A DL-based Approach for Detecting Semantic Relations in Geo-Ontology Matching

Zongyi He, Junli Li, Qiaoli Zhu
School of Resource and Environmental Science, Wuhan University
129 Luoyu Road, Wuhan 430039 P.R.CHINA

Abstract. Ontology matching produces mapping relations between elements of two ontologies and it is a basic problem in geographical information integration. Currently, most existing studies rely on all kinds of semantic similarity between the semantic entities to measure ontology mapping relations. However, these measures are not sufficient due to only detecting equivalence relation of compared ontologies, these results also fail to consider potential restriction conflicts, syntax analysis and semantics reasoning techniques. In addition, evaluation studies have shown that general precision and recall can merely cover a fraction of the semantic matching ontology involved, and error-prone problems inevitably exist. In this study, we presented a DL-based method for detecting semantic relations in geo-ontology matching. First, we use Protege4.1 to build geo-ontologies based mainly on the fundamental geographic information data related to land use from the second national land survey of China. Second, we improve an existing matching algorithm of ontologies and identified the lexical matching relation by using the professional dictionary. Finally, we exploit the reason process for detecting semantic relations in geo-ontology in Eclipse environment and implement inconsistency check. Experiments show that the method is feasible and effective in the mapping of geo-ontologies.

Keywords: ontology matching; description logics; semantic relations; reasoning

1. Introduction

Nowadays, geographic information is extensively used to exchange and share within different application domains, and integrating different geospatial information between interoperating systems has become a hot topic in various scientific disciplines. However, owing to different data standards and incompatible terminologies for expressing spatial information in geographic information science, it is not easy to establish shared explicit formal vocabularies in different sources. One of the main obstacles is semantic heterogeneity, which often occurs when terms are used differently. Ontologies are widely proposed as s a powerful tool to solve the heterogeneity problem (Buccella et al. 2009), and geographic ontology is no exception, which is an explicit formal specification of the conceptual model in the geographic area. Existing geo-ontologies building relies on different experts, different tools and different techniques. Geo-ontologies may differ and even conflict, although ontologies exist in the same domain. As a consequence, detecting semantic relations from multi-source geo-ontologies need new technologies and methods

To detect relations among geo-ontology entities, such as equivalence, subsumption, overlap and disjunction, has been the challenge of ontology research. These problems can possibly be avoided if ontology matching is applied. At present, most studies rely on all kinds of semantic similarity between the semantic entities to measure ontology mapping relations. For instance, if the similarity between T_i and T_j is in close proximity to one, namely T_i and T_j are same in essence at relatively

high probability and satisfy synonymy relation; if the similarity between T_i and T_j is approaching zero, namely T_i and T_j must not satisfy the synonymy relation and inheritance relation, may be considered as two different concepts(Kang et al. 2012). Moreover, these measures are not sufficient due to only detecting equivalence relation of compared ontologies， these results also fail to take into account potential restriction conflicts, and syntax analysis and semantics reasoning techniques. In addition, evaluation studies have shown that general precision and recall can only cover a fraction of the semantic matching ontology involved and error-prone problems inevitably exist (Koll et al. 2010). Recently, there has been some considerable effort in descriptions logic for matching ontologies that can be seen as practical approaches in some literatures. Sheng et al. (Sheng et al. 2007) proposed a DLs-based algorithm to achieve ontologies matching, and RacerPro was selected as DLs reasoner to deduce ontology mapping. Atencia et al. (Atencia et al. 2012) presented a formal semantics for weighted mappings between different ontologies based on the classificational interpretation of mapping, meanwhile, an extension of Distributed Description Logics (DDL) to express concepts and relations mappings between heterogeneous ontologies (Ghidini et al. 2006), a virtual enterprise ontology mapping is achieved by using description logic and bridging axioms (Kumar et al 2013). In addition, CtxMatch (Giunchiglia et al. 2004) and S-Match (Bouquet et al. 2006) were also tried to detect semantic matching with inconsistency deduction based on DL axioms. The applications mentioned above mainly focused on mapping ontologies in the same field, and few application were based on complex reasoning methods by using DL.

In this study, we present a DL-based method for detecting semantic relations in geo-ontology matching. First, a description logic knowledge base K is constituted by the TBox T and the ABox A. Second, an algorithm of determining ontology correspondence (such as equivalence, more general, less general, disjunction and overlapping) is improved based on the algorithm (Sheng et al. 2007). Finally in Eclipse environment, with the aid of Pellet reasoning machine, via Jena to call Pellet function to implement inconsistency check by combining fundamental geographic information data and land use data.

