Extension of Ontologies Assisted by Automated Reasoning Systems

Joaquín Borrego-Díaz and Antonia M. Chávez-González

Departamento de Ciencias de la Computación e Inteligencia Artificial,
E.T.S. Ingeniería Informática-Universidad de Sevilla,
Avda. Reina Mercedes s.n. 41012-Sevilla. Spain

Abstract. A method to extend ontologies with the assistance of automated reasoning systems and preserving a kind of completeness with respect to their associate conceptualizations is presented. The use of such systems makes feasible the *ontological insertion* of new concepts, but it is necessary to re-interpret the older ones with respect to new ontological commitments. We illustrate the method extending a well-known ontology about spatial relationships, the called *Region Connection Calculus*.

1 Introduction

Ontology Management has become in a critical issue in fields related with Knowledge Representation and intelligent information processing as the Semantic Web. One of the involved tasks, the most important, is the need of extending or revising ontologies. This task may be, from the point of view of companies, dangerous and expensive: every change in the ontology can affect to the overall knowledge organization of the company. Moreover it is also known that the self process of extension is hard to automatize: the tools are designed to facilitate the syntactic extension or ontology mapping. But the effect of ontology mapping on the logical reasoning may be, in general, unknown, and specially on the use of automated reasoning systems [2].

The aim of this paper is to propose a formal semantics for ontology extension (following the foundational principles given in [2] and suggested by the computer-assisted cleaning of Knowledge Databases [1]) as well as a feasible method, assisted by Automated Reasoning Systems (ARS), to extend ontologies preserving a certain type of *robustness*.

2 Lattice Categorical Extensions

We assume throughout that the conceptualization associated to the ontology is endowed of lattice structure. Actually it is not a constraint: there are methods

* Partially supported by the project TIN2004-03884 of Spanish Ministry of Education and Science, cofinanced by FEDER funds.

to extract lattices of concepts from data (such as the Formal Concepts Analysis, see e.g. [8]), and an ontology is easy to be extended by definition to satisfy it. Although we think about Description Logics as a language (a logical basis for ontology languages like OWL, see `http://www.w3.org/TR/owl-features/`), the framework is useful for First Order Logic (FOL).

A *lattice categorical* theory is a theory that proves the lattice structure of its basic concepts. It is a reasonable requirement: the theory must certify the basic relationships among them. We aim to replace *completeness* by *lattice categoricity* to make feasible the extension of formal ontologies.

Fixed a language, let $\mathcal{C} = \{C_1, \dots, C_n\}$ be a set of concept symbols and let T be a theory (in the general case, definable concepts in T can be assumed) and let us consider the language $L_\mathcal{C} = \{\top, \bot, \leq\} \cup \{c \ : \ c \in \mathcal{C}\}$. Given $M \models T$, we consider the $L_\mathcal{C}$-structure $L(M, \mathcal{C})$, whose universe is the set of the interpretations in M of the concepts ($\top$ is M, $\bot$ is $\emptyset$), and $\leq$ is the subset relation.

The relationship between $L(M, \mathcal{C})$ and the self M is based in two facts. The first one, that the lattice can be characterized by a finite set of equations E, plus a set of formulas $\Theta_\mathcal{C}$ categorizing the lattice under completion, that is, $\Theta_\mathcal{C}$ includes the domain closure axiom, the unique names axioms and aditionally the axioms of lattice theory. Secondly, there is a natural translation Π of lattice equations into FOL formulas such that if E is a set of equations characterizing $L(M, \mathcal{C})$, then $M \models \Pi(E)$.

Definition 1. *We say that a $L_\mathcal{C}$-theory E is a* **lattice skeleton** *(l.s.) for T if*

- *There is $M \models T$ such that $L(M, \mathcal{C}) \models E + \Theta_\mathcal{C}$, and*
- *$E + \Theta_\mathcal{C}$ has an only model (modulo isomorphism).*

Every consistent theory T has a lattice skeleton (it is sufficent to categorically axiomatize the lattice associated to some model of T). Intuitively, the existence of essentially different lattice skeletons difficults the reasoning with the conceptualization associate to T.

Definition 2. *T is called a* **lattice categorical (l.c.) theory** *if whatever two lattice skeletons for T are equivalent modulo $\Theta_\mathcal{C}$.*

It is easy to see that every T consistent has a lattice categorical extension: it is sufficent to consider a model $M \models T$, and next to find a set E of equations such that $\Theta_\mathcal{C} + E$ has $L(M, \mathcal{C})$ as only model. The theory $T + \Pi(E)$ (and any consistent extension of it) is l.c.

