Mereotopological Patterns for Ontology Evolution and Debugging

Gonzalo A. Aranda-Corral[1], Joaquín Borrego Díaz [2] and Antonia M. Chávez-González[2]

[1] Universidad de Huelva, Department of Information Technology, Palos de La Frontera. Spain
[2] Universidad de Sevilla, Department of Computer Science and Artificial Intelligence, Sevilla, Spain
Gonzalo.Aranda@dti.uhu.es ,jborrego@us.es,tchavez@us.es

Abstract In this paper the foundational principles and the application of a mereotopological theory, the *Region Connection Calculus*, for controlling the revision of formal ontologies by means of visual arrangements is presented. The visual representation of logical relationships between concepts of an ontology is defined, and it is computed by means of an automated theorem prover. The user can recognize mereotopological patterns in the visual representation, particularly those representing anomalies in the ontology. An intelligent tool called Paella is designed and implemented for this task. Also, the extension to this formalism for managing uncertainty in concept reasoning is described.

Keywords Mereotopological Patterns; Region Connection Calculus; Ontologies; Description Logics

1 Introduction

The use of intelligent interfaces for the extraction, representation and reasoning with knowledge and information is one of the most important aims in Artificial Intelligence. This aim especially is reached when the complexity of the formal representation of the knowledge is very high. For example, in the incipient Semantic Web (SW) the angular stone of the technology they are the ontologies formalized and represented in formal designed ad hoc languages as the Ontology Web Language (OWL)[1]. The syntactic complexity of the representation supposes an impassable barrier for the adoption of such ontologies as semantic models for users' big networks (for example, in the Web 2.0), since difficultly they can analyze critically their structure and content. From the point of view of the Knowledge Representation (KR) paradigm, ontology revision comes from the fact that the discourse domain may not be faithfully represented by an ontology (a well known working principle in KR). In many cases, end-users need to interact and transform the ontology. Even if the ontology designer thinks that the ontology is final, the end-user may think the opposite, or simply that the ontology is incorrect. In fact, it should be feasible to achieve the agreement designer-user. This agreement is essential for the assimilation of SW technologies into non-academic community portals, for example.

In this article there is described the work realized by the authors in the formalization and development of a tool that, using mereotopological patterns, allows to the not expert users to analyze ontologies formalized in OWL to detect anomalies. Of this form, it is allowed the users to deal and to adapt ontologies for identical application to his interests. Concretely, we will center on the logician - mathematics properties of the use of *Region Connection Calculus* (RCC) [7] like goal-ontology for the representation and the reasoning with concepts in ontologies. RCC's use like basic prop of representation (extending the mereological to mereotopological) [2] has supposed the ontological treatment of the theory, its extension, refinement and specific application [4].

The structure of the paper is as follows. The next section is devoted to present foundational principles on which the use of mereotopological reasoning in Ontology cleaning is based. The formal theory used in the mereotopological interpretation is presented in section 3. The result of the study, Paella tool, is

[1] http://www.w3.org/TR/owl-ref/

succinctly described in section 4. The last section sketches how the tool can be used for reasoning with uncertain concepts by means of a logic-topological re-interpretation and extension of RCC.

2 Visual Ontology cleaning

End-user preferences on visual representation are well known in other related fields such as Formal Concept Analysis or Data Engineering. The spatial metaphor is a powerful tool in human information processing. The user will feel encouraged to repair the anomaly, although there exist some obstacles: on the one hand, visual reparation may not be corresponded by a logical reparation of the ontology source. This occurs if there is no formal semantics for supporting the change; on the other hand, repairs can be logically complex. Domain experts often underestimate the amount of time required to produce an ontology, and consequently they build an ontology based on a large scope. The resulting conceptual ontologies are consequently a mix of both domain and task ontology concepts which are hard to manage [13].

Paraphrasing [14], visual cleaning of ontologies is important for future end-users of ontology debugging systems due mainly to three reasons [3]:

1. It allows the user to summarize ontology contents.
2. User's information is often fuzzily defined. Visualization can be used to help the user to get a nice representation.
3. Finally, visualization can therefore help the user to interact with the information space.

