International Journal of
Geo-Information

MDPI

Article

Geospatial Information Categories Mapping in a Cross-lingual Environment: A Case Study of "Surface Water" Categories in Chinese and American Topographic Maps

Xi Kuai [1], Lin Li [1,*], Heng Luo [1], Shen Hang [1], Zhijun Zhang [2] and Yu Liu [1]

1 School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China; kuaixi@whu.edu.cn (X.K.); luoheng@whu.edu.cn (H.L.); acyrt@163.com (S.H.); liuyu_whu@foxmail.com (Y.L.)

2 Tianjin Institute of Surveying and Mapping, Tianjin 300381, China; zhangzhijun@whu.edu.cn

* Correspondence: lilin@whu.edu.cn; Tel.: +86-138-7150-4963

Academic Editors: E. Lynn Usery, Dalia Varanka and Wolfgang Kainz
Received: 4 April 2016; Accepted: 6 June 2016; Published: 14 June 2016

Abstract: The need for integrating geospatial information (GI) data from various heterogeneous sources has seen increased importance for geographic information system (GIS) interoperability. Using domain ontologies to clarify and integrate the semantics of data is considered as a crucial step for successful semantic integration in the GI domain. Nevertheless, mechanisms are still needed to facilitate semantic mapping between GI ontologies described in different natural languages. This research establishes a formal ontology model for cross-lingual geospatial information ontology mapping. By first extracting semantic primitives from a free-text definition of categories in two GI classification standards with different natural languages, an ontology-driven approach is used, and a formal ontology model is established to formally represent these semantic primitives into semantic statements, in which the spatial-related properties and relations are considered as crucial statements for the representation and identification of the semantics of the GI categories. Then, an algorithm is proposed to compare these semantic statements in a cross-lingual environment. We further design a similarity calculation algorithm based on the proposed formal ontology model to distance the semantic similarities and identify the mapping relationships between categories. In particular, we work with two GI classification standards for Chinese and American topographic maps. The experimental results demonstrate the feasibility and reliability of the proposed model for cross-lingual geospatial information ontology mapping.

Keywords: geographic information systems; semantic interoperability; cross-lingual; lightweight ontology; topographic map

1. Introduction

The vision of a "Digital Earth" articulated by US Vice President Al Gore [1–3] has contributed significantly to the growth in global geospatial information (GI) on physical and social environments. However, how to query, retrieve, and manipulate those data from heterogeneous sources has challenged the GI community [2–5]. Thus, an approach to integrating GI data from various heterogeneous sources has found increased importance [6].

A data integration process is not as simple as joining several systems because any effort at information sharing runs into the problem of semantic heterogeneity [7]. Semantic heterogeneity occurs when enabling interoperability across geographic information systems (GIS) [8–11] because GIS are often designed to address data from highly distributed, multidisciplinary, and cross-lingual data sources with different application demands [12]. Clarifying the semantics of data is therefore a crucial step toward successful data integration [13]. To achieve this, domain ontologies are built as a mediator to exchange information in such a way that the precise meaning of the data (*i.e.*, semantics) is readily retrievable beyond simple keyword matching via knowledge representation languages and reasoning [7,13–15]. Thus, ontology engineering has been regarded as an effective means of providing seamless connection between component GIS at the semantic level [8,12,16].

While the GI community widely acknowledges the utility of ontology technologies, two main problems need to be solved for GI ontology engineering and sharing are as follows: (1) traditional ontology research and technologies focusing on terminology and schema cannot answer the question surrounding how to engineer GI ontologies and integrate them with GIS or Spatial Data Infrastructures (SDI) [6]; and (2) mechanisms still need to be explored for GI ontology mapping in cross-lingual environments to facilitate semantic integration between GI ontologies described in different natural languages [17–20].

The reason for the first problem is that GI features and categories are a product of spatial cognition and social convention; thus, the ontology engineering works in GI domains are different from others, in which the location, topology, mereology, and other spatial relations play a major role in the identification and representation of GI semantics [14]. For example, from a feature-driven ontology perspective, the geographic categories "river" and "bank" should be specified into different classes, and normally, the spatial relation "adjacent-to" between these two categories is missing. Moreover, geographic and non-geographic entities are ontologically distinct in a number of ways [21]. To enhance the semantic expressiveness and overcome the issue of semantic heterogeneity during the GI ontology engineering process, the spatial-related characteristics of GI categories must be considered to enrich the spatial-related semantics of the given ontology.

Although a majority of current GI ontologies have been developed in English with English vocabularies, the amount of multilingual content on the Semantic Web and thus the number of vocabularies/ontologies in multiple languages continue to grow [22]. Thus, methods for matching vocabularies across languages have become increasingly more important for promoting the accessibility of the data in multiple languages by end users [23]. As a motivating scenario, if a user wants to query the water level data along the Mekong River (The seventh longest river in the world, covering six different countries—Cambodia, Laos, Myanmar, Thailand, Vietnam and China—and the official languages of each country are different), there are several data providers offering the related GI data via their national GIS in their native natural languages. This situation has generated a substantial challenge to integrating highly heterogeneous GI data across natural language barriers.

The purpose of this study is to establish a formal ontology model for cross-lingual geospatial information ontology mapping. Starting from two GI classification standards with different natural languages—Chinese and English(for the sake of simplicity and clarity, this study was restricted to the "surface water" categories from these two standards)—a set of semantic primitives are extracted from the free-text definition of the categories in the standards by applying Natural Language Processing (NLP) techniques. Then, an ontology-driven approach is used, and the formal ontology model is established to formally represent these semantic primitives using semantic statements, in which the spatial-related properties and relations are considered as crucial statements for the representation and identification of the semantics of the GI categories. To overcome the natural language barrier, the statements in Chinese are translated into English by using machine translation tools, and the mapping relationships between statements are determined within an English context, which then serve as the basis for the similarity calculation between categories in different GI ontologies. Finally, a similarity calculation algorithm is designed to distance the semantic similarity between GI categories in different ontologies, and the final mapping relationships between pairs of categories are determined based on calculated similarity values. The contributions of the proposed approach include (1) the construction of the spatial-related semantic properties and relations to serve the requirement of the presentation and identification of the spatial characteristics of the GI categories and (2) the algorithms of GI ontology mapping in a cross-lingual environment based on formally represented and comparable semantic statements.

The remainder of this paper is organized as follows. Section 2 presents the related works in the literature. Then, the main procedure of our methodology is presented in Section 3. Next, a case study demonstrating the application of our method is shown in Section 4. Finally, conclusions are drawn, and future works are noted.

2. Related Works

2.1. Semantic Interpretation

Knowledge acquisition (KA) is a broad field that encompasses the processes of extracting, creating, and structuring knowledge from heterogeneous resources [24]. Semantic interpretation (SI) for KA is defined as the composition of two sub-processes: the extraction of semantic primitives from the free-text definition in ontologies and semantic enrichment based on the extracted semantic primitives.

The research on semantic primitive extraction builds on a large body of works within the fields of Natural Language Processing (NLP) [25]. NLP and text mining are research fields aimed at exploiting rich knowledge resources with the goal of understanding, extracting and retrieving semantic information from unstructured written text. Knowledge resources that have been used for these purposes include the entire range of terminologies, including lexicons, controlled vocabularies, thesauri, and ontologies [26,27]. Although numerous methods and algorithms have been developed recently (such as symbolic, statistical, and hybrid approaches) [26], a fully automated algorithm for semantic extraction using NLP techniques seems unachievable, and a manual process as an assistant is normally inevitable.

For semantic enrichment, authors in [28,29] proposed a systematic methodology to explore and identify semantic information provided by categories in geographic ontologies, in which the semantic representations of categories are enriched with a set of semantic properties and relations to reveal similarities and heterogeneities between these categories. Authors in [30] presented an axiomatic formalization of a theory of top-level relations (parthood relations, sub-universal relations, and cross-categorical relations) between three categories of geospatial-related entities, namely, individuals, universals, and collections. In addition, they demonstrated how a more exact understanding helps to overcome the semantic heterogeneous problems in the information integration process. In [13], the semantics of a concept in GI ontologies were presented using an extendable and structural definition framework composed of a number of RDF triple statements, and a comparison algorithm was designed to determine the semantic relationships of concepts between different domains. The primary objective of these studies was to extract and represent the semantics of concepts/entities based on structural common vocabularies, which make the semantics of the concepts/entities comparable. However, the structural common vocabularies in these works are determined by domain experts manually; thus, the objectivity and automation of the algorithms (avoiding *ad hoc* manual procedures and subjective experts' knowledge) remain quite limited.

2.2. Ontology Mapping

Mapping relationship discovery for ontologies has attracted considerable attention in recent years. Various approaches based on processes to find similarities between different but related ontologies have emerged [31]. With respect to the literature specifically oriented toward geospatial information (GI) ontologies, authors in [32] performed an analysis of the different models of semantic similarity measurement and evaluated these models with respect to the particular requirements of geospatial data. Authors in [7,33] systematically surveyed several of the most recent and often-referenced works on integrating GI and GI ontology mapping by applying comparison criteria, such as logical inference, mapping approaches, degree of automation, and geospatial relativity. In addition, a general conclusion is proposed that, for the ontology mapping task, the use of formal ontologies and, consequently, the use of reasoners should be mandatory.

