

Article

A Heterogeneous Geospatial Data Retrieval Method Using Knowledge Graph

Junnan Liu ¹, Haiyan Liu ^{1*}, Xiaohui Chen ¹, Xuan Guo ², Qingbo Zhao ¹, Jia Li ¹, Lei Kang ¹ and Jianxiang Liu ¹

¹ School of Data and Target Engineering, Strategic Support Force Information Engineering University, Zhengzhou 450001, China; lpjunnan.love@163.com (J.L.); cxh_vrlab@163.com (X.C.); q13526647215@163.com (Q.Z.); lijia_kk@163.com (J.L.); xxgcdxkl@163.com (L.K.); liu_2012111@163.com (J.L.)

² Institute of Geospatial Information, Strategic Support Force Information Engineering University, Zhengzhou 450001, China; gx930707@163.com

* Correspondence: liuharry2020@163.com; Tel.: +86-371-6368-5366

Abstract: Information resources have increased rapidly in the big data era. Geospatial data plays an indispensable role in spatially informed analyses, while data in different areas are relatively isolated. Therefore, it is inadequate to use relational data in handling many semantic intricacies and retrieving geospatial data. In light of this, a heterogeneous retrieval method based on knowledge graph is proposed in this paper. There are three advantages of this method: (1) the semantic knowledge of geospatial data is considered; (2) more information required by users could be obtained; (3) data retrieval speed can be improved. Firstly, implicit semantic knowledge is studied and applied to construct a knowledge graph, integrating semantics in multi-source heterogeneous geospatial data. Then, the query expansion rules and the mappings between knowledge and database are designed to construct retrieval statements and obtain related spatial entities. Finally, the effectiveness and efficiency are verified through comparative analysis and practices. The experiment indicates that the method could automatically construct database retrieval statements and retrieve more relevant data. Additionally, users could reduce the dependence on data storage mode and database Structured Query Language syntax. This paper would facilitate the sharing and outreach of geospatial knowledge for various spatial studies.

Citation: Liu, J.; Liu, H.; Chen, X.; Guo, X.; Zhao, Q.; Li, J.; Kang, L.; Liu, J. A Heterogeneous Geospatial Data Retrieval Method Using Knowledge Graph. *Sustainability* **2021**, *13*, 2005. <https://doi.org/10.3390/su13042005>

Keywords: data retrieval; knowledge graph; geospatial data; semantics; data integration; ontology

Received: 17 December 2020

Accepted: 9 February 2021

Published: 12 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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/>).

1. Introduction

With the advent of the big data era, there has been a significant increase in the amount of multi-source heterogeneous data in various fields, of which more than 50% are relevant to geospatial information [1–3]. The types of geographic entities can be distinguished with the use of semantics contained in specialized terms and names [4,5]. For example, “River” in “Yellow River” and “Railway Station” in “Zhengzhou Railway Station” can represent entity categories.

At present, retrieval technology has been rapidly developed and is widely used. Hence, the intellectualization is remarkably improved in various application scenarios (e.g., data fusion, geospatial analysis, early-warning, etc.) [6,7]. However, existing methods seldom consider the spatial and semantic characteristics comprehensively, leaving great room for improving geospatial data retrieval speed. On the one hand, the multi-source geospatial data is heterogeneous and cannot be retrieved directly [8,9]. On the other hand, it is difficult to consider implicit semantic knowledge and complex spatial relationships, leading to information disorientation and information overload [10,11]. In addition, although ontology-based data access (OBDA) is a popular paradigm for accessing data, it is usually based on a commonsense knowledge base, lacking geospatial semantic knowledge [12,13]. Therefore, it is urgent to introduce a knowledge graph (KG) to

consider the relationship between geospatial data and semantics, so as to organize and retrieve multi-source heterogeneous data. KG consists of tremendous knowledge, which is of great significance to bridge the semantic gap and improve intelligent retrieval methods [14]. The combination of KG and Geographic Information System (GIS) is a great attempt to acquire related geospatial information intelligently.

In order to improve retrieval quality and efficiency, a new method considering geospatial semantics is put forward in this paper. KG is constructed to organize and integrate multi-source heterogeneous data by extracting implicit semantics from geospatial data. Then, conceptual query expansions rules and mapping rules between KG and spatial database are defined to construct retrieval statements automatically. Associated geographic entities are expanded to realize data retrieval based on the implicit semantics of search terms. Finally, the feasibility is verified based on application cases and comparative analysis.

The main contributions of this paper are as follows:

- A KG construction method is proposed to integrate heterogeneous geospatial data. In this paper, KG is firstly constructed through mined knowledge to integrate semantics and relationships. Furthermore, the knowledge graph constructed from a bottom-up way can narrow down the retrieval domain and improve retrieval quality.
- A query expansion method considering relationships between concepts and entities is proposed, by which entities belonging to related concepts are returned. Moreover, for entities with low conceptual similarity, their associated entities are obtained to expand retrieval results coverage.
- A retrieval method automatically building Structured Query Language (SQL) statements is proposed, by which SQL retrieval statements are built through the semantic knowledge of search terms.

The rest of this paper is structured as follows. Section 2 reviews the related work regarding data retrieval methods and geospatial data. In Section 3, a data retrieval technology framework is proposed, which consists of geospatial data integration, semantic query expansion, mapping rules designing, and entity query expansion. Section 4 provides several application cases and some comparative analysis. Section 5 discusses the limitations of this study and puts forward further work.

2. Related Work

In this section, several research directions and related background is reviewed and discussed.

2.1. Geospatial Data Analysis

Geospatial data contains unique spatial identifiers and geographic coordinates in real or virtual space. Geospatial data retrieval is considered as a branch of traditional information retrieval methods [15,16]. In the big data era, an in-depth understanding of the geospatial data model is the basis for achieving knowledge expression and retrieval. There are two main representative geospatial data models, i.e., geographical vector model and OpenStreetMap (OSM) data model.

2.1.1. Geographical Vector Model

The geographical vector model is stored in the form of geographic layers, features, and property fields to display data. It consists of two separate files, which are inter-associated through feature ID, to manage spatial data (i.e., geometric data) and attribute data, respectively [9]. As shown in Figure 1, the road is displayed by a geographical vector model, in which spatial data files store spatial information, while attribute file stores attribute information. The spatial data files record ground objects' locations based on map projection and represents geographic features in the form of point, line, polygon, etc.

Therefore, the features could represent spatial distribution and topological structure better in a two-dimensional Cartesian coordinate system.

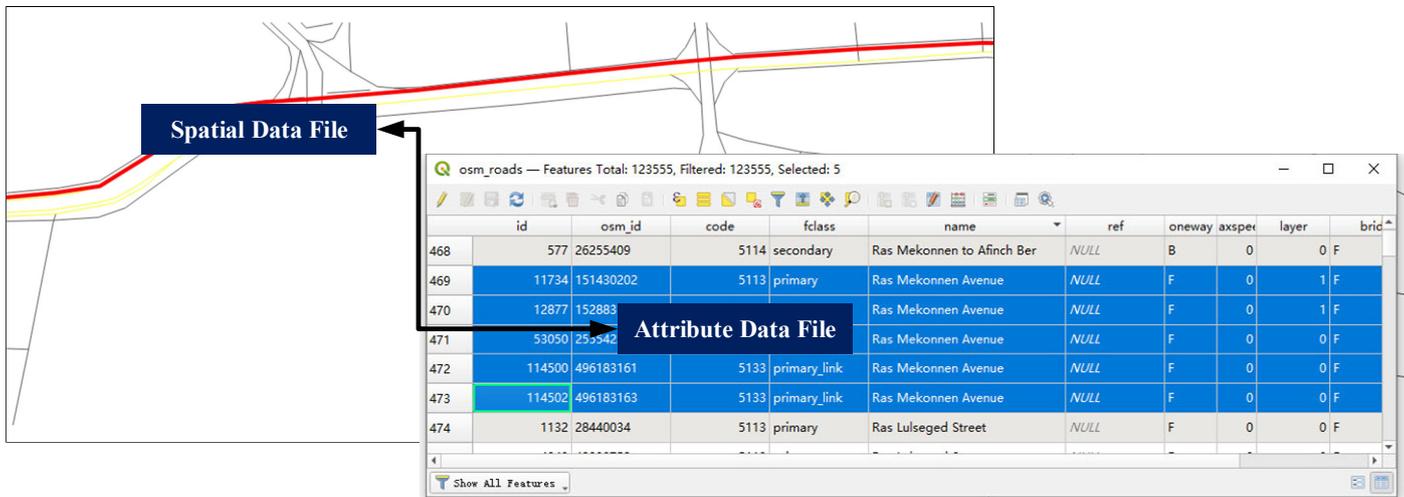


Figure 1. Geographical vector model.

A file-based vector data model frequently links spatial with attribute data, reducing the efficiency of GIS operations [17]. In light of this, spatial database is used to store vector data through several tables, in which spatial coordinates and attributes are uniformly represented. As shown in Figure 2, geometries are stored in a Binary Large Object (BLOB) property field, and attributes are stored in other property fields. Hence, query languages such as SQL can be used to retrieve both spatial and attribute information simultaneously.

