

Research Article

Lizhi Miao, Chengliang Liu, Li Fan, and Mei-Po Kwan*

An OGC web service geospatial data semantic similarity model for improving geospatial service discovery

<https://doi.org/10.1515/geo-2020-0232>

received July 21, 2020; accepted February 05, 2021

Abstract: Open Geospatial Consortium (OGC) Web Services (OWS) are highly significant for geospatial data sharing and widely used in many scientific fields. However, those services are hard to find and utilize effectively. Focusing on addressing the big challenge of OWS resource discovery, we propose a measurement model that integrates spatio-temporal similarity and thematic similarity based on ontology semantics to generate a more efficient search method: OWS Geospatial Data Semantic Similarity Model (OGDSSM)-based search engine for semantically enabled geospatial data service discovery that takes into account the hierarchy difference of geospatial service documents and the number of map layers. We implemented the proposed OGDSSM-based semantic search algorithm on United States Geological Survey mineral resources geospatial service discovery. The results show that the proposed search method has better performance than the existing search engines that are based on keyword-based matching, such as Lucene, when recall, precision, and F-measure are taken into consideration. Furthermore, the returned results are ranked based on semantic similarity, which makes it easier for users to find the most similar geospatial data

services. Our proposed method can thus enhance the performance of geospatial data service discovery for a wide range of geoscience applications.

Keywords: ontology, OGC web service, geospatial semantic, semantic similarity, geospatial query

1 Introduction

Geographic information is widely used in many areas of natural and social science research, such as environmental protection, geological survey, land resource security, disaster warning, emergency response and population management, and urban economic research. Over the past few decades, with the development of data acquisition technology and sensor software and hardware, billions of gigabytes of geospatial data can be produced by various data producers and research institutions through satellite remote sensing, ground measurement, a variety of sensor acquisition, and other means. To enhance the wide sharing and application of geographic information resources and realize their maximum value, a number of advanced techniques, such as geobrowsing [1], spatial web portal [2], distributed geographic information processing [3], and volunteered geographic information [4], have been developed to operationalize the digital Earth concept. Meanwhile, some government agencies and organizations have funded a number of projects, such as Geospatial One Stop [5], Canada's GeoConnections Plan [6], the Australian Spatial Data Infrastructure [7], and the European Spatial Information Infrastructure [8]. However, it is often difficult for users to find and interpret the most appropriate web services to meet their objectives without significant manual intervention [9]. How to find the target geographic information data service automatically, quickly, and accurately is still a challenge. For example, Li et al. [10] highlighted the datasets that are semantically related to a user's query but described differently from the query keyword will be

* **Corresponding author: Mei-Po Kwan**, Department of Geography and Resource Management, and Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Shatin, Hong Kong; Department of Human Geography and Spatial Planning, Utrecht University, Utrecht, The Netherlands, e-mail: mpk654@gmail.com

Lizhi Miao: Department of Surveying, Mapping and Geoinformation, College of Geographical and Biological Information, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210023, China; Jiangsu Engineering Laboratory for Smart Analysis of Healthy Big Data and Location Based Services, Nanjing, Jiangsu 210023, China

Chengliang Liu, Li Fan: Department of Surveying, Mapping and Geoinformation, College of Geographical and Biological Information, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210023, China

considered irrelevant and excluded from the search results. Hence, improving the effectiveness of a geospatial search engine and making available datasets reachable by scientists is becoming even more significant.

At present, the query and retrieval engine oriented to information service, whether based on Catalog Service for the Web or other standards, is based on keyword-matching search technology. Apache Lucene is used by geospatial catalogs and Web portals, which is a full-text keyword-matching technique such as GeoNetwork [11,12]. However, the use of different vocabulary in different application domains might lead to semantic heterogeneity issues [13], which makes it difficult for a service query based on keyword matching to satisfy the real needs of users. The information description method in the service may not correspond well with the query keywords and leads to situations in which some information is considered irrelevant to the query but its semantic meaning is actually related to the query words. For example, when searching the geospatial services about mineral resources of precious metals, the geospatial services of gold or silver mineral resources, which are parts of precious metals, will not be returned. Similarly, when using keyword queries, the query words may be linked to many descriptive labels in the geographic information service metadata, but these labels cannot reflect the real content of the geographic information services. Thus, unrelated results may turn up in the search results. This may lead to low query recall and precision, affecting the quality of the query results.

Fortunately, the emergence of semantic-based search methods is a good solution to the above two issues that adversely affect the quality of search results. Semantic similarity measurement is a useful methodology for supporting geographic information retrieval [14–16]. Similarity is essential for handling data queries and is the basis for semantic information retrieval [17]. The concepts of spatial, temporal, and descriptive attributes could be used to measure semantic similarity between spatial objects by simulating the knowledge acquisition process of humans based on a knowledge base and artificial neural networks [18]. Improving geospatial data retrieval recall and precision will benefit different domains by facilitating the interoperability and sharing of geospatial data and knowledge.

In fact, different geospatial services contain various map layers to publish the related geographic information. The semantic similarity measure is also determined by map layer attribute information as well as geospatial service attribute description. However, existing research mostly considers the attributes of map layers as important as service description when calculating semantic similarity without considering the number of map layers

and the hierarchy differences between geospatial service documents. Besides service description information, the attributes of map layers are also included as child nodes, which are very important parts for calculating semantic similarity values in geospatial service metadata files. For instance, the NASA Socioeconomic Data and Applications Center (SEDAC)¹ map service published more than 240 Web Map Services (WMS) layers, which include many themes, such as agriculture, climate, health, water, and so on, whereas the World Copper Smelters² map service only contains two layers. It cannot reflect the real similarity between the query terms and the geospatial service when aggregating the semantic similarity of each item of all the map layers directly or expressing semantic similarity simply by using the average values of all the map layers.

To solve these limitations, this study develops a semantic similarity measurement model and presents a workflow of the search engine for improving Open Geospatial Consortium (OGC) Web Services (OWSs) retrieval using a case study of geology and mineral OWSs. This semantic similarity model calculates similarity value at two granularity levels: geospatial service and map layer. An ontology is utilized to implement automatic processing and logical representation of knowledge to support geospatial datasets retrieval. To validate the feasibility and effectiveness of the proposed model, a geospatial service prototype was built and implemented. This system can automatically match the query condition with geospatial data semantically and rank the similarity according to the semantic measurement values.

The remainder of this article is structured as follows. In Section 2, we present a literature review. Section 3 presents a domain ontology of geology and mineral for undertaking queries via OGC geographic information service and proposes an oriented OWS Geospatial Data Semantic Similarity (OGDSS) model, whose essential components include semantic distance, semantic structure, attribute information, and the importance of attribute tags in OWS service scheme files. Section 4 presents some experimental results and performs comparisons in performances of discovery based on OGDSS and other methods. Section 5 illustrates the experiments based on the geology and mineral domain, and the results support the feasibility of using semantic search and knowledge reasoning to improve the discovery of geospatial service records. The article ends with conclusions and future research discussions.

¹ <http://sedac.ciesin.columbia.edu/geoserver/wms>

² <https://mrdata.usgs.gov/services/copper-smelters>

2 Related work

As various data collection and sensing techniques developed, more and more Earth Observing Systems have been established and deployed to observe the Earth for many application fields. Consequently, a massive amount of geospatial data has been captured, generated, and distributed to facilitate many research fields, such as geoscience, environmental science, energy studies, and many more. For instance, the NASA Earth Observing System Data and Information System (EOSDIS) alone distributed data to end users amounting to approximately 2 TB a day [19]. To make the enormous amount of geospatial data easy to share and maximize their value, government agencies and organizations have developed many standards and specifications to enable the broad interoperability of geospatial datasets. The OGC initiated and developed new geospatial web service specifications to improve the sharing of the world's geospatial data. These standards include the WMS [20] that deliver geospatial data as georeferenced map images, the Web Feature Services [21] that provide geospatial data as geographical features, and the Web Coverage Services [22] that respond to queries for coverage data in forms representing space/time-varying phenomena. For supporting the ability to publish and search the metadata of geospatial data, services and related information objects, the OGC proposed Catalog Service for the Web [23] that specifies the web interfaces, bindings, and a framework to locate and access digital catalogs of metadata for geospatial data, services, and related resource information.

