of “1368”, location element L_{nj} of “location descriptions in 1368” and attribute element A_{nj} of “administrative region”. The state S_{nj1} of Nanjing can be defined as follows:

$$S_{nj1} = \left\{ \begin{array}{l} T_{nj} \sqsubseteq S_{nj1}, L_{nj} \sqsubseteq S_{nj1}, A_{nj} \sqsubseteq S_{nj1} \\ T_{nj} = \{T_{nj1} | T_{nj}.number \leq 1, T_{nj}.number \geq 1\}, \\ L_{nj} = \{L_{nj1} | L_{nj}.number \leq 1, L_{nj}.number \geq 1\}, \\ A_{nj} = \{A_{nj1} | A_{nj}.number \leq 1, A_{nj}.number \geq 1\} \end{array} \right\}$$

Different states could contain changes indicating different kinds of changes from one state to another one. For example, there are three main changes $\{C_{nj11}, C_{nj12}, C_{nj13}\}$ from the state S_{nj1} of Nanjing in 1368 to the state S_{nj2} of Nanjing in 1949: the change C_{nj11} between time elements, the change C_{nj12} between location elements and the change C_{nj13} between the attribute elements of “administrative region”. Note that all these changes belong to developing change type which indicates the change do not create a new geographic object. The changes can be defined as follows:

$$\begin{aligned} C_{nj11} &= \left\{ \begin{array}{l} St, act, CE, type \sqsubseteq C_{nj11} \\ St = \{S_{nj1}, S_{nj2}\}, act = \{"time change"\}, CE = \{T_{nj1}, T_{nj2}\}, type = Ch_d \end{array} \right\} \sqsubseteq O_{nj} \\ C_{nj12} &= \left\{ \begin{array}{l} St, act, CE, type \sqsubseteq C_{nj12} \\ St = \{S_{nj1}, S_{nj2}\}, act = \{"location change"\}, CE = \{L_{nj1}, L_{nj2}\}, type = Ch_d \end{array} \right\} \sqsubseteq O_{nj} \\ C_{nj13} &= \left\{ \begin{array}{l} St, act, CE, type \sqsubseteq C_{nj13} \\ St = \{S_{nj1}, S_{nj2}\}, act = \{"attribute change"\}, CE = \{A_{nj1}, A_{nj2}\}, type = Ch_d \end{array} \right\} \sqsubseteq O_{nj} \end{aligned}$$

Relation is an indispensable element which exists in geographic objects referring to the relationships between different elements. In this example, there are three relations $\{R_{nj11}, R_{nj12}, R_{nj13}\}$ relate to Nanjing in 1368: the spatial relation R_{nj11} between Nanjing O_{nj} and Yangzi River O_{yz} , the spatial relation R_{nj12} between Nanjing O_{nj} and Zhongshan Mountain O_{zm} , and the attribute relation R_{nj13} between Nanjing O_{nj} and Centre District O_{cd} , where L_{yz1} is the location of Yangzi River O_{yz} in 1368, L_{zm1} is the location of Zhongshan Mountain O_{zm} in 1368 and A_{cd1} is the “administrative region” attribute of Centre District O_{cd} in 1368. The relations can be defined as follows and the diagram of these relations was shown in Figure 6.

$$\begin{aligned} R_{nj11} &= \left\{ \begin{array}{l} E, Sem, type \sqsubseteq R_{nj11} \\ E = \{L_{nj1}, L_{yz1}\}, Sem = \{"Nanjing is south of the Yangzi River"\}, type = Re_l \end{array} \right\} \sqsubseteq O_{nj} \\ R_{nj12} &= \left\{ \begin{array}{l} E, Sem, type \sqsubseteq R_{nj12} \\ E = \{L_{nj1}, L_{zm1}\}, Sem = \{"Nanjing is east of the Zhongshan Mountain"\}, type = Re_l \end{array} \right\} \sqsubseteq O_{nj} \\ R_{nj13} &= \left\{ \begin{array}{l} E, Sem, type \sqsubseteq R_{nj13} \\ E = \{A_{nj1}, A_{cd1}\}, Sem = \{"Centre District is part of Nanjing"\}, type = Re_a \end{array} \right\} \sqsubseteq O_{nj} \end{aligned}$$

Correspondingly, Yangzi River contains the relation $R_{yz1} = R_{nj11}^-$, Zhongshan Mountain contains the relation $R_{zm1} = R_{nj12}^-$, and Centre District contains the relation $R_{cd1} = R_{nj13}^-$:

$$\begin{aligned} R_{yz1} &= \left\{ \begin{array}{l} E, Sem, type \sqsubseteq R_{yz1} \\ E = \{L_{nj1}, L_{yz1}\}, Sem = \{"Nanjing is south of the Yangzi River"\}, type = Re_l \end{array} \right\} \sqsubseteq O_{yz} \\ R_{zm1} &= \left\{ \begin{array}{l} E, Sem, type \sqsubseteq R_{zm1} \\ E = \{L_{nj1}, L_{zm1}\}, Sem = \{"Nanjing is east of the Zhongshan Mountain"\}, type = Re_l \end{array} \right\} \sqsubseteq O_{zm} \\ R_{cd1} &= \left\{ \begin{array}{l} E, Sem, type \sqsubseteq R_{cd1} \\ E = \{A_{nj1}, A_{cd1}\}, Sem = \{"Centre District is part of Nanjing"\}, type = Re_a \end{array} \right\} \sqsubseteq O_{cd} \end{aligned}$$

