

Article

An Entropy-Based Weighted Concept Lattice for Merging Multi-Source Geo-Ontologies

Junli Li ^{1,2,*}, Zongyi He ¹ and Qiaoli Zhu ¹

¹ School of Resources and Environment Science, Wuhan University, No.129 Luoyu Road, Wuhan 430079, China; E-Mails: zongyihe56@yahoo.com (Z.H.); zhuqiaoli2008@163.com (Q.Z.)

² Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, No.129 Luoyu Road, Wuhan 430079, China

* Author to whom correspondence should be addressed; E-Mail: lijunli866@whu.edu.cn; Tel.: +86-27-8715-2312; Fax: +86-27-6877-8381.

Received: 26 March 2013; in revised form: 14 May 2013 / Accepted: 1 June 2013 /

Published: 7 June 2013

Abstract: To deal with the complexities associated with the rapid growth in a merged concept lattice, a formal method based on an entropy-based weighted concept lattice (EWCL) is proposed as a mechanism for merging multi-source geographic ontologies (geo-ontologies). First, formal concept analysis (FCA) is used to formalize different term-based representations in relation to the geographic domain, and to construct a merged formal context. Second, a weighted concept lattice (WCL) is applied to reduce the merged concept lattice, based on information entropy and a deviance analysis. The entropy of the attribute set is exploited to acquire the intent weight value, and the standard deviation contributes to computing the intent importance deviance value, according to the user preferences and interests. Some nodes of the merged concept lattice are then removed if their intent weights are lower than the intent importance thresholds specified by the user. Finally, experiments were conducted by combining fundamental geographic information data and spatial data in the hydraulic engineering domain from China. The results indicate that the proposed method is feasible and valid for reducing the complexities associated with the merging of geo-ontologies. Although there are still some problems in the application, the manuscript offers a new approach for the merging of geo-ontologies.

Keywords: geo-ontology merging; formal concept analysis; weighted concept lattice; information entropy; deviation analysis

1. Introduction

Currently, geo-ontology is widely used for representing and sharing spatial information in various application domains, and integrating different geospatial information between interoperating systems has become a hot topic in many scientific disciplines [1]. However, due to the different data standards and incompatible terminologies for expressing spatial information in geographic information science, it is easy to produce semantic heterogeneity. For instance, semantic differences regarding rivers often occur in the distributed systems of the Ministry of Water Resources and the Ministry of Land and Resources in China. The former emphasizes the natural morphology of rivers, and the latter stresses the negotiability of waterways. These problems could possibly be avoided if ontologies were applied. At present, ontology is widely used as a tool to address heterogeneity problems in many areas, such as knowledge representation, information retrieval, and the semantic web [2,3]. Furthermore, ontology has been defined as “a formal explicit specification of a shared conceptualization” [3], and geo-ontology is no exception in real applications, which is used to define a common vocabulary that will facilitate interoperability and handle some problems with data integration in various systems. [4]. The existing geo-ontology building frameworks different experts, different tools, and different techniques. Geo-ontologies may differ and even conflict, even though the ontologies exist in the same domain. As a consequence, the problem of merging geo-ontologies from multi-source geospatial data is still a big challenge.

Up till now, some significant progress in merging geo-ontologies has been achieved. Kokla *et al.* [5] combined semantic factoring and a concept lattice for integrating multiple ontologies. They used an example of integrating the concept type “*stream*”, as defined by three different ontologies: CYC top-level ontology, WordNet, and SDTS. This method can detect the possible implicit relations between concepts which are not predefined. Similarly, Zhu [6] presented a formal method based on concept lattices to form a more general semantic level, and an algorithm was designed to reduce the redundant concept relations.

Meanwhile, Buccella *et al.* [7] proposed a merging method by using a set of matching functions and inferences over the ontologies in order to find the more suitable correspondences. This method minimizes the redundant information and improves the understandability of data by applying the ISO19109 and ISO19107 standards to normalize the geographic ontologies. Torre *et al.* [4] provided a conceptualized framework of geographic application ontologies for sharing and integrating geospatial information. This method is based on abstract classes to cognitively classify geographic concepts, and directly translates the relationships between mapping concepts by a set of axiomatic relations. In addition, Stumme *et al.* [8] proposed the formal concept analysis merge (FCA-MERGE) method as a semi-automatic method for translating concept lattices into a merged ontology, but the method still requires revision by experts. Chen *et al.* [9] provided a new method for combining WordNet and the fuzzy formal concept analysis technique for merging ontologies with the same domain. Two ontologies, including a base ontology and a revision ontology, can be converted into a novel fuzzy ontology by using the revision ontology to update the base ontology, but the method solely utilize partial semantic factors to determine the relationship between elements, and much detailed information need be considered in the future.

The applications mentioned above mainly focus on top-level ontologies, a conceptualized framework, matching functions, and formal concept analysis. However, the previous works have two drawbacks: (1) Each method must extract the formal representation of semantics from the definitions of geographic entities by FCA, and the complexity in the different levels does not get full attention, due to the rapid growth of the merged concept lattice. (2) These methods generally not only assume that all the intents are equally important during the process of constructing the concept lattice, but also do not fully consider the requirements and preferences of the user. Hence, new techniques must be added to handle these shortcomings. However, due to the high space and time complexity, the performance of most algorithms for constructing a concept lattice for dense and large contexts is not desirable [10]. Accordingly, reducing the size of the merged concept lattice in most of the approaches is essential for the first problem. Since each node is composed of the intent and extent in the structure of a concept lattice [11,12], we propose the term of reduction from the viewpoint of finding an appropriate object set and attribute set, respectively. From the point of view of the object reduction in our previous work [13], we applied a fuzzy equivalence relation matrix to construct the equivalent characteristic components of the extent of the concept lattice. We then selected an appropriate threshold value to receive sets of concept extension in different granulations, and measured the similarity of any two extents in the granulation. Finally, the experimental results indicated that the merging process is a stepwise refinement process corresponding to different levels of granularity, and conceptual similarity in the fine-grained levels was higher than in the coarse-grained levels.