If geospatial data cannot be discovered, accessed, and made sense of by users, it has little value [24]. Developing methods for maximizing the values of geospatial data service to accelerate related scientific research has become a great challenge. Many agencies and organizations have developed systems, gateways, or clearinghouses to collect and gather available geospatial datasets published on the Web to improve accessibility. For example, the NASA established EOSDIS as a distributed system to make geospatial data easily accessible to users. The United States Geological Survey (USGS) developed a geoportal for searching and downloading geospatial data sets interactively to support scientific research in many areas, such as water resources, energy sources, and environmental protection. The Intergovernmental Group on Earth Observation [25] was founded to develop a Global Earth Observation System of Systems to facilitate easy access, search, and share of the global Earth Observation data to benefit the geospatial community. The national spatial data infrastructures of the European

Union, which is called the Infrastructure for Spatial Information in Europe (INSPIRE), proposed to overcome major barriers affecting the availability and accessibility of geospatial data by the components of the INSPIRE project including metadata to describe geospatial information resources, harmonization of geospatial data, and policy agreements on geospatial sharing and access [26,27]. The NGCC (National Geomatics Center of China) established and hosted the National Catalogue Service for Geographic Information of China that is comprised of 2 main sites (the NGCC site and the Land Satellite Remote Sensing Application Center site), 31 subsites maintained by provincial bureaus of surveying and mapping geographic information, and many sites hosted by the industrial sector and geographic information companies. The purpose of the National Catalogue Service is to facilitate the ease of publishing catalogs, data discovery, and service integration for big geospatial data resources [28]. These work have greatly improved the ability to manage, access, discover, and share geospatial data.

However, improving the effectiveness of data discovery poses technical challenges for better accessibility and precision of geospatial data and information. Almost every existing geospatial catalog or web portal is implemented using a geospatial search engine based on a full-text keyword-matching technique [11]. In addition, different geospatial web services have semantic problems due to the lack of meaningful descriptions of the actual content [29]. The geographic information that is semantically related to a user's query but described differently from the query terms will be regarded as unrelated and irrelevant because the knowledge hidden behind the semantic relations is not taken into account by using keyword-based search methods [18,30].

In recent decades, many researchers paid increasing attention to this problem by aiming at improving search effectiveness by using semantic techniques and ontology. Besides, domain semantic knowledge graphs are used for automatic concept extraction from big data [31]. For overcoming the disadvantages of keyword-based matching for Web-based information retrieval, semantics has received much attention in many studies [32–39], where the methods focused on the retrieval of contextual information. In those studies, semantic relationships are neglected when constructing semantic networks. For example, only one semantic relationship is-a is considered for measuring the similarity between two concepts in ref. [32]. However, OGC geospatial service retrieval contains not only context retrieval but also spatiotemporal search. As to geospatial search domain, Arpinar et al. [40] developed the Geospatial Semantics Analytics framework to

support queries and analysis using thematic, spatial, and temporal ontologies. In order to improve the discovery and use of Earth science data, the Semantic Web for Earth and Environmental Terminology (SWEET) [41] was developed to make the software available for understanding the semantics of geospatial information distributed over the Internet. Many studies integrate the SWEET to enrich their knowledge base for improving descriptive properties such as semantic retrieval and matching, including water bodies [18], Arctic hydrology [42], and climate hazards [13,17]. In addition, ontology is very useful for semantic-based searching of geospatial data/service. For instance, Li *et al.* [10] integrated ontology techniques to develop a semantic search tool built on a latent semantic analysis method, which can implement and support the intelligent discovery of polar datasets. As the Semantic Web emerges [43], similarity is combined with semantic methods to improve information retrieval, such as edge counting measures [44], feature-based similarity assessment methods [45,46], information content semantic similarity [47–52], and hybrid methods [53,54]. In geospatial information retrieval research, semantic similarity is utilized to enhance geospatial data and service search by ranking the measured similarity values between query terms and geospatial datasets as well. For instance, Bakillah *et al.* [55] employed a clustering algorithm with semantic similarity to deal with the complex social graphs extracted from Twitter to obtain spatial clusters at different temporal snapshots and detect geo-located communities within the discovered thematic communities. Zhang *et al.* [56] used a semantic similarity matching algorithm based on the ontology to measure the degree of semantic similarity between the geo-event and related geospatial resources. However, there are still some limitations in these methods of semantic similarity computation. In particular, the hierarchical relationships among geospatial service and map layers are not included in semantic similarity metrics.

3 OGDSS model: OGDSSM

Geospatial data contain three types of information: attribute, spatial, and temporal information. For example, a WMS metadata document³ mainly includes the aforementioned three types of information in service granularity and map layer granularity as shown in Figure 1. Here, we

integrate three equations into the proposed model: attribute semantic similarity, spatial similarity, and temporal similarity.

To implement geospatial service semantic retrieval, an OGDSSM-based framework is established integrating the techniques of ontology, Natural Language Processing (NLP), and the semantic web into a unified workflow (Figure 2).

The OGDSS-based framework is briefly described as follows:

- (1) OGC geospatial data service metadata are requested and stored using GetCapabilities operation, which contains descriptions and parameters of the geospatial service, such as service description, temporal dimension, and bounding box of the map layer.
- (2) Once the service metadata are stored into the database, the system will parse each text file, extract the key metadata fields, and filter the stop words and useless words using the WordNet lexical database package to get the meaningful description words (nouns) of the related OWS.
- (3) Geographic extents are extracted when parsing the metadata documents by the ISO19139 standard tag pairs `<BoundingBox>`, `</BoundingBox>`, and `<EX_GeographicBoundingBox>` `</EX_GeographicBoundingBox>`.
- (4) Temporal information is parsed by the ISO19139 standard tag pair `<Dimension name = "time">`, `</Dimension>`.
- (5) Based on the ontology of the knowledge base, geospatial service retrievals are implemented by measuring the semantic similarity (including attribute similarity, spatial similarity, and temporal similarity) of the extracted information from metadata documents. It utilizes logical and hierarchical relationships defined in the geographic ontology base, as discussed in Section 3.1.
- (6) The geospatial service whose semantic measurement value is higher than certain threshold will be returned to users and shown sequentially according to the semantic similarity ranking.

3.1 Construction of a mineral resource knowledge base

When constructing a mechanism to help users find the most suitable data to perform automated analyses for their study, we found some lexical ambiguities about

³ http://schemas.opengis.net/wms/1.3.0/capabilities_1_3_0.xml

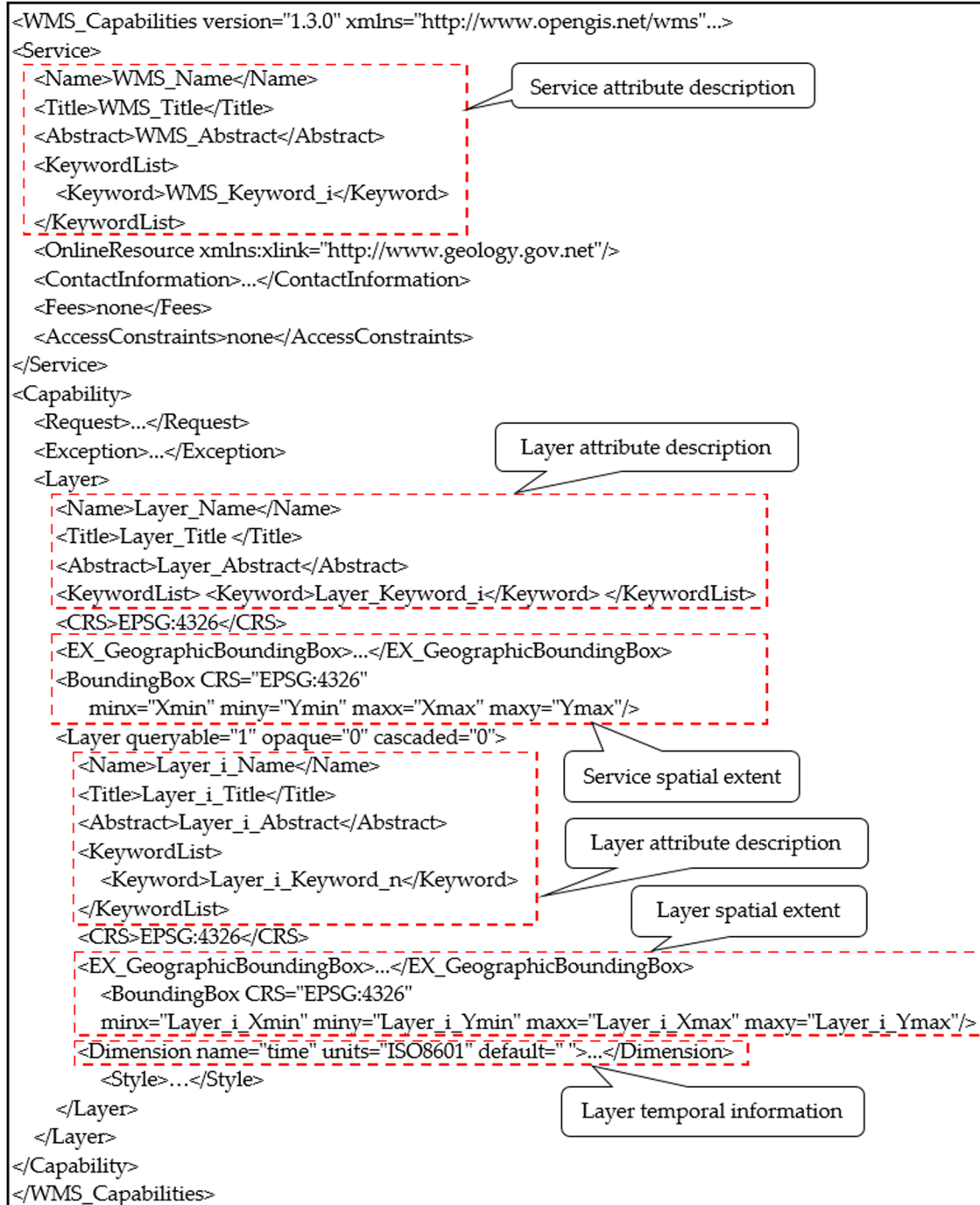


Figure 1: WMS metadata document structure.

the semantic heterogeneity of the data and services. Different terms can refer to the same concept in the same or different services [57], which is a synonymic issue. Also, the same term can be used to describe different contexts for different services [57], which is a polysemantic issue. In addition to these issues, some terms inherently have latent semantic relations. For example, the term “fuel” may refer to the material for supplying an industrial

plant, vehicle, or machine. Users may describe the fuel with detailed linguistic terms using “gas” (gaseous fuel) or “gasoline” (liquid fuel), which have inherent relationships with “fuel” but are described by different terms. To solve this problem and advance geological research, we developed a domain ontology to describe the knowledge base and provided a semantic-enabled search model for geospatial service.