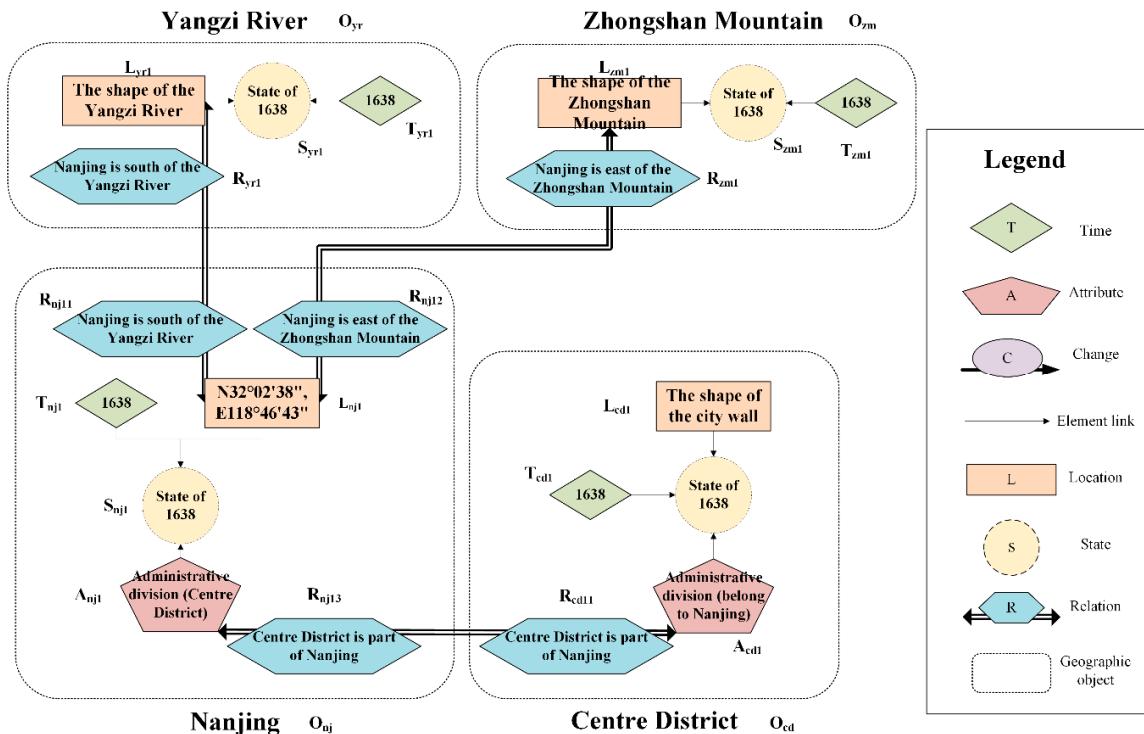


Figure 6. The diagram of relation elements of Nanjing in 1368.

The whole evolution case of administrative divisions of Nanjing can be shown in Figure 7. Corresponding to Figure 4, each geographic object contains one to three states. For instance, Yangzi River and Zhongshan Mountain have three stages of 1368, 1949, and 2018 and Jiangning and Gaochun have two stages of 1949 and 2018. As inner changes are not considered, Centre District only represented one stage. Between different stages, different kinds of changes were considered. For example, different stages of Yangzi River and Zhongshan Mountain include time change $\{C_{yz11}, C_{yz21}, C_{zm11}, C_{zm21}\}$, different stages of Nanjing include time change $\{C_{nj11}, C_{nj21}\}$, location change $\{C_{nj12}, C_{nj22}\}$, and attribute change $\{C_{nj13}, C_{nj23}\}$, and different stages of Jiangning and Gaochun include time change $\{C_{jn11}, C_{gc11}\}$ and attribute change $\{C_{jn12}, C_{gc12}\}$. Additionally, relations link different elements among both different geographic objects and same geographic object. For example, Nanjing in 1368 has relations to Yangzi River R_{nj11} , Zhongshan Mountain R_{nj12} , and Centre District R_{nj13} . Then, Nanjing in 1949 has relations to Yangzi River R_{nj21} , Zhongshan Mountain R_{nj22} , Centre District R_{nj23} , and Jiangning R_{nj24} . In 2018, Nanjing has relations to Yangzi River R_{nj31} , Zhongshan Mountain R_{nj32} , Centre District R_{nj33} , Jiangning R_{nj34} , and Gaochun R_{nj35} .