Following this premise, our current work is motivated by the need to address the two problems mentioned above. The main contributions of this manuscript can be summarized as follows: First, FCA is introduced for the merging of multi-source geo-ontologies. Second, EWCL is applied to consider the importance of the different intents of geo-ontologies in GIScience, in which the entropy of the attribute set is used to acquire the single attribute intent weight value, and the standard deviation is used for the intent importance deviance value between the multiple attributes. Third, the merged concept lattice is simplified from the point of view of attribute reduction, in terms of the importance threshold specified by the user.

The rest of this paper is organized as follows: Section 2 briefly reviews some notions of FCA and analyses some of the semantic representations of geo-ontologies; Section 3 describes the basic contents of the merging process of EWCL; Section 4 describes the experiments undertaken to investigate the merging of geo-ontologies, and presents an analysis of the results; and Section 5 draws conclusions and discusses the future work.

2. A Semantic Representation of Geospatial Information Based on FCA

2.1. Basic Notions of FCA

FCA, a branch of applied mathematics based on lattice theory, is a conceptual framework proposed by Wille in 1982, and has been applied to many different fields, such as data analysis, knowledge discovery, software engineering, and information retrieval [14–17]. To demonstrate the relations between objects and attributes in a given application domain, a concept in FCA is defined within a formal context. Here, we only briefly implement some basic semantic analysis by FCA. For a more extensive introduction refer to [18].

Definition 1. A formal context is defined as a triple $K = (G, M, I)$, where G and M are two non-empty sets called objects (extent) and attributes (intent), respectively, and $I \subseteq G \times M$ is a binary relation. if gIm , for $g \in G$ and $m \in M$, this indicates that the object g has the attribute m .

As a matter of fact, the domain ontology is usually approximately defined as a relation group. In a formal context $K = (G, M, I)$ for a set $A \subseteq G$ and $B \subseteq M$. The formal concepts of the context are derived in terms of the following operations:

$$A = B' = \{a \in G \mid b \in B, aIb\} \quad (1)$$

$$B = A' = \{b \in M \mid a \in A, aIb\} \quad (2)$$

$$C_i = (A'', A') \cup (B', B'') \quad (3)$$

For simplicity, we write a' instead of $\{a\}'$ for all $a \in G$, and write b' instead of $\{b\}'$ for all $b \in M$. A pair (A, B) is called a formal concept if $A = B'$ and $B = A'$. A' is the set of the attributes representing each object in A , whereas B' is the set of objects possessing attributes in B . C_i are the union sets $(A'', A') \cup (B', B'')$ describing the final classes or concept lattice nodes.

A partial order relation between the two formal concepts $((A_1, B_1), (A_2, B_2))$ in a formal context K is defined as follows:

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow B_1 \supseteq B_2 \quad (4)$$

In the above condition, " $\leq$ " is called the hierarchical order of concepts. (A_1, B_1) is called a sub-concept of (A_2, B_2) , and (A_2, B_2) is called a super-concept of (A_1, B_1) . In addition, in a formal context $K = (G, M, I)$, $\forall A_1, A_2, A \subseteq G, \forall B_1, B_2, B \subseteq M$, the following properties hold: $A_1 \subseteq A_2 \Rightarrow A_2' \subseteq A_1'$, $B_1 \subseteq B_2 \Rightarrow B_2' \subseteq B_1'$, $A \subseteq A'', B \subseteq B'', A = A''', B = B''', A \subseteq B' \Leftrightarrow B \subseteq A'$.

2.2. A Sematic Representation of Geospatial Information

In the current research on concept lattices, it is usually assumed that the extent and intent in the formal concept are of equal importance. The extent represents the entities, whereas the intent includes its intrinsic characteristics. However, the semantic basis for spatial concept types in the hydrological domain usually means intent, and the intent is determined by essential geographic properties [4]. The extent and intent of the spatial concept may be extracted by the common understanding, taxonomic structure, and recognized vocabulary of the domain knowledge derived from professional dictionaries and standards. Based on the fundamental philosophical notions of *identity*, *unity*, *essence*, and *dependence*, Guarino and Welty [19] presented a set of meta-properties to represent the behavior of the essential properties, including the rigid property, non-rigid property, anti-rigid property, semi-rigid property, carrier identity, and external dependence. For example, we normally think that reservoirs in the hydrological domain possess properties such as “store water”, “storage capacity”, “name”, etc. The “store water” property is normally rigid for each individual reservoir. The “name” property cannot support identity while being the same individual. The “storage capacity” has a non-rigid property because different reservoirs have different storage capacities. Therefore, “store water” may represent the ontological property of the reservoir.

Table 1. Partial semantics of inland hydrological concepts in GB/T 20258.1-2007.

Object	Normative Description	Ontological Properties
lake	A body of water surrounded by low depressions, has wide areas and slowly changes the yield of water	material/water, spatial morphology/ depressions on the earth.(time/perennial, metrics/ $\geq 10^5 \text{ m}^3$, cause/nature, function/store water)
pond	A pool of water storage	material/water, function/store water. (cause/artificial, spatial morphology/ depressions on the earth, time/perennial, metrics/ $\leq 10^5 \text{ m}^3$)
seasonal lake	A kind of lake which possesses water under seasonal conditions	material/water, time/seasonal ,cause/nature. (function/store water, spatial morphology/ depressions on the earth)
ground river	A kind of natural river on the ground which possesses water	material/water, spatial location/on the earth , material state/flow.(cause/nature, spatial morphology/long strip slot, time/perennial)
seasonal river	A kind of natural river which possesses water under seasonal conditions	material/water, cause/nature, time/seasonal , material state/flow.(spatial morphology/long strip slot, spatial location/on the earth)
reservoir	A body of water or buildings generated from constructing all kinds of dam, gate, dike, and weir, which retain river runoff	material/water, cause/artificial, spatial adjacency/dam, gate, dike and weir, function/ prevent flood. (function/ store water ,spatial morphology/ depressions on the earth)