2.1 Mereotopology and First Cognitive Principle

The thesis that supports the use of the representation and mereotopological reasoning as a tool of ontological purification was described in [3], in the shape of *cognitive principles*. These beginnings are those which would govern the use of the Qualitative Spatial Reasoning (QSR) in this context, as well as they would base the correct application of this one. The main principle is the following one:

Main Cognitive Principle (MCP): If we aim to use spatial reasoning for cleaning ontologies, we have to provide a theory on spatial entities for translating the impact of spatial arrangements into revisions of the ontology source.

In order to satisfy the MCP, a theory on Qualitative Spatial Reasoning (QSR) has to be selected. In this way, the following sub-principle is chosen:

First Cognitive Principle (CP1): The concepts of a conceptualization associate to a clear ontology can be topologically represented by means of regular non-empty regions.

That is, there is a model of the ontology whose universe is the bidimensional or tridimensional space, and that model interprets concept symbols as regions. It is evident that the represented knowledge will depend of topological relations. The starting-up of CP1 needs of a robust theory to reason with spatial regions. Additionally, the theory must facilitate the knowledge interchange between the ontology and spatial models.

3 Region Connection Calculus

The selected theory is the well known *Region Connection Calculus* (RCC) [7]. RCC is a mereotopological approach to QSR; it describes topological features of spatial relationships. It has been used in several subfields of AI, included in Qualitative Reasoning in ontologies for the SW [11], Spatial databases [15] [10].

In RCC, the *spatial entities* are non-empty regular sets. The ground relation is the *connection*, $C(x, y)$, with intended meaning: *"the topological closures of* x *and* y *intersect"*. The basic axioms of RCC are

$$\forall x[C(x,x)] \quad \forall x,y[C(x,y) \rightarrow C(y,x)]$$

$DC(x, y) \leftrightarrow \neg C(x, y)$	(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)

Figure 1: Axioms of RCC

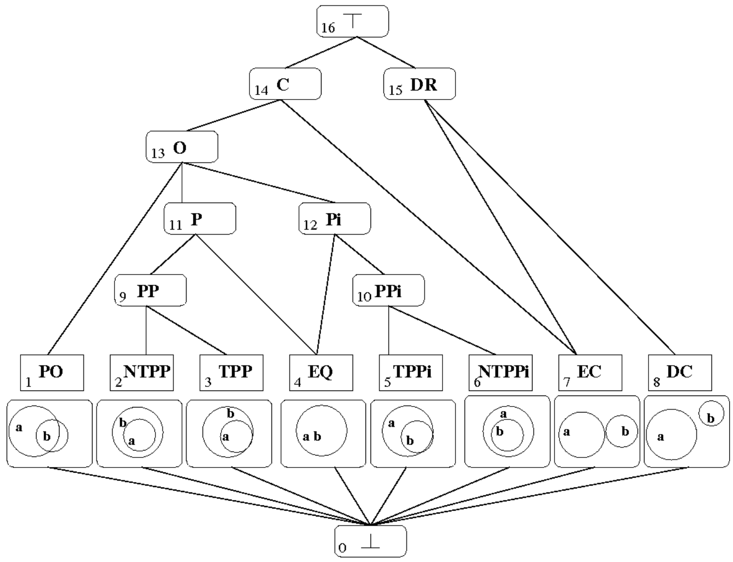

Figure 2: The lattice of spatial relations of RCC and the interpretation of RCC8

and a set of definitions on the main spatial relations (fig. 1), jointly with another set of auxiliary axioms (see [7]). It has been proved that the set of non-empty regular closed (NERC) sets of any T_3 connected space -other than a space containing only one NERC- is a model of RCC [9].

The set of binary relations formed by the eight jointly exhaustive and pairwise disjoint (JEPD) relations given in figure 2 is denoted by RCC8. The lattice structure of RCC relationships is depicted in 2. If this set is thought as a calculus for Constraint Satisfaction Problems (CSP), every set of basic relations is considered. This calculus has been deeply studied by J. Renz and B. Nebel in [16]. Other interesting calculus is RCC5, based on the set $\{DR, PO, PP, PPi, EQ\}$. Roughly speaking, the main difference between RCC5 and RCC8 is that the latter one allows one to represent knowledge that depends on topological frontiers, while the former one does not allow. The cognitive impact of this distinction on the spatial representation of a concept has to be discussed (as we will do, in fact). Nevertheless, it has been empirically constated that RCC8 is more adequate than RCC5 as a tool for representing topological relations discriminated by humans [12].