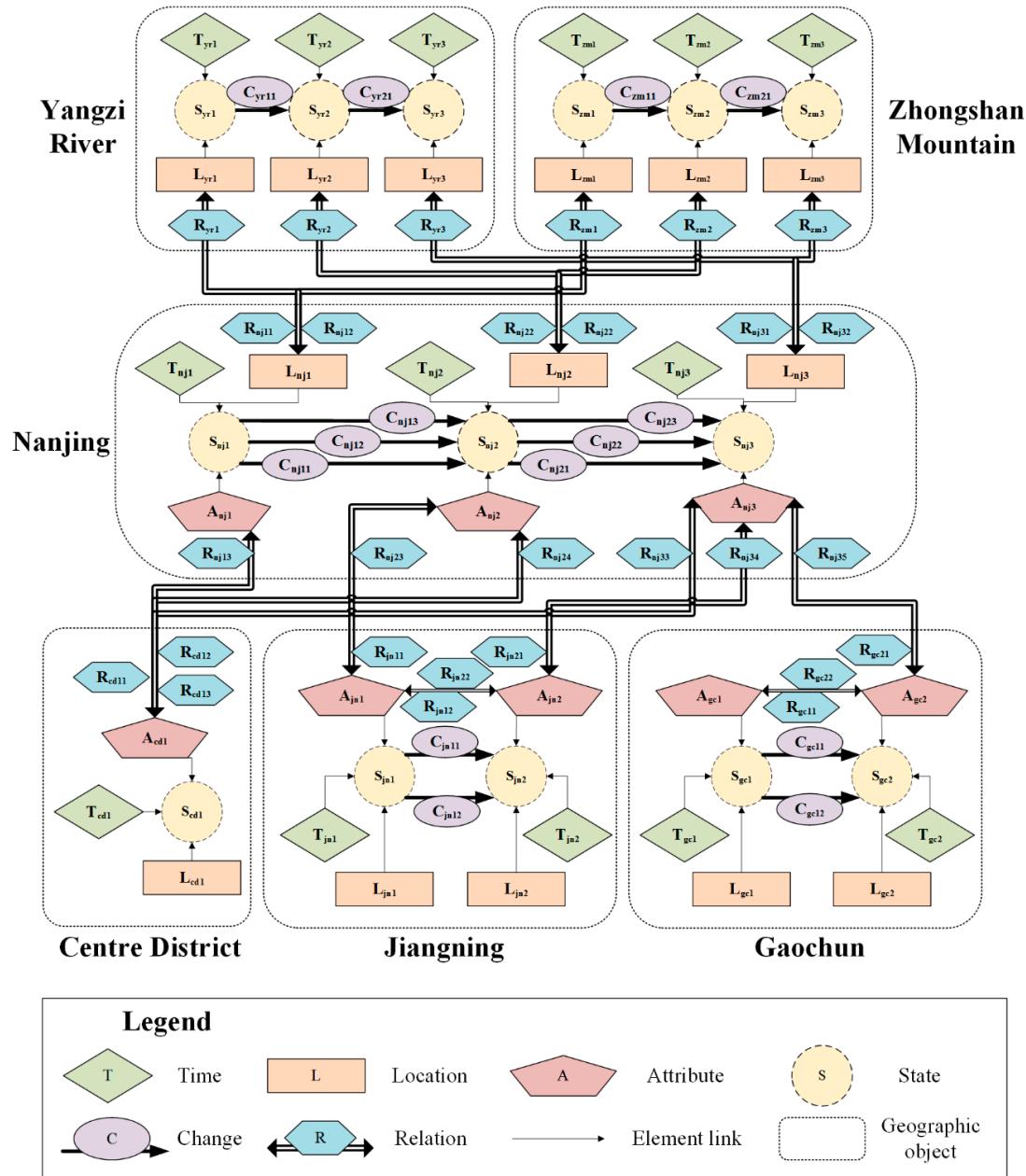


Figure 7. An overview of evolution case of administrative divisions of Nanjing and relevant geographic objects.

Note that there are also inner relations between elements. In this case, the administrative division of Jiangning in 1949 has an attribute relation R_{jn12} of “inheritance relationship” to the administrative division of Jiangning in 2018. Gaocun has the same attribute relation R_{gc11} . All these relations have the inverse relations in the opposite sides.

5. Discussion

In this section, the case study of administrative division evolution of Nanjing was constructed by using the GeoKG model and the YAGO model. YAGO is a representative open source knowledge graph with different versions. Note that we compared our model with YAGO2, a spatially and temporally enhanced version from <https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/>. Then, three kinds of core geographic questions were posted and the results

were analyzed to evaluate the knowledge representation ability of these two models. Finally, a user evaluation was given to verify the comparisons objectively.

5.1. The GeoKG and the YAGO

5.1.1. Structures

The structures of the GeoKG and the YAGO are different. Although Section 2 briefly introduced the characteristics of the YAGO, the comparison between two different structures needs to be analyzed in order to understand the following comparisons of queries and the results in next section. Figure 8 shows the examples structured by different models.