In recent years, Volunteered Geographic Information (VGI) has been proposed for GI ontology mapping in a web environment. Authors in [34,35] devised a mechanism for computing the semantic similarity of the Open Street Map (OSM) geographic classes using volunteered lexical definitions to alleviate the semantic gap between different VGI producers. Another set of studies focused on introducing an artificial neural network approach to simulate the human perception and measure the semantic similarity between spatial entities for the purpose of improving the automaticity of the ontology mapping process [36,37].

All these proposals, combining the use of different models of semantic similarity measurement, have emerged to provide solutions to existing GI ontology mapping problems in English environments. However, the semantic web and ontology engineering have experienced significant advancements in standards and techniques, and increasingly more domain ontologies and localization content in the semantic web are described using native natural languages [23]. There is a pressing need for cross-lingual ontology mapping mechanisms in the GI community that are designed to reconcile semantics of different ontologies in multilingual environments and to improve the accessibility of various GI ontologies across language barriers [38].

Author in [39] groups the existing cross lingual ontology mapping(CLOM) algorithms into following categories: manual processing [40–42], corpus-based approach [43], linguistic enrichment [44], indirect alignment composition [45], and translation-based approach [39,46]. Compared to these CLOM approaches, translation-based CLOM is currently a very popular approach that is exercised by several researchers [47–50], which is enabled by translations achieved through the use of machine translation (MT) tools, bilingual/multilingual thesauri, dictionaries *etc.* Typically, these approaches rely only on string-based lexical comparisons of entity names and descriptions [51–54], while the comparisons between semantic interpretation, e.g., model-theoretic semantics of entities are missing.

3. Methodologies

The main procedure for our methodologies is divided into two sub-processes, as shown in Figure 1. In the semantic interpretation process, two GI formal ontologies, namely, O_A and O_B, are established from the free-text definition of the corresponding classification standards with different natural languages. In the ontology mapping process, all the category names and semantic statements in O_A are translated from L_A into L_B, and the mapping relationships between category names and semantic statements are determined within the same language context, which then serve as the basis for the similarity calculation between categories in different GI ontologies. Finally, a similarity calculation algorithm is designed, and the final mapping results between pairs of categories in different classification standards are determined.

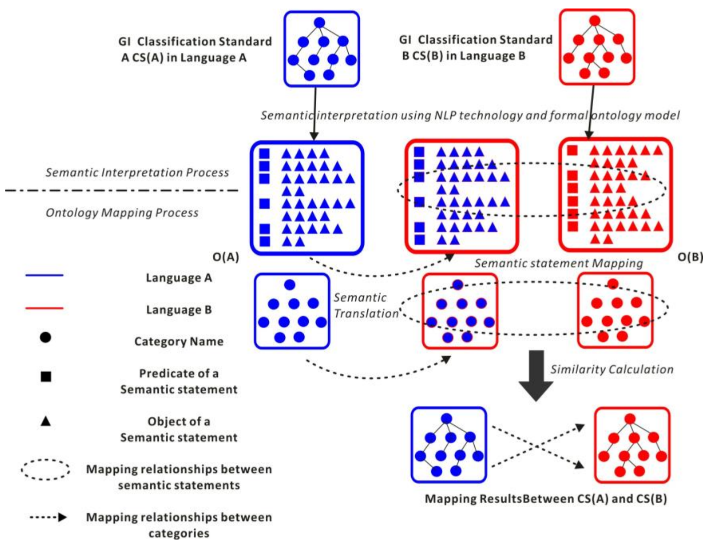


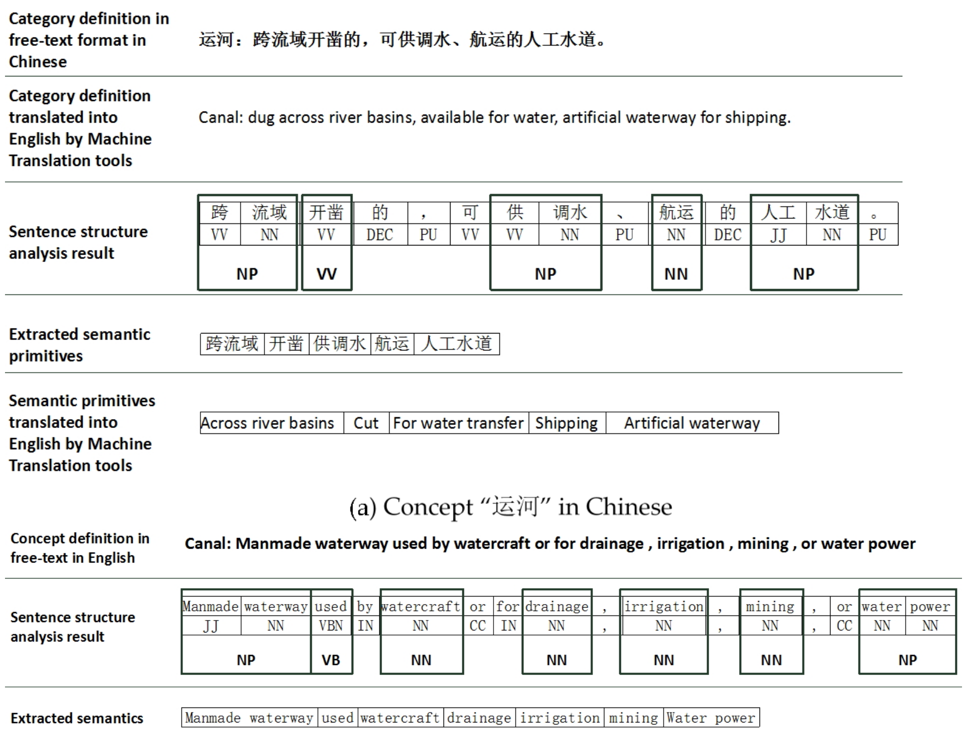
Figure 1. Main procedure for our methodologies.

3.1. Semantic Interpretation

3.1.1. Semantic Primitive Extraction

In geospatial information repositories, free-text definitions are often the primary and only available objective descriptions of categories. Semantic primitives are syntactic and lexical patterns in the free-text definition and can be extracted using NLP tools [55]. The fields of studies on NLP have developed methods and algorithms for information retrieval and extraction from free-text knowledge resources. The methodology adopted here for analyzing definitions and extracting semantic primitives was introduced by [56]. In this research, the lexical patterns of nominal phrases and verb phrases are considered as semantic primitives. An example is illustrated in Figure 2, and the main steps of the process are as follows:

1. One category definition in free-text format is chosen as the input natural language material;
2. Word segmentation is performed to split the whole sentence into individual words;
3. Words are categorized and tagged into their parts-of-speech tag sets (see Tables 1 and 2) and labeled accordingly;
4. The nominal phrases and verb phrases are chunked, and the sentence structure is analyzed to extract lexical patterns as the semantic primitives.



(b) Concept “Canal” in English

Figure 2. Extract the semantic primitives from the free-text definitions by applying NLP tools in Chinese and English.

Table 1. Summary of the Penn Treebank Part-of-Speech Tag sets in English.

Part of Speech	Abbr	Part of Speech	Abbr	Part of Speech	Abbr
Adjective	JJ	Exclamation	UH	Possessive wh-pronoun	WP$
Adjective comparative	JJR	Existential	EX	Predeterminer	PDT
Adjective superlative	JJS	Foreign word	FW	Proper noun plural	NNPS
Adverb	RB	Gerund	VBG	Proper noun	NNP

Table 1. *Cont.*

Part of Speech	Abbr	Part of Speech	Abbr	Part of Speech	Abbr
Adverb comparative	RBR	List item marker	LS	Symbol	SYM
Adverb superlative	RBS	Modal verb	MD	to	TO
Article	DT	Participle past	VBN	Verb base form	VB
Cardinal number	CD	Particle	RP	Verb present tense	VBP
Common noun plural	NNS	Past tense verb	VBD	Verb 3rd person singular	VBZ
Common noun singular or mass	NN	Personal pronoun	PRP	Wh-determiner	WDT
Conjunction coordinating	CC	Possessive ending	POS	Wh-pronoun	WP
Conjunction subordinating	IN	Possessive pronoun	PRP$	Wh-adverb	WRB

Table 2. Summary of the Penn Treebank Part-of-Speech Tag sets in Chinese.