3.1 RCC as a meta-ontology

The use of RCC to visually represent the concepts turns RCC8 into an ontology on conceptual relations. The idea can be translated in different ways.

The straightforward approach consists in interpreting the concepts as regions in some model of the theory. Thus, in the *strong interpretation*, the intended meaning of $\mathtt{C}(\mathtt{x},\mathtt{y})$ is: *there exist a common element in the concepts* $\mathtt{x},\mathtt{y}$ *in some model I of the ontology*. That is, it is consistent with the information represented by the ontology that both concepts have common elements. Formally:

Definition. 3.1 (Strong Interpretation of RCC as a metaontology) *Two concepts* $\mathsf{C}_1, \mathsf{C}_2$ *of an ontology* Σ *are* ***Σ-connected*** *(denoted by* $\mathtt{C}_\Sigma(\mathsf{C}_1,\mathsf{C}_2)$*) if*

$$\Sigma \not\models \mathsf{C}_1 \sqcap \mathsf{C}_2 \equiv \bot$$

The remaining RCC relations can be interpretated by means of its corresponding definition (depicted in fig. 1). Note that the strong interpretation works on abstract spatial encodings of Σ. That is, it does not work on a concrete spatial interpretation of concepts. The following result states a logical limitation of the strong interpretation of RCC as meta-ontology. The limitation comes from the fact of the set of 1-types[2] of a first order theory is a Stone space which is compact, Hausdorff, and totally disconnected. (equipped with the topology induced by the types) [5]. The notation $S_\Sigma(F)$ stands for the set of maximal types w.r.t. the theory Σ containing the formula F.

Theorem. 3.2 *The strong interpretation does not discriminate RCC8 as ontological relations between concepts. Concretely, it has the following characterizations:*

1. $C_\Sigma(\mathsf{C}_1,\mathsf{C}_2) \Longleftrightarrow S_\Sigma(\mathsf{C}_1)\cap S_\Sigma(\mathsf{C}_2) \neq \emptyset$
2. $DC_\Sigma(\mathsf{C}_1,\mathsf{C}_2) \Longleftrightarrow S_\Sigma(\mathsf{C}_1)\cap S_\Sigma(\mathsf{C}_2) = \emptyset$
3. $P_\Sigma(\mathsf{C}_1,\mathsf{C}_2) \Longleftrightarrow S_\Sigma(\mathsf{C}_1)\subseteq S_\Sigma(\mathsf{C}_2)$
4. $PP_\Sigma(\mathsf{C}_1,\mathsf{C}_2) \Longleftrightarrow S_\Sigma(\mathsf{C}_1)\subsetneq S_\Sigma(C_2)$
5. $EQ_\Sigma(\mathsf{C}_1,\mathsf{C}_2) \Longleftrightarrow \Sigma \models C_1 \equiv C_2$.
6. $O_\Sigma(\mathsf{C}_1,\mathsf{C}_2) \Longleftrightarrow C_\Sigma(\mathsf{C}_1,\mathsf{C}_2)$
7. $PO_\Sigma(\mathsf{C}_1,\mathsf{C}_2) \Longleftrightarrow \begin{cases} S_\Sigma(\mathsf{C}_1)\cap S_\Sigma(\mathsf{C}_2)\neq\emptyset \wedge\ S_\Sigma(\mathsf{C}_1)\not\subseteq S_\Sigma(\mathsf{C}_2)\wedge \\ \wedge\, S_\Sigma(\mathsf{C}_2)\not\subseteq S_\Sigma(\mathsf{C}_1) \end{cases}$
8. $DR_\Sigma(\mathsf{C}_1,\mathsf{C}_2) \Longleftrightarrow DC_\Sigma(\mathsf{C}_1,\mathsf{C}_2)$
9. *If* $\mathsf{C}_1,\mathsf{C}_2$ *and* $R\in\{EC,TPP,NTPP,TPPi,NTPPi\}$*, then* $\neg R_\Sigma(\mathsf{C}_1,\mathsf{C}_2)$.