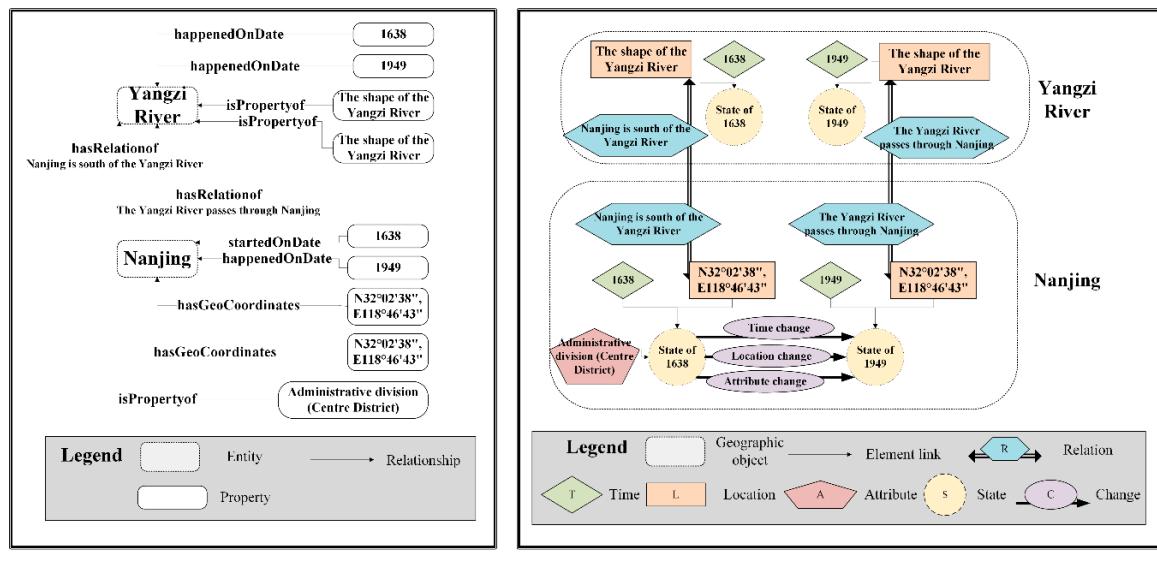


Figure 8. The examples with structures of the YAGO model and the GeoKG model. (a) the entities, properties and relationships in YAGO structure; (b) the elements in GeoKG structure.

In Figure 8a, there are only three kinds of elements: entity, property, and relationship. Each property links to a related entity by a relationship with a predicate. For example, “Nanjing” and “1638” have a relationship named “startedOnDate”. Note that the YAGO structure does not contain the relationships between the properties. Thus, there are no semantic relationships between properties. In other words, the massive descriptive properties of an entity link to the entity independently. For example, two relationships happened on Nanjing and the Yangzi River: “Nanjing is south of the Yangzi River” and “The Yangzi River passes through Nanjing”. It is difficult to understand this knowledge with no links between properties, whereas the GeoKG in Figure 8b sets six core elements and links these elements. With more integrated elements, the relationship of “Nanjing is south of the Yangzi River” can illustrate more clearly because this relationship links two locations in two different states of the two geographic objects. The different states providing this relationship happened on 1638 and the linked locations provide this relationship relate to different location descriptions. This knowledge cannot be provided without these links between the properties.

5.1.2. Construction

Both the GeoKG and YAGO were constructed manually by using the information about the case study of the administrative division evolution of Nanjing. The case study organized by the YAGO model was the classic SPO triple sets which has an open source ontology template. Additionally, the case study organized by the GeoKG model also stored by SPO triple sets that contain more predicates. The main supplement predicates include “isStateof”, “isTimeof”, “isLocationof”, “isAttributeof”,

“isChangeof”, “isRelationof”, “isChangeto”, and “isRelateto”. All these predicates were applied to complete the semantic structure of the GeoKG model. From this perspective, the underlying storage mechanisms of GeoKG and YAGO are the same.

5.2. The Comparison of Knowledge Representation Ability between the GeoKG and the YAGO

5.2.1. Questions

Time, space, and attribute are three indispensable aspects on geoscience. These three kinds of questions can be defined as standard questions to evaluate whether the stored geographic knowledge is good. According to the differences between factual knowledge and inferential knowledge, each question was a set of two parts. To this case study, the questions are shown in Table 2.

Table 2. Questions to the GeoKG model and The YAGO model.

| Question Types | Factual Question | Inferential Question |
|----------------|------------------------------------|--|
| Time | When was Nanjing named? | When does Jiangning belong to Nanjing? |
| Space | Where is Nanjing? | What is the spatial relationship between Nanjing and Yangzi River? |
| Attribute | Which city does Gaochun belong to? | What administrative divisions belong to Nanjing? |

5.2.2. Queries

Questions cannot be directly queried from the GeoKG and YAGO database. Thus, they need to be translated into SPARQL queries, because of either GeoKG or YAGO stored as triples in RDFs. For example, the factual question of time can be translated into SPARQL queries, as shown in Table 3.

Table 3. The SPARQL query of “When was Nanjing named?”

| Steps | SPARQL Query | Semantic Meaning |
|-------|---|-------------------------------------|
| 1 | PREFIX rdfs: < http://www.w3.org/2000/01/rdf-schema# >. | protocol |
| 2 | SELECT ?sTime WHERE { | Query content “?sTime” (start time) |
| 3 | ?s rdfs:type :City. | Type is “City” |
| 4 | ?s :cityName ‘Nanjing’. | Get “Nanjing” geographic object |
| 5 | ?s :hasName ?o. | Get time when named ‘Nanjing’ |
| 6 | ?o :startedOnDate ?sTime. | Get started time |
| 7 | ?o :usedName ?uName. | Constraint condition |
| 8 | FILTER regex(?uName, “^Nanjing”) | Constraint condition setting |
| | } | |

5.2.3. Comparison and Analysis

The collected items of YAGO and GeoKG on six questions are listed in Table 4. The comparisons will be conducted in terms of accuracy, completeness, and repetition.

