## TOWARDS A TOOL FOR ONTOLOGY ENGINEERING[1]

J.A.ALONSO-JIMÉNEZ, J.BORREGO-DÍAZ, A.CHÁVEZ-GONZÁLEZ
Dept. of Computer Science and Artificial Intelligence-University of Sevilla
{jalonso, jborrego, tchavez}@us.es

## ABSTRACT

A tool based on a spatial representation of provisional ontologies is designed. The tool allows the cleaning of Knowledge Bases, as well to induce new concepts in early steps of the building of an ontology.

**KEYWORDS:** Qualitative-Reasoning, Knowledge-Based Systems, Ontology Engineering

## 1 INTRODUCTION

Ontologies provide us with a common understanding in fields as Knowledge Management and e-commerce [1]. For a satisfactory transition from the actual WWW to the Semantic Web it will be need to deal with evolving ontologies, because of the key role they play in the reasoning services for Knowledge Bases (KB) in the Semantic Web [2] [1].

The aim of this paper is to show the foundational issues for a tool to repair KB in provisional ontologies. It is based on a spatial representation of incomplete specifications of the concepts of the ontology using RCC [3], a sound theory for qualitative spatial reasoning. Then we use two types of actions, topological and reticular arrangements on the spatial representation, in order to repair some anomalies. Additional steps require the interaction with the beliefs of the user. This work is based on the analysis given in [4] on RCC knowledge databases. We aim to analyze a certain type of anomalies that arose from a *cleaning-cycle* applied to KB's associated to complex ontologies [6].

The advantage of using this semantics lies in the fact that it is necessary to improve the current data cleaning systems [7]: with a clear separation between the logical specification of data transformation and their physical transformation, an explanation of the reasoning behind cleaning results, and the possibility of interactive facilities to tune data cleaning programs.

Description Logic (DL) is used to represent metadata. DL is a sound formalism to give a clear semantics to several tools for Knowledge Representation (KR) (see.e.g. [2] [5]).

Formally, Description Logics (`http://dl.kr.org`) is a subset of first order logic. Thus it inherits a formalized semantics, as opposite to some early formalisms for KR. DL deals with the representation of structured concepts, describing them with a language with specific features, as conjunction, quantifiers on attributes of concepts, etc.

A KB, $\Sigma$, in DL is a pair $(\mathcal{T}, \mathcal{A})$, where $\mathcal{A}$ is a set of facts (the *extensional* component) and $\mathcal{T}$ is a set of relations between concepts (the *intensional* component). In fig. 1, a little KB on the family ontology is given, which will be our running example.

## 2 ANOMALIES IN PROVISIONAL ONTOLOGIES

Knowledge Bases in DL may be affected by classical anomalies. We must bear in mind

[1]Work partially supported by the MCyT project TIC 2000-1368-C03-0 and the project TIC-137 of the *Plan Andaluz de Investigación*.

$$\mathcal{T} = \begin{cases} \text{Woman} \sqsubseteq \text{Person} \sqcap \text{Female} \\ \text{Man} \sqsubseteq \text{Person} \sqcap \neg\text{Woman} \\ \text{Father} \sqsubseteq \text{Man} \sqcap \exists\text{hasChild.Person} \\ \text{Father} \sqsubseteq \text{Parent} \end{cases} \quad \mathcal{A} = \begin{cases} \text{Father(John)} & \text{Man(John)} \\ \text{Female(RuPaul)} & \text{Man(RuPaul)} \\ \text{Woman(Ann)} & \exists\text{hasChild.Person(Ann)} \end{cases}$$

Figure 1: *A provisional KB, $\Sigma = \langle \mathcal{T}, \mathcal{A} \rangle$, on the family ontology*

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

Figure 2: *Axioms of RCC*

the possible dynamic nature of ontologies, and in the first phases of their building they must be considered as *provisional*. Even if an ontology lacked of unacceptable anomalies, there would be several versions we must work on. Incompleteness of a KB must be understood in two ways: the logical incompleteness ( with respect to a kind of queries), and due to the lack of concepts or roles (incompleteness with expressive nature).