In geographic ontologies, definitions contain the rich sources of scientific knowledge of the geographic domain, in general, they are also the key and the only descriptions of category terms, which result in the semantic definition of geographic categories (*e.g.*, purpose, cause, material) [20]. In order to identify a set of semantic properties-relations, Wang *et al.* [21] proposed the property of geospatial ontology from the view of top-level ontology, including space, time, cause, material, function, object, and metrics. On the basis of their work, we identified partial semantic relations-properties of inland hydrological concepts from GB/T 13923-2006 (Specifications for feature classification and codes of fundamental geographic information), GB/T 20258.1-2007 (Data dictionary for fundamental geographic information features), and SL 213-98 (Specification on basic information coding of water conservancy projects) in China (see Table 1 and Table 2). From the given normative descriptions, heterogeneity problems inevitably exist. For example, one consider a lake consisting of the following semantic properties and relations : *material* with value “water”, *spatial morphology* with value “low depressions”, *metrics* with value “wide areas and slowly changes the yield of water”, and the other defines a lake associated with the corresponding semantic properties: *material* with value “water”, *spatial morphology* with value “natural depressions on the earth”, *function* with value “store water”. In terms of the formal definitions of ontologies, geo-ontologies are applied to capture the universal concepts and meanings in the geospatial domain. However, “lake: wide areas and slowly changes the yield of water” has a bit strong subjectivity and a non-rigid property, because different lakes have different areas, flow velocities and water yields. Hence, in order to replace “wide areas and slowly changes the yield of water”, we adopt a more appropriate expert standard: *metrics* with value “ $\geq 10^5 \text{ m}^3$ ” [22], meanwhile, “natural depressions” is a part of “low depressions”. In addition, other

context-specific semantic elements are also identified. For instance, the semantic properties in relation to hydrography are complemented by domain experts, such as *time* (perennial or seasonal) and *cause* (natural or artificial). Similarly, “reservoir: retain river runoff” is a part of “reservoir: prevent flood”, and “in the river, the valley, the depressions and underground permeable layer” is considered as *spatial morphology* with value “depressions on the earth”. Here, ontological properties of geospatial objects in Table 1 and Table 2 have been extracted from the normative descriptions, moreover, in particular, ontological properties in the parentheses are updated and complemented based on domain experts.

Table 2. Partial semantics of hydraulic engineering concepts in SL 213-98.

Object	Normative Description	Ontological Properties
lake	A lake basin and a body of water accommodated in the lake basin, which can store water, and is surrounded by natural depressions on the earth	material/water, spatial morphology/natural depressions .(function/store water, time/perennial , cause/nature, metrics/ $\geq 10^5 \text{ m}^3$)
polder	An enclosed area for production and living activities, which is generated from constructing all kinds of dikes, along with river, lake, islet in a river, and the coastal side of a beach and the vicinity of a water area	function/production and life, cause/artificial, spatial adjacency/dam, gate, dike and weir. (material/soil and stone, spatial location/on the earth)
water gate	A kind of low/head building is constructed in the rivers and channels for controlling flow and adjusting water level	function/control flow and adjusting water level, cause/ artificial, material/soil and stone (spatial morphology/building, spatial adjacency/river, channel, lake and reservoir)
flood storage and detention basin	Some areas, such as lakes along with rivers, low depressions or specially designated areas, are originated from constructing dikes and ancillary buildings for defense from abnormal floods and storing floods	function/prevent flood and store flood, material/soil or stone, cause/artificial, spatial adjacency/dam, water gate, dike and weir (spatial morphology/depressions on the earth)
reservoir	A kind of artificial lake that possesses a catchment basin area originated from constructing a dam, dike or weir in the river, the valley, the depressions and underground permeable layer	material/water, cause/artificial, spatial morphology/depressions on the earth (function/ store water and prevent flood, spatial adjacency/ dam, gate, dike and weir)
dike	A kind of retaining water building, along with the edge of a lake, channel, flood flowing area, flood diversion area and reclamation area, which controls the flow of water	function/prevent flood and protection against the tide, material/soil or stone, cause/artificial, spatial morphology/retaining water building

In reality, extracting semantic information from the normative descriptions might encounter some differences and conflicts, such as inconsistencies in normative definitions, differences in different spatial locations and incomplete characteristics or overlapping functions. Furthermore, a certain vagueness, caused by different languages, also existed in the literal description of the concepts of geospatial objects. In order to deal with these heterogeneities from the merging of multi-source geospatial data, the formal method based on a top-level ontology should be considered. The formal conceptualization of geographical concepts consisting of two parts: the extent and the intent. The

former includes the entities or objects, which belong to the concept, whereas the latter represents its intrinsic meaning or properties. Each row and column of a formal context represents an extent and intent of a geographical concept, respectively. Due to the extensive contents in the two above-mentioned domains, we only selected partial elements to construct the formal contexts by FCA in Table 3.

Table 3. Parts of the formal contexts of two different ontologies.

Mark	Object	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
s1	seasonal lake	*		*			*		*		*					*
s2	ground river	*			*		*		*		*					*
s3	seasonal river	*		*			*		*			*				*
s4	lake	*		*		*			*		*		*	*		
s5	pond	*		*		*			*			*	*	*		
s6	reservoir	*			*		*		*						*	*
s7	spillway	*			*	*				*					*	
s8	dike		*		*			*	*						*	

Where each letter from “a” to “o”, represents material/water, material/soil or stone, cause/nature, cause/artificial, spatial morphology/long strip slot, spatial morphology/depressions, spatial morphology/buildings, spatial location/on the earth, spatial location/underground, time/perennial, time/seasonal, material state/flow, function/shipping, function/prevent flood and function/store water, respectively. “*” stands for criterion satisfied.

For running example in Table 3, A' and B' take the following form:

$$A' = \{\{\text{seasonal lake}\}', \{\text{ground river}\}', \{\text{seasonal river}\}', \dots, \{\text{dike}\}'\} = \{\{s1\}', \{s2\}', \{s3\}', \dots, \{s8\}'\}$$

$$B' = \{\{a\}', \{b\}', \{c\}', \dots, \{o\}'\}$$

Where $\{\text{seasonal lake}\}' = \{a, c, f, h, j, o\}$ is the set of attributes corresponding to the semantic factor *seasonal lake*, whereas $\{c\}' = \{\text{seasonal lake}, \text{seasonal river}, \text{lake}, \text{pond}\}$ is the set for semantic factors denoted by the attribute c (cause with value “nature”).

