Alignment Patterns based on Unified Foundational Ontology

Natalia F. Padilha[1], Fernanda Baião[1], Kate Revoredo[1]

[1]Federal University of the State of Rio de Janeiro (UNIRIO)
– Rio de Janeiro – RJ – Brazil

{natalia.padilha,fernanda.baiao, katerevoredo}@uniriotec.br

Abstract. *Ontology alignment is the process of finding related entities in different ontologies. In this context, precise and explicit representation of conceptualizations is essential for reaching semantic integration to ensure that only data related to the same (or sufficiently similar) real-world entity are merged. Foundational ontologies describe general concepts independent of a domain and precisely define meta-properties so as to make the semantics of each concept in the ontology explicit. In this paper we show how the use of OntoUML, a conceptual modeling language based on Unified Foundational Ontology, allows the application of alignment patterns and exemplify how this approach may improve precision, recall and refine the type of the alignment.*

1. Introduction

Ontologies are explicit specifications of a conceptualization (Gruber 1995). Many domain ontologies have been developed in recent years and linking conceptualizations covering an area of common or related knowledge is a recent research problem that motivated the development of several techniques for aligning ontologies.

Ontology alignment is the process of finding related entities in different ontologies. Euzenat (2007) presents a classification of elementary alignment techniques based on the kind of data input the algorithms work on: strings (terminological), structure (structural), models (semantics) or data instances (extensional). The first two are found in the ontology descriptions. The third one requires some semantic interpretation of the ontology. The last one constitutes the actual population of an ontology.

The most difficult integration problems are caused by semantic heterogeneity (Ziegler and Dittrich 2007). Semantic integration has to ensure that only data related to the same (or sufficiently similar) real-world entity is merged. This requirement is still a challenge in the process of ontology alignment since most of the techniques discussed and implemented in automated tools so far are based on terminological or structural analyses.

On the other hand, foundational ontologies describe general concepts independent of a domain and if the domain ontologies specialize the terms introduced in a foundational ontology (Guarino 1998), it may be used as external source of common knowledge for exploiting the semantics. However, despite the benefits for building conceptual models of a domain, foundational ontologies are still insufficiently explored in the ontology alignment literature.

An essential issue for reaching semantic integration is the precision of an explicit conceptualization representation. Guizzardi (2005) addresses this issue as ontological adequacy, defined as a measure of how close a model is to the situation in reality it represents. The author presents OntoUML, a modeling language that considers the ontological distinctions and axiomatic theories put forth by the Unified Foundational Ontology (UFO) he proposes.

In this paper we show how the use of OntoUML, based on some design patterns explored in Guizzardi et al. (2011), allows the application of some alignment patterns to improve semantic integration.

Another contribution of this paper is a review in the classification of alignment approaches proposed by Euzenat (2007) concerning the technique called "Upper level, domain specific ontologies" to better organize the works that address foundational ontologies in the process of ontology alignment.

This paper is structured as follows. Section 2 presents some OntoUML design patterns that explore constraints underlying UFO. In section 3 we discuss the ontology alignment process and present a review in the classification of ontology alignment approaches. Section 4 introduces the alignment patterns based on OntoUML design patterns. In section 5 we exemplify the application of these alignment patterns. Section 6 reviews related works, followed by the conclusions in the section 7.

2. Unified Foundational Ontology

Foundational ontologies (also called upper-level or top-level ontologies) describe very general concepts, which are independent of a particular problem or domain (Guarino 1998).

UFO is one example of foundational ontology that has been developed based on a number of theories from Formal Ontology, Philosophical Logics, Philosophy of Language, Linguistics and Cognitive Psychology (Guizzardi 2005). It is composed by three main parts. UFO-A is an ontology of endurants (objects). UFO-B is an ontology of perdurants (events, processes). UFO-C is an ontology of social entities (both endurants and perdurants) built on the top of UFO-A and UFO-B.

OntoUML is a conceptual modeling language designed to comply with the ontological distinctions and axiomatic theories put forth by UFO that results from a redesign process of the Unified Modeling Language (UML). The OntoUML classes, for example, make explicit the distinctions between an object and a process, types of things from their roles, among others.

A fundamental distinction in UFO is between particulars and universals. Particulars are entities that exist in reality possessing a unique identity, while universals are patterns of features, which can be realized in a number of different particulars.