To simplify, we deal with a pair (T, E) -where T is lattice categorical and E is a lattice sekeleton for T- that we call a *lattice categorical core* (l.c.c.). Thus, (T, E) is a l.c.c. if $T + \Pi(E)$ is a l.c. theory.

Definition 3. *Given two l.c.c. $(T_1, E_1), (T_2, E_2)$ with respect to the sets of concepts $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively, we say that (T_2, E_2) is a* **lattice categorical extension** *of (T_1, E_1) if $L(T_1, \mathcal{C}_1) \subseteq L(T_2, \mathcal{C}_2)$ and $L(T_2, \mathcal{C}_2) \models E_1$.*

3 Extending Ontologies

In order to obtain a practical method, some of the basic (theoretical) logical principles required by the *definitional methodologies* of building of formal ontologies must be weakened [3]. Such principles, in their original forms, are:

1. *Ontologies should be based upon a small number of primitive concepts.*
2. *These primitives should be given definite model theoretic semantics.*
3. *Axioms should only be given for the primitive concepts.*
4. *Categorical axiom sets should be sought.*
5. *The remaining vocabulary of the ontology (which may be very large), should be introduced purely by means of definitions.*

The three first principles are assumed, but, in order to a feasible management, the last two ones (two strong logical constraints) are weakened. The fourth one will be replaced by *lattice categoricity*, more manageable than logical categoricity or completeness. With respect to the last one, if we start with a basic theory, it can be hard to define any new concept/relation by means of the basic elements of the ontology. Thus, we must consider that there are *ontological insertions*, that is, additions of new concepts/relations not ontologically defined on the former ontology. This may produce a deep readdress of the domain analysis.

The method consists of four steps, assisted by an automated theorem prover (in our case, OTTER, `http://www-unix.mcs.anl.gov/AR/otter/`), a model finder (MACE4, `www-unix.mcs.anl.gov/AR/mace4/`), and a last stage for ontological reconsideration. Starting from a lattice categorical theory:

1. First, one extends the lattice of the basic concepts of the ontology by extending the selected skeleton.
2. Next, one applies MACE4 on a possible axiomatization of the new lattice in order to obtain the new lattices. In general, the characterization of the lattice is a theory weaker than the initial ontology.
3. The third step consists of the refinement of the skeleton in order to MACE4 exhibits one only model (that is, the theory is lattice categorical).
4. Finally, it is necessary to certify (by means OTTER or hand-made) the unicity of above model.

The final stage of the method is not algorithmical. It consists of an ontological interpretation of the new element, by re-interpreting (generally by refining) if necessary, the older ones. This task, nonalgorithmical in essence, is responsability of experts in the domain represented by the ontology. In fact, such re-interpretation can force us to reconsider the initial ontological commitments.

4 An Example in Qualitative Spatial Reasoning

We shall apply the method for extending an ontology on Qualitative Spatial Reasoning by means of the insertion of relations on imperfect spatial information, concretely the well-known *Region Connection Calculus* (RCC) [6]. The

$DC(x,y) \leftrightarrow \neg C(x,y)$	(x is disconnected from y)
$P(x,y) \leftrightarrow \forall z[C(z,x) \rightarrow C(z,y)]$	(x is part of y)
$PP(x,y) \leftrightarrow P(x,y) \wedge \neg P(y,x)$	(x is proper part of y)
$EQ(x,y) \leftrightarrow P(x,y) \wedge P(y,x)$	(x is identical with y)
$O(x,y) \leftrightarrow \exists z[P(z,x) \wedge P(z,y)]$	(x overlaps y)
$DR(x,y) \leftrightarrow \neg O(x,y)$	(x is discrete from y)
$PO(x,y) \leftrightarrow O(x,y) \wedge \neg P(x,y) \wedge \neg P(y,x)$	(x partially overlaps y)
$EC(x,y) \leftrightarrow C(x,y) \wedge \neg O(x,y)$	(x is externally connected to y)
$TPP(x,y) \leftrightarrow PP(x,y) \wedge \exists z[EC(z,x) \wedge EC(z,y)]$	(x is a tangential prop. part of y)
$NTPP(x,y) \leftrightarrow PP(x,y) \wedge \neg\exists z[EC(z,x) \wedge EC(z,y)]$	(x is a non-tang. prop. part of y)

Fig. 1. Axioms of RCC

$\top \equiv C \sqcup DR$	$PO \sqsubseteq \neg P \sqcap \neg Pi \sqcap \neg DR$	$DR \equiv EC \sqcup DC$
$NTPP \sqsubseteq \neg TPP \sqcap \neg Pi \sqcap \neg DR$	$C \equiv O \sqcup EC$	$TPP \sqsubseteq \neg Pi \sqcap \neg DR$
$O \equiv PO \sqcup P \sqcup Pi$	$EQ \sqsubseteq \neg PPi \sqcap \neg DR$	$Pi \equiv EQ \sqcup PPi$
$TPPi \sqsubseteq \neg NTPPi \sqcap \neg DR$	$P \equiv EQ \sqcup PP$	$NTPPi \sqsubseteq \neg DR$
$PPi \equiv TPPi \sqcup NTPPi$	$EC \sqsubseteq \neg DC$	$PP \equiv TPP \sqcup NTPP$