3. The Entropy-Based Weighted Concept Lattice

Although all the intents are generally of equal importance during the construction of a concept lattice [10–12], in some practical applications, a user is usually interested in some certain attribute characteristics, according to his/her preference and requirement. For example, we may pay more attention to the “*shipping property*” of the canal rather than its “*water storage*”, to some extent. Hence, we add some weights into the intent to capture its importance, and we do not need to investigate all the nodes, but only those nodes according to our needs. Motivated by an incremental updating algorithm used to effectively construct a weighted concept lattice [10–12], a proposed approach of EWCL is outlined to resolve the above-mentioned problem in the following definitions below. In parallel, we briefly recall some basic notions and judgment methods of each intent weight value, with regard to weighting a concept lattice. Refer to [10–12] for a more extensive introduction.

In a general formal context $K = (G, M, I)$, a set of attributes is expressed by $M = \{m_1, m_2, \dots, m_n\}$. We demonstrate the weights of attributes $W = \{w_1, w_2, \dots, w_n\}$, where $w_i \in W$ ($0 \leq w_i \leq 1$) denotes an importance degree of the attribute m_i .

Definition 2. A weighted formal context is defined as a quadruple $K_w = (G, M, I, W)$, where G and M are two non-empty sets called objects and attributes, respectively. W is a set of the weight value, which indicates the importance of a single attribute in M , and $I \subseteq G \times M$ is a binary relation between G and M . $n_w = (A, B, w)$ is a triple, for a set $A \subseteq G, B \subseteq M, w = \text{weight}(B)$ and $0 \leq w \leq 1$. The following two conditions are satisfied:

$$f(A) = \{b \in M \mid \forall a \in A, aIb\}$$

$$f(B) = \{a \in G \mid \forall b \in B, aIb\}$$

If $f(A)=B, f(B)=A$, i.e., then the triple $n_w = (A, B, w)$ is called the weighted concept of K_w . A and B are the extent and intent of n_w , respectively.

Definition 3. Let $n_w = (A, B, w)$ be a weighted concept of a weighted formal context K_w . $B = b_1 \cup b_2 \cup \dots \cup b_n$ ($\cup$ is the combination operator of b). If $n=1$, then B is denoted as a single attribute intent, otherwise B is denoted as a multi-attribute intent. Here, the weight of the multi-attribute intent ($\text{weight}(B)$) is defined as the arithmetic average of the corresponding attributes computed as follows:

$$\text{weight}(B) = \frac{1}{n} \sum_{i=1}^n w_i \quad (5)$$

where $A = \emptyset$, or $B = \emptyset$, we assume that $\text{weight}(B) = 1$.

In general, the weight of the single attribute intent is determined by domain experts. However, the current spatial objects stem from different domains, and it is difficult to determine the weight by experts in a specific domain. Therefore, under the condition of a lack of existing knowledge, we adopt an objective probability method to quantify the related weight by using axiomatic characterizations of information entropy, according to Shannon [11,23–25].

Definition 4. Let any object a_j and $a_j \in G$ ($1 \leq j \leq n$), $p(b_i / a_j)$ is called the probability of a_j possessing the corresponding attribute b_i , and $E(b_i)$ is called the average information of weight of G providing the attribute b_i . In a formal context K_w , if $n_w = (A, B, w)$ and $B = b_i$ ($1 \leq i \leq n$), then w_i is denoted as the weight value of single attribute intent b_i . The probability and the weight value are computed as follows [11,25]:

$$E(b_i) = - \sum_{j=1}^n p(b_i / a_j) \log_2^{p(b_i / a_j)} \quad (6)$$

$$w_i = \frac{E(b_i)}{\sum_{i=1}^n E(b_i)} \quad (7)$$

The above w_i is generally a normalized form. Here, we regard a weighted concept lattice, based on the weighted value produced by the information entropy, as an entropy-based weighted concept lattice. However, in practical applications, $\text{weight}(B)$ does not take into account the importance of the deviation among all the intents. The result is not conducive to the sensitive extraction of a user interested in knowledge. Therefore, in order to explore w_i deviating from $\text{weight}(B)$, we introduce a

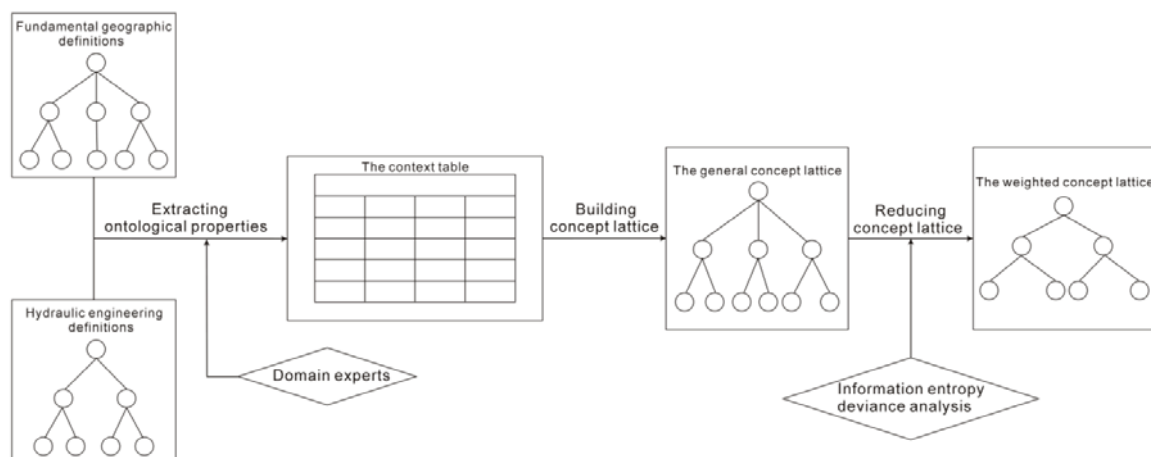
deviation analysis to evaluate the importance of the multi-attribute intent weight value and select an appropriate threshold value to further meet the needs of the user. The deviation analysis is computed as follows [11,12]:

$$D(B) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (w_i - \text{weight}(B))^2} \quad (8)$$

where $D(B)$ is denoted as the deviation of the multi-attribute intent weight value. In particular, if $n = 1$, then $D(B) = 0$.