Part of Speech	Abbr	Part of Speech	Abbr	Part of Speech	Abbr
adverb	AD	determiner	DT	proper noun	NR
aspect marker	AS	for words "dengdeng"("等等")	ETC	temporal noun	NT
in ba-construction	BA	foreign words	FW	ordinal number	OD
coordinating conjunction	CC	interjection	IJ	onomatopoeia	ON
cardinal number	CD	other noun-modifier	JJ	preposition excl. "bei"("被") and "ba"("把")	P
subordinating conjunction	CS	"bei"("被") in long bei-const	LB	pronoun	PN
"de"("的")in a relative-clause	DEC	localizer	LC	punctuation	PU
Associative "de"	DEG	measure word	M	"bei"("被") in short bei-const	SB
"de"("得") inV-deconst. and V-de-R	DER	other particle	MSP	sentence-final particle	SP
"di"("地") before VP	DEV	common noun	NN	predicative adjective	VA
"shi"("是")	VC	"you"("有") as the main verb	VE	other verb	VV

3.1.2. Construction of the Formal Ontology Model

From Wikipedia an "ontology in information science" is a formal naming and definition of the types, properties, and interrelationships of the concepts that really or fundamentally exist for a particular domain of discourse. It is thus a practical application of philosophical ontology, with taxonomy. In addition, a domain ontology (or domain-specific ontology) represents concepts that belong to a general domain. Thus, for a formal representation [57,58], the domain ontology (denoted by O_{Domain}), and concepts in the domain could be summarized by Equations (1)–(3).

$$O_{Domain} = \{S(C_{Domain}), S(R_C), S(H_C), S(P_C)\} \tag{1}$$

$$C_{Domain} = \{T_C, D_C\} \tag{2}$$

$$D_C = \{R_C, H_C, P_C\}, R_C \in S(R_C), H_C \in S(H_C), P_C \in S(P_C) \tag{3}$$

In Equation (1), $S(C_{Domain})$ represents the set of concepts in a domain, and the semantics of each concept in the domain are categorized into different groups, namely, $S(H_C)$, $S(R_C)$, and $S(P_C)$; $S(H_C)$ represents the set of the hierarchical relations about the taxonomic information in O_{Domain}, $S(R_C)$ represents the set of other interrelations between these concepts, and $S(P_C)$ represents the set of the semantic properties belong to the concepts in this domain.

In Equation (2), the semantics of a concept in the domain are considered as the composition of terminology of this concept (denoted by T_C) and structural definition of this concept (denoted by D_C). Unlike the free-text format of definition, D_C commonly consists of the semantic properties of the concept (P_C), the hierarchical relation (H_C) and other interrelations (R_C) between this concept and other concepts in the domain. Thus, from Equations (2) and (3), a certain concept in the domain, C_{Domain} can be deduced as a function of T_C, R_C, H_C, and P_C in Equation (4)

$$C_{Domain} = \{T_C, R_C, H_C, P_C\}, R_C \in S(R_C), H_C \in S(H_C), P_C \in S(P_C) \tag{4}$$

in which R_C, H_C, P_C are used to represent the semantics of this concept, and belong to $S(R_C)$, $S(H_C)$, $S(P_C)$, respectively.

Considering the situation in the GI domain, we use the word "category" instead of "concept". Because the semantic characteristics of the GI category are highly correlated in space and time [59], the spatial- and temporal-related semantic properties and relations should be included in the model as crucial vocabularies for the representation and identification of the semantics of the GI categories. Thus, the GI ontology O_{GI} and the semantics of a certain category C_{GI} in O_{GI} can be represented as Equations (5) and (6):

$$O_{GI} = \{S(T_C), S(R_S), S(R_T), S(R_C), S(H_C), S(P_S), S(P_T), S(P_C)\} \tag{5}$$

$$\begin{aligned} C_{GI} = \big\{ & (T_C = V_{T_C}) \cap (R_S = V_{R_S}) \cap (R_T = V_{R_T}) \cap (R_C = V_{R_C}) \cap (H_C = V_{H_C}) \\ & \cap (P_S = V_{P_S}) \cap (P_T = V_{P_T}) \cap (P_C = V_{P_C}) \big\} \end{aligned} \tag{6}$$

In Equation (5), $S(T_C)$ represents the set of the category names in O_{GI}; $S(R_S)$, $S(R_T)$ represent the set of the spatial-related and temporal-related semantic relations between categories; $S(H_C)$represents the set of hierarchical relations; $S(P_S)$, $S(P_T)$ represent the set of the spatial-related and temporal-related semantic properties belong to the categories in O_{GI}; and $S(R_C)$, $S(P_C)$ represent the set of other semantic properties and relations in O_{GI}. And in Equation (6), V_x represents the values of certain semantic properties/relations of C_{GI}; T_C, R_S, R_T, R_C, H_C, P_S, P_T, P_C are used to represent the semantics of C_{GI}, and belong to $S(T_C)$, $S(R_S)$, $S(R_T)$, $S(R_C)$, $S(H_C)$, $S(P_S)$, $S(P_T)$, $S(P_C)$, respectively.

In order to solve the problems of geographic representation, authors in [60] distinguished three main theoretical tools that are required for the purposes of developing an overall formal theory of spatial representation, namely, mereology, location, and topology, these theoretical tools are selected as the basis for defining spatial-related semantics in our formal ontology model. In addition, geographic entities in reality is essentially dynamic, authors in [61] pointed out that a good ontology must be capable of accounting for spatial reality both synchronically (as it exists at a time) and diachronically(as it unfolds through time), thus the "time point" and "time period" properties should be used to describe dynamic characteristics of the geographic entities in our model. Moreover, in order to specify semantic relations and properties used in geographic definitions, authors in [28] analyzed several geographic ontologies and identified patterns which were systematically used to express specific semantic relations and properties, including hierarchical relations, part-whole relations and neighborhood relations, and semantic properties such as purpose, nature, material, size, and so on.

Based on previous researches and our formal ontology model in Equation (5), the semantic property and relation types in our model are subdivided and shown in Figure 3.

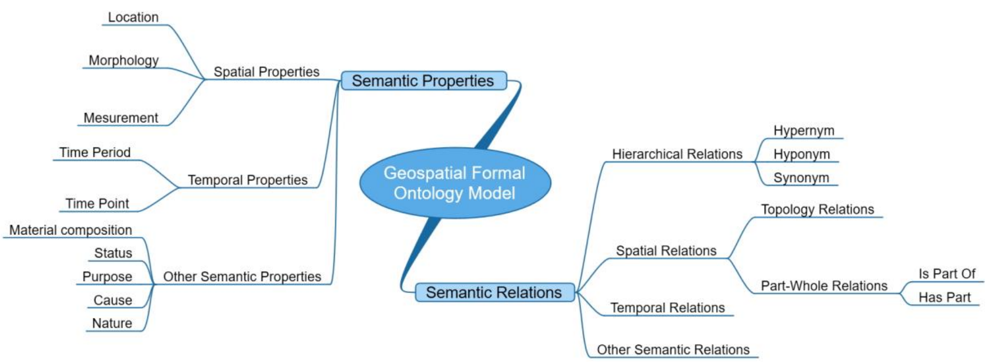


Figure 3. Semantic property and relation groups in the Geospatial Formal Ontology Model.

3.1.3. Transformation from Semantic Primitives to Formal Ontology Model

In order to make the semantic primitives structural and comparable, domain experts are responsible for analyzing these semantic primitives and transforming them into different groups of semantic properties/relations in our geospatial formal ontology model. The famous triple statement

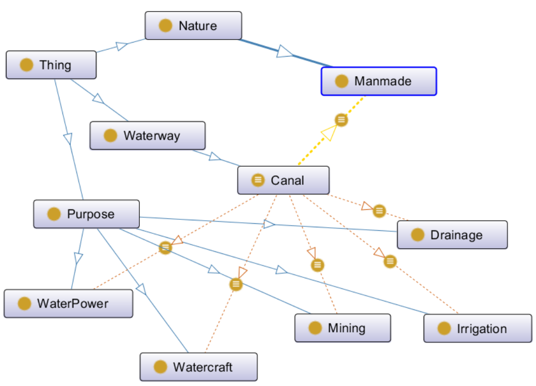
Subject-Predicate-Object and the web ontology language (OWL) are selected as the basis for presenting the semantic properties/relations and their values in a machine-readable manner. The *Subject* represents a C_{GI} in O_{GI}; the *Predicate* is a certain semantic property or semantic relation type illustrated in Figure 3, in which all of the semantic relations are presented by object property and the *Object* in these semantic relations is another C_{GI} in O_{GI} or an "owl:class" object type, while most of the semantic properties are presented by object property too, and a few of them are presented by datatype property in OWL syntax, and the *Object* in these semantic properties is a "rdfs:literal" datatype. The following rules are adopted to handle the formalization process:

(1) The GI category can be represented by a number of semantic relations/properties; however, the number of semantic relations/properties involved should be minimized to avoid redundancy.
(2) Not every GI category must cover all semantic relations/properties in the model. The situation whereby two different categories use the same set of semantic relations/properties to represent their semantics cannot be guaranteed.
(3) The semantic information of a certain category in our model is the combination of different semantic relations-properties and their values. This combination should represent all the semantic information of the category and be able to distinguish the different geospatial categories within and beyond domain ontologies to avoid ambiguity.
(4) The hypernym, hyponym, and synonym relations should be included in the hierarchical relation group. If category A is a hyponym of category B, A must inherent all the semantic properties/relations of B to retain semantic consistency.