Fig. 2. The skeleton E for the lattice of concepts of RCC

need of such extension arose, for example, when we applied RCC as a meta-ontological tool for analysing and repairing anomalies in ontologies [1] [5]. RCC is a mereotopological approach to spatial reasoning; the *spatial entities* are non-empty regular sets. The primary relation is the *connection*, $C(x,y)$, with intended meaning: *"the topological closures of* x *and* y *intersect"* and basic axioms $\forall x[C(x,x)]$ and $\forall x,y[C(x,y) \rightarrow C(y,x)]$ jointly with a set of definitions on the main spatial relations (fig. 1). Actually the theory has other axioms (see [6]), but these are not necessary to prove the lattice structure of the set of relations (shown in fig. 3). Thus, RCC is lattice categorical.

4.1 Isolating a Skeleton for RCC

In order to isolate a skeleton without redundant formulas, we start with the lattice equations induced by the Hasse diagram of the RCC-relations. Next we sequentially remove equations of this set when such elimination does not produce other new lattices modelling the final set. The set of equations E, see the figure 2, we obtain has an only model (is a skeleton). categorizes under completion the lattice of the RCC-spatial relationships (given in fig. 3) [5].

4.2 Inserting New Elements

The Jointly Exhaustive and Pairwise Disjoint set (JEPD) of atomic relations of the lattice (fig. 3) is denoted by RCC8. It represents the set of the most specific spatial relations in RCC8. Our aim is to insert a new relation representing undefinition, such relation must be disjoint with RCC8.

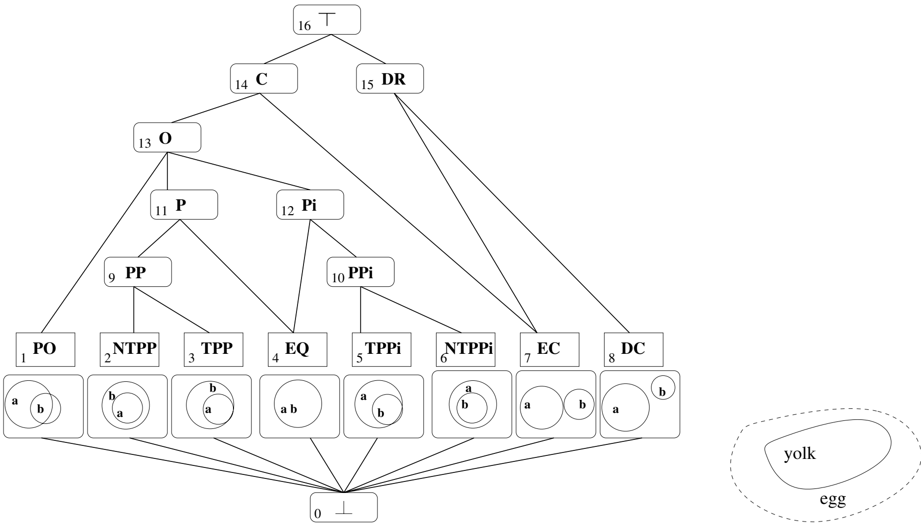

Fig. 3. The lattice of RCC-relations (left) and the egg-yolk representation of vague regions (right)

Theorem 1. *There are only eight E-conservative extensions of the lattice of RCC by insertion of a new relation D such that* $RCC8 \cup \{D\}$ *is a JEPD set.*

In the proof of the theorem we use MACE4 for listing the lattice extensions, taking as input the lattice axioms, the skeleton, the unique names principles and the closure domain axiom. The system outputs eight extensions. Since MACE4 has not been formally verified to work correctly, it is necessary to certify that such models are correct, and, by finding OTTER's proofs, to show that the list of models is exhaustive. The analysis of the extensions (fig. 4) suggests us that the new relations represent *undefinition up to a degree.*

4.3 Refining the Skeleton for the New Extension

We are not specially interested here in a determinated extension, although there exist situations where it is necessary to select one of them (by example, when we intend to classify unaccurate data [4]). However, the refinement of the skeleton is easy: once an extension is selected. For instance, with respect to the first extension, it is sufficent to substitute the formula $PP \equiv TPP \sqcup NTPP$ by the new formula $PP \equiv TPP \sqcup I_1 \sqcup NTPP$ to obtain a skeleton E' for the extension. Every extension of (RCC, E') is a l.c. extension of (RCC, E).