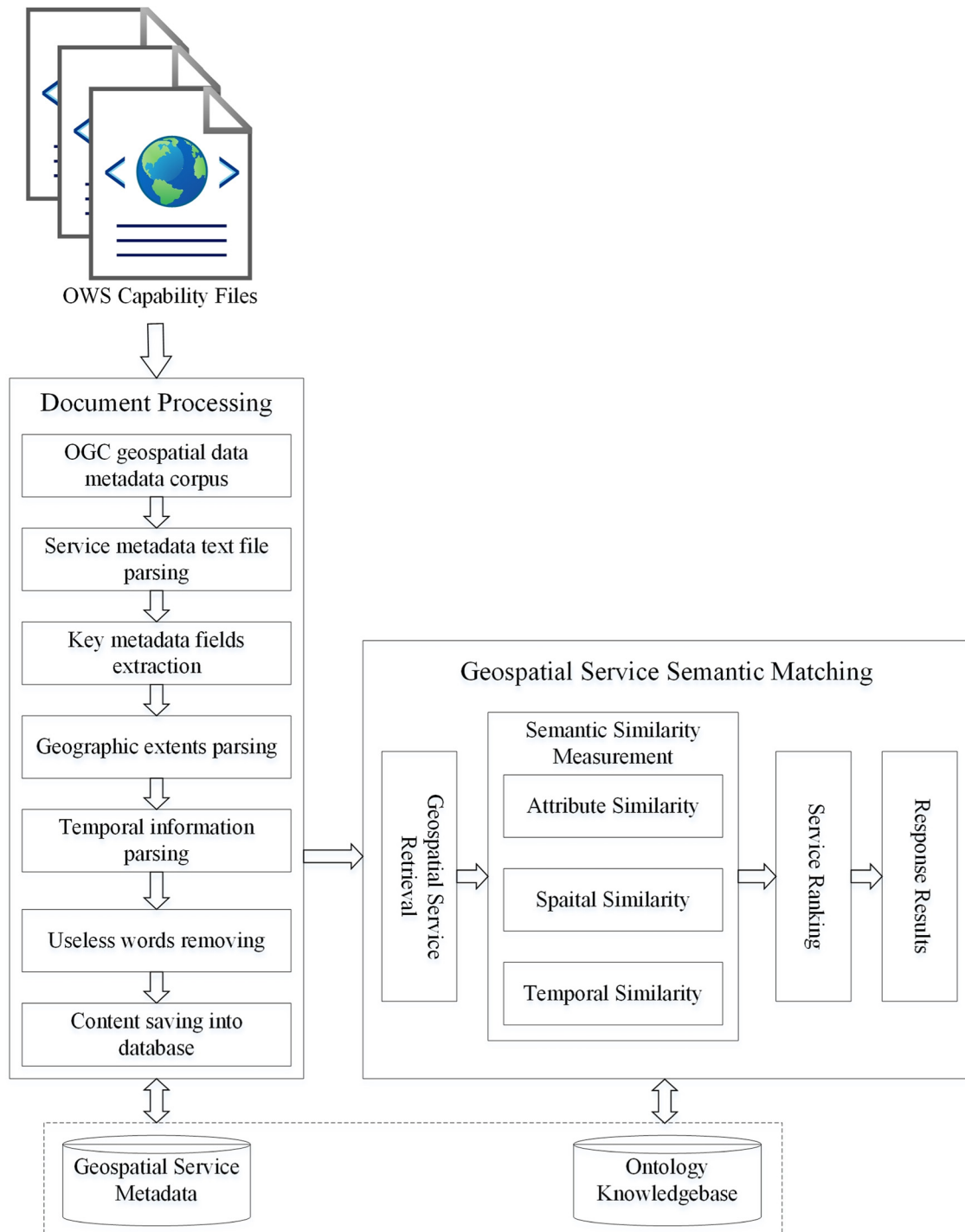


Figure 2: The workflow of the OGDSSM-based framework for geospatial service data.

The ontology, defined as “an explicit specification of a conceptualization” [58], can support semantic retrieval and organization of related geospatial services and present the logical definition of a group of concepts. It can reveal the hidden relationships among concepts that are not generally encoded in metadata XML files [59]. We

used the geospatial metadata sets from the USGS Mineral Resources Program (MRP)⁴ as our test corpus in this study.

⁴ <https://www.usgs.gov/energy-and-minerals/mineral-resources-program>

The ontology of the geology field is constructed using Protégé editor, which was developed and published by Stanford University [60]. Figure 3a shows the structure of the place name keywords where the continent is a

universal terminology as a top ontology. The continent is divided into seven subclasses as shown in the left panel of Figure 3a: Asia, North America, South America, Oceania, Europe, and Africa. The nodes in the trees are examples of

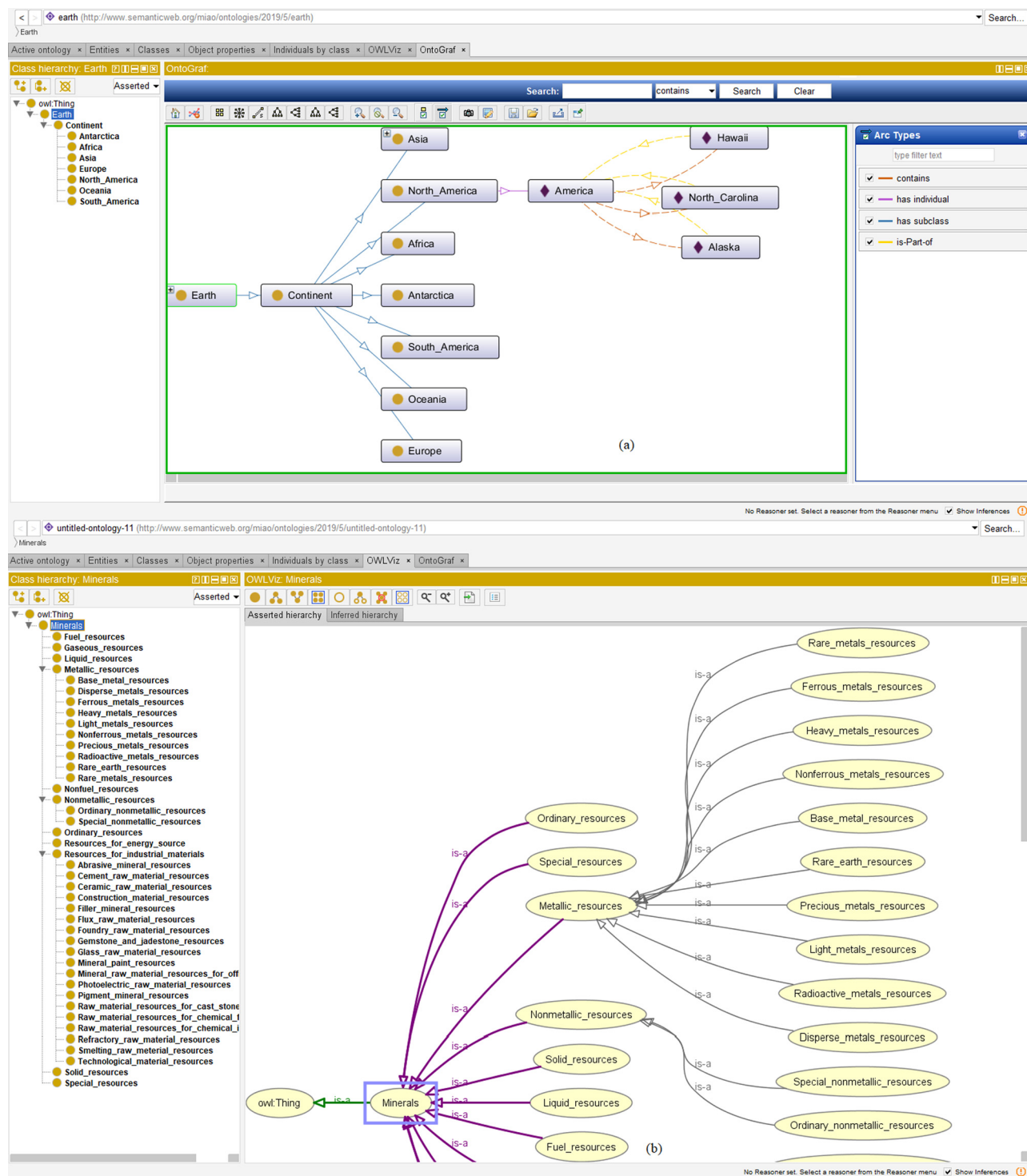


Figure 3: Ontology construction. (a) Fragment of ontology and (b) sublevel ontology on the mineral.