There are anomalies due to lack of an exact profile of several concepts. When it occurs, the user works on beliefs not even explicited in the KB. Such concepts will be called *notions*. The existence of notions in an ontology implies that two concepts covered by the same notion can not be distinguished, namely the ontology is *coarse*.

On the other hand, neglected development of the ontology leads to a problem in the validation field, different from classical validation task in Knowledge Based Systems: the ontology does not fit in with the user beliefs about his/her framework, or it is both hard to use and to be understood by the others. Messy ontologies definitively are a risk for the management of large KB in the Semantic Web.

## 3 SPATIAL REPRESENTATION OF ONTOLOGIES

The Region Connection Calculus (RCC) [3] is a topological approach to qualitative spatial representation and reasoning on *spatial entities*, which are non-empty regular sets. The basic relation between them is the connection relation $C(x, y)$, which is interpreted as *"the closures of* x *and* y *intersect"*. The axioms of RCC are two basic axioms on $C$, $A_1 := \forall x[C(x, x)]$ and $A_2 := \forall x, y[C(x, y) \to C(y, x)]$, plus several axioms/definitions on the main spatial relationships, see fig. 2. The eight jointly exhaustive and pairwise disjoint relations forming the relational calculus RCC-8, have been deeply studied by J. Renz and B. Nebel [8].

We will use later two kinds of *motions* on this relational calculus: *reticular motions* and *topological motions*. They are cognitively adequate motions and have *continuous* nature. Reticular motions are refinements of relationships (downward motions in the

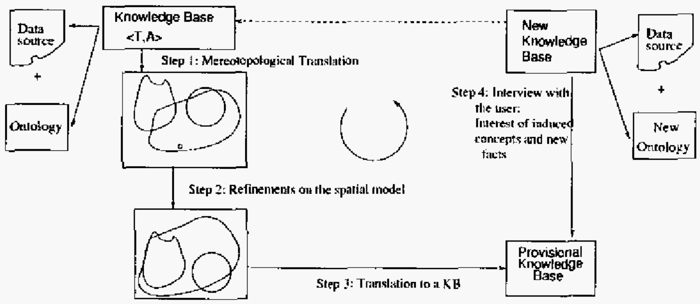


Figure 3: *The ontology transformation process*

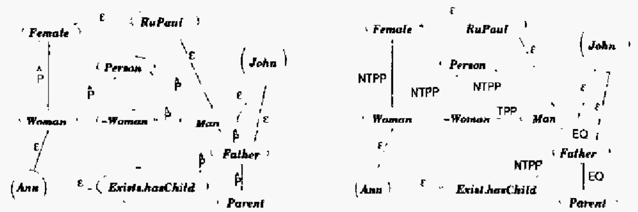

Figure 4: *Initial graph (left) and solution (right)*

lattice of the RCC relationships). Topological motions are motions of *least topological distance*, as the substitution of a relation by other one *cognitively near*. It will be used too the *reticular projection* on RCC-8, $R \mapsto \hat{R} := \{R' \in RCC8 \ : \ R' \subseteq R\}$.

The initial KB is a DL knowledge base. The cleaning process is shown in figure 3.

The result of the process will be a new KB consistent with the beliefs of the user. Indeed, the process must be a cycle, because it is possible that the ontology changes, new data have been induced, and they lead to a new revision.