UML class diagrams are intended to represent the static structure of a domain, in which classes typically represent endurant universals. The UML profile proposed by Guizzardi (2005) is a finer-grained distinction between different types of classes that represent each of the leaf ontological categories (gray entities in figure 1) specializing substantial universal types of UFO-A.

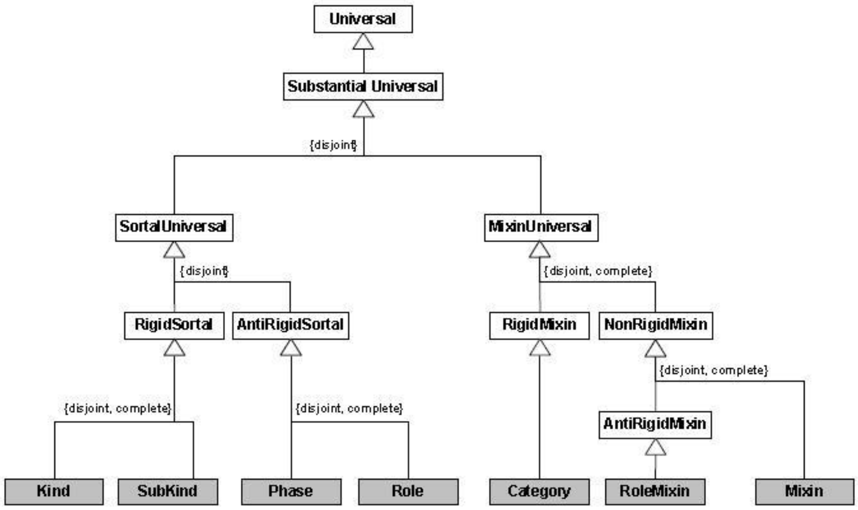

Figure 1. Ontological Distinctions in a Typology of Substantial Universals (Guizzardi 2005)

Substantials are entities that persist in time while keeping their identity (as opposed to events such as a business process or a birthday party). Constructs that represent Sortal Universals can provide a principle of identity and individuation for its instances. Mixin Universal is an abstract metaclass that represents the general properties of all mixins, i.e., non-sortals (or dispersive universals). A type is rigid iff for every instance of that type, it is necessarily an instance of that type. In contrast, a type is anti-rigid iff for every instance of the type, there is always a possible world in which it is not an instance of this type.

A kind (and subkinds) represent rigid sortals that applies necessarily to its instances, i.e., in every possible world (such as a Person, Man or Woman). A phase represents an anti-rigid sortal instantiated in a specific world or time period, but not necessarily in all of them (such as Child, Adolescent and Adult phases of a Person). A role defines an anti-rigid sortal which may be assumed in a world, but not necessarily in all possible worlds (such as a Student or a Professor role played by a Person), but once it is, this depends on its participation in a specific relation or event. Due to space restrictions, we will not define all other OntoUML categories. The design patterns presented in the next section are limited to these primitives: kind/subkind, phases and roles.

2.1. Design Patterns

The design patterns presented in this section were explored by Guizzardi et al. (2011) and are derived from the ontological foundations of OntoUML.

Subkinds can be manifested as a part of a generalization set which has as a common superclass a Kind S. In this case, the subkind classes are disjoint and complete. The Subkind Design Pattern is illustrated in figure 2(a).

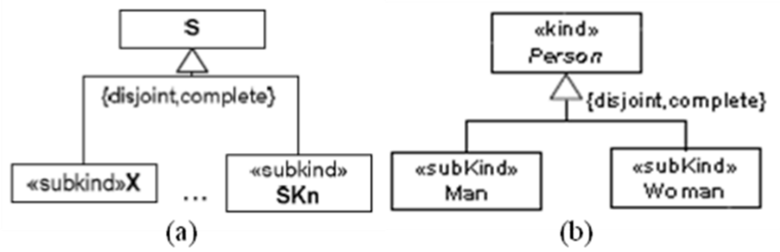

Figure 2. The Subkind Design Pattern (a) and an example of use (b) (Guizzardi et al. 2011)

Phases are always manifested as part of a phase partition (PP). In a PP there is always one unique root common supertype which is necessarily a Kind S. As well as subkinds, phases are manifested as a part of a generalization set of type S. The Phase Design Pattern is illustrated in figure 3(a).