locations with semantic paths. For example, through the ontology graph, it can be derived that the America contains Alaska, which is a part of America, which is an individual of North America, which is a subclass of Continent. And, we constructed the ontology using the terminological classification of geology and mineral resources based on the national standard of China (GB/T 9649.32-2009) as the geological and mineral terminology in this study. As shown in Figure 3b, the mineral class is a top ontology, which has 11 subclasses denoted with yellow circle as shown in TOC (Table of Content): fuel resources, gaseous resources, liquid resources, metallic resources, nonfuel resources, nonmetallic resources, ordinary resources, resources for energy sources, resources for industrial materials, solid resources, and special resources.

Here, the place names in the descriptive information of the OWS metadata are considered as the location classes in the same way as any other keywords as shown in Figure 3a. This will improve the flexibility and usability of spatial taxonomy searches without drawing a bounding box or inputting the geographic coordinates. Figure 3b shows part of the ontology on the mineral resources, which is an important division in the geology field.

Figure 3a shows a sublevel ontology on the place name. The place names usually contain inherit and potential relationships among each other. Here, America is taken as an example to build the place name ontology and embedded semantic relationship considering experiment data. Alaska, North Carolina, and Hawaii have is-Part-of relationships with America. When a query containing the place term “America,” the geographic information related to those three place names will be returned according to their semantic relatedness. Figure 3b shows the ontological fragment of mineral, mainly referring to metallic resources. Each subclass can be further divided, such as into rare metal resources, rare earth resources, radioactive metal resources, precious metal resources, nonferrous metal resources, light metal resources, heavy metal resources, ferrous metal resources, disperse metal resources, and base metal resources. The leaf nodes denoted with violet diamond in Figure 4 are the descendant individuals of certain subclass.

For example, *Precious_metals_resources* subclass has four individuals including palladium ore, platinum ore, silver ore, and gold ore. *Heavy_metals_resources* subclass has five individuals including gold ore, silver ore, iron ore, zinc ore, and copper ore. It may happen that some individuals belong to different subclasses. For example, silver ore and gold ore are not only precious metal resources but also heavy metal resources.

In our ontology, there are four kinds of relationships: is-a (hypernymy), is-parent-class-of (hyponymy), is-part-of (meronymy), and equal-to (synonymy). The is-part-of is customized to describe the relationship between two entities, wherein entity A is a part of entity B, but is not a subclass of B. The equal-to relationship denotes that two concepts are equivalent. For distinguishing the difference in the semantic distance among them, the distance weight of each semantic relation is set by extending the method set by Sycara *et al.* [61] and Zhang *et al.* [56], as shown in Table 1.

3.2 Description term extraction

The metadata files are encoded in ISO19115 (2003) in XML format with unstructured text. Therefore, we need to parse the metadata to extract and filter the attribute and spatiotemporal information. In one metadata file, there are on average 600 metadata tags in each metadata record [10], and there are many unnecessary tags that do not describe the actual content, such as “ContactInformation” or “AccessConstraints.” Only “Name,” “Title,” “Abstract,” “Keyword,” “BoundingBox,” and “Dimension name = ‘time’” were extracted from each metadata file. These tags contain not only the geographic information of the service but also the map layer information.

For eliminating meaningless terms, we integrated WordNet [62] package into the processing framework to filter the stop words such as “is” and “the.” WordNet is a useful tool for computational linguistics and NLP [63,64]. After using WordNet, the geographic attributes, spatial extents, temporal information of OWS, and map layer are extracted and converted to structured words, which are then stored in the database.

3.3 OGDSSM

After the extraction of the metadata file with their spatiotemporal information, the related OWS is retrieved by using an OGDSSM based on the ontology to calculate the semantic similarities between the query terms and geospatial service metadata. OGDSSM is composed of three parts: attribute similarity, spatial similarity, and temporal similarity. The final semantic similarity is the sum of the weighted results.

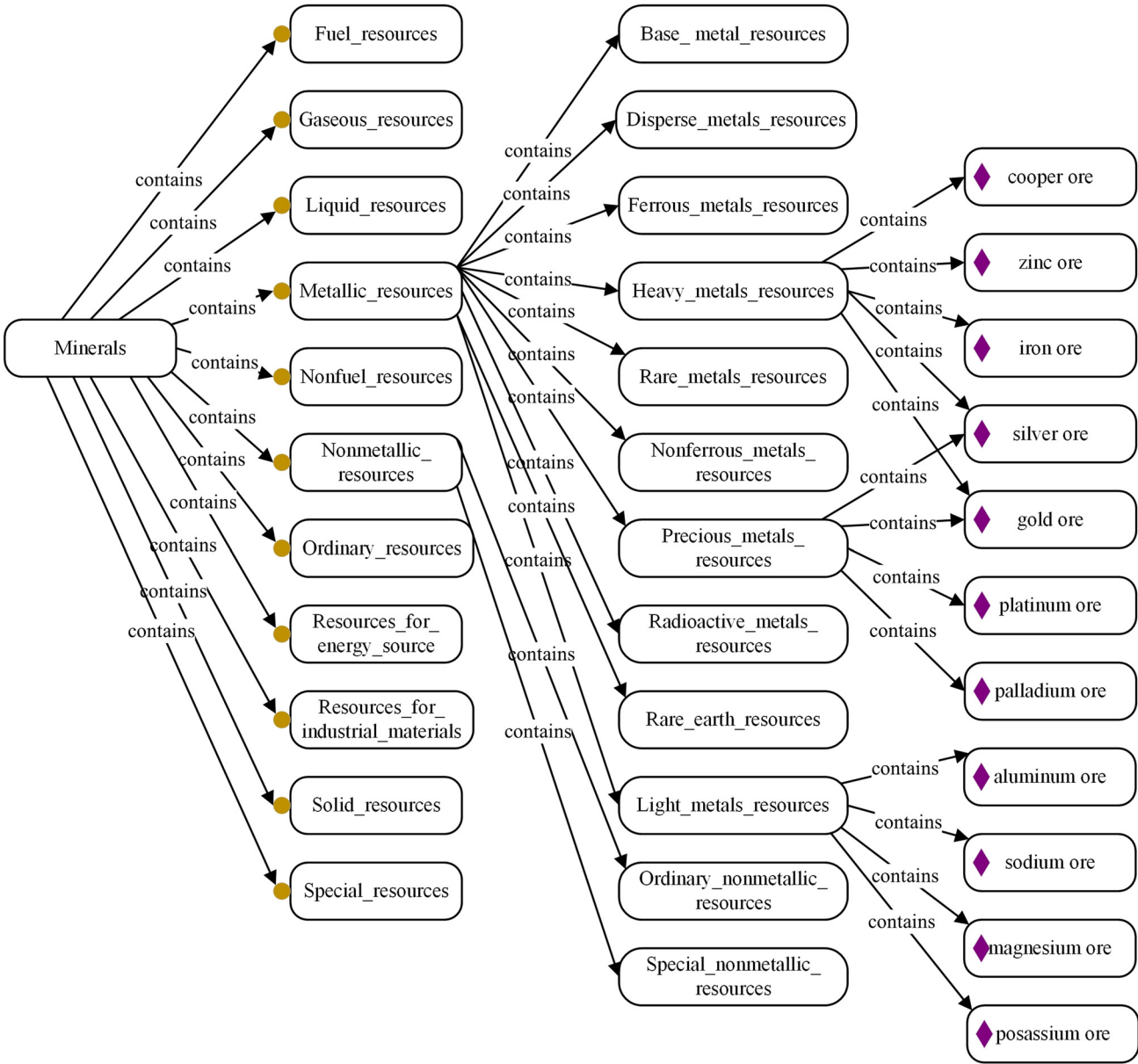


Figure 4: Fragment of mineral ontology.

Table 1: Distance weight of each semantic relation

Semantic relations	Characteristic	Weight
Equal-to	Symmetric, transitive	0
Is-a	Monodirectional, transitive	4/15
Is-part-of	Monodirectional, transitive	2/5
Is-parent-class-of	Monodirectional	1/3

3.3.1 Attribute similarity

Attribute similarity is utilized to reflect the matching between service themes and query terms. The similarity

of two concepts is mainly determined by the distance between them in the knowledge base. Many studies have discussed and measured semantic similarity using different methods from a perspective of ontology concepts and their relationships [30,65–68]. The directed graph is generally used to represent ontology, and the distance of two concepts is calculated by the edge of the graph.