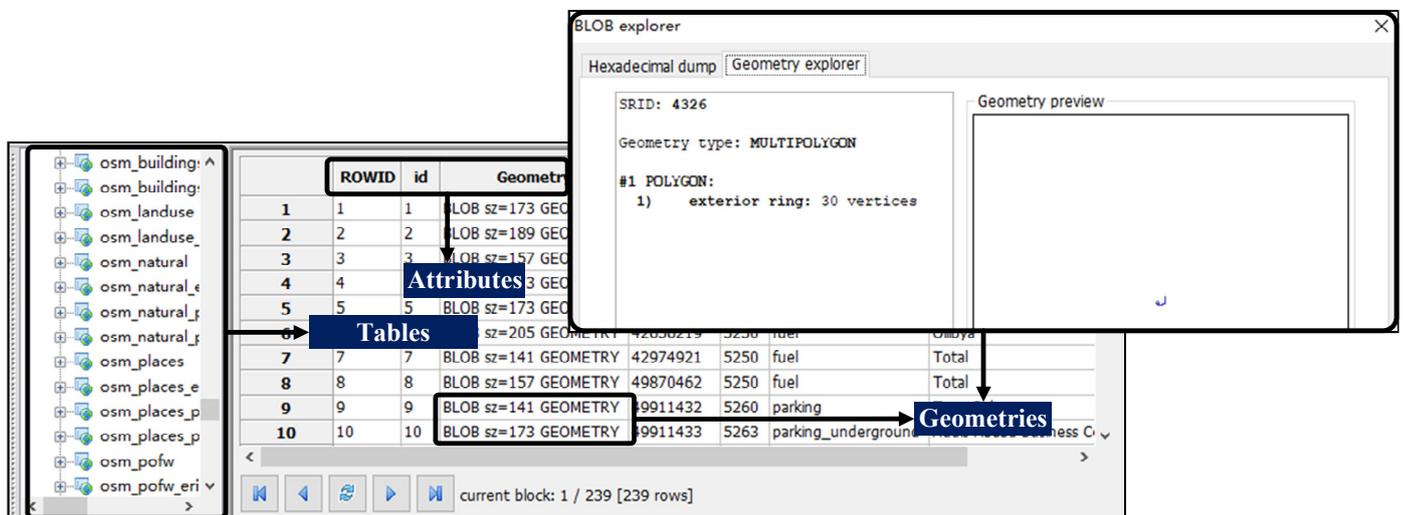


Figure 2. Geographical vector model in the spatial database.

2.1.2. OSM Data Model

OpenStreetMap (OSM), one of the most influential applications in volunteered geographic information projects, is an editable, free, and crowdsourced world map [8]. It has been widely used in earthquake disaster relief [18], real-time navigation [19], travel cost calculation [20], because of its advantages like high time-efficiency and rich semantics. OSM provides two free download services (i.e., vector file and Extensible Markup Language (XML) file) in Geofabrik (<http://download.geofabrik.de>). The OSM XML file repre-

sents the spatial information through three fundamental items (i.e., Node, Way, and Relation) [21]. These items contain tags describing the attributes, making them richer in semantics.

The item Node represents a point, taken as a basic unit of line and polygon. As shown in Figure 3, tags “lat” and “lon” record latitude and longitude coordinates, respectively, while tags “k” and “v” identify other attributes of ground objects (such as name).

```
<Node id="244080486" version="26" lat="34.0000001" lon="113.9999999">
  <tag k="name" v="Henan Province"/>
  <tag k="gns:DSG" v="ADM1"/>
  ...
  <tag k="wikipedia" v="en:Henan"/>
</Node>
```

Figure 3. An example of item Node.

The item Way is another XML element, which displays ground objects abstracted as line and polygon. As shown in Figure 4, item Nd and tag “ref” are combined to represent referenced Nodes in Way. The item Way represents a line if the referenced Nodes are different at the beginning and the end, otherwise it represents a polygon.

```
<Way id="745962721" version="4" timestamp="2020-10-16T16:39:01Z">
  <Nd ref="1912964710"/> <!--Nd is regarded as an item, and ref is a tag in OSM file -->
  ...
  <tag k="passenger_lines" v="2"/>
</Way>
```

Figure 4. An example of item Way.

The item Relation, as an XML element, associates Nodes and Ways. As shown in Figure 5, tags “type” and “ref” are used to represent referenced Nodes or Ways, and “role” represents the relation between Ways and Nodes.

```
<Relation id="193064" version="490" timestamp="2020-10-29T00:10:30Z">
  <member type="node" ref="269806844" role="contain"/>
  ...
  <member type="way" ref="860521887" role="">
</Relation>
```

Figure 5. An example of item Relation.

2.1.3. Free Tagging Mechanism of OSM

The XML file records attributes of ground objects through item Tag, which includes Key and Value abbreviated as “k” and “v”, respectively (as shown in Figures 3–5). In detail, tag Key describes attribute type, and tag Value refers to the value of Key. For example,

“k = ‘route’ v = ‘railway’” represents a parent-child relationship between concepts, and “k = ‘maxspeed’ v = ‘200’” indicates the limitation of the maximum speed. Prescribing the manifestation through tags without restricting attribute content is defined as Free Tagging mechanism of OSM. Based on this mechanism, the semantics of geospatial objects can be freely contained, edited, and inserted. Hence OSM is employed as an essential data source to extract semantic knowledge in this paper.

The vector model and OSM model have become the primary organization mode of geospatial data. However, there are some practical application problems: (1) the vector model is rich in spatial knowledge but weak in semantic knowledge; (2) both data models are heterogeneous, making it retrieved difficultly. Therefore, it is urgent to organize semantic knowledge to fuse multi-source geographic data and improve data retrieval speed.

2.2. Traditional Data Retrieval Method

2.2.1. Attribute Information Retrieval

The attribute information retrieval method of geospatial data is consistent with the traditional method based on character features. In early researches, name retrieval methods have become the main research direction, and researchers proposed many search algorithms based on domain dictionaries such as Hash index [22], Trie index [23], Double-word Hash index [24], etc. The Hash index retrieval method contains a three-level structure (i.e., original text, the index table of word, and the Hash table of the first character) to narrow down retrieval results by dichotomy. However, it mainly relies on global matching, leading to low efficiency. The Trie index method introduces a tree structure (including first-character Hash table and tree index) to reduce the number of matching operations, thus obtaining results quickly. Nevertheless, there are still some limitations, such as complex indexes and high memory consumption. To improve data retrieval speed, the double-word Hash index method combines the advantages of Hash and Tries, in which the Hash index is used to retrieve words with more than three characters and the Trie index retrieves words with less than or equal to two characters. On this basis, researchers started to research the characteristics of geographic names. For instance, Zhang et al. [25] proposed a geographic name retrieval method considering character features, in which individual characters are taken as basic units to construct indexes.

2.2.2. Spatial Information Retrieval

Spatial information retrieval is mainly used to acquire, display, and analyze ground objects. Spatial index plays a crucial role in filtering unrelated ground objects and improving retrieval speed. Recently, researchers have designed several spatial indexes (such as quadtree [26], R-tree [27], and R*-tree [28]) to reduce time and resource cost in retrieving two-dimensional geographic data. Quadtree divides space into four subspaces in a recursive way, until the tree level reaches a preset depth or meets other preset conditions. It has been widely used in GIS because of its simple structure and high spatial data query efficiency. In R-tree and R*-tree indexes, virtual rectangles describing ground objects with close distance are used to form multi-level indexes. Besides, researchers have designed other indexes for retrieving different kinds of data, such as trajectory data and complex polygons [29,30].

All methods mentioned above can improve the efficiency of geospatial data retrieval. However, in terms of retrieving attributes, it is difficult to consider implicit semantics by only processing search terms as character strings. Moreover, in terms of retrieving spatial information, spatial relationships are rarely considered because spatial indexes are constructed in a “mechanical” way. Therefore, traditional data retrieval methods need to be transformed into semantic information retrieval methods.

2.3. Semantic-Oriented Data Retrieval Method

In the 21st century, most researchers mainly focused on semantics to improve information retrieval quality with Semantic Web technology developing [31]. For example, Guha R et al. [32] argued that semantics could be used to acquire more results and improve result accuracy by expanding retrieval scope and understanding the implicit semantics of search terms. The University of Maryland designed a semantic web search engine to provide document retrieval service by calculating metadata similarity [33]. The Open University launched Watson, a semantic web search engine, to provide document retrieval service, which calculated entities similarity in document [34]. Falcons [35] and Hermes [36] systems were developed to provide data retrieval services by calculating similarity of entities and relationships. Besides, geospatial data retrieval is an essential branch of traditional semantic retrieval and studied from two perspectives: data integration and query expansion. The related researches are reviewed as follows.

2.3.1. Semantic-Oriented Data Integration

Geospatial data integration is one of the key technologies for multi-source heterogeneous data retrieval, and its ability to overcome semantic heterogeneity has attracted more and more attention in GIS [37]. To represent geospatial information formally, Semantic Web technology is used to materialize original data into Resource Description Framework (RDF) triples. In order to integrate geospatial data and overcome semantic heterogeneity barriers, Semantic Web technology has been introduced in the application areas of trajectory data mining, earthquake disaster emergency response, and ocean data discovery [38,39]. Moreover, KG is taken as another semantic technology to integrate heterogeneous knowledge in spam detection and movie recommendation [40]. However, the processing cost will be high if data changes frequently. Ontology-based Data Access (OBDA), a popular data integration approach, enables users to access original data through semantic information in an ontology [41]. For instance, Zhang et al. [42] defined a general ontology and provided a uniform interaction paradigm for retrieving geospatial information. Furthermore, Ontology-based Data Integration (OBDI), generating from OBDA [3], can not only retrieve data from multiple sources but also import all of them into a unique geospatial database [43]. To realize OBDI, Relational Database to RDF Mapping Language (R2RML) is adopted to associate semantics with original databases and publish geographic features as RDF graph [44].