According to the above-mentioned rules, the semantic primitives can be specified into these properties/relations types as structure statements for identification and representation of the GI categories. For example, the free-text definition of the "canal" category in English is "manmade waterway used by watercraft or for drainage, irrigation, mining, or water power". In addition, the semantic primitives of the "canal" category are extracted by applying NLP tools to the set of phrases including "manmade waterway", "used", "watercraft", "drainage", "irrigation", "mining", and "water power". Then, transforming these semantic primitives into the proposed formal ontology model, the semantics of the category "canal" can be represented as a set of several semantic statements as follows:

$$\begin{aligned} C_{Canal} = \{T_C = \text{“}canal\text{”} \cap H_C = \text{“}Hypernym : waterway\text{”} \cap P_C = \text{“}(Purpose : watercraft) \cup \\ (Purpose : drainage) \cup (Purpose : irrigation) \cup (Purpose : mining) \cup \\ (Purpose : waterpower)\text{”} \cap P_C = \text{“}Nature : Manmade\text{”}\} \end{aligned} \tag{7}$$

In addition, the representation in OWL format is illustrated in Figure 4.



(a)

```
:Canal rdf:type owl:Class ;

       rdfs:label "Canal"@en ;

       owl:equivalentClass [ rdf:type owl:Restriction ;
                             owl:onProperty :Purpose ;
                             owl:someValuesFrom :Irrigation
                           ] ,
                           [ rdf:type owl:Restriction ;
                             owl:onProperty :Nature ;
                             owl:allValuesFrom :Manmade
                           ] ,
                           [ rdf:type owl:Restriction ;
                             owl:onProperty :Purpose ;
                             owl:someValuesFrom :WaterPower
                           ] ,
                           [ rdf:type owl:Restriction ;
                             owl:onProperty :Purpose ;
                             owl:someValuesFrom :Drainage
                           ] ,
                           [ rdf:type owl:Restriction ;
                             owl:onProperty :Purpose ;
                             owl:someValuesFrom :Mining
                           ] ,
                           [ rdf:type owl:Restriction ;
                             owl:onProperty :Purpose ;
                             owl:someValuesFrom :Watercraft
                           ] ;

       rdfs:subClassOf :Waterway .
```

(b)

Figure 4. Representation of the category "canal" in OWL format: (**a**) The OntoGraf view in Protégé and (**b**) the semantic statement presentation in turtle file format.

3.2. Ontology Mapping Algorithms

3.2.1. Semantics Translation

Assume that we have formal ontologies O_A, O_B presented in different natural languages, namely, language A (L_A) and B (L_B), respectively. According to the geospatial formal ontology model introduced in Section 3.1.2, the semantics of ontologies O_A and O_B consist of category name sets $S\ (CN_A)$ and $S\ (CN_B)$ and semantic statement sets $S\ (SS_A)$ and $S\ (SS_B)$, labeled in different natural languages, in which the semantic statement consists of semantic property/relation types (as illustrated in Figure 3 in Section 3.1.2) and their corresponding values. In order to cross the natural language barrier between O_A and O_B, algorithm 1 illustrates the process of semantics translation between L_A and L_B:

Algorithm 1. Semantics Translation.

1: **Input:** Formal ontologies $O_A(S(CN_A), S(SS_A))$ in L_A
2: **Output:** Translation candidate result set of the semantics in O_A, $O'_A\ (S(TC(CN_A)), S(TC(SS_A\text{-}object)))$ in
3: L_B.
4: **Symbols:**
5: $S(TC(CN_A))$—Translation candidate result set of $S(CN_A)$ in L_B.
6: $S(TC(SS_A))$—Translation candidate result set of $S(SS_A)$ in L_B.
7: ss_A-object—The *Object* part of the semantic statement ss_A.
8: 1:**for each category name** cn_A in $S(CN_A)$, translate cn_A in L_A into $cn_A{}'$ in L_B by using different Machine Translation (MT) web services (Google Translator API at" http://translate.google.cn/", Bing Translator API at" http://www.bing.com/translator/?ref=SALL&mkt=zh-CN", and Baidu Translator API at "http://fanyi.baidu.com/?aldtype=16047#zh/en/"), collect all of the translation results about cn_A, into the translation candidate results $TC(cn_A)$, and store all of the category name translation candidate results into the translation candidate set $S(TC(CN_A))$;
9: 2:**for each semantic statement** ss_A in $S(SS_A)$, according to the OWL triple statement syntax, it can be subdivided into three part, *Subject*, *Predicate*, and *Object*, translate ss_A*-object* in L_A into $ss_A{}'$*-object* in L_B by using different Machine Translation (MT) web services, collect all of the translation results about ss_A*-object*, into the translation candidate results $TC(ss_A\text{-}object)$, and store all of the semantic statements translation candidate results into the translation candidate set $S(TC(SS_A\text{-}object))$.
10: Take the "运河" category in Chinese as an example, the semantic primitives of the "运河" category are extracted by applying NLP tools to the set of phrases including "跨流域", "开凿", "供调水", "航运", "人工水道". Then, transforming these semantic primitives into the proposed formal ontology model, the semantics of the category "运河"can be represented as a set of several semantic statements as follows:
11:

$$C_{运河} = \Big\{T_C = ''\ 运河'' \cap H_C = ''\ Hypernym : 水道'' \cap P_C = ''\ (Purpose : 调水) \cup (Purpose : 航运)'' \cap P_C = ''\ Nature : 人工'' \cap R_S = ''\ Topology : 跨流域''\Big\} \tag{8}$$

12: And the semantics translation result of $C_{运河}$ in English is as follows:
13:

$$\begin{aligned} C_{运河} = \{T_C = \text{“(Canal)”} \cap H_C = \text{“}Hypernym : \text{(Waterway, Aqueduct)”} \cap \\ P_C = \text{“}(Purpose : \text{(Water transfer, Diversion)}) \cup \\ (Purpose : \text{(Shipping)})\text{”} \cap P_C = \text{“}Nature : \text{(Manual,Artificial)”} \cap \\ R_S = \text{“}Topology : \text{(Inter-basin, Across river basins)”}\} \end{aligned} \tag{9}$$

3.2.2. Semantic Statement Mapping

To determine the mapping relationships between categories in different GI ontologies, the mapping relationships at the semantic statement level should be determined first because the semantic statement presents the most detailed semantic characteristics of the compared categories. Once their relationships are determined, the similarity between categories can be determined quantitatively. Algorithm 2 shows the comparison process for category names and semantic statements between O_A and O_B. In addition, all the mapping results $M(O_A, O_B)$ are stored as the basis for the similarity calculation between the concepts in different GI ontologies.

Algorithm 2. Semantic Statement Mapping.

Input: O'_A $(S(TC(CN_A)), S(TC(SS_A\text{-}object)))$ in L_B, Formal ontologies $O_B(S(CN_B), S(SS_B))$ in L_B
Output: Mapping result set $M(O_A, O_B)$ about category names and semantic statements between
O_A and O_B.
Symbols:
$T(ss)$—semantic property/relation types for a certain semantic statement ss.
$M(O_A{}', O_B)$—mapping relationships about category names and semantic statements between $O_A{}'$
and O_B.
1: **for each category name** cn_A in $S(CN_A)$, find the translation candidate results of cn_A, $TC(cn_A)$,
  2: **for each translation candidate** $tc(cn_A)$ in $TC(cn_A)$, search $S(CN_B)$ *in* O_B, find the matched
  category name cn_B in $S(CN_B)$ by applying Equation(10),
    3: If there is a translation candidate $tc(cn_A)$ has the mapping relationship "exact match"
    with cn_B, store the mapping result $m(cn_A, cn_B$, 'exact match') in $M(O_A, O_B)$;
    4: else If there is a translation candidate $tc(cn_A)$ has the mapping relationship
    "close match" with cn_B, store the mapping result $m(cn_A, cn_B$, 'close match') in $M(O_A, O_B)$;
5: **for each semantic statement** ss_A in $S(SS_A)$, find the translation candidate results of ss_A-*object*,
$TC(ss_A$-*object*),
  6: **for each translation candidate** $tc(ss_A$-*object*$)$ in $TC(ss_A$-*object*), search $S(SS_B$-*object*$)$ *in* O_B,
  find the matched semantic statement *Object*, ss_B-*object* in $S(SS_B)$ by applying Equation(10),
    7: If there is a translation candidate $tc(ss_A$-*object*$)$ has the mapping relationship "exact
    match" with ss_B-*object*, and $T(ss_A)$ equals $T(ss_B)$, store the mapping result $m(ss_A, ss_B$, 'exact
    match') in $M(O_A, O_B)$;
    8: else If there is a translation candidate $tc(ss_A$-*object*$)$ has the mapping relationship "close
    match" with ss_B-*object*, and $T(ss_A)$ equals $T(ss_B)$, store the mapping result $m(ss_A, ss_B$, 'close
    match') in $M(O_A, O_B)$.

$$m(A,B) = \begin{cases} exactly\ match, & A\ is\ the\ same\ word\ or\ synonym\ of\ B \\ close\ match, & A\ is\ the\ near\ synonym\ of\ B \\ not\ match, & otherwise \end{cases} \tag{10}$$

3.2.3. Similarity Calculation

Given two categories, C_a and C_b in the formal ontologies O_A and O_B, respectively, based on the $M(O_A, O_B)$, the semantic similarity between C_a and C_b can be calculated using algorithm 3.