In most previous studies, the distance between two concepts was measured using the shortest path without considering the weight of the semantic relationships between the child node and parent node. To address the difference between two concepts in the ontology,

weighted semantic distance is a good solution to express different semantic relationships [61]. In our work, we proposed our own method that different semantic relationships of two concepts are set to the corresponding weights that can describe the real parent–child relationship and the real distance between the two concepts.

Considering the relationships between a pair of concepts, the equation of semantic distance can be expressed as follows:

$$\text{Dist}(O_1, O_2) = (k + 1) \times \sum_{i=1}^n W_i, \quad (1)$$

where W_i is the weight of the edge i in the path, $\sum_{i=1}^n W_i$ indicates the weighted sum of the path between concept nodes O_1 and O_2 , and k is the number of interruption times of transitive relationships in the path between concept nodes O_1 and O_2 .

While using distance to calculate semantic similarity, we not only consider the maximum semantic distance between any pair of nodes in the ontology but also consider the weighted distance from the lowest common ancestor (LCA) [69] node to the root node. The equation of semantic similarity can be expressed as follows:

$$\begin{aligned} \text{Sim}(O_1, O_2) \\ = 1 - \frac{\text{Dist}(O_1, O_2)}{\text{Max}\{\text{Dist}(O_i, O_j)\} + \text{Dist}(\text{root}, \text{LCA}(O_1, O_2))}, \end{aligned} \quad (2)$$

where $\text{Dist}(O_1, O_2)$ is obtained from equation (1); $\text{Max}\{\text{Dist}(O_i, O_j)\}$ is the maximum semantic distance between any pair of nodes in the ontology with a fixed value for the knowledge base of a certain domain. $\text{Dist}(\text{root}, \text{LCA}(O_1, O_2))$ is the weighted distance from the lowest joint ancestor node to the root node. $\text{LCA}(O_1, O_2)$ denotes the LCA between two nodes O_1 and O_2 .

To illustrate how to use equation (2) for calculating the attribute similarity, let the ontology of Figure 4, denoted by C_g , C_a , and C_s the concepts “graphite ore,” “aluminum ore,” and “silver ore.” By applying equation (2) measure, the similarity value is calculated as follows:

$$\begin{aligned} \text{Sim}(O_g, O_s) &= 1 \\ &- \frac{(1 + 1) \times \left(\frac{4}{15} + \frac{4}{15} + \frac{4}{15} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} \right)}{3.6 + 0} = 0, \\ \text{Sim}(O_a, O_s) &= 1 \\ &- \frac{(1 + 1) \times \left(\frac{4}{15} + \frac{4}{15} + \frac{1}{3} + \frac{1}{3} \right)}{3.6 + \frac{4}{15}} = 0.38. \end{aligned}$$

The values obtained by equation (2) measure show that the neighboring concepts C_a and C_s are more similar than the concepts C_g and C_s located in the same hierarchy.

The final attribute similarity equation is

$$\begin{aligned} \text{aSim} &= \max\{\text{Sim}_{\text{aService}}(O_q, O_{s-i}), \\ &\text{Sim}_{\text{aLayer}_j}(O_q, O_{l-k})\}, \end{aligned} \quad (3)$$

$$i = 1, 2, \dots, m; \quad j = 1, 2, \dots, n; \quad k = 1, 2, \dots, p,$$

where O_q is the ontology of the query term; O_{s-i} is the ontology of the attribute term describing the service; O_{l-k} is the ontology of the attribute term describing one layer; m is the number of attribute concepts describing the service theme; n is the layer count in the geospatial service; and p is the number of attribute concepts describing layer j . In addition, $\text{Sim}_{\text{aService}}(O_q, O_{s-i})$ is the attribute similarity between query terms and service attributes; $\text{Sim}_{\text{aLayer}_j}(O_q, O_{l-k})$ is the attribute similarity between query terms and layer attributes.

3.3.2 Spatial similarity

A metadata file of geospatial service data usually contains the spatial extent of the service and several spatial extents of its map layers. Each spatial extent is defined by the <BoundingBox> tag in the metadata file. The similarity of two spatial areas is a scalar. Therefore, the spatial similarity can be expressed by using bounding boxes of query condition and service as follows:

$$\text{sSim}(R_q, R_s) = \frac{\frac{\text{Area}(R_q \cap R_s)}{\text{Area}(R_q)} + \frac{\text{Area}(R_q \cap R_s)}{\text{Area}(R_s)}}{2}, \quad (4)$$

where R_q is the query spatial condition, R_s is the target data spatial extent, and $R_q \cap R_s$ is the overlapped area of R_q and R_s . Spatial similarity is calculated from two ratios of the overlapped area to the area of R_q and R_s .

As to geospatial service metadata, it contains many spatial extents including the bounding box of the service and the bounding boxes of its map layers. Considering the two categories of spatial extents, the final spatial similarity equation is as follows:

$$\begin{aligned} \text{sSim} &= \max\{\text{sSim}(R_q, R_s), \text{sSim}(R_q, R_i)\} \\ &i = 1, 2, \dots, n, \end{aligned} \quad (5)$$

where R_i is the bounding box of the map layer i and n is the layer count in the geospatial service.

3.3.3 Temporal similarity

Besides thematic and spatial property, temporal property is also a very important dimension in geospatial services and includes the timestamps or timelines of the service. The dynamic change of spatiotemporal trends can be analyzed according to the time series of the geospatial service data.

Temporal information in the geospatial service can be categorized into two types: time point and time period. In fact, the time point is one special kind of time period with the same start time and end time. In the temporal similarity calculation, all time formats comply with ISO 8601:2004 “yyyy-MM-ddThh:mm:ss,” i.e., 2004-05-03T17:30:08. Thus, temporal similarity can be calculated as follows:

$$\text{tSim} = \frac{\frac{D(T_q \cap T_s)}{D(T_q)} + \frac{D(T_q \cap T_s)}{D(T_s)}}{2}, \quad (6)$$

where T_q is the temporal condition of query, T_s is the temporal value of the geospatial service, and $D(T_q \cap T_s)$ is the duration of the overlapped time between T_q and T_s .

3.3.4 Final semantic similarity

OGDSSM contains the weighted sum of attribute, spatial and temporal similarities, and the final semantic similarity is expressed as follows:

$$\text{totalSim} = \begin{cases} \omega_1 \times \text{aSim} + \omega_2 \times \text{sSim} + \omega_3 \times \text{tSim} & \text{aSim} \neq 0, \text{sSim} \neq 0, \text{tSim} \neq 0 \\ 0 & \text{aSim} = 0 \text{ or } \text{sSim} = 0 \text{ or } \text{tSim} = 0 \end{cases} \quad (7)$$

where ω_1 , ω_2 , and ω_3 indicates the weights of the attribute, spatial, and temporal similarities, respectively, and

the weight for each similarity is normalized to a value between 0 and 1, $\omega_1 + \omega_2 + \omega_3 = 1$. For indicating the same significance of each similarity, ω_1 , ω_2 , and ω_3 are set the same value 1/3. The weights can be adjusted accordingly to indicate which subsimilarity is emphasized. For example, when focusing on the spatial similarity, ω_2 could be increased.

4 Experimental results

In the experiments, we selected the mineral subset (114 OGC: WMS) to be our test geospatial service for the following reasons. First, the USGS MRP is a special project for a comprehensive understanding of mineral resource potential, production, and consumption, and it is easy to build a domain knowledge base for mineral resource projects especially. Second, it is simple to identify the actual amount of geospatial service data related to the defined query condition from a small corpus. Therefore, using mineral resources as the domain area, it will be easy to verify the availability and feasibility of the OGDSSM-based search method.

Seven queries, listed in Table 2, were conducted on the sample data. All queries were performed on a standard laptop Intel Core i7-7500U CPU-2.70 GHz with 8 GB RAM and the Windows 10 (64-bit) operating system. Here, we use their parameters to measure the effectiveness of the retrieval method: precision, recall, and F-measure. Precision is the ratio between the number of relevant records retrieved from a query and the total number of records retrieved. Recall is the ratio of the number of relevant records retrieved from a query to the total number of relevant records within the corpus. The F-measure is the weighted harmonic mean of precision and recall. The precision and recall rates of the seven queries from our proposed model OGDSSM were

Table 2: Selected queries for evaluating search effectiveness

Query	Query condition		
	Attribute terms	Spatial extent	Temporal property
Q1	Heavy metal resources	—	—
Q2	American mineral	—	—
Q3	Copper mineral	—	—
Q4	—	Rect (24.5, -25, 49.4, -66.9)	—
Q5	—	—	[1980, 1989]
Q6	Mineral deposits	Rect (24.5, -25, 49.4, -66.9)	—
Q7	Mineral deposits	Rect (24.5, -25, 49.4, -66.9)	[1980, 1989]

compared with those obtained from LSATTR [10], a geospatial semantic search method and GeoNetwork with Lucene-based search engine, one of the most popular catalog applications of metadata management.