From the above analysis, the basic steps of merging multi-source geo-ontologies based on EWCL are shown in Figure 1. We can see that the approach consists of three stages: extracting ontological properties, building a general concept lattice and reducing the general concept lattice based on information entropy and a deviance analysis.

Figure 1. The workflow of merging multi-source geo-ontologies based on EWCL.



4. Case Study and Discussion

Our method is focused on the interdisciplinary merging of geo-ontologies, which is quite different from other approaches in the same field. For example, we have employed the shared concept related to the hydrographic ontology between the fundamental geographic information (GB/T 13923-2006) and hydraulic engineering (SL 213-98) to implement the merging of ontologies. We selected several extents from these two domains, respectively, and constructed a synthetic formal context in Table 3, in which *lakes*, *reservoirs*, *spillways*, and *dikes* are interdisciplinary common objects. The intents of the two domain ontologies were expanded unanimously, based on domain experts (see Table 3). The process of building ontologies was not discussed in detail. Refer to [5,21] for a more extensive introduction. Here, we reduced the merged concept lattice from the intent direction by EWCL.

First, by using Equations (5)–(7), we obtained the single intent weight value of the merged concept lattice in Table 4, by using information entropy based on Table 3. Then, by using the incremental updating algorithm of the concept lattice [11,12,26,27], we drew a general weighted concept lattice (WCL) for representing the merged ontologies, which was induced from the formal context in Table 3.

Despite semantic heterogeneities between multi-source geospatial data, the integrated concept lattice comprised a common and equally perceived part of geospatial objects in the two domains. As shown in Figure 2, all the intents were of equal importance, and all concept lattice nodes C_i were the union of sets $(A'', A') \cup (B', B'')$. The operations result in the following concepts:

$C0 = (\{s1, s2, s3, s4, s5, s6, s7, s8\}; \emptyset; 1)$ “largest concept”

$C1 = (\{s1, s2, s3, s4, s5, s6, s7\}; \{a\}; 0.026)$

$C2 = (\{s2, s6, s7\}; \{d\}; 0.076)$

$C3 = (\{s1, s2, s3, s4, s5, s6, s8\}; \{h\}; 0.026)$

$C4 = (\{s4, s5, s7\}; \{a, e\}; 0.054)$

$C5 = (\{s1, s2, s3, s4, s5, s6\}; \{a, h\}; 0.026)$

$C6 = (\{s2, s6, s7\}; \{a, d\}; 0.051)$

$C7 = (\{s1, s3, s4, s5\}; \{a, ch\}; 0.043)$

$C8 = (\{s6, s7\}; \{a, d, h\}; 0.061)$

$C9 = (\{s6, s8\}; \{d, h, n\}; 0.061)$

$C10 = (\{s1, s3\}; \{a, c, f, h, o\}; 0.056)$

$C11 = (\{s2, s6\}; \{a, d, f, h, o\}; 0.056)$

$C12 = (\{s7\}; \{a, d, e, i, n\}; 0.064)$ “spillway”

$C13 = (\{s8\}; \{b, d, g, h, n\}; 0.059)$ “dike”

$C14 = (\{s1\}; \{a, c, f, h, j, o\}; 0.060)$ “seasonal lake”

$C15 = (\{s2\}; \{a, d, f, h, j, o\}; 0.060)$ “ground river”

$C16 = (\{s3\}; \{a, c, f, h, k, o\}; 0.059)$ “seasonal river”

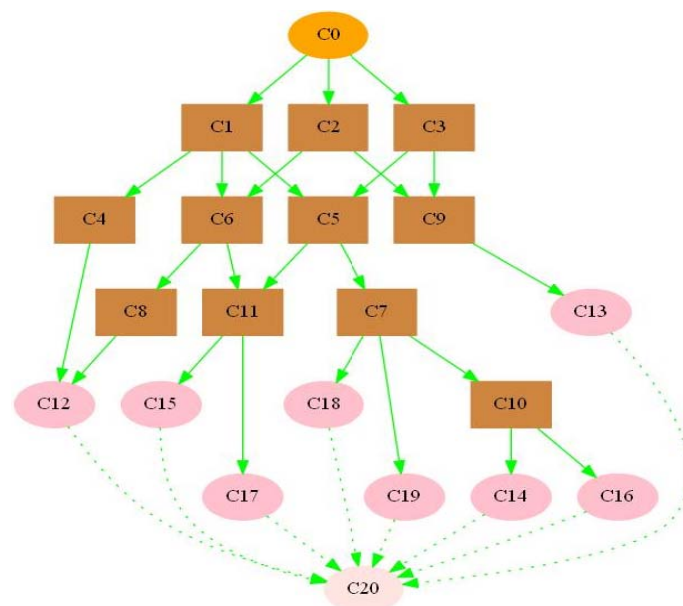
$C17 = (\{s6\}; \{a, d, f, h, n, o\}; 0.060)$ “reservoir”

$C18 = (\{s4\}; \{a, c, e, h, j, l, m\}; 0.063)$ “lake”

$C19 = (\{s5\}; \{a, c, e, h, k, l, m\}; 0.063)$ “pond”

$C20 = (\emptyset; \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o\}; 1)$ “least concept”

Figure 2. The merged concept lattice based on Table 3.



Specifically, the last term of each node C_i is denoted as the weight value in the above-mentioned parentheses. $C_{12}, C_{13}, C_{14}, C_{15}, C_{16}, C_{17}, C_{18}$ and C_{19} are the origin concepts by examining the formal context in Table 3. The least concept C_{20} does not possess the corresponding meanings in the geospatial domain, due to the completeness to form the bottom of the concept lattice. $C_0, C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9, C_{10}$ and C_{11} are the new generated concepts by the algorithm. Although not every new concept does correspond to a meaningful concept or a specialized term, it is beneficial to reveal the hierarchical relationships between geographic categories. The manuscript will not discuss further the hierarchical semantic classification of the integrated concept lattice, since our work does not contribute to those aspects. Instead, we focus on reducing the merged concept lattice, based on information entropy and a deviance analysis. Figure 2 is only a concept lattice generated by the partial objects in relation to hydrography, which is not a complete ontology category.