The result which allows us to build a visual representation of the relationships between the concepts of the ontology is based on the one of J. Renz who ensures the equivalence between logical consistency and topological consistency of these relations [16].

Theorem. 3.3 *[6] The constraint satisfaction problem associate to mereotopological relationships between concepts of an ontology Σ is spatially consistent iff Σ is consistent.*

Moreover, it is possible to obtain a spatial model on the plane formed by polygonal regions [16].

4 Paella Tool

The above result provides a means of building build the visual representation of the relationship between concepts based on RCC5. The tool Paella builds such a representation [1]. Roughly speaking, Paella is an ontology reviewer through spacial metaphors. Specifically, this tool uses a visual/topological interpretation based on the logical/mathematical properties of ontologies, where non-expert users can transform ontologies as they see, but keeping safe the formal properties of the ontology source. Although there are many tools for visual representation (for example Jambalaya), our tool also allows transformations. Therefore, this prototype represents:

[2] A 1-type in Σ is a set of one variable formulas finitely consistent with the theory Σ

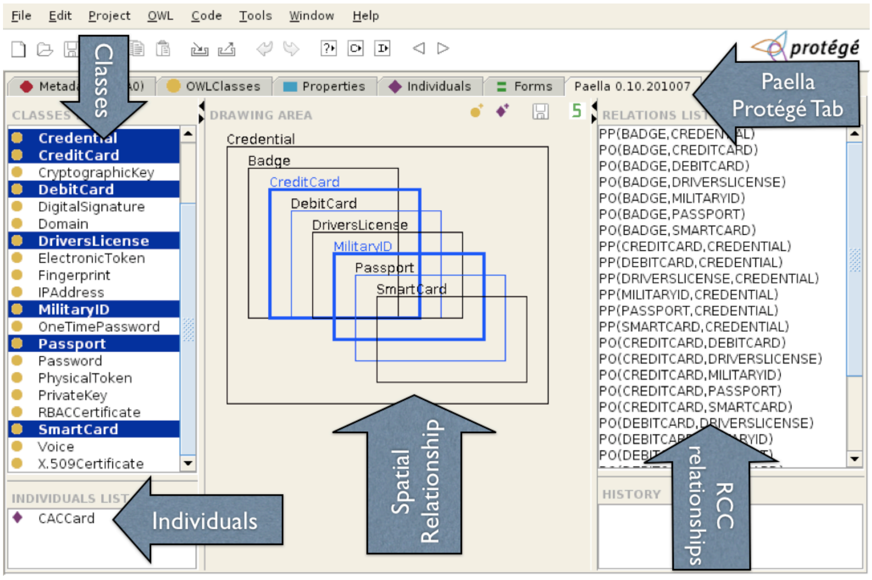

Figure 3: Screenshot of Paella

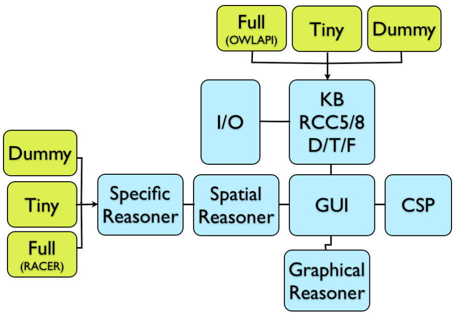

Figure 4: Architecture of Paella

- A very useful tool for socially appropriate management of formal ontologies; in the past restricted to expert users only.
- A tool to uncover hidden relationships in the ontology code between concepts, which would discover intentionally hidden relationships or even harmful to personal data, data security,etc, which are referenced to them.

Paella uses SWI-Prolog (and the CHR library for constraints reasoning) and is integrated by JAVA-SWI. The tools provide three different spatial interpretations, according to nature of data which have associated different debugging methods: Dummy Paella, Tiny Paella and Full Paella (see Fig. 4). The latter is the one used for visual analysis of OWL ontologies, and it uses RACER[3] as automated reasoning system for computing the spatial relationship between classes.