For Q1, one query expects to retrieve all the heavy metal-related geospatial service data by inputting “heavy metals resources.” In the ontology base, heavy metal is one subclass of metal resources other than a certain metal. So, even though geospatial service metadata do not contain the aforementioned keywords, the service is still considered to be relevant if the subject is related to the instance of heavy metal. The results show that GeoNetwork and LSATTR got 0 response and 10 responses on Q1 separately. LSATTR method retrieved all relevant eight records in the corpus, and has 100% recall rate and 80% precision rate. However, our OGDSS method yielded eight records. By examining the returned metadata, we found that seven of these eight records are relevant to the query. The other metadata returned by the proposed method is on the subject of “A compendium of previously published databases and database records that describe PGE, nickel, and chromium deposits and occurrences.” This metadata includes terms such as “nickel and chromium deposits” that are related to “heavy metal.” However, this metadata subject focuses on the “database” instead of “metal” and so it is considered irrelevant. Other response metadata include copper/nickel/chromium/silver/lead/mercury deposits and resources that are categorized to heavy metal. By examining the test geospatial services, we identified that 8 of the 46 records are relevant to Q1 query totally where the OGDSS method did not get one metadata which is really related to Q1 query. That is because a compound word “zinc-lead deposits” is used to describe the metal term. Currently, the OGDSS method cannot

identify compound words beyond the WordNet dictionary. So, the precision and recall rates of OGDSS are both 87.5%, whereas these rates for GeoNetwork are 0%. A significant reason for the different performance of the proposed method and GeoNetwork searching is that “copper/nickel/chromium/silver/lead/mercury” are instances of “heavy metal” (query words) class, but the query words are not present in the metadata file. The OGDSS model is able to find these instances based on this semantic relationship, whereas the keyword-based search of GeoNetwork cannot.

For Q4, a spatial query expects to find out all the geospatial services related to the conterminous US by setting its spatial extent with WGS84 minx = “24.5”, miny = “-125”, maxx = “49.4”, and maxy = “-66.9”. LSATTR method returned 65 records including all the relevant 39 metadata with 100% recall rate and 60% precision. And, the results show that GeoNetwork obtained 39 responses that are arranged randomly and these responses intersect with the input spatial extent. However, the OGDSS method also obtained the same 39 records, but they are aligned orderly. The higher the similarity is, the higher the ranking is. In this experiment, three records are listed first with the same similarity value 1 as shown in Table 3. As well known, the higher the similarity value is, the greater the relatedness is. The most relevant record to the user’s query will be listed in the forefront. It will be very helpful for users to find the most similar or equal geospatial service easily. It will be very helpful for users to find the most similar or equal geospatial service easily. As to recall rates of Q4, all three methods OGDSS, LSATTR, and GeoNetwork achieve a rate of 100%. As to precision of them, only LSATTR did not reach 100%, which is 60%. For Q6, GeoNetwork

Table 3: Spatial similarity ranking based on OGDSS

Id	USGS mineral resources WMS	BBOX of WMS	Spatial similarity
1	State_Geologic_Map_Compilation	(24.5, -125, 49.4, -66.9)	1
2	Soil_Geochemical_Landscape	(24.5, -125, 49.4, -66.9)	1
3	Geology of the conterminous US (King and Beikman)	(24.5, -125, 49.4, -66.9)	1
4	Prospect- and mine-related features from US Geological Survey 7.5 and 15 min topographic quadrangle maps of the western United States	(29, -125, 49.1, -65)	0.89
5	1998 assessment of undiscovered deposits of gold, silver, copper, lead, and zinc in the United States	(24, -165, 73, -66)	0.65
:	:	:	:
39	Mica deposits of the Blue Ridge in North Carolina	(35, -83, 37, -81)	0.5

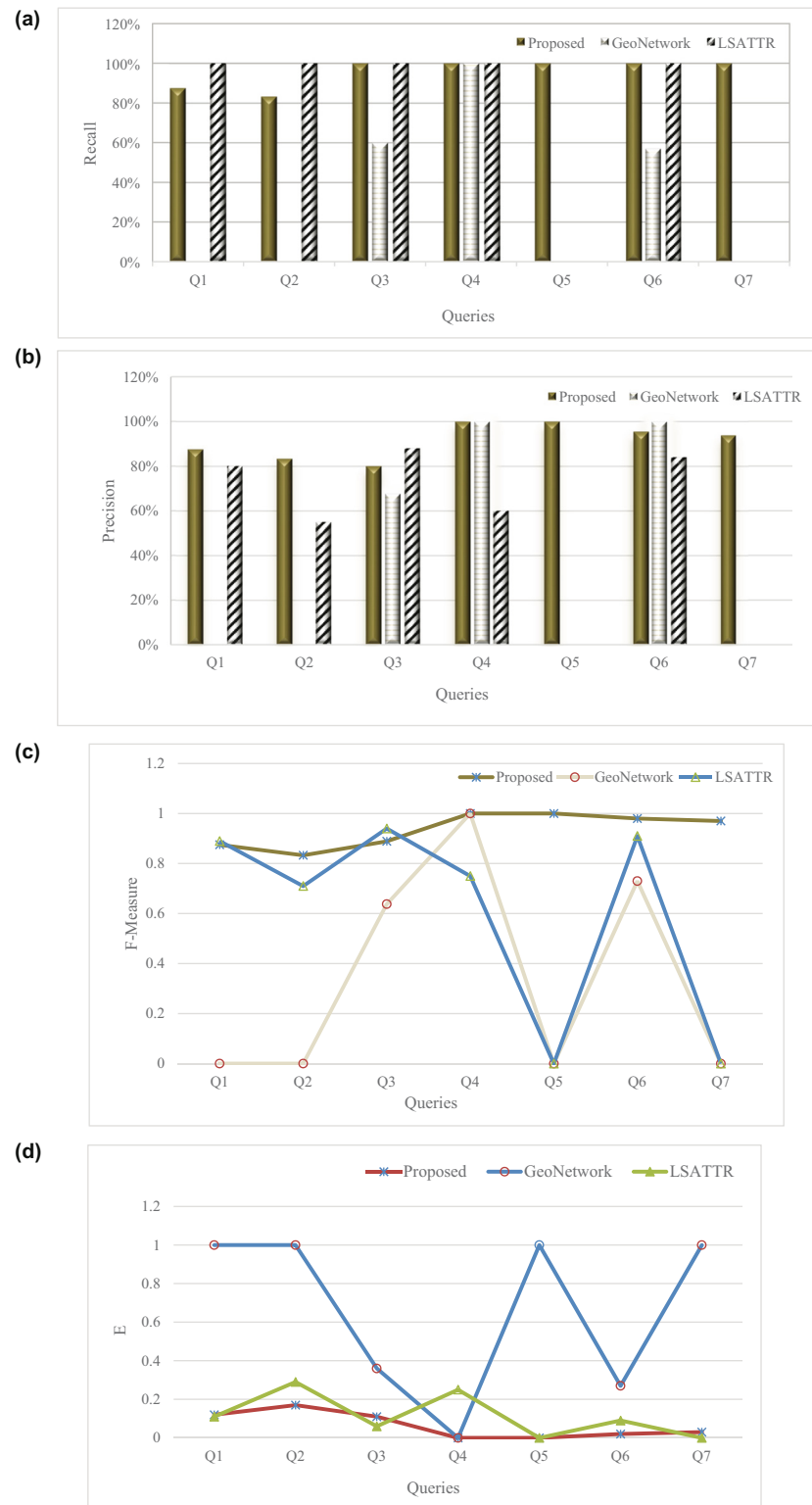


Figure 5: (a) Recall comparison among OGDSSM-based method, LSATTR-based method, and GeoNetwork. (b) Precision comparison among OGDSSM-based method, LSATTR-based method, and GeoNetwork. (c) F-Measure comparison among OGDSSM-based method, LSATTR-based method, and GeoNetwork. (d) E-value comparison among OGDSSM-based method, LSATTR-based method, and GeoNetwork.

maintains a higher precision rate than the proposed method and LSATTR because it returns fewer records and all of them are relevant.

Many geospatial service data contain temporal property to describe the time dimension of map layers. For example, one geospatial service can collect the mineral resource distribution developed from 1970s to 1990s. GeoNetwork can search the date when the geospatial services were produced and the date when metadata records were created in GeoNetwork. However, it cannot search the time dimension in geospatial data, which is a very important component of geospatial service, especially in the earth observation field. Besides thematic and spatial searching abilities, our OGDSS model also can perform temporal queries to prompt the applications of geospatial data service in related research fields. To implement temporal search experiments, simulated temporal information is appended to each Getcapabilities XML file by `<Dimension name = "time" units = "ISO8601" default = " "></Dimension>` fragment description.