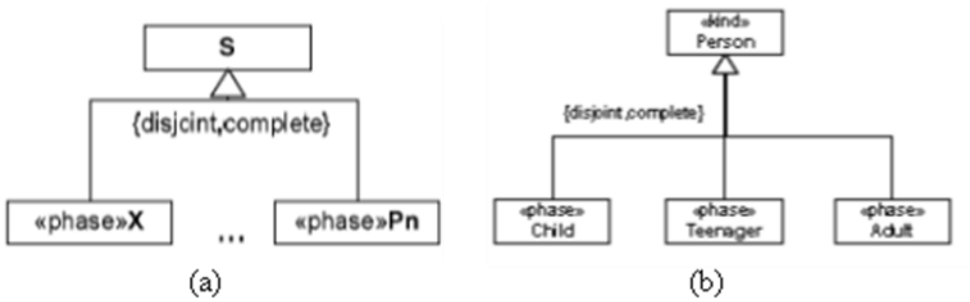

Figure 3. The Phase Design Pattern (a) and an example of use (b) (Guizzardi et al. 2011)

Roles represent (possibly successive) specializations of a Kind S by using a relational specialization condition R with another type T of the model. The Role Design Pattern is illustrated in figure 4(a).

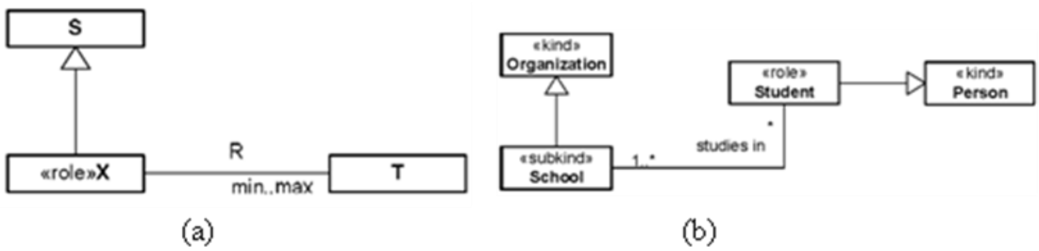

Figure 4. The Role Design Pattern (a) and an example of its use (b). Source: (Guizzardi et al. 2011)

3. Ontology Alignment

Ontology alignment is the process of finding corresponding entities (concept, relation, or instance) in two ontologies describing the same domain. A general ontology alignment function based on the vocabulary, E, of all terms e Є E, based on the set of possible ontologies, O, and based on possible alignment relations, M, is a partial function: align: E *x* O *x* O → E *x* M. Apart from one-to-one equality alignments, mostly investigated in existing work, one entity often has to be aligned not only to equal entities, but based on another relation (e.g., subsumption). Further, there are complex composites such as a concatenation of terms (e.g., name equals first plus last name) (Ehrig 2007).

Precision and recall are commonplace measures in information retrieval and are also applied to evaluate alignment results. Precision measures the correctness of the method by the ratio of correctly found correspondences over the total number of returned correspondences. Recall is a completeness measure and considers the ratio of correctly found correspondences over the total number of expected correspondences.

3.1. Classification of ontology alignment approaches

The classification of Euzenat (2007), reproduced in figure 5, if read from the bottom up, focuses on how the techniques interpret the input information. Element-level alignment techniques compute correspondences by analyzing entities in isolation, ignoring their relations with other entities. Structure-level techniques compute correspondences by analyzing how entities appear together in a structure. Syntactic techniques interpret the input with regard to its sole structure following some clearly stated algorithm. External techniques exploit auxiliary (external) resources of a domain and common knowledge in order to interpret the input. Semantic techniques use some formal semantics to interpret the input.

If the classification is read in ascending it focus on the kinds of manipulated objects: strings (terminological), structure (structural), models (semantics) or data instances (extensional).