## 4 FIRST STEP: GRAPHICAL INTERPRETATION

Firstly, a constraint satisfaction problem (CSP) on the spatial relational calculus RCC8 (or RCC5) is produced by a cognitively sound translation of the TBox to RCC formulas, obtaining a consistent scenario, which is represented in 2D. Facts of the Abox are added as points. Each elementary concept A of $\Sigma$ is interpreted as a region in the plane, $\mathsf{A}^I \subseteq \mathbb{R}^2$ (we will write $\mathsf{A}^I = A$). In order to carry out this interpretation, a translation of $\Sigma$ to RCC8 is applied, translating the formulas of TBox to a set of RCC formulae as follows: $(\mathsf{C} \sqsubseteq \mathsf{D})^* = \{P(C,D)\}$, $(\mathsf{C} \sqsubseteq \mathsf{D}_1 \sqcap \mathsf{D}_2)^* = \{P(C,D_1), P(C,D_2)\}$, $(\mathsf{C} \sqsubseteq \mathsf{D}_1 \sqcup \mathsf{D}_2)^* = \{O(C,D_1), O(C,D_2)\}$. (Notice that in this case, if $C \cap D_1 = \emptyset$, the user will discard $O(C,D_1)$ later). Each fact of the Abox is translated as $\mathsf{A}(\mathsf{a}) \mapsto a \in A$. For our example, the graph of constraints is given in fig. 4 (left), if we make reticular projection on RCC8. A solution is on the right in fig. 4.

The consistent scenario [9] is spatially represented by a set of (not necessarily connected) regular regions (see fig 5 left).

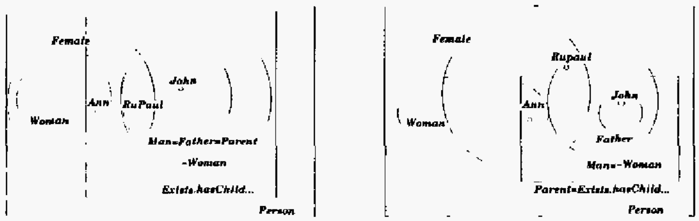


Figure 5: *Spatial representation of the solution (left) and after the arrangements (right)*

| | Female | Woman | Man | Father | Parent | Person | ¬ Woman | ∃hasChild.P |
|---|---|---|---|---|---|---|---|---|
| Female | EQ | PO | PO | DC | PO | NTPP | PO | PO |
| Woman | PO | EQ | DC | DC | PO | NTPP | EC | PO |
| Man | PO | DC | EQ | NTPPi | PO | NTPP | EQ | PO |
| Father | DC | DC | NTPP | EQ | NTPP | NTPP | NTPP | NTPP |
| Parent | PO | PO | PO | NTPPi | EQ | NTPP | PO | EQ |
| Person (P) | NTPPi | NTPPi | NTPPi | NTPPi | NTPPi | EQ | NTPPi | NTPPi |
| ¬ Woman | PO | EC | EQ | NTPPi | PO | NTPP | EQ | PO |
| ∃hasChild.P | PO | PO | PO | NTPPi | EQ | NTPP | PO | EQ |

Figure 6: *Spatial relationships among the concepts*

## 5 SECOND STEP: SPATIAL ARRANGEMENTS

In this step the user is requested to make reticular and/or topological arrangements on the graphical representation. By introduction of new regions, new concepts might be introduced too. For our example, the new picture is on the right of fig. 5.

## 6 THIRD STEP: A NEW KB

When the user believes that the current spatial scenario is a sound representation, we must translate it to a new KB. It is necessary to make some remarks on the spatial scenario.

The spatial relationships may be inadequate with respect to the *mental* ontology believed by the user. This anomaly is detected when the user refuses the translated KB, producing a new graphical refinement. Actually, the above translation must be applied to a representation of the map by a graph that the user thinks as fair. Next we define the translation $R \in RCC \mapsto R^*$ of each relation on RCC to a set of DL formulae by recursion in the order of the RCC axiomatization (fig. 2). The relation $a \in A$ is translated to $\mathsf{A}(\mathsf{a})$. From now, "element" means "spatial interpretation of a constant symbol". One selected translation of rule are: $PP(A,B)^* = P(A,B)^*$ if it exists a region $D$ such that $C(B,D) \wedge \neg C(A,D)$. In other case, a new concept name is introduced $\mathsf{N}_{\mathsf{B}\setminus\mathsf{A}}$, and $PP(A,B)^* = P(A,B)^* \cup \{\mathsf{N}_{\mathsf{B}\setminus\mathsf{A}} \sqsubseteq \mathsf{B}, \mathsf{N}_{\mathsf{B}\setminus\mathsf{A}} \sqsubseteq \neg\mathsf{A}\}$. The names for the new concepts does not refer to any intended feature of the concept. For example, it is not initially true that $\mathsf{N}_{\mathsf{B}\setminus\mathsf{A}} \equiv B \setminus A$. For example, the table of relations is in fig. 6 and the KB obtained is shown in 7.