4.1 Analysis with Paella of ontologies on security

Paper [1] it describes how Paella is useful in detecting potential anomalies, in a particularly case, in three security ontologies, with different logical complexity. The selected ontologies are serviceSecurity.owl,

[3]http://www.racer-systems.com/products/tools/index.phtml

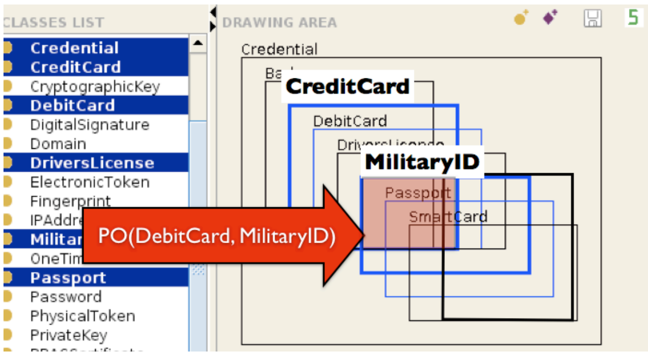

Figure 5: Representational anomaly in serviceSecurity.owl

SecurityOntology_min.owl[4] and MemoryProtection.owl[5]. Using Paella, the experiments show that there exist several representational anomalies. The graphical representation allows to visualize anomalies of type (2) and (4).

The most common anomaly is about the vague relation between critical concepts. To illustrate a problem of type (2), it is interesting to analyze a specific example in serviceSecurity.owl (part of NRL security ontology). In Fig. 5 a piece of screenshot of Paella is depicted, showing that C$reditCard$ partial overlaps to M$ilitaryID$ (under Strong Interpretation). That is -by the strong interpretation- this ontology is potentially dangerous if a population of data considers a credit card as military identification for military installations where the access is restricted. Paella also provides graphical movements to make both classes disjoint, translating this spatial configuration of ontology source (that is, it repairs the anomaly). Note that this kind of anomalies does not imply logical inconsistency, only warns of potential non-intended models of the ontology. In Fig. 5, the reparation in Paella consists in an axiom stating that the classes are disjoint.

5 Extending ontologies with uncertain concepts

To be able to handle concepts with uncertainty, it is necessary to extend the theory RCC to be able to work with the new concepts. RCC's extension, considered as extension of an ontology, presupposes certain guarantees, respecting the beginning of the methodology.

Strong interpretation works with precise regions. In both cases, the interpretation of C is a subset. The *vague interpretation* deals with the spatial interpretation of the concepts by *vague* regions. A vague region can be represented by means of two regular regions although there are other options such as egg-yolk [8], topological spaces with pulsation [6], rough sets, etc.

In order to work with vague regions, a robust extension of RCC [2, 6] is needed. Such an extension is obtained by insertion of a new mereotopological relation. Also, it requires the re-interpretation of the ontology for interpreting old and new relations. In figure 6 we present one of the seven possible robust extensions of the ontology RCC given in [2]. The interpretation is based on *pulsation*. A pulsation in a topological space $\Omega = (\mathcal{X}, \mathcal{T})$ is a map that associates to each regular set X a set $\sigma(X)$ such that its closure contains the closure of X; $\overline{X} \subseteq \overline{\sigma(X)}$. All the RCC relationships must be re-interpreted. In Fig. 6, the topological interpretation of the new relation $I(a, b)$ is $PP(a, b) \wedge EQ(\sigma(a), \sigma(b))$ (see Fig. 6, right). Note that this kind of extension extends both RCC5 and RCC8 preserving JEPD properties. The reasoning of vague regions is based on the following principle:

Third Cognitive Principle (CP3): Given a spatial interpretation I, the region $\sigma(I(C))\setminus I(C)$ represents the set of individuals with doubtful membership to C

In order to apply this principle for visual encoding, it considers the concept and its *approximate definition* in the ontology.