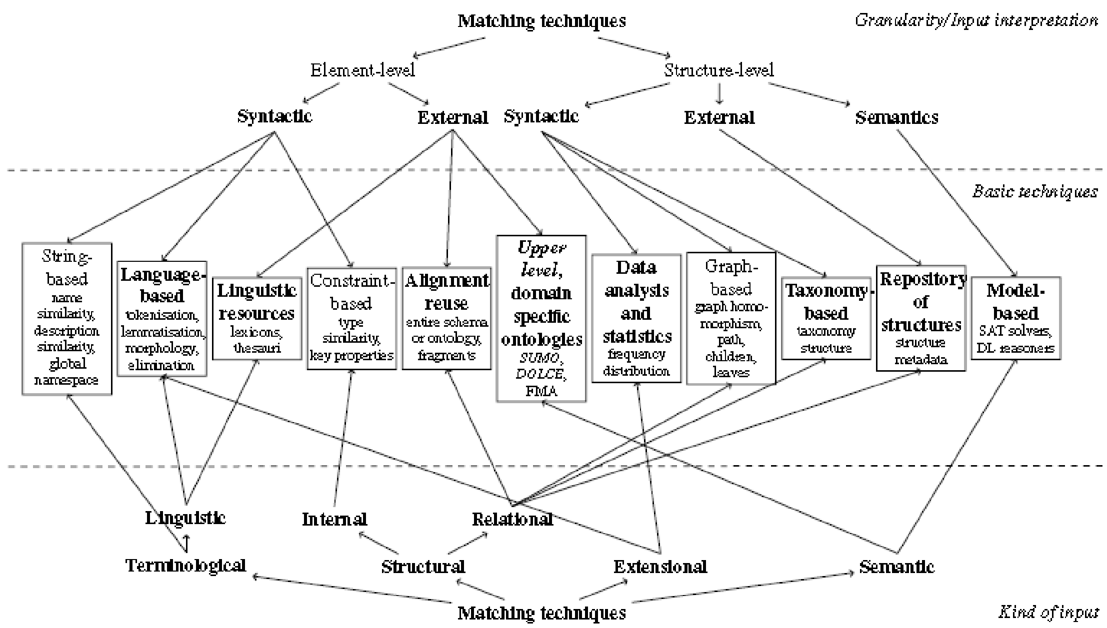

Figure 5. The retained classifications of elementary matching approaches (Euzenat 2007)

The approach proposed in this paper fits the classification described as "Upper level, domain specific ontologies", which is an element-level technique based on external semantic input. This classification groups approaches based on domain specific ontologies used as external sources of background knowledge of the particular domain being aligned and those ones that actually exploit foundational ontologies as external sources of common knowledge. Although both cases involve the use of an external ontology, their role in the alignment process is very different. In this paper we propose a dissociation of these inputs in two techniques, as illustrated in figure 6.

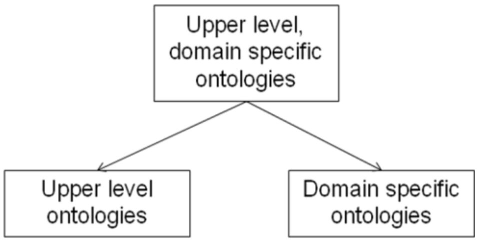

Figure 6. Dissociation of the classification "Upper level, domain specific ontologies" in "Upper level" and "Domain specific ontologies"

Considering this new classification, our approach is instantiated in the "Upper level ontologies" technique. In section 6 we will present the related work instantiated in this classification.

4. Alignment patterns based on foundational ontologies

Considering the design patterns presented in section 2 it is possible to derive some alignments patterns given below.

4.1. Subkind Alignment Pattern

This pattern consists of three rules:

Rule 1: The alignment (equivalence) of a <<kind>> class C1 of an ontology A to a <<kind>> class C2 of an ontology B is possible if all <<subkind>> classes of C1 have a corresponding class that is also a <<subkind>> of C2 in ontology B (a one-to-one equivalence is not required).