Algorithm 3. Similarity Calculation.

Input: Categories $C_a(CN_a, SS_a)$ in O_A, $C_b(CN_b, SS_b)$ in O_B and mapping relationship set $M(O_A, O_B)$
about category names and semantic statements between O_A and O_B.
Output: Semantic similarity value between C_a and C_b, $Sim(a, b)$.
Symbols:
$Cot(SS_a)$—the number of semantic statements in SS_a.
$Cot(SS_b)$—the number of semantic statements in SS_b.
$m(CN_a, CN_b)$—mapping relationship between CN_a and CN_b.
$m(SS_a(i), SS_b(j))$—mapping relationship between $SS_a(i)$ in C_a and $SS_b(j)$ in C_b.
$Pt(SS_{ab})$—the sum of the match point value between SS_a and SS_b.
$Pt(CN_{ab})$—the match point value between CN_a and CN_b.
1: **for each semantic statement** $SS_a(i)$ in SS_a, find the matched semantic statement $SS_b(j)$ in SS_b
based on the mapping relationship set $M(O_A, O_B)$;
If $m(SS_a(i), SS_b(j))$ = "exact match", then the match point value between $SS_a(i)$ and $SS_b(j)$ is assigned 1;
Else if $m(SS_a(i), SS_b(j))$ = "close match", then the match point value between $SS_a(i)$ and $SS_b(j)$ is
assigned 0.5;
2: Record the sum of the match point values between SS_a and SS_b as $Pt(SS_{ab})$ and the number of
matched statements between SS_a and SS_b as $Cot(SS_{ab})$;
3: find the mapping relationship between CN_a and CN_b based on $M(O_A, O_B)$,
If $m(CN_a, CN_b)$ = "exact match", then the match point value between CN_a and CN_b is assigned 1;
Else if $m(CN_a, CN_b)$ = "close match", then the match point value between CN_a and CN_b is
assigned 0.5;
4: Record the match point value between CN_a and CN_b as $Pt(CN_{ab})$;
5: the similarity of categories C_a and C_b can be calculated using the following equation:

$$Sim(a,b) = \begin{cases} \frac{1}{2} * \frac{Pt(SS_{ab})}{Cot(SS_a)} + \frac{1}{2} * \frac{Pt(SS_{ab})}{Cot(SS_b)}, & if\ m(CN_a, CN_b) = \text{"not match"} \\ \frac{1}{3} * \frac{Pt(SS_{ab})}{Cot(SS_a)} + \frac{1}{3} * \frac{Pt(SS_{ab})}{Cot(SS_b)} + \frac{Pt(CN_{ab})}{3}, & if\ m(CN_a, CN_b) = \text{"exact match"/"close match"} \end{cases} \quad (11)$$

In addition, the mapping relationships between category pairs C_a and C_b, namely, $MR(a, b)$, can
be determined using the following equation:

$$MR(a,b) = \begin{cases} exact\ match, & if\ \ Sim(a,b) = 1 \\ close\ match, & if\ 0.5 <= Sim(a,b) < 1 \\ related, & if\ \ 0 < Sim(a,b) < 0.5 \\ not\ match, & if\ \ Sim(a,b) = 0 \end{cases} \quad (12)$$

4. A Case Study

4.1. Study Material

To illustrate the methodologies, two different classification standards in two corresponding natural languages have been selected for use in the mapping process. CS_C is developed based on the national topographic map standards in China (Standards of "Cartographic symbols for national fundamental scale maps" and "Specifications for feature classification and codes of fundamental geographic information"). CS_A is developed by the U.S. Geological Survey in America (http://cegis.usgs.gov/ttl/USTopographic.ttl). Both standards are digital literature materials; the category names and their free-text definitions are provided as source information for our experiment. In addition, for the sake of simplicity and clarity, our study was restricted to the "surface water" categories from these two classification standards. Table 3 briefly lists the characteristics of these two selected dataset, with detailed explanations as follows:

(1) Both standards have their own classification system to address the categories of "surface water". The categories in CS_C are organized using a four-level hierarchy with six major categories. By contrast, the categories in CS_A are organized by a four-level hierarchy with 81 major categories, which means that the hierarchical structure of CS_A does not closely match that of CS_C.
(2) The free-text definitions in both standards are used as category definitions.
(3) The number of categories in CS_C is 74, and the number of concepts in CS_A is 92; thus, the CS_A covers more category types than does CS_C.
(4) The natural language in CS_C is Chinese, whereas the natural language in CS_A is English, which means that there is a natural language barrier between these two GI classification standards.

Table 3. Characteristics of CS_C and CS_A.

Characteristic	CS_C	CS_A
Number of categories	74	92
Classification system	Taxonomy (without overlap)	Taxonomy (without overlap)
Levels of hierarchy	4	4
Number of major categories	6	81
Definition	Free-text, unstructured	Free-text, unstructured
Attribute	Id, Category name	Category name, Source of the definition
Language	Chinese	English

4.2. Results

The well-defined category definitions in both CS_C and CS_A serve as the basis for our study. The Web Ontology Language (OWL) API is integrated to facilitate the implementation of the proposed algorithm in Eclipse with the JAVA language, and the experiment results are as follows.

4.2.1. Semantic Statement Mappings

The semantic primitives are extracted using the Stanford Natural Language Processing Tools (http://nlp.stanford.edu/software/) and are transformed into the formal ontologies O_C and O_A with the set of category names and semantic statements by domain experts and encoded by the OWL via Protégé. Using the semantic statement mapping algorithm introduced in Section 3.2.2, the number of mapping relationships between the statements in O_C and O_A is recorded, and the mapping results for different semantic property/relation types are shown in Table 4.

Table 4. Condition of the mapping statements between O_C and O_A.

		Number of Semantic Statements in O_C	Number of Semantic Statements in O_A	Number of Mapping Statements	Mapping Rate
Property Types		80	111	44	29.93%
Spatial Properties	Location	7	8	1	7.14%
	Morphology	4	23	3	12.50%
	Measurement	2	1	0	0.00%
Temporal Properties	Time Period	2	4	3	75%
	Time Point	3	1	1	33%
Other Semantic Properties	Material Composition	15	22	11	42.31%
	Nature	2	2	2	100.00%
	Status	19	22	8	24.24%
	Cause	2	6	1	14.29%
	Purpose	24	22	14	43.75%
Relation Types		62	70	28	26.92%
Hierarchical Relations		25	24	9	22.50%
Spatial Relations	Topology Relations	27	29	14	33.33%
	Part-Whole Relations	9	17	5	23.81%
Temporal Relations		0	0	0	0.00%
Other Related Relations		1	0	0	0.00%
Total		142	181	72	28.69%

The total number of semantic statements in O_C is 142, and the total number of such statements in O_A is 181. In addition, the total mapping rate of the semantic statements between O_C and O_A is 28.69%. The details of the mapping relationships between semantic statements in each type can be found in Appendix 5.

For the semantic statement about the semantic property types, the most matched type is "purpose". This is because the semantic property type of "purpose" is used to represent the manmade category, which includes "ditch", "canal", and "dam", and the free-text definitions in both the Chinese and American classification standards for these types of categories are very similar. The semantic information about purpose and functionality are considered as the crucial characteristics of the categories. It is easy to understand that the semantic property type "nature" has the highest mapping rate, namely, 100%, because there are only two values for this type of semantic statement, namely, "natural" and "manmade", in both O_C and O_A. Considering the semantic property type "location", there are seven semantic statements in O_C, and eight in O_A, but the mapping rate of this type is extremely low(only one semantic statement is mapped with mapping rate 7.14%). That's because the semantic property type "location" is used to describe the region environment where certain geographic category is at, and a lot of the categories in O_A are bay-related or glacier-related, such as "glacier", "ice cap" and "iceberg tongue" with semantic property value of "location", "mountainous area", "regions of perennial frost", and "coast", respectively, and there are no such categories in O_C. For the semantic statement about the relation types, the most matched type is "spatial relation", which is also the type with the highest mapping rate, indicating that the spatial-related relations play a major role in the identification and representation of GI semantics.

4.2.2. Similarity Calculation and Category Mappings

The similarities between concepts are calculated using the semantic statement mapping relationships and Algorithm 3 proposed in Section 3.2.3. Three typical examples of the mapping results between categories are chosen for further discussion. Table 5 shows the names and free-text definitions of the compared category pairs. In addition, the corresponding semantic statements, calculated similarity values and final mapping relationships between these category pairs are presented in Table 6.

Table 5. Names and free-text definitions of the compared concept pairs.