Table 4. Acquisition method for the single intent weight value.

Intent	$p(X)$	$E(X)$	w_i
a	0.875	0.169	0.026
b	0.125	0.375	0.057
c	0.500	0.500	0.076
d	0.500	0.500	0.076
e	0.375	0.531	0.081
f	0.500	0.500	0.076
g	0.125	0.375	0.057
h	0.875	0.169	0.026
i	0.125	0.375	0.057
j	0.375	0.531	0.081
k	0.250	0.500	0.076
l	0.250	0.500	0.076
m	0.250	0.500	0.076
n	0.375	0.531	0.081
o	0.500	0.500	0.076

Second, according to Table 4, we computed the average weight of the multi-attribute intent and the deviation of the intent importance using Equations (5) and (8). The results are shown in Table 5. In the construction process of the WCL, we took into account the user preference and interest by combining the average weight of the multi-attribute intent and the deviation of the intent importance to set thresholds for the intent importance. For any weighted concept $n_w = (A, B, w)$ in the WCL, $w = \text{weight}(B)$, we defined the quantity θ ($0 \leq \theta \leq 1$) as the threshold of the intent importance. We represented the following: if $w \geq \theta$, n_w is denoted as a frequent weighted node; otherwise n_w is denoted as an infrequent weighted node. In general, the WCL is usually composed of all of the frequent weighted nodes [10,11]. As is shown in Table 6, the quantity θ is denoted in different granulations according to the range of $\text{weight}(B)$ in Table 5. If we set up $\theta = 0.040$ as the threshold, then C_1 , C_3 , and C_5 are removed because of $w < \theta$. This indicates that one concept only involved the intents of *material/water* and *spatial location/on the earth*, which was unsatisfied with the threshold of the intent importance. In order to ensure the completeness of the concept lattice, C_1 and C_3 are temporarily retained, and C_5 is removed. Similarly, if we set up $\theta = 0.052$ as the threshold, C_1 , C_3 ,

C5, C6 and C7 are removed. The results are shown in Figure 2 and Figure 3 in different granularity, obviously, Figure 3 is greatly simplified comparing with Figure 2, whereas Figure 4 is greatly simplified comparing with Figure 3. From the above analysis, we drew a conclusion that Hasse Diagram was gradually simplified with increase of the granularity(θ), and the process of reducing the intergrated concept lattice was a stepwise refinement process corresponding to different levels of granularity.

Table 5. The intent weight value and importance deviation value of the lattice node.

Name	Intent	Intent Average Value	Weight(B)	D(B)
C0	Φ	1	1	0
C1	a	0.026	0.026	0
C2	d	0.076	0.076	0
C3	h	0.026	0.026	0
C4	ae	0.054	0.054	0.055
C5	ah	0.026	0.026	0
C6	ad	0.051	0.051	0.036
C7	ach	0.043	0.043	0.029
C8	adn	0.061	0.061	0.031
C9	dhn	0.061	0.061	0.031
C10	acfho	0.056	0.056	0.028
C11	adfho	0.056	0.056	0.028
C12	adein	0.064	0.064	0.024
C13	bdghn	0.059	0.059	0.022
C14	acfhjo	0.060	0.060	0.027
C15	adfho	0.060	0.060	0.027
C16	acfhko	0.059	0.059	0.026
C17	adfho	0.060	0.060	0.027
C18	acehilm	0.063	0.063	0.026
C19	acehklm	0.063	0.063	0.025
C20	abcdefghijklmno	1	1	0

Figure 3. The reduced weighted concept lattice when $(\theta = 0.40)$.

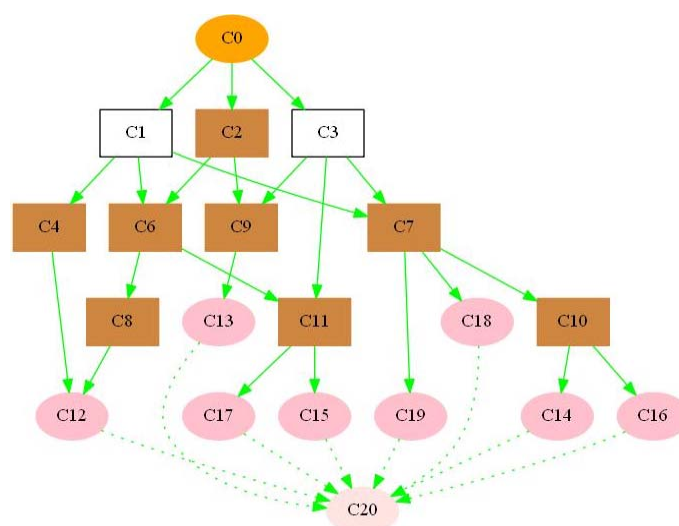
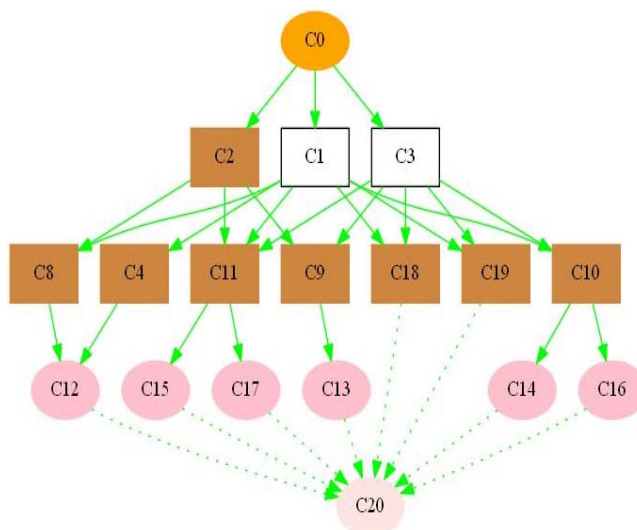


Table 6. The reduced concept lattice nodes in different granulations.