2.3.2. Query Expansion

In the last century, string matching had reached a technical bottleneck and, hence, Van Rijsbergen [45] proposed query expansion techniques to reflect original query intention and improve retrieval performance. Query expansion can parse search terms to form a new and comprehensive collection, reflecting the original query intention. Recently, most researchers expanded search terms by word co-occurrence. For example, Voorhees et al. [46] calculated the co-occurrence probability between words to obtain correlations of original data, so that the words with high probability can be added to enlarge vocabularies. Xu et al. [47] estimated the relevance of retrieval results to expand search terms. Based on query records, Cui et al. [48] added the retrieved search terms to expanded vocabularies, thus reflecting query intention. Besides, many researchers attempted to expand search terms from the semantic level. From the aspect of technology, Voorhees et al. [46] and Navigli [49] calculated conceptual similarity and obtained more similar concepts to expand search terms based on ontology hierarchy. In terms of applied research, the attribute and spatial information of geospatial features are expanded in maritime safety [10] and smart city project [50]. Ji et al. [51] proposed an ontology-based semantic query expansion model to improve semantic retrieval accuracy in agriculture and forestry.

Researchers have accumulated rich experience in data integration and retrieval. However, most of the current methods are based on domain ontologies or general

knowledge bases (such as the artificial intelligence project Cyc (<https://www.cyc.com/>) [12] and the lexical database WordNet [13]), thus lacking semantic knowledge oriented to specific geospatial data set. The application of semantic retrieval is still primary in GIS, and it may be the research focus to find out how to improve the retrieval quality based on KG.

3. Approach

3.1. Retrieval Technology Framework

3.1.1. Basic Abstract Technology Framework

The ontology describes semantic concepts and relationships through graph structure, where nodes represent concepts and edges define relationships [52]. Diego Calvanese [53] proposed the OBDA technology framework in 2007 to integrate, share and access semantic information. OBDA consists of Terminological Box (TBox) and Assertional Box (ABox), which are used to represent the relationships between concepts, entities, and attributes [54].

Generally, KG has schema layer (SLayer) and data layer (DLayer), that is, $KG = \langle SLayer, DLayer \rangle$ [55]. Schema layer constructs concepts and their relationships based on ontology, while data layer constructs entities, attributes, and relationships based on Semantic Web Technology. As shown in Figure 6, virtual ABox is adopted in this paper to implement OBDA, where ABox is taken as an independent syntax object. To achieve data access, mapping (M) is used to associate geospatial relational database (S) with TBox in schema layer (i.e., $SLayer = \langle TBox, M, S \rangle$). Entity (E) and Relationship (R) are constructed to integrate spatial and semantic information of ground objects in data layer (i.e., $DLayer = \langle E, R \rangle$).

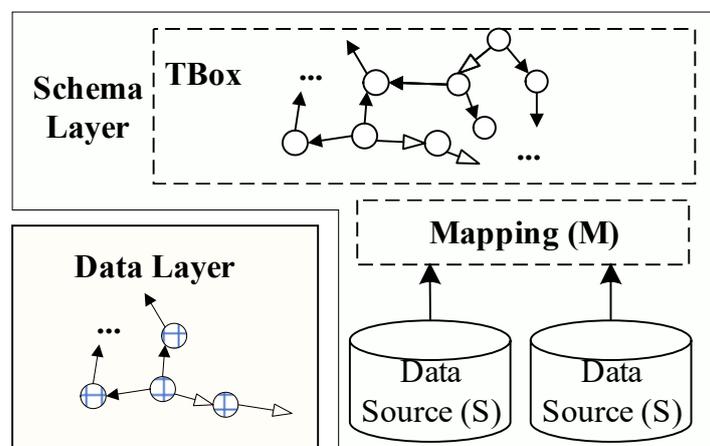


Figure 6. Abstract technology framework of the proposed method.

3.1.2. Retrieval Process

A data retrieval process (Figure 7) is proposed based on the basic abstract technology framework, including KG construction, semantic query expansion, mapping design, entity query expansion, and retrieval result return. The details are as follows:

1. Knowledge graph construction. The correlation is established between schema layer and geospatial database. The data layer relationships are completed by extracting knowledge from the database, including map layers, geographic features, and property fields.
2. Semantic query expansion. The concept of a search term (Q) is matched with the concepts in schema layer. Based on conceptual hierarchical relationships and description logic axioms, query expansion rewrites the search term into related concepts (Q') to reflect query intention.

3. Mapping design. Mapping rules represent the correlation between geospatial databases and concepts. Based on these rules, SQL statements (Q'') are automatically constructed by mapping search terms onto table names, property fields and values.
4. Entity query expansion. Q'' is delegated to geospatial database after designing mapping rules. Moreover, data layer can expand the entities associated with search terms based on the constraints of concept types, administrative divisions, and cognitive styles.
5. Retrieve database and return results. The geospatial database is retrieved through the above steps, and retrieval results are displayed in a multi-view mode. Hence, the method can provide more implicit information and meet query requirements.

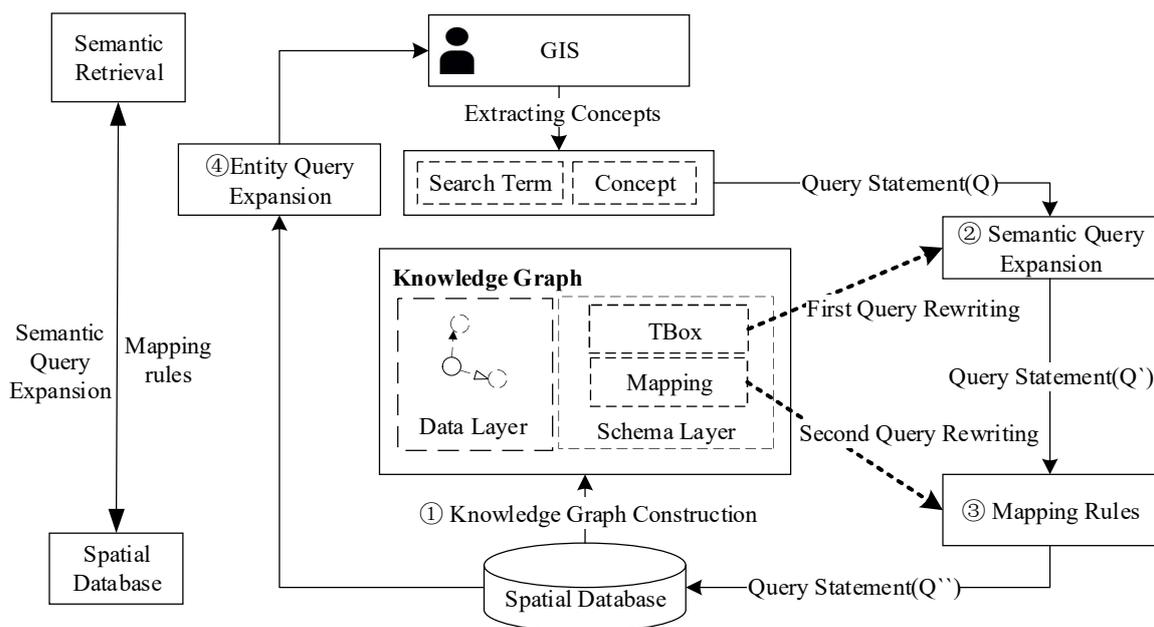


Figure 7. Data retrieval process.

3.1.3. Retrieval Method Characteristics

Compared with relational database, the “point-edge” structure of KG can improve data retrieval flexibility. The characteristics of the method are as follows:

- Data-Centered Knowledge Graph Construction

A domain knowledge graph is generally constructed by extracting experts’ knowledge. However, traditional KG does not exist explicit domain boundaries covering the retrieval data, leading to a lower retrieval result accuracy.

KG construction centers on original data. In detail, schema layer is reversely generated from database to constrain knowledge in data layer. Constructing KG in a bottom-up way can limit the domain scope, complete the conceptual system, and approximate expert knowledge. Furthermore, the close integration between the retrieval process and original data lays a solid foundation for query expansion and rewriting.

- Relationship-Dependent Retrieval Process

The retrieval process mainly relies on the relationships in KG, which are generally represented by edges. These relationships can indicate mapping rules, expand concepts and entities, and construct SQL statements.

- Knowledge-Centered Retrieval Result

Data are a generalization of objective things in the form of number, character, and image. Although it can reflect the real world and human thoughts, only interpreted and processed data can be transformed into knowledge. Associated concepts, attributes and entities are displayed in various ways, such as maps and force-directed graphs.

3.2. Data Integration

A complete and detailed concept system is a prerequisite for retrieval. This paper extracts geospatial knowledge and establishes their relationships to integrate data.

3.2.1. Standard Ontology

Table 1 shows the list of some prefixes and Uniform Resource Locator (URL) used in schema layer. In practice, xml, XML Schemas Definition (xsd), RDF, Resource Description Framework Schema (rdfs), and Web Ontology Language (owl) are basic prefixes, while geo and sf in GeoSparql are the extensional prefixes for GIS. In schema layer, concepts and relationships are expanded as follows:

- Concepts of GeoEntity, GeoGraphicDatasetEntity, and GeoBaikEntity are created to integrate geospatial features and represent geospatial entities' origins.
- Relationship hasFeature inheriting from owl:ObjectProperty is used to represent the association between geographic entities and geographic features.
- Relationship hasProperty inheriting from owl:ObjectProperty can associate database with schema layer.

Table 1. Prefixes in knowledge graph schema layer.