Concept Pairs	Concepts	Names	Free-Text Definitions
Pair 1	Concept 1 in O_C	溢洪道	水库的泄洪水道，用以排泄水库预定蓄水高度以上的洪水。
	Translation of Concept 1 in O_C	Spillway	Reservoir spillway channel to drain reservoir reservation head above the flood.
	Concept 1 in O_A	Spillway	A passage for surplus water to run over or around a dam.
Pair 2	Concept 2 in O_C	干河床（干涸河）	降水或融雪后短暂时间内有水的河床或河流改道后遗留的河道。
	Translation of Concept 2 in O_C	Arroyo (dry river)	Precipitation or snowmelt water within a short time after the river or river diversions left after the river.
	Concept 2 in O_A	Wash	The usually dry portion of a stream bed that contains water only during or after a local rainstorm or heavy snowmelt.
Pair 3	Concept 3 in O_C	水系	江、河、湖、海、井、泉、水库、池塘、沟渠等自然和人工水体及连通体系的总称。
	Translation of Concept 3 in O_C	Water System	River, river, lake, sea, wells, springs and reservoirs, ponds, ditches, and other natural and artificial water bodies and the connected system in general.
	Concept 3 in O_A	Surface Water	The water portion of the Earth's surface, including the surface of sea and inland waters

Table 6. Example of categories definitions and similarity calculation.

Concepts	Semantic Statements	Translation of Semantic Statement in O_C	Mapping Relationships between Statement	Similarity Values	Mapping Results
Concept 1 in O_C	(Hypernym: 水道)⊓ (Is-Part-Of:水库) ⊓ (Purpose: 排泄洪水)	(Hypernym: Waterways) ⊓ (Is-Part-Of:Reservoir) ⊓ (Purpose:Drain flood)	"Spillway" Exact match "Spillway"(Concept Name) "Hypernym:Waterways" Close match "Hypernym:Passage" "Is-Part-Of:Reservoir"Close match "Is-Part-Of:Dam" "Purpose:Drain flood" Exact match "Purpose:Surplus Water"	0.78	Close Match
Concept 1 in O_A	(Hypernym: Passage) ⊓ (Is-Part-Of: Dam) ⊓ (Purpose: Surplus Water)				
Concept 2 in O_C	(Hypernym:河床)⊓(Material:水)⊓(Status: 干涸)⊓(Temporality:降雪或融雪后)	(Hypernym: riverbed) ⊓(Material: water) ⊓(Status: dry) ⊓(Temporality: After the rainfall or snowmelt)	"Hypernym:riverbed" Exact match "Hypernym:Streambed" "Material:water" Exact match "Material:Water" "Status:dry"Exact match "Status:Dry" "Temporality:After the rainfall or snowmelt" Exact match "Temporality:during or after a local rainstorm or heavy snowmelt"	1.0	Exact Match
Concept 2 in O_A	(Hypernym: Streambed) ⊓(Material: Water) ⊓(Status: Dry) ⊓(Temporality: during or after a local rainstorm or heavy snowmelt)				
Concept 3 in O_C	(Hyponym:江)⊓ (Hyponym:河) ⊓ (Hyponym:湖)⊓ (Hyponym:海)⊓ (Hyponym:井) ⊓ (Hyponym:泉) ⊓ (Hyponym:水库) ⊓ (Hyponym:池塘)⊓ (Hyponym: 沟渠) ⊓ (Hyponym:水体) ⊓(Nature:自然⊔ Nature: 人工) ⊓ (Material: 水)	(Hyponym: river) ⊓ (Hyponym: river) ⊓ (Hyponym: lake) ⊓ (Hyponym: sea) ⊓ (Hyponym: well) ⊓ (Hyponym: spring) ⊓ (Hyponym: reservoir) ⊓ (Hyponym: pond) ⊓ (Hyponym: ditch) ⊓ (Hyponym: body of water) ⊓ (Nature: natural ⊔ Nature: artificial) ⊓ (Material: water)	"Hyponym: river" Exact match "Hyponym:River" "Hyponym: river" Exact match "Hyponym:Stream" "Hyponym:lake" Exact match "Hyponym:Lake" "Hyponym:sea" Exact match "Hyponym:Sea" "Hyponym:spring" Exact match "Hyponym:Spring" "Hyponym:reservoir" Exact match "Hyponym:Reservoir" "Hyponym:pond" Exact match "Pond" "Hyponym:ditch" Exact match "Ditch" "Hyponym:body of water" Exact match "Hyponym:Water body" "Nature:natural" Exact match "Nature:Natural" "Nature:artificial" Exact match "Nature:Manmade" "Material:water" Exact match "Material:Water"	0.92	Close Match
Concept 3 in O_A	(Material: Water) ⊓ (Hyponym: Sea) ⊓ (Hyponym: Inland Water) ⊓ (Is-Part-Of: Earth's surface) ⊓ [(Hyponym: River) ⊓ (Hyponym: Stream)⊓ (Hyponym: Lake) ⊓ (Hyponym: Spring)⊓ (Hyponym: Reservoir) ⊓ (Hyponym: Pond)⊓ (Hyponym: Ditch) ⊓ (Hyponym: Water body)⊓ (Nature: Natural)⊓ (Nature: Manmade)] (The semantic statements in "(. . .)" were not the semantic information extracted from the free-text definition and were inferred based on the semantic statements in other concepts, which have a hierarchical relation with the concept. They were added to the concept by the domain expert manually.)				

Example 1: Concept pair of "spillway" in O_C and "spillway" in O_A

These two concepts are comparable because the mapping relationship between their concept names is "exact match". Because their concept names and four semantic statements are matched (detailed mapping relationships are illustrated in Table 6, line 1 and 2), the second condition in Equation (9) is used to calculate the final similarity between "spillway" in O_C and "spillway" in O_A. The similarity value between these two concepts is calculated as 0.78; thus, the mapping relationship between these two concepts is "close match". This example demonstrates the simplest case for the calculation of the semantic similarity between concepts.

Example 2: Concept pair of "arroyo (dry river)" in O_C and "wash" in O_A

In this example, the mapping relationship between the concept name of "arroyo (dry river)" and "wash" cannot be determined based on the mapping algorithm in Section 3.2.1. However, the similarity value between these two concepts is higher than the value in example (1). This is because all the semantic statements used to represent the semantic meaning of these two concepts are correspondingly matched (detailed mapping relationships are illustrated in Table 6, line 3 and 4), and all the mapping relationships between them are "exact match". The first condition in Equation 9 is used to calculate the final similarity between the concepts "arroyo (dry river)" in O_C and "wash" in O_A. The similarity value between these two concepts is calculated as 1.0; thus, the mapping relationship between these two concepts is "exact match". This example demonstrates a common situation in the cross-lingual environment in that two concepts have the same semantic meaning while their names are definitely different. Moreover, the utility of applying our methodologies to the complex application of cross-lingual GI ontology integration has been proven.

Example 3: Concept pair of concept 3 "Water System" in O_C and concept 3 "Surface water" in O_A

At first glance, the semantic statements between the concept "water system" and "surface water" are not matched very well, and the concept names of these two concepts cannot be matched either.

This is because these two concepts are both the top concept in their own taxonomies, and these two concepts are abstract concepts in that they do not represent real-world objects with detailed characteristic entities, for example, rivers, lakes, and oceans. Thus, the definitions of this category in different languages may be very different, even when they are conveying the same meaning. Therefore, the solution for the semantic meaning representation of this type concept is not the same as the solution used in Examples (1) and (2). The sematic meaning of the hyponym-related concepts should be considered to infer the integrated semantic meaning of this abstract concept. After the implicit semantic statements have been inferred out (detailed mapping relationships are illustrated in Table 6, line 5 and 6), the first condition in Equation (9) is used to calculate the final similarity between the concepts "water system" in O_C and "surface water" in O_A. The similarity value between these two concepts is calculated as 0.92; thus, the mapping relationship between these two concepts is "close match".

5. Conclusions and Future Work

The presented research focuses on the determination of semantic mapping relationships between categories in different GI ontologies with natural language barriers. The proposed formal ontology model in this study is used to represent and identify the semantic characteristics of the GI categories with OWL-based semantic statements transformed from free-text definitions of two GI classification standards. A new similarity calculation algorithm based on this formal ontology model is presented to distance the semantic similarities and identify the mapping relationships between categories.

In particular, we work with two classification standards of topographic maps in Chinese and American English. The conducted experiment indicates that the proposed approach successfully determines the mapping relationships between categories in different GI ontologies and facilitates ontology integration in a cross-lingual environment. Due to the usages of the multilingual supported

NLP tools in our experiment, it is easy to replicate our model to determine the mapping relationships between other GI ontologies, which may be described using other native natural languages, in addition to Chinese. However, this model has only been applied to geospatial information (GI) integration at the category level, and research on GI integration at the data level has not been fulfilled. That will form the basis for future study. In addition, publishing the mapping information in a cross-lingual context as linked data in a semantic web environment should also be considered.

Acknowledgments: This research is supported by the National Administration of Surveying, Mapping and Geoinformation, China, under the Special Fund for Surveying, Mapping and Geographical Information Scientific Research in the Public Interest (No. 201412014), and Specialized Research Fund for the Doctoral Program of Higher Education (No. 20120141110048).