a. Accuracy

In general, the results of the GeoKG are slightly better than the YAGO. Both of the two models can respond with accurate results to #Q1, #Q2, #Q3, #Q4, and #Q6. In #Q5, the result of the YAGO model returned two items and the results of the GeoKG model returned four items. Actually, “Zhenjiang” and “Nanjing” from the YAGO model are the misleading answers to the question of “Which city does Gaochun belong to?” Though the results from the GeoKG model: “Zhenjiang(Gaochun, state of 1949)”, “Zhenjiang(Zhenjiang, state of 1949)”, “Nanjing(Gaochun, state of 2018)” and “Nanjing(Nanjing, state of 2018)” are similar to the front, these results contain the geographic object and relevant state information which is a benefit for the users to understand the results. From this perspective, these state information from GeoKG provided more accurate information than the YAGO model.

Table 4. The results of YAGO and GeoKG on SPARQL queries.

| Question Types | Questions | Results | |
|----------------|---|--|---|
| | | YAGO | GeoKG |
| Time | #Q1: When was Nanjing named? | ➤ 1368 | ➤ 1368 (Nanjing, state of 1368) |
| | #Q2: When does Jiangning belong to Nanjing? | ➤ 1949 ➤ 2018 | ➤ 1949 (Jiangning, state of 1949) ➤ 2018 (Jiangning, state of 2018) ➤ 1949 (Nanjing, state of 2018) ➤ 2018 (Nanjing, state of 2018) |
| | #Q3: Where is Nanjing? | ➤ N32°02'38'', E118°46'43'' ➤ N32°02'38'', E118°46'43'' ➤ N32°02'38'', E118°46'43'' | ➤ N32°02'38'', E118°46'43'' (Nanjing, state of 1368) ➤ N32°02'38'', E118°46'43'' (Nanjing, state of 1949) ➤ N32°02'38'', E118°46'43'' (Nanjing, state of 2018) |
| | #Q4: What is the spatial relationship between Nanjing and Yangzi River? | ➤ Nanjing is south of the Yangzi River ➤ The Yangzi River passes through Nanjing ➤ The Yangzi River passes through Nanjing | ➤ Nanjing is south of the Yangzi River (Nanjing, state of 1368) ➤ Nanjing is south of the Yangzi River (Yangzi River, state of 1368) ➤ The Yangzi River passes through Nanjing(Nanjing, state of 1949) ➤ The Yangzi River passes through Nanjing(Yangzi River, state of 1949) ➤ The Yangzi River passes through Nanjing(Nanjing, state of 2018) ➤ The Yangzi River passes through Nanjing(Yangzi River, state of 2018) |
| | #Q5: Which city does Gaochun belong to? | ➤ Zhenjiang ➤ Nanjing | ➤ Zhenjiang(Gaochun, state of 1949) ➤ Zhenjiang(Zhenjiang, state of 1949) ➤ Nanjing(Gaochun, state of 2018) ➤ Nanjing(Nanjing, state of 2018) |
| | #Q6: What administrative divisions belong to Nanjing? | ➤ Centre District ➤ Jiangning ➤ Jurong ➤ Dangtu ➤ Luhe ➤ Pukou ➤ Hexian ➤ Lishui ➤ Gaochun ➤ Qixia | ➤ Centre District(Nanjing, state of 1368) ➤ Centre District(Nanjing, state of 1949) ➤ Centre District(Nanjing, state of 2018) ➤ Jiangning(Nanjing, state of 1949) ➤ Jiangning(Nanjing, state of 2018) ➤ Jurong(Nanjing, state of 1949) ➤ Dangtu(Nanjing, state of 1949) ➤ Luhe(Nanjing, state of 1949) ➤ Luhe(Nanjing, state of 2018) ➤ Pukou(Nanjing, state of 1949) ➤ Pukou(Nanjing, state of 2018) ➤ Hexian(Nanjing, state of 1949) ➤ Lishui(Nanjing, state of 2018) ➤ Gaochun(Nanjing, state of 2018) ➤ Qixia(Nanjing, state of 2018) ➤ more items |

b. Completeness

Although both of these two models can return the complete results, the results of the GeoKG contains more semantic integrity. In #Q6, YAGO returned 10 items: Centre District, Jiangning, Jurong, Dangtu, Luhe, Pukou, Hexian, Lishui, Gaochun, and Qixia. Among these divisions, Centre District belonged to Nanjing since 1368. Jiangning, Luhe and Pukou belonged to Nanjing since 1949. Jurong, Dangtu and Hexian belonged to Nanjing in 1949. Lishui, Gaochun, and Qixia belonged to Nanjing in 2018. As the question does not have an explicit time constraint condition, YAGO returned all the items, whereas GeoKG returned 30 items and each item recorded the target object and its relevant

geographic object and state. It contains the item of “Centre District (Nanjing, state of 1368)” and the item of “Centre District (Centre District, state of 1368)”, because of the relation existed oppositely.

c. Repetition

The results of the GeoKG has more repeat items than the results from the YAGO. The results from the YAGO have repeat items in #Q3 and #Q4, because of the records are repeat. However, the GeoKG model is different. In #Q2, #Q4, #Q5, and #Q6, the results of the GeoKG have many repeat items; for example, the items of “1949 (Jiangning, state of 1949)” and “1949 (Nanjing, state of 2018)” in #Q2. The query target object “1949” is the same. In spite of these two items sourced from different geographic objects (Jiangning and Nanjing), these two items are still quite similar, which pushed more redundant information to the users.