Prefix	URL
xml:	http://www.w3.org/XML/1998/namespace/
xsd:	http://www.w3.org/2001/XMLSchema#
rdf:	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs:	http://www.w3.org/2000/01/rdf-schema#
owl:	http://www.w3.org/2002/07/owl#
sf:	http://www.opengis.net/ont/sf#
geo:	http://www.opengis.net/ont/geosparql#

3.2.2. Semantic Knowledge Extraction

Geospatial data is a compression of three-dimensional space based on layers, features, and fields, and its type code follows the national, industrial, and regional coding standards [24]. In geospatial data, a geographic name includes general name and proper name. General name can distinguish the type of geographic feature and be mapped onto concepts [25]. For example, general names Province and Railway Station refer to the feature types of Henan Province and Zhengzhou Railway Station, respectively.

The conceptual hierarchical relationships are extracted from geospatial database by a data-driven method. Firstly, layer names and type property fields are used to extract relationships between feature types and database layers. As shown in Figure 8a, the layer concepts Transportation Warehousing and Highway are extracted, and the concept Bridge is also obtained from the type property field. Therefore, parent-child relationships of these concepts can be further established in schema layer, such as <Bridge, is-a, Transportation Warehousing>. Then, other concepts are extracted from general names to construct parent-child relationships with the type property field. Figure 8b shows the concepts (including Railway Station, Passenger Railway Station, and Freight Railway Station) and their parent-child relationships (such as <Passenger Railway Station, is-a, Railway Station> and <Freight Railway Station, is-a, Railway Station>). Finally, BaiduBaik, an encyclopedia, is used to complement concepts and relationships extracted from the geospatial database. As shown in Figure 8c, Transportation Facility and Station are extracted and associated with Railway Station through the parent-child relationship.

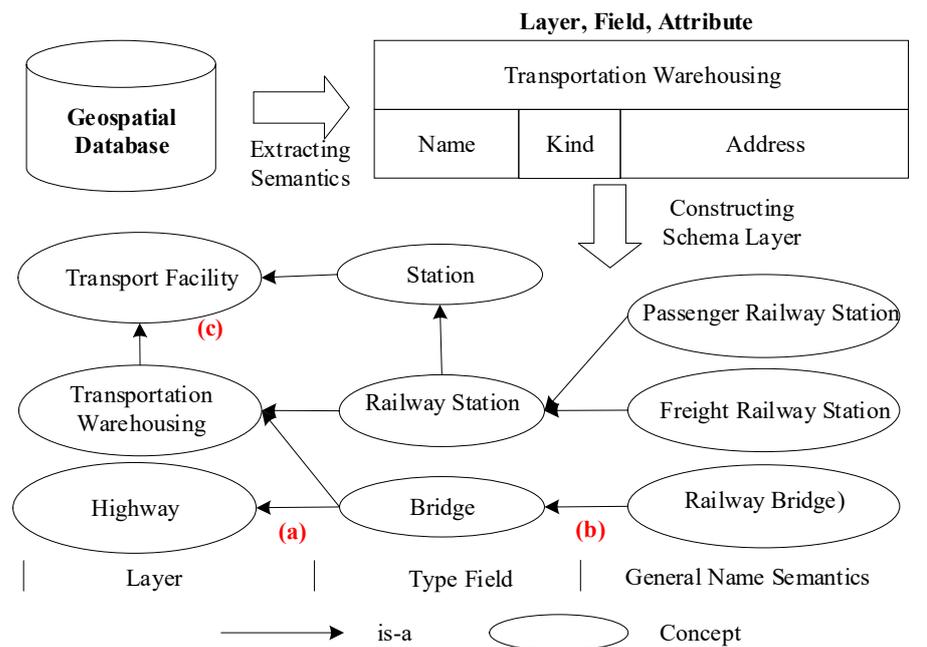


Figure 8. Semantic knowledge extraction from geospatial database. (a) Concepts extracted from layers. (b) Concepts extracted from property fields. (c) Concepts extracted from encyclopedia.

This paper establishes conceptual relationships based on the Rules for the Classification and Coding of Chinese Geographic Names (Figure 9a). Therefore, the schema layer is complemented by extracting concepts in coding rules and establishing conceptual relationships (Figure 9). Moreover, the description of geospatial database property fields is added by the hasProperty relationship, and instances of property fields are used to represent the types of geographic features. As shown in Figure 9b, Transportation Warehousing contains property fields Name and Kind, and they associate with instances of “230103” and “230107”.

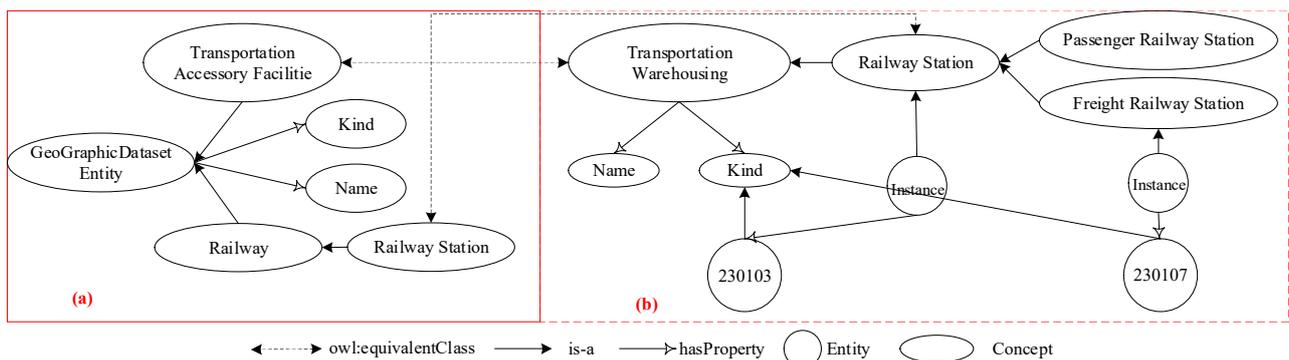


Figure 9. An example of schema layer. (a) Concepts extracted from encoding specification. (b) Concepts extracted from geospatial database.

3.2.3. OSM Semantic Knowledge Extraction

The XML data file of OSM is a list of multiple items (including Node, Way, and Relation) based on the Free Tagging mechanism (Figure 10). By traversing the OSM item list (file size 13.4 GB), concepts are extracted from “key-value” tags to construct parent-child relationships. In detail, the prefix “k_” is added to Key and taken as a parent concept, while the prefix “v_” is added to Value and taken as a child concept. As shown in Figure 10, the tags of Zhengzhou Railway Station (i.e., “<tag k = “public_transport” v = “station”/>” and “<tag k = “railway” v = “station”/>”) are used to extract concepts (including

“k_public_transport”, “k_railway”, and “v_station”) and their parent–child relationships (including <v_station, is-a, k_public_transport> and <v_station, is-a, k_railway>). Additionally, equivalentClass is used to associate equivalent concepts in schema layer to complete semantic relationships based on crowdsourcing data, such as <v_station, owl:equivalentClass, Railway Station>. By integrating multi-source geospatial data, schema layer can clearly define geographic features and their types, thus improving the standardization of semantic knowledge.

```
<node id="269806844" version="10" timestamp="2020-10-20T11:12:44Z"
  lat="34.7475076" lon="113.6536663">

  <tag k="name:en" v="Zhengzhou"/>
  <tag k="public_transport" v="station"/>
  <tag k="railway" v="station"/>
  <tag k="train" v="yes"/>

</node>
```

Figure 10. An example of Zhengzhou Railway Station in the OpenStreetMap file.

3.3. Semantic Query Expansion

3.3.1. Semantic Similarity Calculating

It is crucial to calculate concept similarity in semantic query expansion. Linguistic researchers argued that there is an inverse relationship between word distance and concept similarity [56]. Taking concepts C_1 and C_2 as examples, $Sim(C_1, C_2)$ records their semantic similarity. The greater the distance between them, the lower the similarity, and vice versa. The similarity between concepts can be expressed as follows:

$$Sim(C_1, C_2) = \begin{cases} 0, & D = +\infty \\ P, & D \in (0, +\infty) \\ 1, & D = 0 \end{cases} \quad (1)$$

If the distance between two concepts is zero, C_1 and C_2 are connected by equivalentClass. Therefore, this paper calculates the semantic similarity P , whose distance range is between 0 and infinity in schema layer. Concept depth indicates the degree of semantic specialization in KG. The greater the depth, the more detailed the concept [57]. Semantic similarity considering concept depth is calculated as:

$$Sim(C_1, C_2) = \frac{1}{Dis(C_1, C_2) \times y + 1} \quad (2)$$

where $Dis(C_1, C_2)$ represents the shortest distance between concepts C_1 and C_2 , and y represents the weighting coefficient. Parameter y balances semantic similarity, thus taking as the reciprocal of maximum depth in schema layer. In this paper, concepts have been successively extracted from four levels: database, layer, property field, and general name. Hence, y is taken as 0.25 to quantify conceptual similarity.

3.3.2. Semantic Query Expansion Type

Semantic query expansion mainly uses semantic similarity to expand related concepts in schema layer [10]. Compared with conventional retrieval methods, conceptual relationships are adopted to expand the search term based on various inference rules. Semantic query expansion includes synonym extension, attribute extension, and hierar-

chical extension (Table 2). The most basic type is synonym extension, enabling the proposed method to obtain several synonyms through equivalentClass. For example, concepts of Transportation Warehousing and Transportation Facility can be obtained through synonym extension. Attribute extension can acquire related concepts through object properties. For example, concepts of Name and Type are obtained from concept Transportation Facility through attribute extension. Hierarchical extension determines the concept to which an entity or a concept belongs, and it can expand or narrow down the scope of concepts based on parent-child relationships. For instance, Freight Railway Station can obtain Railway Station and Transportation Warehousing.

Table 2. Semantic query expansion type [58].