[4]Both from NRL ontology, `http://chacs.nrl.navy.mil/projects/4SEA/ontology.html`
[5]A SHOIN(D) ontology, `http://www.ida.liu.se/~iislab/projects/secont/`

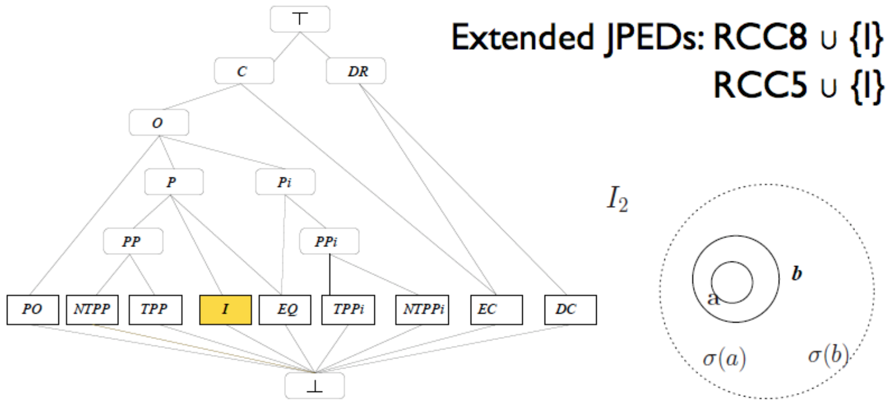

Figure 6: A *robust* ontological extension of RCC by insertion of an uncertain relation (left) and its spatial interpretation (right)

From now on, it is assumed that Σ is an unfoldable DL-ontology, that is, the left-hand sides of the axioms (defined concepts) are atomic and the right hand sides contain no direct or indirect references to defined concepts. The idea is that the properties on a concept C expressed by Σ shapes an approximation to the concept.

Definition. 5.1 *Let Σ be a DL ontology. The approximate definition according to Σ is a map σ that associates to any $\mathsf{C} \in concepts(\Sigma)$ a DL-formula as follows:*

$$\sigma(\mathsf{C}) = \begin{cases} \mathsf{C}, & \textit{if} \quad \mathsf{C} \textit{ is a defined concept} \\ \sqcap\{\mathsf{D} : \mathsf{C} \sqsubseteq \mathsf{D} \in \Sigma\}, & \textit{if} \quad \mathsf{C} \textit{ is a primitive concept} \\ \top, & \textit{if} \quad \mathsf{C} \textit{ is an atomic concept} \end{cases}$$

*Two concepts $\mathsf{C}_1, \mathsf{C}_2 \in concepts(\Sigma)$, will be said **$\Sigma$-connected** under σ, which will be denoted by $\mathsf{C}^{\sigma}_{\Sigma}(\mathsf{C}_1, \mathsf{C}_2)$, if $\mathsf{C}_{\Sigma}(\sigma(\mathsf{C}_1), \sigma(\mathsf{C}_2))$. The formula $\sigma(\mathsf{C}_1)$ will be named the associate* **notion** *to C_1 in Σ.*

The notion is defined for any concept. Nevertheless, in practice, this definition is not used intensively for atomic concepts (in the analysis of anomalies). It is that because the undefinition of the notions of an atomic concept can be deliberated: they are primitive concepts of the ontology (abstract concepts in many cases). Thus, it is not advisable to force the user to refine them. The spatial idea of Σ-connection under σ is obviously that of the topological connection of the pulsation of sets. Now, $\sigma(\mathsf{C})$ represents a DL formula associate to a concept C. Notice that the new connection is related with the previous one:

$$\mathsf{C}^{\sigma}_{\Sigma}(\mathsf{C}_1, \mathsf{C}_2) \Longleftrightarrow \Sigma \not\models \sigma(\mathsf{C}_1) \sqcap \sigma(\mathsf{C}_2) \equiv \bot$$

Therefore, given two concepts $\mathsf{C}_1, \mathsf{C}_2$ in the ontology Σ and $R \in RCC8$,

$$R^{\sigma}_{\Sigma}(\mathsf{C}_1, \mathsf{C}_2) \Longleftrightarrow R_{\Sigma}(\sigma(\mathsf{C}_1), \sigma(\mathsf{C}_2))$$

However, there is no cognitive reason to consider frontiers in vague regions, because the undefinition is represented by $\sigma(I(C)) \setminus I(C)$. Thus, RCC5 is more adequate [8] and topological extensions above mentioned are also useful in this case. Starting with C^{σ}_{Σ}, it is possible both to classify *all* the relative positions between concepts/notions and, in due course, to repair them by using spatial reasoning preserving consistency [6].