In summary, the results of the GeoKG model are more accurate and complete than the YAGO model with the enhancing state information. It can decrease the influence from the fuzziness questions and obtain answers with more semantic meaning (e.g., geographic object and its relevant state). Meanwhile, the GeoKG model could generate more pairs results (e.g., “Nanjing is south of the Yangzi River (Nanjing, state of 1368)” vs. “Nanjing is south of the Yangzi River (Yangzi River, state of 1368)”), because the relation is stored oppositely in a different geographic object.

5.2.4. User Evaluation

An online questionnaire survey is also given in order to verify the results of comparative analyses. The questionnaire is divided into eight parts. The first part is the basic information survey that asks individuals four aspects of information (gender, familiarity to the research area, background, and education level). The statistics of these basic information are shown in Figure 9. The 2nd–7th parts correspond to the questions #Q1–#Q6 and ask the questions about the best answer, accuracy, completeness, and repetition. The 8th part are summary questions including the overall evaluation, scores on YAGO and scores on GeoKG on different aspects. The scores are set as 1–5 corresponding to very bad, bad, normal, good, and very good, and each score group includes an overall score, accuracy score, completeness score, and repetition score. There are 106 valid feedbacks we finally received.

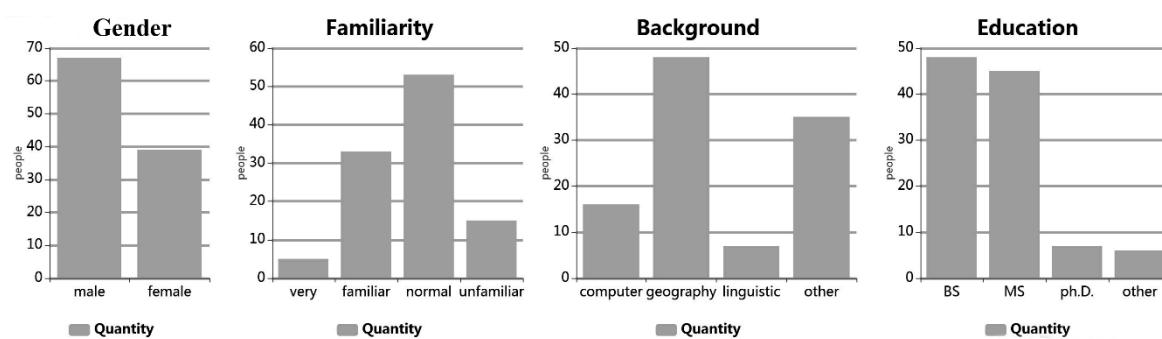


Figure 9. The statistics of the four main types of the basic information about the survey.

Figure 10 shows the best answers on #Q1–#Q6 and the overall scores of the YAGO and the GeoKG. In the best answer histogram, the overall results show 54.72% individuals support the GeoKG, which is 23.59% higher than the YAGO at 31.13%. Specifically, the quantities of #Q1 and #Q2 are quite close but the quantities of #Q3–#Q6 are not. The quantities of the GeoKG are much higher than the YAGO among the last four questions, especially in #Q5. The line charts of overall scores on YAGO and GeoKG also show that the evaluation of GeoKG is better than the YAGO. A 7.8% improvement on the average score from the YAGO (3.15) to the GeoKG (3.49) is obtained.

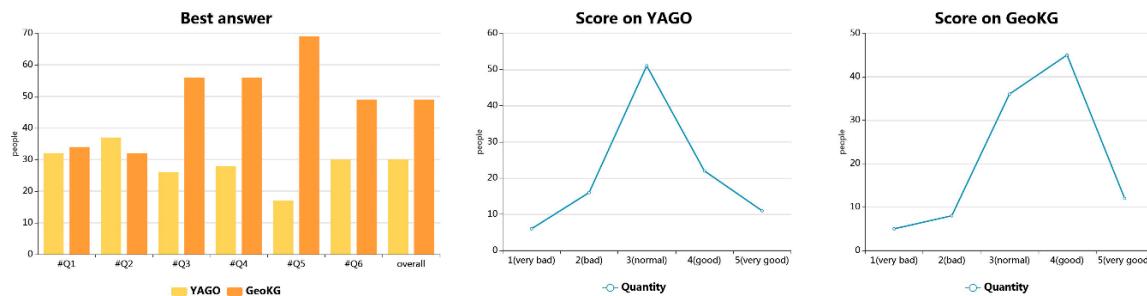


Figure 10. The best answer on #Q1–#Q6 and the overall scores of the YAGO and the GeoKG.

From the sub-aspects (accuracy, completeness, and repetition) of the point of view, different quantities can immediately show the scores from the YAGO and the GeoKG. Different quantities show the ability from the model (details in Figure 11). Nearly all three aspects of the YAGO obtained the score 3, whereas the GeoKG was different: a score of 4 on accuracy, a score of 4–5 on completeness, and a score of 3–4 on repetition. Comparing these scores, it can be seen that there is little promotion on the accuracy from the 3.11 average score in YAGO to the 3.78 average score in GeoKG. An overwhelming improvement shows on the completeness of the answers from a 2.99 average score in YAGO to a 3.87 average score in GeoKG. Additionally, the GeoKG also obtains a higher repetition from a 3.01 average score in YAGO to a 3.42 average score in GeoKG.