Query Expansion Type	Description
Synonymous extension	Obtaining concepts by equivalentClass, which are equivalent to the concept extracted from the search term.
Attribute extension	Obtaining concepts by hasProperty, which are related to the concept extracted from the search term.
Hierarchical extension	Expanding or narrowing the scope of concepts by parent-child relationships.

3.3.3. Semantic Query Expansion Principle

In order to expand concepts in schema layer, the concept set $C=\{C_i | i \in N\}$ matching with the search term is taken as a search condition, and the relationship $R_m(m \in N)$ is regarded as a semantic relationship between concept C and others. Therefore, the expanded concepts can be defined as $C_{ik} = \{C_k | R_m(C_k, C_i) \text{ or } R_m(C_i, C_k), i \in N, k \in N\}$, where R_m includes conceptual equivalent relationship, object property relationship, and conceptual hierarchical relationship. Taking Niulanshan Town Passenger Railway Station as an example, conceptual hierarchical relationships can be used to extract concepts (i.e., Railway Station and Transportation Warehousing) from Passenger Railway Station. Then, Transportation Facility can be obtained by conceptual equivalent relationship. Object property relationships can be used to extract type codes corresponding to Passenger Railway Station and Railway Station (i.e., "230103", as shown in Figure 11).

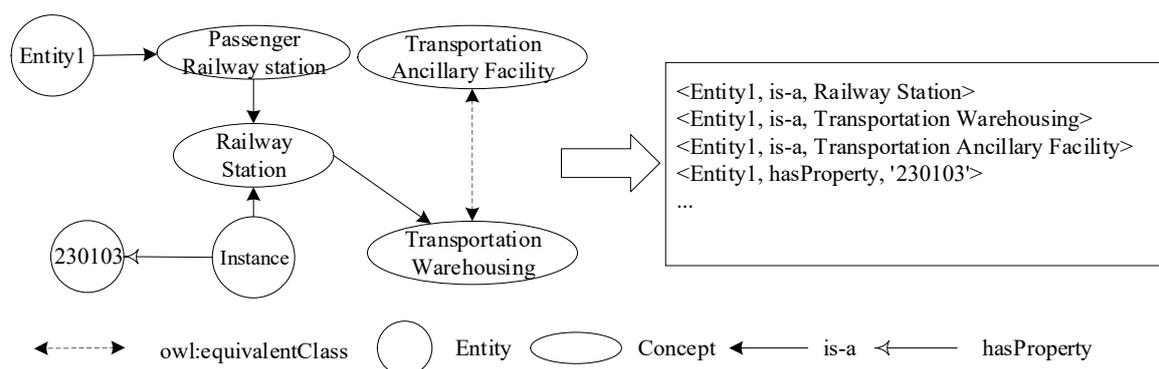


Figure 11. An example of semantic query expansion.

The expanded concepts are filtered and sorted by calculating semantic similarity, and the concepts with high similarity can be regarded as retrieved concepts. By Formula (2), the semantic similarity value between Passenger Railway Station and Transportation Warehousing is 0.6, while the value between Passenger Railway Station and Railway Station is 0.8. Hence, Railway Station is more in line with users' search intention. Semantic query expansion can obtain related concepts, which are the foundation of subsequent mapping between semantic knowledge and relational database.

3.4. Mapping Design

3.4.1. Mapping Rules Type

Mapping Rules (M) represent relationships between spatial database and schema layer and construct SQL statements [31]. They can transform static semantics into dynamic semantics in real-time, thus bridging the distance between man and machine [59]. Each rule is formally defined as $\Psi(O) \rightarrow \Phi(S)$, where $\Psi(O)$ refers to a retrieval statement in schema layer, and $\Phi(S)$ is a SQL statement in database. Table names, property fields, and types can be obtained through mapping rules. As shown in Table 3, mapping rules of MappingToTable, MappingToField and hasProperty are designed based on R2RML.

Table 3. Mapping rules between schema layer and geospatial database.

Mapping Tag	Mapping Relation Description
MappingToTable	Mapping concepts onto tables in a database.
MappingToField	Mapping concepts onto property fields in a table.
hasProperty	Mapping concepts of tables onto the concepts of property fields.

3.4.2. Representing Tables and Property Fields

Database typically contains definitions of table and property field to ensure data integrity. As shown in Figure 12, Transportation Warehousing includes geographic features (such as railway stations), property fields (such as Name and Kind), and spatial property field (i.e., Geometry). As shown in Table 4, table name and concept are represented in schema layer to map Transportation Warehousing onto Transportation Facility. Moreover, the mapping between property field and concept is constructed. For example, property fields Name and Kind are associated with concepts Name and Type, and spatial property field is also associated with concept sf:Geometry.

Name:String	Kind:String	Geometry
Niulanshan Town Passenger Railway Station	230103	POINT(.....)
Dawan Town Railway Station	230107	POINT(.....)

Figure 12. An example of geospatial database.

Table 4. Semantic information representation between schema layer and geospatial database.

Mapping Type	Database Representation	Schema Layer Representation
Table name and concept	{x Transportation Warehousing(x)}	{x Transportation Facility(x)}
Spatial field and concept	{x Geometry(x)}	{x sf:Geometry(x)}
	{x Name(x)}	{x Name(x)}
Property field and concept	{x Kind(x)}	{x Type(x)}

3.4.3. Mapping Relationships Construction

Mapping can show the correlation between schema layer and geospatial database based on the representation of tables and property fields. Description logic rules (such as “ \subseteq ” and “ \exists ”) are introduced to represent relationships of parent–child and conceptual existent. The mappings representing description logic statements are shown as follows:

m1: Transportation Facility $\subseteq \exists$ MappingToTable.Transportation Warehousing.

m2: Name $\subseteq \exists$ MappingToFiled.Name.

m3: Type $\subseteq \exists$ MappingToFiled.Kind.

m4: sf:Geometry $\subseteq \exists$ MappingToFiled.Geometry.

m5: Transportation Facility $\subseteq \exists$ hasProperty. Name.

m6: Transportation Facility $\subseteq \exists$ hasProperty. Type.

m7: Railway Station $\subseteq \exists$ hasProperty.Type \cap (Type("230103") \cup).

m8: v_station $\subseteq \exists$ hasProperty.Type \cap (Type("230103") \cup).

The tag MappingToTable is used to transform concepts to map layers; therefore, m1 maps concept Transportation Facility onto Transportation Warehousing. The tag MappingToFiled refers to the transformation from concepts to property fields. For example, m3 maps concept Type onto property field Kind, while m4 maps the concept Geometry onto a geometric field. Moreover, the tag hasProperty is used to correlate concepts and property fields. For instance, m5 and m6 indicate that the concept Transportation Facility contains property fields of Name and Type. Geographic entity type is represented by tag hasProperty and Type code. For example, m7 and m8 indicate that the Type code of Railway Station and v_station is "230103".

3.4.4. SQL Statement Construction

Semantic query information can be converted into SQL statements through the proposed method. The table names and property fields are automatically obtained based on mappings. SQL statements are then constructed from instances of Type and the search term. As shown in Table 5, the table name Transportation Warehousing and its property fields (i.e., Name and Kind) are obtained based on tags MappingToTable and MappingToFiled. The type code (i.e., "230103") is also obtained from tag hasProperty. Hence, a SQL statement (i.e., "Select * From 'Transportation Warehousing' Where Kind = '230103' AND Name LIKE '%Niulanshan Town Passenger Railway Station%'") is automatically constructed to retrieve Niulanshan Passenger Railway Station. In schema layer, concepts used to build SQL statements can be returned as additional information to explain retrieval results.

Table 5. Structured Query Language statements construction. The wildcard "*" means select all property fields.

Concept	Relationship	SQL Statement
Transportation Facility (x)	Transportation Facility $\subseteq \exists$ MappingToTable. Transportation Warehousing	Select * From 'Transportation Warehousing'
Railway Station(x) or v_station(x)	Railway Station \subseteq Transportation Facility v_station \subseteq Transportation Facility Transportation Facility $\subseteq \exists$ MappingToTable. Transportation Warehousing Railway Station $\subseteq \exists$ hasProperty.Type \cap (Type("230103"))... v_station $\subseteq \exists$ hasProperty.Type \cap (Type("230103"))...	Select * From 'Transportation Warehousing' Where Kind = '230103'...

3.5. Geographic Entity Query Expansion

Geographic entities are spatially related to others in the real world [60]. Zhang et al. [61] argued that the similarity of two geospatial objects should be quantified from spatial relationships. Hence, it is necessary to expand the retrieval result through spatial relationships: type, administrative region, and affiliation.

From the perspective of administrative regions, geographic entities are obtained through GIS operations because they adhere to the same concept or the same administrative region. Additionally, entities associated with the search term are also retrieved in the OSM file. In terms of affiliation, although a large number of geographic entities belong to different concepts, they are closely related in the real world and are usually combined for recognition. For example, railway contains affiliations of railway station and railway bridge. Therefore, railway can be regarded as a central geographic entity, and other entities (such as railway station and railway bridge) along it could be extracted to expand retrieval results in data layers. As shown in Table 6, some affiliation relationships are designed to realize entity query expansion. In addition, a list of relationships associated with the search term could be returned to explain retrieval results in data layers.

Table 6. Affiliation relationships.

Relationship Description	Data Types
Road affiliation	Roads contain service areas, toll stations, gas stations, etc.
Railway affiliation	Railways contain railway stations, railway bridges, etc.
River affiliation	Rivers contain bridges, ferries, etc.