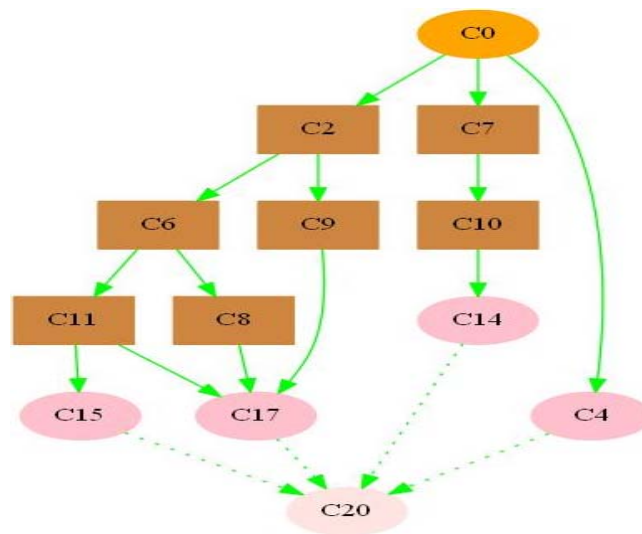
Weight(B)	θ	Removed Nodes
1	$0.76 < \theta \leq 1$	C1,C2 C3,C4, C5,C6,C7,C8,C9,C10,C11,C12,C13,C14,C15,C17,C18,C19
0.076	$0.64 < \theta \leq 0.76$	C1, C3,C4, C5,C6,C7,C8,C9,C10,C11,C12,C13,C14,C15,C17,C18,C19
0.064	$0.63 < \theta \leq 0.64$	C1, C3,C4, C5,C6,C7,C8,C9, C10,C11,C13,C14,C15,C17,C18,C19
0.063	$0.61 < \theta \leq 0.63$	C1, C3,C4, C5,C6,C7,C8,C9,C10,C11,C13,C14,C15,C17
0.061	$0.60 < \theta \leq 0.61$	C1, C3,C4, C5,C6,C7,C10,C11,C13,C14,C15,C17
0.06	$0.59 < \theta \leq 0.60$	C1, C3,C4, C5,C6,C7,C10,C11,C13
0.059	$0.56 < \theta \leq 0.59$	C1, C3,C4, C5,C6,C7,C10,C11
0.056	$0.54 < \theta \leq 0.56$	C1, C3,C4, C5,C6,C7
0.054	$0.51 < \theta \leq 0.54$	C1, C3, C5, C6, C7
0.051	$0.43 < \theta \leq 0.51$	C1, C3, C5,C7
0.043	$0.26 < \theta \leq 0.43$	C1, C3, C5
0.026	$0 \leq \theta \leq 0.26$	$\emptyset$

Figure 4. The reduced weighted concept lattice when ($\theta = 0.52$).

Finally, we defined the quantity δ ($0 \leq \delta \leq 1$) as the threshold of the deviation of the intent importance. In terms of $D(B)$ in Table 5, if we set up $\delta = 0.27$ as the deviation threshold when ($\theta = 0.40$), then C12, C13, C16, C18, and C19 are moved because of $D(B) < \delta$, namely, their intent weights were lower than the intent importance thresholds specified by the user. At the same time, we might encounter situations when although the intent weight values of these nodes (C12, C13, C16, C18, and C19) are greater than the predefined threshold (θ), the deviation values of the intent importance of these nodes are lower than the predefined threshold (δ), as can be seen the strong weighted concept lattice in Figure 5, then these nodes should also be removed. The reduced WCL when ($\theta = 0.40$ and $\delta = 0.27$) is shown in Figure 5, which is greatly simplified compared to Figure 3. In addition, according to the meaning of the deviance analysis, the smaller is the standard deviation δ , the less are the single-attribute intents deviating from $weight(B)$, namely, if the greater is the difference between the single-attribute intent weight values, the greater is the deviation value $D(B)$. For example, the deviation value C6 is greater than C4 in Table 5. That is to say that the potential weight value difference between “a” and “d” is greater than that between “a” and “e”. By ascending

the deviance value in the process of reducing the multi-attribute concept lattice, in general, the method should give priority to retain some nodes existed a greater difference between the single-attribute intent weight values, and remove other nodes existed a small difference. The example shows that the proposed method is feasible and effective for reducing the merged concept lattice in the geo-ontological domain, and is appropriate for different user requirements. Consequently, we can select an appropriate threshold to reduce the complexity of the merging of the geo-ontologies by combining information entropy and a deviance analysis.

Figure 5. The reduced weighted concept lattice when ($\theta = 0.40$ and $\delta = 0.27$).



5. Conclusion and Future Work

In this study, we present a novel method for the merging of multi-source geo-ontologies by EWCL. Firstly, to deal with the semantic heterogeneity of the geospatial information, FCA was used to extract the formal semantics, and we constructed the formal contexts from two different domains of the fundamental geographic information and hydraulic engineering domain. Secondly, in order to address the complexity of the merged concept lattice with the rapid growth in the ontology size, we have proposed a merging method for the geo-ontologies in different granulations. According to the user preference and interest, we reduced the intent of the merged concept lattice by the WCL based on information entropy and deviance analysis. We can then select an appropriate threshold value to reduce the merged concept lattice according to the specific need. Finally, experiments were conducted by combining fundamental geographic information data and spatial data in the hydraulic engineering domain. The results showed that the proposed method is both feasible and valid. As a matter of fact, the merging of multi-source geo-ontologies is still a challenge, and error-prone problems inevitably exist. WCL is a known theory in other disciplinary areas, but the proposed EWCL in this paper is a crucial application related to information entropy theory in the field of the merging of geo-ontologies, and the research is a new attempt in this direction.