Figure 5a shows the comparison of overall recall rate for our OGDSS model, LSATTR, and the keyword-based search using GeoNetwork. It is apparent that the recall rates of LSATTR-based algorithm reach 100% except for Q5 and Q7, as it did not support temporal query. The recall rate by using the OGDSSM-based method is higher than that by using GeoNetwork. Only three of the seven queries returned relevant records in GeoNetwork search. Each recall rate by the OGDSSM-based method is higher than 80%, which means that the most relevant metadata was retrieved. Sometimes, GeoNetwork has a higher precision rate than OGDSSM-based search as shown in Figure 5b (e.g., Q4, Q6). The main reason is that fewer records returned by GeoNetwork are relevant. To evaluate the effectiveness of the three methods, F-measure and E-value are computed as shown in Figure 5c and d, which demonstrates obviously that the OGDSSM-based method has better performance than LSATTR and GeoNetwork.

5 Conclusions and future work

This article discussed an integrated measurement model of spatiotemporal similarity to improve the effectiveness of geospatial data service discovery for supporting semantic-based search from massive geographic catalogs. The experiments show that the OGDSSM-based method significantly improved geospatial service discovery. For the seven queries, almost all of the F-measures are closer to 1. Although the precision of the

OGDSSM-based method for Q6 is lower than that of GeoNetwork, the OGDSSM-based method returned all the records discovered by GeoNetwork with a 100% recall rate. Although the recall rate of OGDSSM-based method for Q1 and Q2 is lower than that of LSATTR-based method, the OGDSSM-based method has higher precision. Besides the ability to respond to thematic and spatial queries, the proposed method can also handle temporal queries by measuring temporal similarity.

Utilizing the proposed methodology in geospatial service discovery has the following advantages: (1) conducting semantic analysis to discover relevant records instead of keyword-based matching; (2) enabling temporal queries as well as spatial and thematic queries simultaneously; and (3) ranking the returned results according to the semantic similarity between geospatial service and query condition.

There are several areas that call for future research, and they might promote additional improvements in geospatial service discovery, access, and usage, including the following aspects. First, multiple granularities of the descriptive terms should be further considered and implemented when parsing OWS metadata files. In the current method, besides keywords are parsed as a phrase in `<Keywords>` tag, the granularity is still a single word when parsing `<Name>`, `<Title>`, and `<Abstract>` tags. In the future, the parsing method and the vocabulary database should be improved for handling phrase-based attribute information extraction by extending the multiple granularities of domain lexicon.

Second, the relationship between place name and spatial extent can be constructed to improve the effectiveness of the region name-oriented search. In the proposed model, the region/place name or spatial taxonomy is identified as attribute terms for implementing attribute queries. That is why Q2 in Figure 3 returns a wrong record with the phrase "outside the United States." In the future, the linkage between each region/place name and a bounding box [10] needs to be integrated into our method to improve the subject-based search containing region/place name by performing a spatial query.

Third, the ontology base to integrate the existing ontologies, such as SWEET, needs to be extended. Our current ontology is based on the National Standard of China. The ontology should not be limited to the standard. Otherwise, some professional terms cannot be distinguished. For example, a combined term "zinc-lead deposits" cannot be returned when executing a query with "heavy mineral" search. Definitely, zinc-lead deposit is one kind of heavy mineral. Integrating the existing knowledge base into the current ontology not only can

improve OGDSS-based query recall and precision but also can benefit more related research domains of the geospatial community.

Fourth, a large number of queries need to be supported when using OGDSSM-based method. Massive volume is a characteristic of geospatial data. We plan to develop a middleware integrating Ajax and distributed geospatial computing techniques to facilitate OGDSSM model to respond to massive queries automatically.

Acknowledgments: This research was funded by China Scholarship Council (201808320014), NSFC incubation project of Nanjing University of Posts and Telecommunications (NY218084). The authors would like to thank the anonymous reviewers for their helpful comments and suggestions that greatly improved this article.

Author contributions: All authors contributed substantially to the conception of this article. Lizhi Miao proposed the OWS Geospatial Data Semantic Similarity Model, developed the workflow of the OGDSSM-based framework, and wrote the main parts of the manuscript. Chengliang Liu and Li Fan implemented the experiments and developed the program to validate the results. Mei-Po Kwan enhanced the manuscript substantially and gave guidance and advice during the development process of the article.

References

- [1] Peuquet DJ, Kraak M. Geobrowsing creative thinking and knowledge discovery using geographic visualization. *Inf Vis*. 2002;1:80–91.
- [2] Yang P, Evans J, Cole M, Marley S, Alameh N, Bambacus M. The emerging concepts and applications of the spatial web portal. *Photogramm Eng Remote Sens*. 2007;73:691–98.
- [3] Yang C, Li W, Xie J, Zhou B. Distributed geospatial information processing-sharing distributed geospatial resources to support the digital earth. *Int J Digital Earth*. 2008;1:259–78.
- [4] Goodchild MF. Citizens as sensors: the world of volunteered geography. *GeoJournal*. 2007;69:211–21.
- [5] Goodchild MF, Fu P, Rich P. Sharing geographic information: an assessment of the geospatial one-stop. *Ann Assoc Am Geogr*. 2007;97:250–66.
- [6] Johnson BD, Singh J. Building the national geobase for Canada. *Photogramm Eng Remote Sens*. 2003;69:1169–73.
- [7] Nairn AD. Australia's developing GIS infrastructure-achievements and challenges from a federal perspective. The fifth international seminar on GIS, Seoul, Korea, 28–29 September; 2000.
- [8] Salvemini M. The infrastructure for spatial information in the European community vs. regional SDI: the shortest way for reaching economic and social development. Ninth united nations regional cartographic conference for the Americas, New York, USA, 10–14 August; 2009.
- [9] Stock K. Determining semantic similarity of behavior using natural semantic metalanguage to match user objectives to available web services. *Trans GIS*. 2008;12:733–55.
- [10] Li W, Goodchild MF, Raskin R. Towards geospatial semantic search: exploiting latent semantic relations in geospatial data. *Int J Digital Earth*. 2014;7:17–37.
- [11] McCandless M, Hatcher E, Gospodnetić O. *Lucene in action*. 2nd edn. Greenwich, CT: Manning; 2010.
- [12] Giannecchini S, Tajariol E. GeoNetwork, the open source solution for the interoperable management of geospatial metadata. Available at: https://demo.geo-solutions.it/share/profile/geonetwork/geonetwork_2014.pdf (verified 08 February 2020); 2014.
- [13] Klien E, Lutz M, Kuhn W. Ontology-based discovery of geographic information services – an application in disaster management. *Comput Environ Urban Syst*. 2006;30:102–23.
- [14] Martins B, Silva M, Ribeiro L. Indexing and ranking in Geo-IR systems. Proceedings of the 2005 workshop on geographic information retrieval, November 4, 2005, Bremen, Germany; 2005. p. 31–4.
- [15] Martins B, Silva M, Chaves M. Challenges and resources for evaluating geographical IR. Proceedings of the 2005 workshop on geographic information retrieval, November 4, 2005, Bremen; Germany; 2005. p. 65–9.
- [16] Janowicz K, Raubal M, Schwing A, Kuhn W. Semantic similarity measurement and geospatial applications. *Trans GIS*. 2008;12:651–59.
- [17] Schwing A. Approaches to semantic similarity measurement for geo-spatial data. *A Surv Trans GIS*. 2008;12:5–29.
- [18] Li W, Raskin R, Goodchild MF. Semantic similarity measurement based on knowledge mining: an artificial neural net approach. *Int J Geogr Inf Sci*. 2012;26:1415–35.
- [19] Ramapriyan HK, Behnke J, Sofinowski E, Lowe D, Esfandiari MA. Evolution of the earth observing system (EOS) data and information system (EOSDIS). In: Di L, Ramapriyan H, eds., *Standard-based data and information systems for earth observation*. Lecture notes in geoinformation and cartography. Berlin; Heidelberg: Springer; 2010.
- [20] Beaujardiere J. OpenGIS web map service. Client implementation specification version 1.3.0. OGC:06042; 2006. Available at: http://portal.opengeospatial.org/files/?artifact_id=14416 (verified 08 February 2020)
- [21] Vretanos PA. OpenGIS web feature service. Implementation specification version 1.1.0. OGC: 04-094; 2005. Available at: http://portal.opengeospatial.org/files/?artifact_id=8339 (verified 08 February 2020)
- [22] Whiteside A, Evans JD. Web coverage service implementation standard version 1.1.2. OGC 07-067r5; 2008. Available at: <https://portal.opengeospatial.org/files/07-067r5> (verified 08 February 2020)
- [23] Nebert D, Whiteside A, Vretanos P. OpenGIS catalogue service implementation specification; 2007. Available at: <https://www.opengeospatial.org/standards/cat> (verified 08 February 2020)
- [24] Foster I. Service-oriented science. *Science*. 2005;308:814–17.
- [25] GEO (The Group on Earth Observations). The global earth observation system of systems 10-year Implementation plan.