4. Experiment and Analysis

4.1. Evaluation Index

The result evaluation is an activity to assess whether the method satisfies users' needs, thus improving the geospatial data retrieval result. In 1995, Saracevic [62] proposed the Cranfield evaluation system, including sample set, correct answers, and evaluation indexes. On this basis, retrieved relevant entities, retrieved entities, and relevant entities are used to calculate Recall (R), Precision (P), and F -Score, among which F -Score is taken as a compromise between R and P [63]. These evaluation indexes are calculated as follows:

$$R = \frac{\text{The number of retrieved relevant entities}}{\text{The total number of relevant entities in the data}} \quad (3)$$

$$P = \frac{\text{The number of retrieved relevant entities}}{\text{The total number of retrieved entities}} \quad (4)$$

$$F = \frac{2 \times R \times P}{R + P} \quad (5)$$

where R is the ratio of the number of retrieved relevant entities to the total number of relevant entities in original data. P is the ratio of the number of retrieved relevant entities to the total number of retrieved entities. Besides, F -Score is the harmonic mean of R and P .

4.2. Retrieval Result Analysis

4.2.1. Retrieval Efficiency

LIKE and WHERE are two operators in SQL, which can be used to retrieve data in relational database, thus comparing the efficiency in terms of SQL statements. LIKE statement retrieves geographic features by iterating data and, therefore, its time complexity is $O(N)$, where N is the number of geographic features in a map layer. In this paper, type codes are extracted from KG to refine a filtering statement. Therefore, the time complexity of the proposed method is $O(\log(Q)) + O(M)$, where Q is the number of feature types in a map layer and M is the number of features with the same type code. Q and M are much smaller than N in raw data, so that the proposed method takes less time. In detail, a WHERE statement constructed through a conventional method is "WHERE Name LIKE '%...%'", while the proposed method automatically adds type codes to a filter statement (i.e., "WHERE Kind = '...' AND Name LIKE '%...%'"). As shown in Table 7, Transportation Warehousing contains bridge, flyover, toll station, charging station, gasoline station, etc., classified by type codes. The conventional method retrieves an entity by iterating data in Transportation Warehousing, taking about 430 ms. The type codes are extracted to refine the filter statement in the proposed method, thus reducing time cost. For example, retrieving an entity of Freight Railway Station consumes 5 ms.

Table 7. The time cost.

No.	Retrieval Concept	Type Code	Features Number	Time (ms)	
				Conventional Method	Proposed Method
1	Bridge	230201, 230202	202,112	429	183
2	Flyover	230202	21,944	439	34
3	Toll station	230209	19,223	428	14
4	Charging station	230218	14,884	462	23
5	Gasoline station	230215, 230217	104,036	428	113
6	Gas station	230216, 230217	7489	420	32
7	Station	230100, 230103, 20107	13,522	429	30
8	Railway station	230103, 230107	10,771	426	22
9	Freight railway station	230107	1413	421	5
10	Parking lot	230212, 230225, 230211	258,211	424	333

4.2.2. Retrieval Quality

Further experiments are performed for geographic entities in multiple layers or a single layer. SQL statements are used to retrieve entities and, thus, the precision is closely related to the search term. If the search term is unique in the database, the precision can be 100%. $Count_A$ and $Count_B$ are defined as the numbers of identical type features in layers A and B, respectively and, hence, the recall of retrieving a single layer is a ratio of $Count_A$ or $Count_B$ to the sum of them. The conventional method needs to specify layers, while the proposed method can retrieve entities in multiple layers based on KG, increasing the recall to 100%. As shown in Table 8, taking Hong Kong-Zhuhai Bridge as an example, recall increases from 97.5 to 100% in layer A and from 2.5 to 100% in layer B. Besides, the conventional method cannot obtain results without specifying layers. In comparison, SQL statements are automatically constructed to retrieve geographic entities through the proposed method (as shown in Table 8, retrieving Zhengzhou Railway Station and Lianhuo Highway in Layers A and B).

Table 8. Comparison of experiment results. Layers A and B represent Highway and Transportation Warehousing, respectively.

Experiment Type	Search Term	Features (A)	Features (B)	Proposed Method			Method (A)			Method (B)		
				R (%)	P (%)	F	R (%)	P (%)	F	R (%)	P (%)	F
Entities in multiple layers	Hong Kong-Zhuhai Bridge	39	1	100	100	1	97.5	100	0.98	2.5	100	0.04
	Zhengzhou Railway Station	0	1	100	100	1	0	0	0	100	100	1
Entities in a single layer	Lianhuo Highway	10,344	0	100	100	1	100	100	1	0	0	0

4.3. Geospatial Data Retrieval Example

In this paper, the map layers of China Navigation Map (<https://www.navinfo.com>, accessed on 1 September 2019) are used as the experiment data (i.e., transportation warehousing, highway, and railroad), containing 37 geographic entity types such as railway station and passenger railway station.

4.3.1. Conventional Retrieval Method

The conventional retrieval method constructs SQL statements based on specified map layers and property fields. However, it is hard to build SQL statements automatically.

4.3.2. Proposed Method

In the proposed method, the concepts are extracted, obtained, and expanded from the search term to build SQL statements, and then results are retrieved automatically. Taking Zhengzhou Railway Station as an example, semantic query expansion can obtain concepts (i.e., Railway Station, Transportation Facility, Name, and Type) and type codes (i.e., “230103” and “230107”) from general name Railway Station. The mapping rules can get a table name (i.e., Transportation Warehousing) and field names (i.e., Name and Kind) to build a SQL statement (i.e., “Select * From ‘Transportation Warehousing’ Where (Kind = ‘230103’ OR Kind = ‘230107’) AND Name LIKE ‘%Zhengzhou Railway Station%’”). Besides, Figure 13 shows the retrieval result in a multi-view GIS platform.

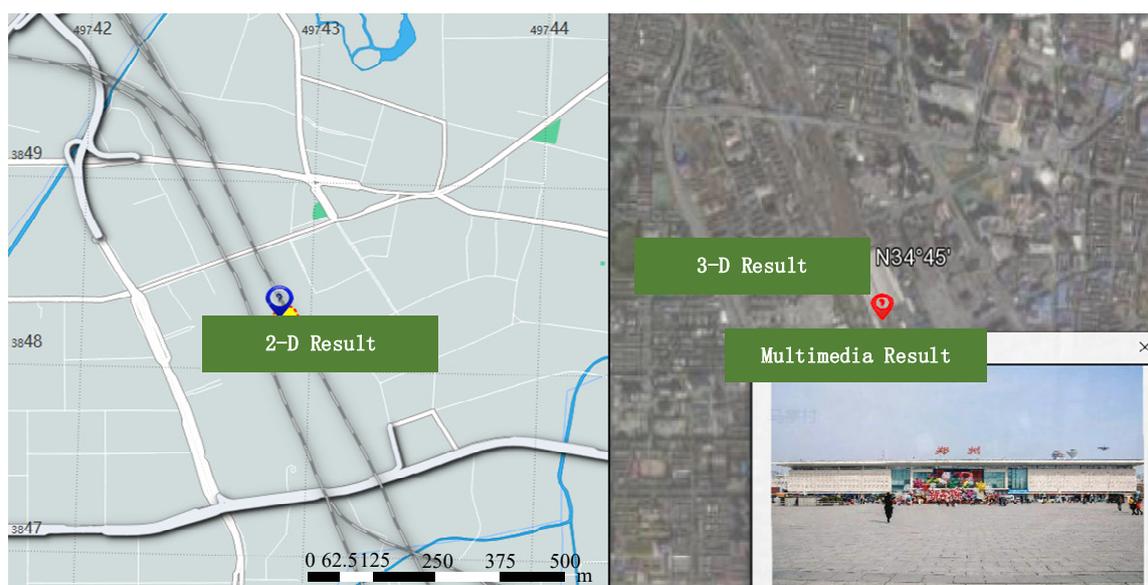
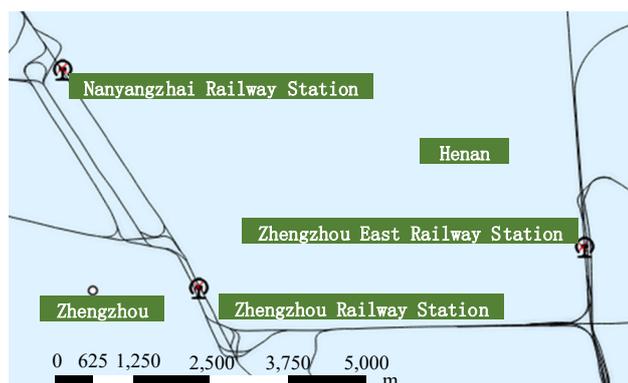


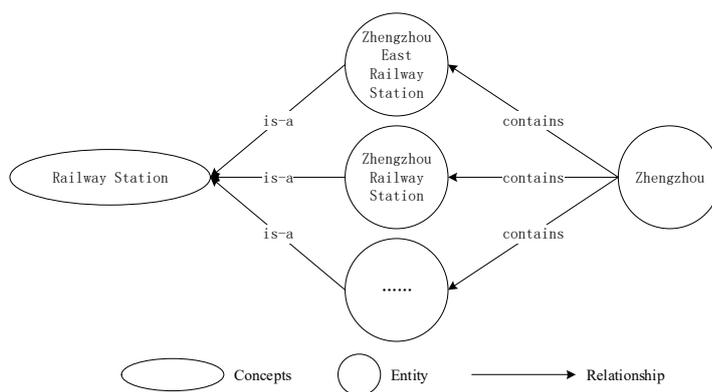
Figure 13. Retrieval result in the proposed method.