2. Method

Description logics, a family of state-of-the-art knowledge representation language, is the backbone of OWL. OWL is used to develop the semantic web by the Word Wide Web Consortium (W3C) (Baader et al. 2003). A DL knowledge base often called ontology involving the explicit and implicit information. In general, DL knowledge consists of two parts by the terminological box (TBox), which is an axiom set to induce the terminology, i.e., the vocabulary of an applicastion domain, and Assertion box (ABox), which contains extensional knowledge, i.e., some assertions and relations about named individuals in terms of this vocabulary. An example of TBox expressing that the Paddy is a kind of cultivated land which can plant rice farmland is:

$$T=\{PaddyField \equiv \exists hasPlantRice.CultivatedLand\}$$

An ABox describe a specific state of the world represented by TBox. For instance, we can express with an assertion that Beijing is the capital of China is:

$$A=\{Capital(Beijing);Country(China);locatedAt(Beijing,China)\}$$

In many applications, DL is equipped with a Reasoner (such as Racer-Pro, FaCT++, Pellet) to acquire new knowledge from TBoxes and ABoxes. To represent the knowledge given by TBox and ABox, a DL knowledge is denoted by K = {T , A}, where T and A are two nonempty sets called TBox and ABox, respectively (Nikkilä et al. 2013). In parallel, the formal meaning of DL axioms is given by their semantics, moreover, axioms are formulas of the form $C \subseteq D$ or $C \equiv D$, where C and D are possibly complex concept descriptions. According to the expressivity of DLs, some symbols can be used to express axioms and how these symbols can be combined. For example, where C is a concept name and R is a role, a general atomic concept uses a set of constructors such as top ($\top$), bottom concept ($\perp$), union ($\cup$) , intersection($\cap$), existential quantifier ($\exists R.C$),universal quantifier($\forall R.C$), cardinality restriction ($\geq R.C, \leq R.C$), (Baader et al. 2003) etc.

```
enum SemanticRelation ={ equivalence, generalization, specialization,disjunction,overlapping };
boolean general=subConcept( C_i, C_j );

boolean special= subConcept( C_j , C_i );

String ConceptRelationItem= Null;
ConceptRelationSet= Φ
Input:  C_i ,  C_j  SemanticRelationSet

Ouput: ConceptRelationSet

while(i<sizeof( O_i ) and j<sizeof( O_j )) do

If (general=T) and (special=T) then
ConceptRelationItem= equivalence;
ConceptRelationSet= ConceptRelationSet+ ConceptRelationItem;
endif
If (general=T) and (special=F) then
ConceptRelationItem= generalization;
ConceptRelationSet= ConceptRelationSet+ ConceptRelationItem;
endif
if (general=F) and (special=T) then
ConceptRelationItem=specialization;
ConceptRelationSet= ConceptRelationSet+ ConceptRelationItem
endif
else
    if ( C_i ∧ C_j ≠ Φ) then
    ConceptRelationItem= overlapping;
    else
    ConceptRelationItem= disjunction;
    ConceptRelationSet= ConceptRelationSet+ ConceptRelationItem;
endif
return ConceptRelationSet
```

based on the elementary descriptions. Here, we will give some basic syntax and semantics of the description logic Table. The interpretation of complex concepts is defined according to the literature (Baader et al. 2003) as follows:

$$\mathrm{T}^I = \Delta^I \qquad (C \bigcap D)^I = C^I \bigcap D^I$$

$$\perp^I = \Phi \qquad (C \bigcup D)^I = C^I \bigcup D^I$$

$$\neg C^I = \Delta^I \setminus C^I \qquad (\exists R.C)^I = \{x \in \Delta^I \mid (x, y) \in R^I \ \& \ y \in C^I\}$$

On the one hand , based on (Koll et al. 2010), we exploited a set of formal rules, in relation to the basic reasoning mechanism provided by logics systems, in order to detect the semantic relations among semantic entities which includes subsumption, equivalence, disjunction and overlapping ($\cong$).

On the other hand, we turn the complex reasoning into simple reasoning for some complex semantic relation. In the practical applications, we usually combine the complex reasoning and basic reasoning mechanism to detect the semantic relation by improving an existing matching algorithm (Figure 1).

Experiment and Discussion

The process of implementation between ontologies can be summarized as follows.

Fist, we used protege4.1 to build geo-ontologies based mainly on the fundamental geographic information data related to land use from the second national land survey of China, which are from GB/T 13923-2006 (Specifications for feature classification and codes of fundamental geographic information) and GB/T13923-92 (Classify and codes for the national land information) in China(Description Logic is regarded as a family of knowledge representation languages which can be used to represent the terminological knowledge of a geographical application domain in a structured and formally well-understood way, see Table 1), moreover, partial land use type of experimental data are from Jiangxia District in Wuhan. The second step identifies the lexical matching relation by using the professional dictionary and the final step uses the DL reasoners (e.g. Jena) to deduce the semantic relation, namely, we implement the reason process in Eclipse environment, with the aid of Pellet reasoning machine, via Jena to call Pellet function to implement inconsistency check.