Author Contributions: This research was mainly performed and prepared by Xi Kuai and Lin Li. Xi Kuai and Lin Li contributed with ideas, conceived and designed the study. Xi Kuai wrote the paper. Heng Luo, Hang Shen and Yu Liu contributed the tools, and analyze the results of the experiment. Zhijun Zhang reviewed and edited the manuscript.

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

Appendix: Detail of the Mapping Statements between O_C and O_A

Table A1. Detail of the Mapping Statements between O_C and O_A.

Property Types	Semantic Statements in O_c in Chinese	Translation of Semantic Statements in O_c in English	Semantic Statements in O_A in English	Mapping Relations	Semantic Statements in O_c in Chinese	Translation of Semantic Statements in O_c in English	Semantic Statements in O_A in English	Mapping Relations
Material	水	water	water	Exact match	石	stone	stones	Exact match
	水蒸气	water vapor	vapors	Exact match	木桩	wooden stake	wood	Close match
	泥	mud	mud	Exact match	草地	grassland	grassy	Close match
	砖	brick	brick	Exact match	砾石	gravel	gravel	Exact match
	沙	sand	sand	Exact match	礁石	reef	reef	Exact match
	水泥	cement	concrete	Exact match				
Nature	自然的	natural	natural	Exact match				
	人造的	manmade	manmade	Exact match				
Status	流动	flow	flowing	Exact match	倾泻	pour	moving outward an downslope	Close match
	独立	stand along	free standing	Exact match	高潮时被水体淹没，低潮时露出	submerged at high tide the water, exposed at low tide	alternately covered and left bare by the tide	Exact match
	有水潮浸	tide water immersion	washed by waves or tides	Close match	涌出	emission	issue from the ground	Close match
	干涸	dried up	dry	Exact match	洪水泛滥	flood	subject to flooding	Exact match
Temporality	长期	long-term	permanent	Close match	降水或融雪后短时间内	within a short time after rainfall or snowmelt	during or after a local rainstorm or heavy snowmelt	Exact match
	终年	all year round	permanent	Exact match	季节性	seasonal	occasionally	Close match
Location	沙地	sandy	desert	Close match				
Purpose	引水	water diversion	run water	Exact match	减缓水流流速	slow water flow rate	restrain current or tide	Close match
	输水	water delivery	conveying water	Exact match	保护港口	protection of harbor	protect harbor	Exact match
	贮水	water storage	contain water	Exact match	护岸	bank protection	sustain an embankment	Exact match
	将水位升高或降低，使船能在不同高低水位的水道间通行	To raise or lower the water level, at different high and low water level so the ship channel traffic	raise and lower vessels as they pass from one level to another.	Exact match	抬高水位	raising of water level	raise the level of water	Exact match
	控制流量	control flow	control the flow of water	Exact match	通行船只	passage vessel	route for watercraft	Exact match
	灌溉	irrigation	irrigation	Exact match	拦截河流	blocked rivers	Across the course of a stream	Close match
	调节水流方向	adjusting to the flow direction	direct current or tide	Exact match	扬水	pump up water-	Pump	Exact match

Table A1. *Cont.*

Property Types	Semantic Statements in O_c in Chinese	Translation of Semantic Statements in O_c in English	Semantic Statements in O_A in English	Mapping Relations	Semantic Statements in O_c in Chinese	Translation of Semantic Statements in O_c in English	Semantic Statements in O_A in English	Mapping Relations
Morphology	陡坡	steep slope	a vertical or near vertical descent	Close match	坝式	dam type	dam	Exact match
	虹吸式	siphon	siphon	Exact match				
Cause	堆积	accumulation	accumulate	Exact match				
Relation Types								
Hierarchical Relation	源头	source	source	Exact match	设施	facilities	facility	Exact match
	河床	riverbed	channel bottom	Exact match	构筑物	structure	construction	Exact match
	区域	regional	region	Exact match	通道	channel	path	Exact match
	地带	zone	zone	Exact match	水道	waterways	waterway	Exact match
	设备	device	device	Exact match				
Spatial Relation	地面上	on the ground	on the surface of the land	Close match	水体平均大潮高潮面与水体最低低潮面之间	mean high water springs of water and water between the lowest low water	Between high water and low water marks	Exact match
	水体内	in body of water	in water	Exact match	沿河流	along the river	alongside a stream	Exact match
	海域内	within the sea	in the sea	Exact match	水陆间	between land and water	contact between a body of water and the land	Exact match
	水下	underwater	below the surface of water	Exact match	洼地内	in the depressions	surrounded by land	Close match
	跨流域	across river basins	across the course of a stream	Exact match	陆地上	on the land	Covered with the earth	Close match
	跨道路	cross roads	crossing road or trail	Exact match	海岸线与干出线之间	between the coastline and the dry line	Between high water and low water lines	Exact match
	海岸边	coastal	adjacent to the shore	Exact match	海岸边	the coast	offshore	Close match
Is-part-of	网状水系	network drainage	network of interlacing channels	Exact match	水库	reservoir	dam	Close match
	网状水系	network drainage	a drainage network	Exact match	河渠	canal	a river system	Close match
	闸室	chamber	lock chamber	Exact match				

References

1. Gore, A. The digital earth: Understanding our planet in the 21st century. *Photogramm. Eng. Remote Sens.* **1999**, *65*. [CrossRef]
2. Craglia, M.; Goodchild, M.F.; Annoni, A.; Camara, G.; Gould, M.; Kuhn, W.; Mark, D.; Masser, I.; Maguire, D.; Liang, S.; *et al.* Next-generation digital earth: A position paper from the vespucci initiative for the advancement of geographic information science. *Int. J. Spat. Data Infrastruct. Res.* **2008**, *3*, 146–167.
3. Craglia, M.; de Bie, K.; Jackson, D.; Pesaresi, M.; Remetey-Fülöpp, G.; Wang, C.; Annoni, A.; Bian, L.; Campbell, F.; Ehlers, M.; *et al.* Digital Earth 2020: Towards the vision for the next decade. *Int. J. Digit. Earth* **2012**, *5*, 4–21. [CrossRef]
4. Yue, P.; Di, L.; Yang, W.; Yu, G.; Zhao, P. Semantics-based automatic composition of geospatial Web service chains. *Comput. Geosci.* **2007**, *33*, 649–665. [CrossRef]
5. Janowicz, K.; Hitzler, P. The digital earth as knowledge engine. *Semant. Web* **2012**, *3*, 213–221.
6. Janowicz, K. Observation-driven geo-ontology engineering. *Trans. GIS* **2012**, *16*, 351–374. [CrossRef]
7. Buccella, A.; Cechich, A.; Gendarmi, D.; Lanubile, F.; Semeraro, G.; Colagrossi, A. Building a global normalized ontology for integrating geographic data sources. *Comput. Geosci.* **2011**, *37*, 893–916. [CrossRef]
8. Bishr, Y. Overcoming the semantic and other barriers to GIS interoperability. *Int. J. Geogr. Inf. Sci.* **1998**, *12*, 299–314. [CrossRef]
9. Lemmens, R.L. Semantic Interoperability of Distributed Geoservices. Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, 2006.
10. Fallahi, G.R.; Frank, A.U.; Mesgari, M.S.; Rajabifard, A. An ontological structure for semantic interoperability of GIS and environmental modeling. *Int. J. Appl. Earth Obs. Geoinf.* **2008**, *10*, 342–357. [CrossRef]
11. Ma, X.; Wu, C.; Carranza, E.J.M.; Schetselaar, E.M.; van der Meer, F.D.; Liu, G.; Wange, X.; Zhang, X. Development of a controlled vocabulary for semantic interoperability of mineral exploration geodata for mining projects. *Comput. Geosci.* **2010**, *36*, 1512–1522.
12. Kuhn, W. Geospatial semantics: Why, of what, and how? *J. Data Semant. III* **2005**, *3534*, 1–24.
13. 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*, 1–25. [CrossRef]
14. Fonseca, F.T.; Egenhofer, M.J.; Davis, C.A., Jr.; Borges, K.A.V. Ontologies and knowledge sharing in urban GIS. *Comput. Environ. Urban Syst.* **2000**, *24*, 251–272. [CrossRef]
15. Pundt, H.; Bishr, Y. Domain ontologies for data sharing–An example from environmental monitoring using field GIS. *Comput. Geosci.* **2002**, *28*, 95–102. [CrossRef]
16. Yang, C.; Raskin, R.; Goodchild, M.; Gahegan, M. Geospatial Cyberinfrastructure: Past, present and future. *Comput. Environ. Urban Syst.* **2010**, *34*, 264–277. [CrossRef]
17. Stoimenov, L.; Stanimirovic, A.; Djordjevic-Kajan, S. Discovering mappings between ontologies in semantic integration process. In Proceedings of the 9th AGILE Conference on Geographic Information Science, Visegrád, Hungary, 20–22 April 2006.
18. Janowicz, K.; Raubal, M.; Kuhn, W. The semantics of similarity in geographic information retrieval. *J. Spat. Inf. Sci.* **2011**, *2*, 29–57. [CrossRef]
19. Schwering, A.; Raubal, M. Spatial relations for semantic similarity measurement. In *Perspectives in Conceptual Modeling*; Springer-Verlag: Heidelberg, Germany, 2005; pp. 259–269.
20. Hakimpour, F. *Using Ontologies to Resolve Semantic Heterogeneity for Integrating Spatial Database Schemata*; Zurich University: Zurich, Switzerland, 2003.
21. Mark, D.M.; Skupin, A.; Smith, B. Features, objects, and other things: Ontological distinctions in the geographic domain. In *Spatial Information Theory*; Springer: New York, NY, USA, 2001; pp. 489–502.
22. Stadler, C.; Jens, L.; Konrad, H.; Sören, A. Linkedgeodata: A core for a web of spatial open data. *Semantic Web* **2012**, *3*, 333–354.
23. Trojahn, C.; Fu, B.; Zamazal, O.; Ritze, D. *State-of-the-Art in Multilingual and Cross-Lingual Ontology Matching*; Springer: Heidelberg, Germany, 2014.
24. Liu, K.; Hogan, W.R.; Crowley, R.S. Natural Language Processing methods and systems for biomedical ontology learning. *J. Biomed. Inform.* **2011**, *44*, 163–179. [CrossRef] [PubMed]
25. Buitelaar, P.; Cimiano, P.; Magnini, B. Ontology learning from text: Methods, evaluation and applications. *Comput. Linguist.* **2006**, *32*, 569–572.