4.3.3. Entity Query Expansion Example

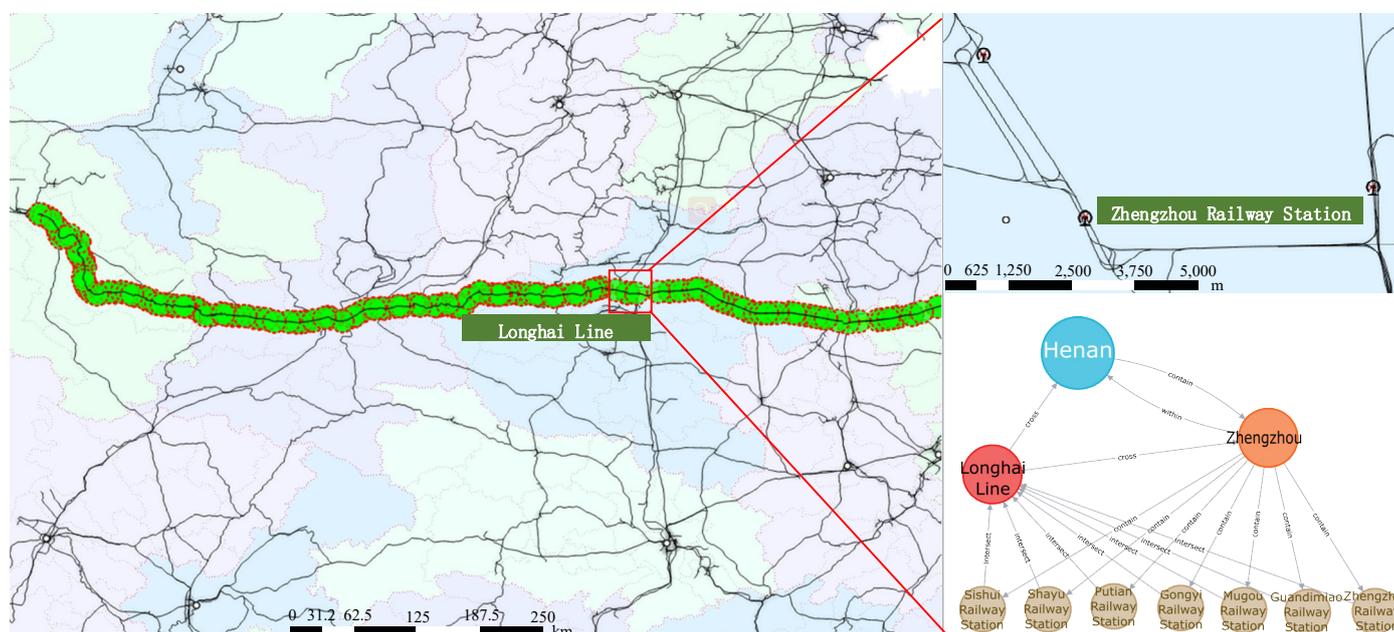
Entities can be expanded through their types and administrative regions in data layer. As shown in Figure 14a, there are some railway station entities in Zhengzhou, such as Zhengzhou Railway Station, Zhengzhou East Railway Station, and Nanyangzhai Railway Station. In order to retrieve related entities of Zhengzhou Railway Station, administrative region and type are used to get entity Zhengzhou East Railway Station (Figure 14b). The entity can be expanded through its affiliations. Taking Longhai Line as a search term, Figure 14c shows railway stations along Longhai Line. Besides, a large number of affiliated geographic entities along the railway can be extracted by GIS spatial operators in real-time.



(a) Map



(b) An example of a knowledge graph



(c) Entity query expansion based on affiliations

Figure 14. Examples of entity query expansion. (a) Map; (b) An example of a knowledge graph; (c) Entity query expansion based on affiliations.

5. Conclusions

The number of geospatial datum has increased rapidly in the big data era. However, traditional geospatial data retrieval methods require the understanding of database storage structure, and OBDA is difficult to obtain entities with large semantic differences. In this paper, a new retrieval method is put forward to retrieve geospatial data based on KG, which is constructed from heterogeneous geospatial data and encyclopedias. In order to retrieve data, semantic expansion, SQL statement construction, and entity expansion are, respectively, implemented based on schema layer, mapping rules, and data layer. The geospatial data retrieval method is verified through practices and comparison analysis. The experimental results indicate that the retrieving process could be simplified and the quality and efficiency in retrieving geospatial data could be improved. Furthermore, the proposed method can integrate semantics and spatial knowledge, build SQL statement automatically, and retrieve more implicit knowledge.

In the future, the authors will focus on how to illustrate results based on explainable reasoning to meet users' needs better. Besides, how to realize raster data retrieval will be a crucial research direction.

Author Contributions: Conceptualization, J.L. (Junnan Liu); Methodology, J.L. (Junnan Liu), H.L., and X.C.; Software, J.L. (Junnan Liu), Q.Z., J.L. (Jia Li), L.K., J.L. (Jianxiang Liu) and X.G.; Resources, H.L.; Writing—Original Draft Preparation, J.L. (Junnan Liu) and X.G.; Writing—Review and Editing, J.L. (Junnan Liu), H.L., X.C., and X.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 41801313 and the Natural Science Foundation of Henan Province, grant number 182300410005.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data of the current study are available from the corresponding author based on a reasonable request.

Acknowledgments: We would like to thank the anonymous reviewers for their insightful comments and substantial help on improving this article.

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

References

- Hahmann, S.; Burghardt, D. How Much Information is Geospatially Referenced? Networks and Cognition. *Int. J. Geogr. Inf. Sci. Ijgis* **2013**, *27*, 1171–1189.
- Aloteibi, S.; Sanderson, M. Analyzing Geographic Query Reformulation: An Exploratory Study. *J. Assoc. Inf. Sci. Technol.* **2014**, *65*, 13–24.
- Ding, L.; Xiao, G.; Calvanese, D.; Meng, L. A Framework Uniting Ontology-Based Geodata Integration and Geovisual Analytics. *ISPRS Int. J. Geo Inf.* **2020**, *9*, 474.
- Zhang, C.; Zhang, X.; Ji, L.; Wang, H. Relation Mapping between Generic Terms of Place Names and Geographical Feature Types. *Geomat. Inf. Sci. Wuhan Univ.* **2011**, *36*, 857–861.
- Zhang, X.; Liu, J.; Wang, Y.; Luo, A. An Semantics Extended Framework for Spatial Direction Relation Query Based on Natural Language. *Geogr. Geo Inf. Sci.* **2018**, *34*, 7–14.
- Lim, S.C.J.; Liu, Y.; Lee, W.B. Multi-Facet Product Information Search and Retrieval Using Semantically Annotated Product Family Ontology. *Inf. Process. Manag.* **2010**, *46*, 479–493.
- Yoo, D. Hybrid Query Processing for Personalized Information Retrieval on the Semantic Web. *Knowl. Based Syst.* **2012**, *27*, 211–218.
- Haklay, M.; Weber, P. Openstreetmap: User-Generated Street Maps. *IEEE Pervasive Comput.* **2008**, *7*, 12–18.
- Xie, M.; Zhou, G.; Li, D.; Gong, J. Design and Implementation of Attribute Database Management System in a Gis System: Geostar. *Geogr. Inf. Sci.* **2000**, *6*, 170–180.
- Brüggemann, S.; Bereta, K.; Xiao, G.; Koubarakis, M. Ontology-Based Data Access for Maritime Security. In *European Semantic Web Conference*; Springer: New York, NY, USA, 2016; pp. 741–757.
- Giese, M.; Soylu, A.; Vega-Gorgojo, G.; Waaler, A.; Haase, P.; Jiménez-Ruiz, E.; Lanti, D.; Rezk, M.; Xiao, G.; Özçep, Ö. Optique: Zooming in on Big Data. *Computer* **2015**, *48*, 60–67.
- Lenat, D.B. Cyc: A Large-Scale Investment in Knowledge Infrastructure. *Commun. ACM* **1995**, *38*, 33–38.
- Miller, G.A. Wordnet: A Lexical Database for English. *Commun. ACM* **1995**, *38*, 39–41.
- Wu, T.; Qi, G.; Li, C.; Wang, M. A Survey of Techniques for Constructing Chinese Knowledge Graphs and their Applications. *Sustainability* **2018**, *10*, 3245.
- Larson, R.R. Geographic Information Retrieval and Spatial Browsing. In *Geographic Information Systems and Libraries: Patrons, Maps, and Spatial Information*; Library Applications of Data Processing: Urbana-Champaign, IL, USA, 1996.
- Jensen, J.; Saalfeld, A.; Broome, F.; Cowen, D.; Price, K.; Ramsey, D.; Lapine, L.; Usery, E.L. *A Research Agenda for Geographic Information Science*; CRC Press: Boca Raton, FL, USA, 2004; pp. 17–60.
- Li, L.; Wang, Q.; Wang, H. GIS Data Management Based on Spatialite Database. *Geomat. World* **2010**, *8*, 71–75.
- Soden, R.; Palen, L. From crowdsourced mapping to community mapping: The post-earthquake work of openstreetmap Haiti. In Proceedings of the 11th International Conference on the Design of Cooperative Systems (COOP 2014), Nice, France, 27–30 May 2014; pp. 311–326.
- Luxen, D.; Vetter, C. Real-time routing with openstreetmap data. In Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, IL, USA, 1–4 November 2011; pp. 513–516.