## 7 FOURTH STEP: EVALUATION BY THE USER

One of the goals of this step is to give a name for new concepts. This implies that the user must make an effort for interpreting the result. Moreover, the user must decide if

$$
\mathcal{T}' = \left\{ \begin{array}{l}
\mathsf{Father} \sqsubseteq \mathsf{Person} \sqcap \mathsf{Parent} \sqcap \mathsf{Man} \\
\mathsf{Father} \sqsubseteq \neg\mathsf{Woman} \sqcap \neg\mathsf{Female} \\
\mathsf{Female} \sqsubseteq \mathsf{Person} \\
\mathsf{Man} \equiv \neg\mathsf{Woman} \\
\mathsf{Woman} \sqsubseteq \mathsf{Person} \\
\mathsf{Man} \sqsubseteq \mathsf{Person} \\
\mathsf{Parent} \sqsubseteq \mathsf{Person} \\
\exists\mathsf{hasChild.Person} \equiv \mathsf{Parent} \\
\mathsf{N}_{\mathsf{Woman}\sqcap\mathsf{Female}} \sqsubseteq \mathsf{Woman} \sqcap \mathsf{Female} \\
\mathsf{N}_{\mathsf{Woman}\backslash\mathsf{Female}} \sqsubseteq \mathsf{Woman} \sqcap \neg\mathsf{Female} \\
\mathsf{N}_{\mathsf{Female}\backslash\mathsf{Woman}} \sqsubseteq \mathsf{Female} \sqcap \neg\mathsf{Woman} \\
\mathsf{N}_{\mathsf{Parent}\sqcap\mathsf{Woman}} \sqsubseteq \mathsf{Parent} \sqcap \mathsf{Woman} \\
\mathsf{N}_{\mathsf{Parent}\backslash\mathsf{Woman}} \sqsubseteq \mathsf{Parent} \sqcap \neg\mathsf{Woman} \\
\mathsf{N}_{\mathsf{Woman}\backslash\mathsf{Parent}} \sqsubseteq \mathsf{Woman} \sqcap \neg\mathsf{Parent} \\
\mathsf{N}_{\mathsf{Man}\sqcap\mathsf{Female}} \sqsubseteq \mathsf{Man} \sqcap \mathsf{Female} \\
\mathsf{N}_{\mathsf{Man}\backslash\mathsf{Female}} \sqsubseteq \mathsf{Man} \sqcap \neg\mathsf{Female} \\
\mathsf{N}_{\mathsf{Female}\backslash\mathsf{Man}} \sqsubseteq \mathsf{Female} \sqcap \neg\mathsf{Man} \\
\mathsf{N}_{\mathsf{Parent}\sqcap\mathsf{Female}} \sqsubseteq \mathsf{Parent} \sqcap \mathsf{Female} \\
\mathsf{N}_{\mathsf{Parent}\backslash\mathsf{Female}} \sqsubseteq \mathsf{Parent} \sqcap \neg\mathsf{Female} \\
\mathsf{N}_{\mathsf{Female}\backslash\mathsf{Parent}} \sqsubseteq \mathsf{Female} \sqcap \neg\mathsf{Parent} \\
\mathsf{N}_{\mathsf{Parent}\backslash\mathsf{Man}} \sqsubseteq \mathsf{Parent} \sqcap \neg\mathsf{Man} \\
\mathsf{N}_{\mathsf{Man}\backslash\mathsf{Parent}} \sqsubseteq \mathsf{Man} \sqcap \neg\mathsf{Parent}
\end{array} \right.
\qquad
\mathcal{A}' = \left\{ \begin{array}{l}
\mathsf{Female(Rupaul)},\ \mathsf{Man(Rupaul)} \\
\mathsf{Father(John)},\ \mathsf{Man(John)} \\
\mathsf{Parent(John)},\ \mathsf{Woman(Ann)} \\
\exists\mathsf{hasChild.Person(Ann)},\ \mathsf{Female(Ann)} \\
\mathsf{Woman}(c_{\mathsf{Woman},\neg\mathsf{Woman}}) \\
\neg\mathsf{Woman}(c_{\mathsf{Woman},\neg\mathsf{Woman}}) \\
\mathsf{N}_{\mathsf{Man}\sqcap\mathsf{Female}}(\mathsf{Rupaul}) \\
\mathsf{Ab}_{\mathsf{Woman},\neg\mathsf{Woman}}(c_{\mathsf{Woman},\neg\mathsf{Woman}})
\end{array} \right.
$$