- The Group on Earth Observations; 2015. Available at: <http://www.preventionweb.net/english/professional/publications/v.php?id=8631> (verified 08 February, 2020)
- [26] Craglia M. Building INSPIRE: the spatial data infrastructure for Europe; 2010. Available at: <https://www.esri.com/news/arcnews/spring10articles/building-inspire.html> (verified 08 February 2020)
- [27] Craglia M, Annoni A. INSPIRE: An innovative approach to the development of spatial data infrastructures in Europe; 2007. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.556.3798&rep=rep1&type=pdf> (verified 08 February, 2020)
- [28] National Catalogue Service for Geographic Information of China (NCSGIC). Introduction of national catalogue service for geographic Information of China; 2016. Available at: <https://www.gov.cn/xinwen/2016/10/13/5118709/files/6aa6f08-c41aa4986b8b6071b0cfbedb2.pdf> (verified 08 February 2020)
- [29] Janowicz K, Schade S, Bröring A, Keßler C, Maué P, Stasch C. Semantic enablement for spatial data infrastructures. *Trans GIS*. 2010;14:111–29.
- [30] Li W, Yang C, Raskin R. A semantic enhanced model for searching in spatial web portals. *Proceedings of semantic scientific knowledge integration AAI/SSKI symposium*, 26–28 March 2008. Palo Alto, CA: Association of American Artificial Intelligence; 2008. p. 47–50.
- [31] Zhao Q, Wang C, Wang P, Zhou M, Jiang C. A novel method on information recommendation via hybrid similarity. *IEEE Trans Syst Man Cybern Syst*. 2018;48(3):448–59.
- [32] Charbel N, Sallaberry C, Laborie S, Tekli G, Chbeir R. LinkedMDR: a collective knowledge representation of a heterogeneous document corpus. *database and expert systems applications: 28th international conference, DEXA 2017, France: Lyon; August 28–31; 2017*. p. 362–77.
- [33] Baziz M, Boughanem M, Aussenac-Gilles N. Conceptual indexing based on document content representation. In: Crestani F, Ruthven I, eds., *Context: nature, impact, and role. CoLIS 2005. Lecture notes in computer science*, vol. 3507. Berlin, Heidelberg: Springer; 2005. p. 171–86.
- [34] Finkelstein L, Gabrilovich E, Matias Y, Rivlin E, Solan Z, Wolfman G, et al. Placing search in context: the concept revisited. *ACM Trans Inf Syst*. 2002;20:116–31.
- [35] Buscaldi D, Zargayouna H. YaSemLR: yet another semantic information retrieval system. *International conference on information and knowledge management, proceedings; 2013*. p. 13–16. doi: 10.1145/2513204.2513211.
- [36] D’Amato C, Fanizzi N, Esposito F. A semantic similarity measure for expressive description logics. *arXiv preprint*; 2009. Available at: <http://arxiv.org/abs/0911.5043> (verified 10 May 2020)
- [37] Gracia J, Mena E. Web-based measure of semantic relatedness. In: Bailey J, Maier D, Schewe KD, Thalheim B, Wang XS, eds., *Web information systems engineering – WISE 2008. WISE 2008. Lecture notes in computer science*, vol. 5175. Berlin, Heidelberg: Springer; 2008.
- [38] Gui Z, Yang C, Xia J, Liu K, Xu C, Li J, et al. A performance, semantic and service quality enhanced distributed search engine for improving geospatial resource discovery. *Int J Geogr Inf Sci*. 2013;27(6):1109–132. doi: 10.1080/13658816.2012.739692.
- [39] Hu K, Gui Z, Cheng X, Qi K, Zheng J, You L, et al. Content-based discovery for web map service using support vector machine and user relevance feedback. *PLoS One*. 2016;11(11):e0166098. doi: 10.1371/journal.pone.0166098.
- [40] Arpinar IB, Sheth A, Ramakrishna C, Usery EL, Azania M, Kwan M-P. Geospatial ontology development and semantic analytics. *Trans GIS*. 2006;10:551–57.
- [41] Raskin RG, Pan MJ. Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). *Comput Geosci*. 2005;31:1119–125.
- [42] Li Z, Yang C, Wu H, Li W, Miao L. An optimized framework for seamlessly integrating ogc web services to support geospatial sciences. *Int J Geogr Inf Sci*. 2011;25:595–613.
- [43] Berners-Lee T, Handler J, Lassila O. The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Sci Am*. 2001;284:28–37.
- [44] Li Y, Bandar ZA, McLean D. An approach for measuring semantic similarity between words using multiple information sources. *IEEE Trans Knowl Data Eng*. 2003;15(4):871–82.
- [45] Retzer S, Yoong P, Hooper V. Inter-organisational knowledge transfer in social networks: a definition of intermediate ties. *Inf Syst Front*. 2012;14(2):343–61.
- [46] Jiang Y, Zhang X, Tang Y, Nie R. Feature-based approaches to semantic similarity assessment of concepts using Wikipedia. *Inf Process Manage*. 2015;51(3):215–34.
- [47] Buscaldi D, Bessagnet M, Royer A, Sallaberry C. Using the semantics of texts for information retrieval: a concept- and domain relation-based approach. *Adv Intell Syst Comput*. 2014;241:257–66.
- [48] Sanchez D, Batet M. Semantic similarity estimation in the biomedical domain: an ontology-based information-theoretic perspective. *J Biomed Inform*. 2011;44(5):749–59.
- [49] Sanchez D, Batet M, Isern D. Ontology-based information content computation. *Knowl Syst*. 2011;24(2):297–303.
- [50] Jiang Y, Bai W, Zhang X, Hu J. Wikipedia-based information content and semantic similarity computation. *Inf Process Manage*. 2017;53(1):248–65.
- [51] Hu X, Feng Z, Chen S, Huang K, Li J, Zhou M. Accurate identification of ontology alignments at different granularity levels. *IEEE Access*. 2017;5:105–20.
- [52] Qiu J, Chai Y, Tian Z, Du X, Guizani M. Automatic concept extraction based on semantic graphs from big data in smart city. *IEEE Trans Comput Soc Syst*. 2019;2019:1–9.
- [53] Batet M, Sanchez D, Valls A, Gibert K. Semantic similarity estimation from multiple ontologies. *Appl Intell*. 2013;38(1):29–44.
- [54] Zhang Q, Haglin D. Semantic similarity between ontologies at different scales. *IEEE/CAA J Autom Sin*. 2016;3(2):132–40.
- [55] Bakillah M, Li RY, Liang SH. Geo-located community detection in twitter with enhanced fast-greedy optimization of modularity: the case study of typhoon haiyan. *Int J Geogr Inf Sci*. 2015;29:258–79.
- [56] Zhang Y, Wu W, Wang Q, Su Fe A. Geo-event-based geospatial information service: a case study of typhoon hazard. *Sustainability*. 2017;9:534. doi: 10.3390/su9040534.
- [57] Nauman M, Khan S, Amin M, Hussain F. Resolving lexical ambiguities in folksonomy based search systems through commonsense and personalization. In *Proceedings of the*

- workshop on semantic search at the fifth european semantic web conference. Spain: Tenerife; 2008. p. 2–13.
- [58] Gruber T. A translation approach to portable ontology specification. *Knowl Acquis.* 1993;5:199–220.
 - [59] Latre MÁ, Lacasta J, Mojica E, Nogueras-Iso J, Zarazaga-Soria FJ. An approach to facilitate the integration of hydrological data by means of ontologies and multilingual thesauri. In: Sester M, Bernard L, Paelke V, eds., *Advances in GIScience. Lecture notes in geoinformation and cartography.* Berlin: Heidelberg: Springer; 2009.
 - [60] Jain V, Singh M. Ontology development and query retrieval using Protégé tool. *IJ Intell Syst Appl.* 2003;9:67–75.
 - [61] Sycara K, Widoff S, Klush M, Lu J. Larks: dynamic matchmaking among heterogeneous software agents in cyberspace. *Auton Agents Multi-Agent Syst.* 2002;5:173–203.
 - [62] Miller GA. WordNet: a lexical database for English. *Commun ACM.* 1995;38:39–41.
 - [63] Fellbaum C. WordNet: an electronic lexical database. Cambridge, MA: MIT Press; 1998.
 - [64] Vockner B, Mittlböck M. Geo-enrichment and semantic enhancement of metadata sets to augment discovery in geo-portals. *ISPRS Int J Geo-Inf.* 2014;3:345–67.
 - [65] Rodriguez MA, Edenhofer MJ. Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure. *Int J Geogr Inf Sci.* 2004;18:229–56.
 - [66] Bhattacharjee S, Mitra Ghosh SK. Spatial interpolation to predict missing attributes in GIS using semantic kriging. *IEEE Trans Geosci Remote Sens.* 2014;52:4771–80.
 - [67] Al-Bakri M, Fairbairn D. Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources. *Int J Geogr Inf Sci.* 2012;26:1437–456.
 - [68] Chen N, He J, Yang C, Wang C. A node semantic similarity schema-matching method for multi-version web coverage service retrieval. *Int J Geogr Inf Sci.* 2012;26:1051–72.
 - [69] Wu Z, Palmer M. Verb semantics and lexical selection. *Proceedings of the 32nd annual meeting of the association for computational linguistics, New Mexico; 1994.* p. 133–38.