In the previous work, although FCA acquired some new implicated concepts induced concepts from the given multi-source geospatial data set, it generated some redundant concepts and relations. The intent weight value, based on information entropy and a deviance analysis, was regarded as a kind of

constraint in order to reduce the merged concept lattice. Sometimes, however, the reduction achieved by EWCL is not enough; in such situations, a reasonable threshold relying on domain experts might be useful to control the simplicity. For example, extracting the semantic relations and concepts is not just a technical issue but requires a revision by domain experts to make sure that the resulting concepts make sense. Consequently, in the future, we will concentrate on multiple ways of acquisition for the intent weights of the concept lattice. Other important aspects must also be taken into account, in order to achieve semantic integration. For instance, we will further apply the formal reasoning mechanism to improve the merging integration process, and we will evaluate the quality of the merged ontology.

Acknowledgement

The authors gratefully acknowledge the reviews for constructive suggestions and the support of all the members of our research group. We are grateful to Yong Cao for his valuable comments. This research was financially supported by the National Natural Science Foundation of China (Grant NO. 41071290, 41201463).

Conflict of Interest

The authors declare no conflict of interest.

References

1. Renear, A.H.; Palmer, C.L. Strategic reading, ontologies, and the future of scientific publishing. *Science* **2009**, *326*, 230–230.
2. Buccella, A.; Cechich, A.; Fillottrani, P. Ontology-driven geographic information integration: A survey of current approaches. *Comput. Geosci.* **2009**, *35*, 710–723.
3. Gruber, T.R. A translation approach to portable ontology specifications. *Knowl. Acquis.* **1993**, *5*, 199–220.
4. Torres, M.; Quintero, R.; Moreno-Ibarra, M.; Menchaca-Mendez, R.; Guzman, G. GEONTO-MET: An approach to conceptualizing the geographic domain. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 1633–1657.
5. Kokla, M.; Kavouras, M. Fusion of top-level and geographical domain ontologies based on context formation and complementarity. *Int. J. Geogr. Inf. Sci.* **2001**, *15*, 679–687.
6. Zhu, J.W. A formal method for integrating distributed ontologies and reducing the redundant relations. *Kybernetes* **2009**, *38*, 1872–1879.
7. Buccella, A.; Cechich, A.; Gendarmi, D.; Lanubile, F.; Semeraro, G.; Colagross, A. GeoMergeP: Geographic information integration through enriched ontology matching. *New Generat. Comput.* **2010**, *28*, 41–71.
8. Stumme, G.; Adche, M.A. FCA-Merge: bottom-up merging of ontologies. In Proceedings of the Seventeenth International Conference on Artificial Intelligence (IJCAI'01), Seattle, WA, USA, 2001; pp. 225–230.
9. Chen, R.C.; Bau, C.T.; Yeh, C.J. Merging domain ontologies based on the WordNet system and fuzzy formal concept analysis techniques. *Appl. Soft Comput.* **2011**, *11*, 1908–1923.
10. Zhang, S.L.; Guo, P.; Zhang, J.F.; Wang, X.X.; Pedrycz, W. A completeness analysis of frequent weighted concept lattices and their algebraic properties. *Data Knowl. Eng.* **2012**, *82*, 104–117.

11. Zhang, S.L.; Guo, P.; Zhang, J.F. Intent weight value acquisition of weighted concept lattice based on information entropy and deviance (in Chinese). *T. Beijing Inst. Technol.* **2011**, *31*, 59–63.
12. Zhang, J.F.; Zhang, S.L.; Zheng, L. Weighted concept lattice and incremental construction (in Chinese). *Pattern Recogn. Artif. Intell.* **2005**, *18*, 171–176.
13. Li, J.L.; He, Z.Y.; Zhu, Q.L.; Liu, Y.H. A geographic ontology fusion method based on granular theory (in Chinese). *Geomatics Inf. Sci. Wuhan Univ.* **2013**, *38*, 489–492.
14. Wille, R. Concept lattices and conceptual knowledge systems. *Comput. Math. Appl.* **1992**, *23*, 493–515.
15. Chen, Y.; Yao, Y.Y. A multiview approach for intelligent data analysis based on data operators. *Inform. Sciences* **2008**, *178*, 1–20.
16. Kumar, C.A. Knowledge discovery in data using formal concept analysis and random projections. *Int. J. Appl. Math. Comp.* **2011**, *21*, 745–756.
17. Kwon, O.; Kim, J. Concept lattices for visualizing and generating user profiles for context-aware service recommendations. *Expert Syst. Appl.* **2009**, *36*, 1893–1902.
18. Ganter, B.; Wille, R. *Formal Concept Analysis: Mathematical Foundations*; Springer: Berlin, Germany, 1999.
19. Guarino, N.; Welty, C. A Formal Ontology of Properties. In *Knowledge Engineering and Knowledge Management: Methods, Models and Tools*; Dieng, R., Corby, O., Eds.; In Proceedings of 12th International Conference on Knowledge Engineering and Knowledge Management, (EKAW 2000), Juan-les-Pins, French Riviera, France, 2–6 October 2000; Springer-Verlag: Berlin, Germany, 2000; pp. 97–112.
20. Kavouras, M.; Kokla, M.; Tomai, E. Comparing categories among geographic ontologies. *Comput. Geosci.* **2005**, *31*, 145–154.
21. Wang, H.; Li, L.; Zhu, H.H. *Research on National Fundamental Geographic Information Ontology* (in Chinese); Science Press: Beijing, China, 2011; pp. 74–91.
22. Li, L.; Zhu, H.H.; Wang, H.; Li, D.R. Semantic analyses of fundamental geographic information based on formation ontology—exemplifying hydrological category (in Chinese). *Acta Geodaetica et Cartographica sinica.* **2008**, *37*, 230–235.
23. Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, IL, USA, 1946.
24. Lee, I.; Seo, D.-C.; Choi, T.-S. Entropy-based block processing for satellite image registration. *Entropy* **2012**, *14*, 2397–2407.
25. Csiszár, I. Axiomatic characterizations of information measures. *Entropy* **2008**, *10*, 261–273.
26. Godin, R.; Missaoui, R.; Alaoui, H. Incremental concept formation algorithms based on Galois (concept) lattices. *Comput. Intell.* **1995**, *11*, 246–267.
27. Kourie, D.G.; Obiedkov, S.; Watsona, B.W.; van der Merwe, D. An incremental algorithm to construct a lattice of set intersections. *Sci. Comput. Program.* **2009**, *74*, 128–142.