$$
\mathcal{T}'' = \left\{ \begin{array}{l}
\mathsf{Father} \sqsubseteq \mathsf{Parent} \sqcap \mathsf{Man} \sqcap \neg\mathsf{Female} \\
\mathsf{Female} \sqsubseteq \mathsf{Person} \\
\mathsf{Woman} \sqsubseteq \mathsf{Person} \\
\mathsf{Man} \sqsubseteq \mathsf{Person} \\
\mathsf{Parent} \sqsubseteq \mathsf{Person} \\
\mathsf{Man} \sqsubseteq \neg\mathsf{Woman} \\
\exists\mathsf{hasChild.Person} \equiv \mathsf{Parent} \\
\mathsf{Crossdresser} \sqsubseteq \mathsf{Female} \\
\mathsf{Crossdresser} \sqsubseteq \neg\mathsf{Woman} \\
\mathsf{Mother} \sqsubseteq \mathsf{Parent} \sqcap \mathsf{Woman}
\end{array} \right.
\qquad
\mathcal{A}'' = \mathcal{A}' \cup \left\{ \begin{array}{l}
\mathsf{Mother(Ann)} \\
\mathsf{Crossdresser(Rupaul)}
\end{array} \right.
$$

Figure 7: *KB from spatial representation and KB after the last step*

the elements belong to the new concepts, if they are topologically close. For example, the user must decide if $Rupaul \in N_{Female\backslash Woman}$.

It is possible that a new concept had been discarded. It might occur if the graphic representation the user has done became inadequate for her/his beliefs.

In our case, the user gives a name to the expression $\mathsf{N}_{\mathsf{Female}\backslash\mathsf{Woman}}$ that is $\mathsf{Crossdresser}$; the expression $\mathsf{N}_{\mathsf{Parent}\backslash\mathsf{Father}}$ is named $\mathsf{Mother}$, and it is identified with $\mathsf{N}_{\exists\mathsf{hasChild.Person}\backslash\mathsf{Father}}$. The final KB is on figure 7.

If the user had believed that $DR(Female, Man)$, the translation would produce the concept name $\mathsf{Ab}_{\mathsf{Female},\mathsf{Man}}$ and the fact $\mathsf{Ab}_{\mathsf{Female},\mathsf{Man}}(\mathsf{RuPaul})$ would be included into the Abox, but no relation between $\mathsf{Ab}_{\mathsf{Female},\mathsf{Man}}$ and $\mathsf{Man}$ or $\mathsf{Female}$ is added. It is preferable to be still a notion.

## 8 CONCLUSIONS, RELATED AND FUTURE WORK

We showed how to translate, with plausibility cognitive, the analysis of KB to a graphical refinement with a sound qualitative spatial reasoning tool, RCC.