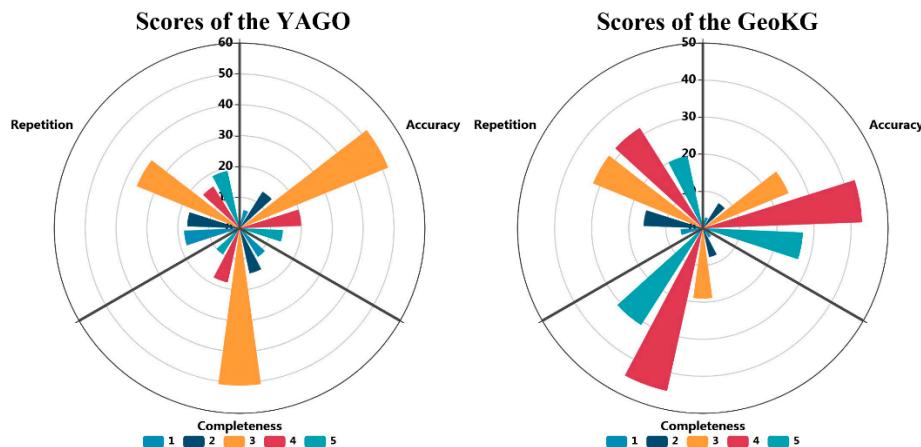


Figure 11. Rose maps of scores of different aspects on the YAGO and the GeoKG.

In summary, the answers from GeoKG makes an improvement to those of YAGO's. The user evaluation objectively verified the analyses in Section 5.2.3 and specifically showed clear answers. It can be seen that the main improvements of the GeoKG are on the #Q3–#Q6, which are spatial and attribute questions. These answers to these questions require more related state information and temporal information, which need the links between the elements (Figure 8). This is the reason why the GeoKG is better than the YAGO. In addition, the GeoKG contains more redundancy information than the YAGO because of the bi-directionality of the relation element. This could be a focus of continuous further research on the index and applications in the future.

6. Conclusions

Given that much attention has been paid to the representation of geographic knowledge, this paper is focused on the development of current geographic knowledge representations. We analyzed the problems of current geographic knowledge representation and found that two issues must be improved: the elements of geographic knowledge representation and the supplement of the construction operators of DL.

Following the basic idea of the six core geographical questions, we designed a conceptualized model called GeoKG based on the six elements around the geographical questions, then supplemented the construction operators of DL and finally provided the formalizations of the model with these operators. Additionally, an evolution case of administrative divisions of Nanjing was formalized and illustrated. Then, the knowledge graphs were constructed by both the GeoKG model and the YAGO model by using the case study. After setting a group of standard geographic questions, the query results were finally compared. The results showed that the results of GeoKG are more accurate and complete than the YAGO results, which are verified by the following user evaluation. This comparison indicates the GeoKG model displays its ability to organize geographic knowledge in computers and is a promising and powerful model for geographic knowledge representation.

Author Contributions: Conceptualization: S.W., X.Z., and P.Y.; data curation: S.W.; formal analysis: S.W.; funding acquisition: X.Z.; investigation: S.W., M.D., Y.L., and H.X.; methodology: S.W., X.Z., and M.D.; supervision: X.Z.; validation: P.Y., M.D., and Y.L.; visualization: M.D.; writing—original draft: S.W.; writing—review and editing: S.W. and P.Y.