To detect the semantic relation between the above basic ontologies, we also used the reasoner (pellet and SPARQLDL Java API) to find the relationships. In particular, a starting point was provided as follow :

Table 1. Partials of two different geo-ontologies based on DL

Ontology Name	Logic semantic
O2006 (GB/T 13923-2006)	$Thing \supset CultivatedLand$ $CultivatedLand \supset PaddyField$ $CultivatedLand \supset Upland$ $CultivatedLand \supset VegetableField$ $Thing \supset GardenLand$ $GardenLand \supset Orchard$ $GardenLand \supset MulberryGarden$
O1992 (GB/T13923-92)	$Thing \supset CultivatedLand$ $CultivatedLand \supset PaddyField$ $CultivatedLand \supset RiceField$ $CultivatedLand \supset AquaticPantField$ $CultivatedLand \supset Upland$ $CultivatedLand \supset VegetableField$ $Thing \supset GardenLand$ $GardenLand \supset Orchard$ $GardenLand \supset TeaGarden$ $Thing \supset ForestLand$ $ForestLand \supset BambooForest$ $ForestLand \supset$ AfforestLand $ForestLand \supset$ Slash

$$\mathrm{O2006}:CultivatedLand = O1992:CultivatedLand \wedge$$
$$\mathrm{O2006}:GardenLand = O1992:GardenLand$$

In addition, some functions in SPARQLDL such as SubClassOf(a,b), DirectSubClassOf(a,b),SameAs(a,b),DisjointWith(a,b),DifferentFrom(a,b) and ComplementOf(a,b) etc. are exploited to ask concepts semantic relations. For instance, we may detect whether class *PaddyField* is a direct subclass of *CultivatedLand* by using SPARQLDL as follow:

ASK { DirectSubClassOf($O2006:PaddyField$, $O2006:CultivatedLand$)}

If the above function return value is true, it means that *CultivatedLand* is more general than *PaddyField* . Similarly, SemanticRelation { *TeaGarden* , *Orchard* }= $\subseteq$ means that *TeaGarden* is less general than *Orchard* , SemanticRelation { AfforestLand , Slash }= $\cong$ means that AfforestLand is overlaid with Slash . SemanticRelation { *BambooForest* , *PaddyField* }= $\perp$ means that *BambooForest* is disjoint from *PaddyField* . Here, we restrict the semantic relation to be one of the relations form the set :{ $=, \perp, \subseteq, \supseteq, \cong$ }, and in this way, we can detect the mapping relationship between the two concepts which orgin from different geo-ontologies .

3.Conclusion

In this study, we proposed a DL-based approach for detecting semantic relations in geo-ontology, which included equivalence, subsumption, overlap and disjunction. During the experimental process, we improved an existing matching algorithm of ontologies, and implemented inconsistency check by using Pellet reasoning machine, SPARQLDL Java AP and Jena API in Eclipse environment. The results show that experiment based on description logic can yield better performance than comparing the similarities measures and expert mappings. Although this method at present can only consider the concept matching, it is also a beneficial attempt to use semantic interpretation system to identify the integrity and accuracy of the semantic relations. In our future work, we will develop an ontological framework based on DL to detect the semantic relationships between geographic entities, and handle the complexity of time and spatial for exploring the semantic relationships comparing with other methods. Another significant area which needs a closer attention to perform mapping of ontologies exploited in DL with different expressivity (OWL-Lite, OWL-DL, fuzzy description logic).

References

Buccella A, Cechich A, Fillottrani P (2009). Ontology-driven geographic information integration: A survey of current approaches. Computers & Geosciences. 35: 710-723.

Kang X P; Li D Y; Wang S G (2012). Research on domain ontology in different granulations based on concept lattice. Knowledge-Based Systems. 27:152–161.

Koll M, Boufaida Z. A Description Logics Formalization for the Ontology Matching. Proc. Comput. SCi. (WCIT2010), 2010,29-35.

Shen G H，Huang Z Q，ZHU X D.Using Description Logics Reasoner for Ontology Matching [c].IITA Proceedings of the Workshop on Intelligent Information Technology Application,2007,30-33.

Atencia M, Borgida A, Euzenat J, et al. A Formal Semantics for Weighted Ontology Mappings[C]. ISWC 2012, Part I, LNCS 7649, 2012,17–33.

Ghidini C, Serafini L. Reconciling concepts and relations in heterogeneous ontologies[C], in: Proceedings of the 3rd European Semantics Web Conference(ESWC), Montenegro, 2006. 50–64.

Kumar S K, Harding J A (2013). Ontology Mapping Using Description Logic and Bridging Axioms[J]. Computers in Industry.64:19-28

Giunchiglia F, Shvaiko P, Yatskevich M. S-Match: an algorithm and an implementation of semantic matching, in: C. Bussler, J. Davies, D. Fensel, R. Studer (Eds.), The Semantic Web: Research and Applications, Springer, Berlin/Heidelberg.2004, 61-75.

Bouquet P, Serafini L, Zanobini S, et al. Bootstrapping semantics on the Web: meaning elicitation from schemas, in: Proceedings of the 15th International Conference on World Wide Web, Edinburgh, Scotland, 2006, 505-512.

Baader F, Calvanese D, McGuinness D(2003). Patel-Schneider(Eds.), The Description Logics Handbook, Cambridge University Press,UK.

Nikkilä R , Nash E , Wiebensohn J , Seilonen I , Koskinen K(2013): Spatial inference with an interchangeable rule format, International Journal of Geographical Information Science, DOI:10.1080/13658816.2012.750323.