The representational principle on we work is that an acceptable small set of concepts must have a clear spatial representation. In other case, this set should be messy. This hypothesis is argued by the practice.

The tool is not for reasoning service based on entailment. Nonmonotonic reasoning (as default) is used in several steps of the cleaning process. The tool is more related with the graphical representation of ontologies and the mereological analysis of concepts given in [10]. In fact, our representation satisfies, in a non temporal setting, the minimal requirements proposed in the cited paper.

There are related works on reasoning about concepts such as Galois lattices. In [11], information on visual tools to represent concepts lattices are given, but their aim is not specifically to transform the ontology because it is supported by real data. On the other hand, the method allows us to use logically consistent reasoning for repairing ontologies.

There exist two research lines we are currently studying. First, although in this paper we do not deal with a spatial representation of the roles of KB, this feature can be added to the tool. On the other hand, the tool may be considered as a learning process (based on generalization and refinement of concepts). Thus it seems interesting to design metrics reflecting such a process by means of adapting techniques for clausal learning [12].

## References


[1] D. Fensel: Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce. Springer-Verlag (2000).

[2] I. Horrocks, D.L. McGuinnes, C. A. Welty. Digital Libraries and Web-Based Information Systems. In F. Baader et al. (ed.) The Description Logic Handbook. pp. 436-459. Cambridge University Press (2003).

[3] A. G. Cohn, B. Bennett, J. M. Gooday and N. M. Gotts. Representing and Reasoning with Qualitative Spatial Relations about Regions. In O. Stock (ed.) Temporal and Spatial Reasoning, pp 97-134. Kluwer (1997).

[4] J.A. Alonso-Jiménez, J. Borrego-Díaz, Antonia M. Chávez-González, M. A. Gutiérrez-Naranjo and Jorge D. Navarro-Marín. Towards a Practical Argumentative Reasoning with Qualitative Spatial Databases. Developments in Applied Artificial Intelligence, (IEA/AIE 2003) (LNAI 2718) Springer-Verlag. pp.789–798 (2003).

[5] G. Antoniou and F. van Harmelen. Web Ontology Language: OWL. In S. Staab and R. Studer (eds.) Handbook on Ontologies in Information Systems, pp.67-92, Springer-Verlag (2003).roc. of 27th Very Large Databases Conference VLDB. Morgan Kauffman (2001).

[6] J.A. Alonso-Jiménez, J. Borrego-Díaz, Antonia M. Chávez-González, M. A. Gutiérrez-Naranjo and Jorge D. Navarro-Marín. A Methodology for the Computer–Aided Cleaning of Complex Knowledge Databases. Proc. of 29th Annual Conference of the IEEE Industrial Electronics Society, IECON 2002. IEEE Press. pp. 1806–1811 (2002).

[7] H. Galhardas, D. Florescu, D. Sasha, E. Simon, C.A. Saita. Declarative Data Cleaning: Language, Model and Algorithms. Proc. of 27th Very Large Databases Conference VLDB. Morgan Kauffman (2001).

[8] J. Renz, B. Nebel. On the Complexity of Qualitative Spatial Reasoning: A Maximal Tractable Fragment of the Region Connection Calculus. Artificial Intelligence 108(1-2): 69-123 (1999).

[9] J. Renz. A Canonical Model of the Region Connection Calculus. Proc. of the 6th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'98) (1998).

[10] N. Guarino and C Welty. Towards a Methodology for Ontology-Based Model Engineering. ECOOP-2000 Workshop on Model Engineering (2000).

[11] B. Ganter, R. Wille. Formal Concept Analysis-Mathematical Foundations. Springer (1999).

[12] M.A. Gutiérrez-Naranjo, J. A. Alonso-Jiménez and J. Borrego-Díaz. A Quasi-metric for Machine Learning. Advances in Artificial Intelligence-IBERAMIA 2002. Lecture Notes in Artificial Intelligence 2527, pp. 193–203 (2002).