26. Bird, S.; Klein, E.; Loper, E. *Natural Language Processing with Python*; O'Reilly Vlg. GmbH & Co.: Sebastopol, CA, USA, 2009.
27. Jurafsky, D.; Martin, J.H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*; Prentice Hall: Upper Saddle River, NJ, USA, 2000.
28. Kavouras, M.; Kokla, M.; Tomai, E. Comparing categories among geographic ontologies. *Comput. Geosci.* **2005**, *31*, 145–154. [CrossRef]
29. Kavouras, M.; Kokla, M. *Theories of Geographic Concepts: Ontological Approaches to Semantic Integration*; CRC Press: Boca Raton, FL, USA, 2008.
30. Bittner, T.; Donnelly, M.; Smith, B. A spatio-temporal ontology for geographic information integration. *Int. J. Geogr. Inf. Sci.* **2009**, *23*, 765–798. [CrossRef]
31. Zheng, J.G.; Fu, L.Y.; Ma, X.G.; Fox, P. SEM+: Tool for discovering concept mapping in Earth science related domain. *Earth Sci. Inform.* **2015**, *8*, 1–8. [CrossRef]
32. Schwering, A. Approaches to semantic similarity measurement for geo-spatial data: A survey. *Trans. GIS* **2008**, *12*, 5–29. [CrossRef]
33. Buccella, A.; Cechich, A.; Fillottrani, P. Ontology-driven geographic information integration: A survey of current approaches. *Comput. Geosci.* **2009**, *35*, 710–723. [CrossRef]
34. Ballatore, A.; Bertolotto, M.; Wilson, D.C. Geographic knowledge extraction and semantic similarity in OpenStreetMap. *Knowl. Inf. Syst.* **2013**, *37*, 61–81. [CrossRef]
35. Ballatore, A.; Wilson, D.C.; Bertolotto, M. Computing the semantic similarity of geographic terms using volunteered lexical definitions. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 2099–2118. [CrossRef]
36. Li, W.; Raskin, R.; Goodchild, M.F. Semantic similarity measurement based on knowledge mining: An artificial neural net approach. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1415–1435. [CrossRef]
37. Xu, Y.; Xie, Z.; Chen, Z. Research on semantics of entity space similarity measure based on artificial neural networks. In Proceedings of the 23rd International Conference on Geoinformatics, Wuhan, China, 19–21 June 2015.
38. Laurini, R. Geographic ontologies, gazetteers and multilingualism. *Future Internet* **2015**, *7*, 1–23. [CrossRef]
39. Fu, B.; Brennan, R.; O'Sullivan, D. A configurable translation-based cross-lingual ontology mapping system to adjust mapping outcomes. *Web Semant. Sci. Serv. Agents World Wide Web* **2012**, *15*, 15–36. [CrossRef]
40. Sini, A.; Sini, M. Mapping AGROVOC and the Chinese Agricultural Thesaurus: Definitions, tools, procedures. *New Rev. Hypermedia Multimed.* **2006**, *12*, 51–62.
41. Wang, S.; Isaac, A.; Schopman, B.; Schlobach, S.; van der Meij, L. Matching multi-lingual subject vocabularies. In *Research & Advanced Technology for Digital Libraries*; Springer: Berlin, Germany, 2009; pp. 125–137.
42. Meilicke, C.; García-Castrod, R.; Freitas, F.; van Hage, W.R.; Montiel-Ponsoda, E.; de Azevedo, R.R.; Stuckenschmidt, H.; Šváb-Zamazal, O.; Svátek, V.; Tamilin, A.; *et al.* MultiFarm: A benchmark for multilingual ontology matching. *J. Web Semant.* **2012**, *15*, 62–68. [CrossRef]
43. Ngai, G.; Carpuat, M.; Fung, P. Identifying concepts across languages: A First step towards a corpus-based approach to automatic ontology alignment. In Proceedings of the 19th international conference on Computational linguistics, Stroudsburg, PA, USA, August 2002.
44. Pazienza, M.T.; Stellato, A. Linguistically motivated ontology mapping for the semantic web. In Proceedings of the 2nd Italian Semantic Web Workshop, Trento, Italy, 14–16 December 2005; pp. 14–16.
45. Jung, J.J.; Håkansson, A.; Hartung, R. Indirect alignment between multilingual ontologies. In *Agent and Multi-Agent Systems: Technologies and Applications*; Springer: Berlin, Germany, 2009; Volume 5559, pp. 233–241.
46. Trojahn, C.; Quaresma, P.; Vieira, R. A Framework for multilingual ontology mapping. In Proceedings of the International Conference on Language Resources and Evaluation, Marrakech, Morocco, 28–30 May 2008; pp. 1034–1037.
47. Wang, S.; Englebienne, G.; Schlobach, S. Learning concept mappings from instance similarity. In *The Semantic Web—ISWC 2008*; Springer: Berlin, Germany, 2008; pp. 339–355.
48. Zheng, Q.; Shao, C.; Li, J.; Wang, Z.; Hu, L. RiMOM2013 results for OAEI 2013. In Proceedings of the 8th International Conference on Workshop on Ontology Matching, Sydney, Australia, 21 October 2013.
49. Zhang, X.; Zhong, Q.; Shi, F.; Li, J.; Tang, J. RiMOM results for OAEI 2009. In Proceedings of the 4th International Conference on Workshop on Ontology Matching, Washington, DC, USA, 25 October 2009.
50. Wang, Z.; Zhang, X.; Hou, L.; Zhao, Y.; Li, J.; Qi, Y.; Tang, J. RiMOM results for OAEI 2010. In Proceedings of the 5th International Conference on Ontology Matching, Shanghai, China, 7 November 2010.

51. Euzenat, J.; Shvaiko, P. *Ontology Matching*; Springer: Berlin, Germany, 2007.
52. Ehrig, M.; Sure, Y. Ontology mapping–An integrated approach. In *The Semantic Web: Research and Applications*; Springer: Berlin, Heidelberg, Germany, 2004; pp. 76–91.
53. Kalfoglou, Y.; Schorlemmer, M. Ontology mapping: The state of the art. *Knowl. Eng. Rev.* **2003**, *18*, 1–31. [CrossRef]
54. Doan, A.H.; Madhavan, J.; Domingos, P.; Halevy, A. Ontology matching: A machine learning approach. In *International Handbooks on Information Systems*; Springer: Berlin, Germany, 2004; pp. 397–416.
55. Kantor, P. Foundations of statistical natural language processing. *Nat. Lang. Eng.* **1999**, *26*, 91–92.
56. MacCartney, B. The Stanford Natural Language Processing Group. Available online: http://nlp.stanford.edu/ (accessed on 18 December 2015).
57. Guarino, N. Formal ontology, conceptual analysis and knowledge representation. *Int. J. Hum. Comput. Stud.* **1995**, *43*, 625–640. [CrossRef]
58. Herre, H. General Formal Ontology (GFO): A foundational ontology for conceptual modelling. In *Theory & Applications of Ontology Computer Applications*; Springer Netherlands: Dordrecht, The Netherlands, 2010; pp. 297–345.
59. Frank, A.U. Ontology for spatio-temporal databases. In *Spatio-Temporal Databases*; Springer: Berlin, Germany, 2003; pp. 9–77.
60. Casati, R.; Smith, B.; Varzi, A.C. Ontological tools for geographic representation. In *Formal Ontology in Information Systems*; IOS Press: Amsterdam, The Netherlands, 1998; pp. 77–85.
61. Grenon, P.; Smith, B. SNAP and SPAN: Towards dynamic spatial ontology. *Spat. Cognit. Comput.* **2004**, *4*, 69–104. [CrossRef]