4.4 Final Stage: Ontological Interpretation

Finally, we need to mereotopologically interpret the new relations. In [5] four different interpretations are offered, we tried to use some of them for supporting

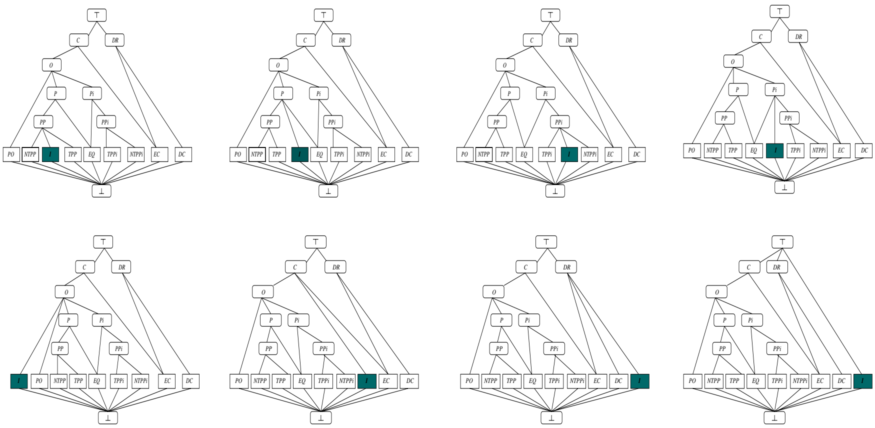

Fig. 4. The eight lattice describing the l.c. extensions of RCC by undefinition

ontological cleaning tasks. In order to show how they can be interpreted in each case, we consider the classical *egg-yolk* interpretation of spatial vague regions [7]. Intuitively, a spatial region a compounded by two subregions, as figure 3 shows, the first, $y(a)$ (*the yolk*) which represents accurate locations in a, and the second one, $e(a)$ bounding the unaccurate locations of a. In [7] the 48 possible spatial relations between two vague regions are shown. If we want to work with I_1, for example, its vague interpretation is $I_1(a, b) \equiv PP(e(a), e(b))$, while RCC relations are interpreted by the natural way.

5 Final Remarks

The method described here is a logical basis for extending ontologies. Since there is a lack of formal notions -feasible in practice- describing features about completeness in the evolution of formal ontologies, we think that our proposal can be useful to add formal semantics to several ontological transformations [5] [1], achieving in this way the logical trust. The feasibility of the method depends of two factors: the use of efficent ARS and the simplicity of the *completeness* notion, related with the conceptualization of the ontology. Future research lines are addressed to embrace the use of roles on spatial reasoning.

References

1. J. A. Alonso-Jiménez, J. Borrego-Díaz and A. M. Chávez-González, Ontology Cleaning by Mereotopological Reasoning Proc. of DEXA Workshop on Web Semantics (WebS'04), (2004). IEEE Press, pp. 132-137.
2. J. A. Alonso-Jiménez, J. Borrego-Díaz, A. M. Chávez-González and F.J. Martín-Mateos, Foundational challenges in Automated and Ontology Cleaning in the Semantic Web. To appear in IEEE Intelligent Systems.

3. B. Bennett, The Role of Definitions in Construction and Analysis of Formal Ontologies, in: P. Doherty, J. McCarthy, and M. Williams (eds.) Logical Formalization of Commonsense Reasoning (2003 AAAI Symposium), pp. 27-35, AAAI Press (2003).
4. J. Borrego-Díaz and A. M. Chávez-González, Management Undefinability in Data Through Automated Ontology Extensions, submitted (2005).
5. A. Chávez-González, Mereotopological Automated Reasoning for Ontology Cleaning, forthcoming Ph.D. thesis.
6. A. G. Cohn, B. Bennett, J. M. Gooday and N. M. Gotts. Representing and Reasoning with Qualitative Spatial Relations about Regions. Chapter 4 in O. Stock (ed.), Spatial and Temporal Reasoning, Kluwer, Dordrecth (1997).
7. A.G. Cohn and N.M. Gotts, The 'Egg-Yolk' Representation of Regions with Indeterminate Boundaries in P. Burrough and A. M. Frank (eds), Proc. GISDATA Specialist Meeting on Geographical Objects with Undetermined Boundaries, GISDATA Series, vol. 3, Taylor and Francis, pp. 171-187 (1996).
8. G. Stumme, R. Taouil, Y. Bastide, N. Pasquier and L. Lakhal, Computing Iceberg Concept Lattices with TITANIC. *Data Knowl. Eng.* 42(2) pp. 189-222 (2002).