In Paella it is possible to visually encode both concept and the notion associated (by defining a new concept defined by such notion). In this way it is possible to relate concepts and notions. However, Paella current version considers the topological frontier of the region as the set of points with doubtful belonging to the concept represented by the spatial region.

References

[1] G. Aranda-Corral and J. Borrego-Díaz. Mereotopological analysis of formal concepts in security ontologies. In *CISIS 2010. To appear in Advances in Soft Computing series (Springer)*, 2010.

[2] J. Borrego-Díaz and A. M. Chávez-González. Extension of ontologies assisted by automated reasoning systems. In R. Moreno-Díaz, Franz Pichler, and A. Quesada-Arencibia, editors, *EUROCAST 2005*, volume 3643 of *Lecture Notes in Computer Science*, pages 247–253. Springer, 2005.

[3] J. Borrego-Díaz and A. M. Chávez-González. Visual ontology cleaning: Cognitive principles and applicability. In Y. Sure and J. Domingue, editors, *ESWC*, volume 4011 of *Lecture Notes in Computer Science*, pages 317–331. Springer, 2006.

[4] J. Borrego-Díaz and A. M. Chávez-González. Using cognitive entropy to manage uncertain concepts in formal ontologies. In P. Cesar G. da Costa, Claudia d'Amato, Nicola Fanizzi, Kathryn B. Laskey, Kenneth J. Laskey, Thomas Lukasiewicz, Matthias Nickles, and Michael Pool, editors, *URSW*, volume 5327 of *Lecture Notes in Computer Science*, pages 315–329. Springer, 2008.

[5] C. C. Chang and H. J. Keisler. *Model Theory*. North-Holland, Amsterdam, 1973.

[6] A. M. Chávez-González. *Automated Mereotopological Reasoning for Ontology Debugging*. PhD thesis, University of Seville, Spain, 2005.

[7] A G Cohn, B Bennett, J Gooday, and N Gotts. Representing and reasoning with qualitative spatial relations about regions. In O Stock, editor, *Temporal and spatial reasoning*. Kluwer, 1997.

[8] A G Cohn and N M Gotts. A theory of spatial regions with indeterminate boundaries. In C Eschenbach, C Habel, and B Smith, editors, *Topological Foundations of Cognitive Science*, 1994.

[9] N M Gotts. An axiomatic approach to topology for spatial information systems. Technical report, Report 96.25, School of Computer Studies, University of Leeds, 1996.

[10] M. Grohe and L. Segoufin. On first-order topological queries. *ACM Trans. Comput. Logic*, 3(3):336–358, 2002.

[11] R. Grütter, Th. Scharrenbach, and B. Bauer-Messmer. Improving an rcc-derived geospatial approximation by owl axioms. In A. P. Sheth, S. Staab, M. Dean, M. Paolucci, D. Maynard, T. W. Finin, and K. Thirunarayan, editors, *International Semantic Web Conference*, volume 5318 of *Lecture Notes in Computer Science*, pages 293–306. Springer, 2008.

[12] M. Knauff, R. Rauh, and J. Renz. A cognitive assessment of topological spatial relations: Results from an empirical investigation. In S. C. Hirtle and Andrew U. Frank, editors, *COSIT*, volume 1329 of *Lecture Notes in Computer Science*, pages 193–206. Springer, 1997.

[13] H. Mizen, C. Dolbear, and G. Hart. Ontology ontogeny: Understanding how an ontology is created and developed. In M. Andrea Rodríguez, I. F. Cruz, M. J. Egenhofer, and S. Levashkin, editors, *GeoS*, volume 3799 of *Lecture Notes in Computer Science*, pages 15–29. Springer, 2005.

[14] F. Murtagh, T. Taskaya, P. Contreras, J. Mothe, and K. Englmeier. Interactive visual user interfaces: A survey. *Artif. Intell. Rev.*, 19(4):263–283, 2003.

[15] C. H. Papadimitriou, D. Suciu, and V. Vianu. Topological queries in spatial databases. *J. Comput. Syst. Sci.*, 58(1):29–53, 1999.

[16] J. Renz. *Qualitative spatial reasoning with topological information*. Springer-Verlag, Berlin, Heidelberg, 2002.