20. Huber, S.; Rust, C. Calculate Travel Time and Distance with Openstreetmap Data Using the Open Source Routing Machine (OSRM). *Stata J.* **2016**, *16*, 416–423.
21. 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; pp. 165–172.
22. Wang, X.; Li, Z.; Jian, Y.; Liu, J. Machine Translation Dictionary Based on Hash Method. *J. Dalian Univ. Technol.* **1996**, *3*, 108–111.
23. Sun, M.; Zuo, Z.; Huang, C. An Experimental Study on Dictionary Mechanism for Chinese Word Segmentation. *J. Chin. Inf. Process.* **2000**, *14*, 1–6.
24. Li, J.; Zhou, Q.; Chen, Z. A Study on Fast Algorithm for Chinese Dictionary Lookup. *J. Chin. Inf. Process.* **2006**, *20*, 31–39.
25. Ye, P.; Zhang, X.; Du, M. Query Method of Chinese Gazetteer Based on the Character Features. *J. Geo Inf. Sci.* **2018**, *20*, 880–886.
26. Tanaka, S. Performance Improvement of MX-CIF Quadtree by Reducing the Query Results. *International Journal of Computer Theory and Engineering.* **2012**, *4*, 902–906.
27. Jin, P.; Xie, X.; Wang, N.; Yue, L. Optimizing R-Tree for Flash Memory. *Expert Syst. Appl.* **2015**, *42*, 4676–4686.
28. Roumelis, G.; Vassilakopoulos, M.; Corral, A.; Manolopoulos, Y. Efficient Query Processing on Large Spatial Databases: A Performance Study. *J. Syst. Softw.* **2017**, *132*, 165–185.
29. Xiang, L.; Gao, M.; Wang, D.; Gong, J. A Quadtree Spatial Index Method with Inclusion Relations for Complex Polygons. *Geomat. Inf. Sci. Wuhan Univ.* **2019**, *44*, 436–442.
30. Wang, H.; Zhou, X. A Quadtree Spatial Index Method with Inclusion Relations for Complex Polygons. *J. Hunan Univ. Nat. Sci.* **2020**, *47*, 99–109.
31. Yang, Y.; Du, J.; Ping, Y. Ontology-Based Intelligent Information Retrieval System. *J. Softw.* **2015**, *26*, 1675–1687.
32. Guha, R.; McCool, R.; Miller, E. Semantic Search, In Proceedings of the 12th International Conference on World Wide Web, Budapest, Hungary, 20–24 May 2003; pp. 700–709.
33. Ding, L.; Finin, T.; Joshi, A.; Pan, R.; Cost, R.S.; Peng, Y.; Reddivari, P.; Doshi, V.; Sachs, J. Swoogle: A Semantic Web Search and Metadata Engine. In Proceedings of the 13th ACM Conference on Information and Knowledge Management, Washington, DC, USA, 8–13 November, 2004; pp. 10–1145.
34. Sabou, M.; Dzbor, M.; Baldassarre, C.; Angeletou, S.; Motta, E. Watson: A gateway for the semantic web. In Proceedings of the Poster Session of the European Semantic Web Conference, ESWC, Innsbruck, Austria, 3–7 June 2007.
35. Cheng, G.; Ge, W.; Qu, Y. Falcons: Searching and Browsing Entities on the Semantic Web. In Proceedings of the 17th international conference on World Wide Web, Beijing, China, 21–25 April 2008; pp. 1101–1102.
36. Zhang, L.; Yu, Y.; Zhou, J.; Lin, C.; Yang, Y. An Enhanced Model for Searching in Semantic Portals. In Proceedings of the 14th international conference on World Wide Web, Chiba, Japan, 10–14 May 2005; pp. 453–462.
37. Hong, J.-H.; Kuo, C.-L. A Semi-Automatic Lightweight Ontology Bridging for the Semantic Integration of Cross-Domain Geospatial Information. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 2223–2247.
38. Hu, Y.; Janowicz, K.; Carral, D.; Scheider, S.; Kuhn, W.; Berg-Cross, G.; Hitzler, P.; Dean, M.; Kolas, D. A Geo-Ontology Design Pattern for Semantic Trajectories. In Proceedings of the International Conference on Spatial Information Theory, Scarborough, UK, 2–6 September 2013; pp. 438–456.
39. Xu, J.; Nyerges, T.L.; Nie, G. Modeling and Representation for Earthquake Emergency Response Knowledge: Perspective for Working with Geo-Ontology. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 185–205.
40. Wilcke, X.; Bloem, P.; De Boer, V. The Knowledge Graph as the Default Data Model for Learning on Heterogeneous Knowledge. *Data Sci.* **2017**, *1*, 39–57.
41. Poggi, A.; Lembo, D.; Calvanese, D.; De Giacomo, G.; Lenzerini, M.; Rosati, R. Linking Data to Ontologies. In *Journal on Data Semantics X*; Springer: New York, NY, USA, 2008; pp. 133–173.
42. Zhang, Y.; Li, C.; Liu, S.; Wen, F.; Du, L.; He, H. A Unified Approach to Automate Geospatial Data Retrieval Using Semantic Web Technologies. In Proceedings of the 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), Okayama, Japan, 26–29 June 2016; pp. 1–6.
43. Calvanese, D.; De Giacomo, G.; Lembo, D.; Lenzerini, M.; Poggi, A.; Rodriguez-Muro, M.; Rosati, R.; Ruzzi, M.; Savo, D.F. The Mastro System for Ontology-Based Data Access. *Semant. Web* **2011**, *2*, 43–53.
44. Sequeda, J.; Priyatna, F.; Villazón-Terrazas, B. Relational Database to Rdf Mapping Patterns. In Proceedings of the 3rd International Conference on Ontology Patterns-Volume 929, Boston, MA, USA, 12 November 2012; pp. 97–108.
45. Van Rijsbergen, C.J. Acm Sigir Forum. In *A New Theoretical Framework for Information Retrieval*; ACM: New York, NY, USA, 1986; pp. 23–29.
46. Voorhees, E.M. SIGIR'94. In *Query expansion using lexical-semantic relations*; Springer: New York, NY, USA 1994; pp. 61–69.
47. Xu, J.; Croft, W.B. Acm Sigir Forum. In *Query Expansion Using Local and Global Document Analysis*; ACM: New York, NY, USA, 2017; pp. 168–175.
48. Cui, H.; Wen, J.; Li, M. A Statistical Query Expansion Model Based on Query Logs. *J. Softw.* **2003**, *14*, 1593–1599.
49. Navigli, R.; Velardi, P. An Analysis of Ontology-Based Query Expansion Strategies. In Proceedings of the 14th European Conference on Machine Learning, Workshop on Adaptive Text Extraction and Mining, Cavtat-Dubrovnik, Croatia, 22 September 2003; pp. 42–49.

50. Lopez, V.; Stephenson, M.; Kotoulas, S.; Tommasi, P. Data Access Linking and Integration with Dali: Building a Safety Net for an Ocean of City Data. In Proceedings of the International Semantic Web Conference, Bethlehem, PA, USA, 11–15 October 2015; pp. 186–202.
51. Ji, P.; Xiao, Y.; Hou, R.; Zhang, N. Application of Data Integration Technology to Forestry in China and Its Progress. *World For. Res.* **2018**, *31*, 49–54.
52. Zhu, J.; You, X.; Xia, Q. A Semantic Similarity Calculation Method for Battlefield Environment Elements Based on Operational Task Ontology. *Geomat. Inf. Sci. Wuhan Univ.* **2019**, *44*, 1407–1415.
53. Calvanese, D.; De Giacomo, G.; Lembo, D.; Lenzerini, M.; Poggi, A.; Rosati, R. Ontology-Based Database Access. In Proceedings of the Fifteenth Italian Symposium on Advanced Database Systems, SEBD 2007, Fasano, Italy, 17–20 June 2007; pp. 324–331.
54. Zhang, Y. *Design and Implementation of an Ontology-based Data Access and Integration System*; Zhejiang University: Hangzhou, China, 2018.
55. Liu, J.; Liu, H.; Chen, X.; Guo, X.; Guo, W.; Zhu, X.; Zhao, Q. The Construction of Knowledge Graph towards Multi-Source Geospatial Data. *J. Geo Inf. Sci.* **2020**, *22*, 1476–1486.
56. Chen, Y.; Wu, C.; Guo, X.; Xie, M.; Long, F. Augmenting Collaborative Recommendation by Grouping Synonymy Tags. *J. Comput. Inf. Syst.* **2011**, *7*, 1350–1357.
57. Aleman-Meza, B.; Halaschek-Weiner, C.; Arpinar, I.B.; Ramakrishnan, C.; Sheth, A.P. Ranking Complex Relationships on the Semantic Web. *IEEE Internet Comput.* **2005**, *9*, 37–44.
58. Díaz-Galiano, M.C.; Martín-Valdivia, M.T.; Ureña-López, L. Query Expansion with a Medical Ontology to Improve a Multi-modal Information Retrieval System. *Comput. Biol. Med.* **2009**, *39*, 396–403.
59. Calvanese, D.; De Giacomo, G.; Lembo, D.; Lenzerini, M.; Rosati, R. Tractable Reasoning and Efficient Query Answering in Description Logics: The DL-Lite Family. *J. Autom. Reason.* **2007**, *39*, 385–429.
60. Cao, H.; Chen, J.; Du, D. Qualitative Extended Description of Spatial Target Orientation Relationship. *Acta Geod. Cartogr. Sin.* **2001**, 162–167.
61. Zhang, Y.; Yang, P.; Li, C.; Zhang, G.; Wang, C.; He, H.; Hu, X.; Guan, Z. A Multi-Feature Based Automatic Approach to Geospatial Record Linking. *Int. J. Semant. Web Inf. Syst. IJSWIS* **2018**, *14*, 73–91.
62. Saracevic, T. Evaluation of Evaluation in Information Retrieval. In Proceedings of the 18th Annual International Acm Sigir Conference on Research and Development in Information Retrieval, Seattle, WA, USA, 9–13 July 1995; pp. 138–146.
63. Lancaster, F.W. *Information Retrieval Systems; Characteristics, Testing, and Evaluation*; John Wiley and Sons: New York, NY, USA, 1979; pp. 57–58.