Acknowledgments: The authors thank Mingguang Wu, Junzhi Liu and Jie Zhu for their critical reviews and constructive comments. This research is supported by the National Natural Science Foundation of China grants no. 41631177 and no. 41671393 and the National Key Research and Development Program of China, no. 2017YFB0503602.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Golledge, R.G. The Nature of Geographic Knowledge. *Ann. Assoc. Am. Geogr.* **2015**, *92*, 1–14. [[CrossRef](#)]
2. Haubrich, H. International Charter on Geographical Education. *J. Geogr.* **1997**, *96*, 33–39.
3. Davis, R. What Is a Knowledge Representation? *AI Mag.* **1993**, *14*, 17–33.
4. Zhang, Y.; Gao, Y.; Xue, L.L.; Shen, S.; Chen, K. A common sense geographic knowledge base for GIR. *Sci. Technol. Sci.* **2008**, *51*, 26–37. [[CrossRef](#)]
5. Kuhn, W. Modeling Vs Encoding for the Semantic Web. *Semant. Web* **2010**, *1*, 11–15.
6. Baader, F.; Sattler, U. An Overview of Tableau Algorithms for Description Logics. *Stud. Log.* **2001**, *69*, 5–40. [[CrossRef](#)]
7. Hoffart, J.; Suchanek, F.M.; Berberich, K.; Weikum, G. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artif. Intell.* **2013**, *194*, 28–61. [[CrossRef](#)]
8. Guarino, N.; Oberle, D.; Staab, S. What Is an Ontology? *Handb. Ontol.* **2009**, 1–17. [[CrossRef](#)]
9. Ding, Y.; Foo, S. Ontology research and development. Part 1: A review of ontology generation. *J. Inf. Sci.* **2002**, *28*, 123–136.
10. Couclelis, H. Ontologies of geographic information. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 1785–1809. [[CrossRef](#)]
11. Siricharoen, W.V.; Pakdeetrakulwong, U. A Survey on Ontology-Driven Geographic Information Systems. In Proceedings of the Fourth International Conference on Digital Information and Communication Technology and Its Applications, Bangkok, Thailand, 6–8 May 2014.
12. Gruber, T.R. Toward principles for the design of ontologies used for knowledge sharing? *Int. J. Hum.-Comput. Stud.* **1995**, *43*, 907–928. [[CrossRef](#)]
13. Fonseca, F.T.; Egenhofer, M.J. Ontology-driven geographic information systems. In Proceedings of the 7th ACM International Symposium on Advances in Geographic Information Systems, Kansas City, MO, USA, 2–6 November 1999; Volume 71, pp. 14–19.
14. Jun, X.U.; Tao, P.; Yao, Y. Conceptual Framework and Representation of Geographic Knowledge Map: Conceptual Framework and Representation of Geographic Knowledge Map. *J. Geo-Inf. Sci.* **2010**, *12*. [[CrossRef](#)]
15. Chen, J.; Deng, S.; Chen, H. Crowdgeokg: Crowdsourced Geo-Knowledge Graph. In Proceedings of the China Conference on Knowledge Graph and Semantic Computing, Chengdu, China, 26–29 August 2017.
16. Arvor, D.; Durieux, L.; Andrés, S.; Laporte, M.-A. Advances in Geographic Object-Based Image Analysis with ontologies: A review of main contributions and limitations from a remote sensing perspective. *ISPRS J. Photogramm. Remote. Sens.* **2013**, *82*, 125–137. [[CrossRef](#)]
17. Brown, S.H. *Knowledge Representation and the Logical Basis of Ontology*; Springer: London, UK, 2012; pp. 11–50.

18. Pittet, P.; Cruz, C.; Nicolle, C. Modeling Changes for Shoin(D) Ontologies: An Exhaustive Structural Model. In Proceedings of the IEEE Seventh International Conference on Semantic Computing, Irvine, CA, USA, 16–18 September 2013.
19. Sattler, U.; Horrocks, I. A description logic with transitive and inverse roles and role hierarchies. *J. Log. Comput.* **1999**, *9*, 385–410. [[CrossRef](#)]
20. Horrocks, I.; Sattler, U.; Tobies, S. Practical Reasoning for Expressive Description Logics. In Proceedings of the International Conference on Logic for Programming and Automated Reasoning, Tbilisi, Georgia, 6–10 September 1999.
21. Horrocks, I.; Sattler, U.; Tobies, S. Practical reasoning for very expressive description logics. *Log. J. IGPL* **2000**, *8*, 239–263. [[CrossRef](#)]
22. Aachen, R.; Informatik, L.T.; Horrocks, I.; Sattler, U.; Tobies, S. Pspace-Algorithm for Deciding Alcnir+-Satisfiability. In *LTCS-Report 98-08*; ACM Digital Library: Aachen, Germany, 1998.
23. Mei, J. From Alc to Shoq(D):A Survey of Tableau Algorithms for Description Logics. *Comput. Sci.* **2005**, *32*, 1–11. [[CrossRef](#)]
24. Singhal, A. *Official Google Blog: Introducing the Knowledge Graph: Things, Not Strings*; Northwestern University: Evanston, IL, USA, 2012.
25. Suchanek, F.M.; Kasneci, G.; Weikum, G. Yago: A Core of Semantic Knowledge. In Proceedings of the 16th International Conference on World Wide Web (WWW), Banff, AB, Canada, 8–12 May 2007; Volume 272, pp. 697–706.
26. Bollacker, K.; Cook, R.; Tufts, P. Freebase: A Shared Database of Structured General Human Knowledge. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 22–26 July 2007.
27. Wu, W.; Li, H.; Wang, H.; Zhu, K.Q. Probase: A Probabilistic Taxonomy for Text Understanding. In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data (SIGMOD'12), Scottsdale, AZ, USA, 20–24 May 2012.
28. Lehmann, J. Dbpedia: A Large-Scale, Multilingual Knowledge Base Extracted from Wikipedia. *Semant. Web* **2015**, *6*, 167–195.
29. Li, J.; Liu, R.; Xiong, R. A Chinese Geographic Knowledge Base for Gir. In Proceedings of the IEEE International Conference on Computational Science and Engineering, Guangzhou, China, 21–24 July 2017.
30. Kauppinen, T.; Espindola, G.M. Ontology-Based Modeling of Land Change Trajectories in the Brazilian Amazon. In Proceedings of the Geoinformatik, Münster, Germany, 15–17 June 2013.
31. Zhu, Y.; Zhou, W.; Xu, Y.; Liu, J.; Tan, Y. Intelligent Learning for Knowledge Graph towards Geological Data. *Sci. Program.* **2017**, *2017*, 1–13. [[CrossRef](#)]
32. William, M. (Ed.) *The American Heritage Dictionary of the English Language*; New College Edition; Houghton Mifflin Company: Boston, MA, USA, 1980.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).