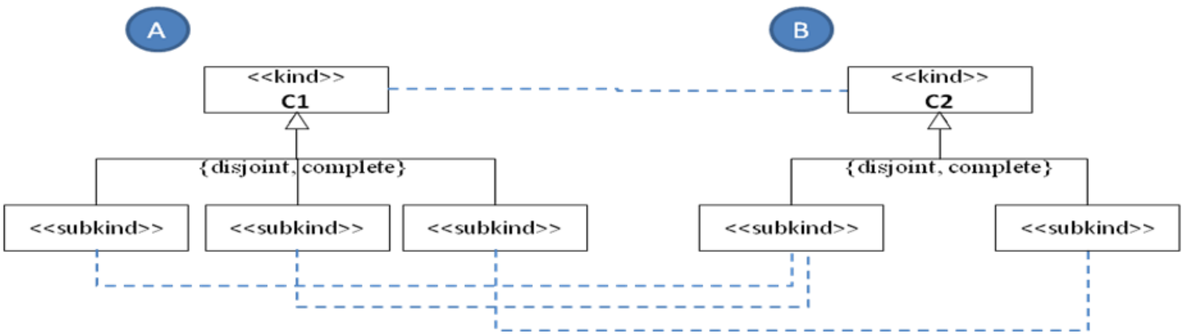

Figure 7. Rule 1 of Subkind Alignment Pattern

Rule 2: The alignment (specialization) of a <<kind>> class C2 of an ontology B to a <<kind>> class C1 of an ontology A is possible if some of the <<subkind>> classes of C1 have equivalent classes in ontology B, and these equivalences cover all the <<subkind>> classes of C2 in ontology B.

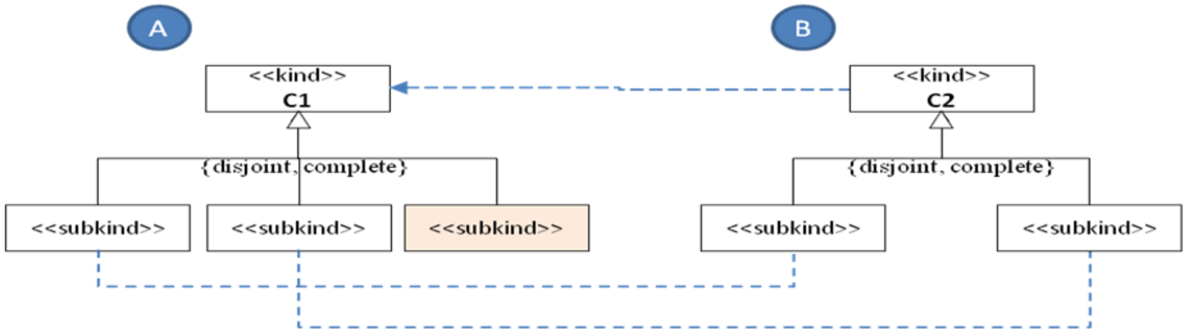

Figure 8. Rule 2 of Subkind Alignment Pattern

Rule 3: The alignment of a <<kind>> class C1 of an ontology A to a <<kind>> class C2 of an ontology B is not possible if at least one <<subkind>> class of C1 in ontology A is not aligned to a <<subkind>> class of C2 in ontology B.

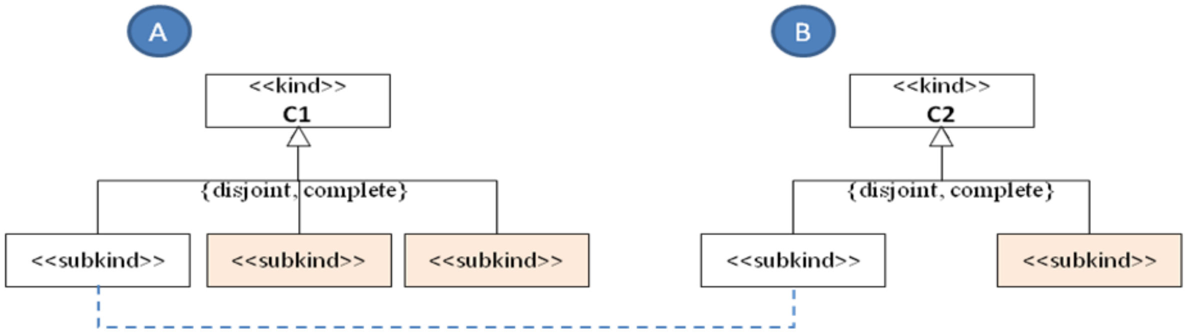

Figure 9. Rule 3 of Subkind Alignment Pattern

4.2. Phase Design Pattern

Besides the distinct semantics of Subkind and Phase classes, since both are manifested as a part of a disjoint and complete generalization set which has as a common superclass a Kind S, the rules of the Phase Design Pattern are analogous to the rules of the Subkind alignment pattern that were previously presented.

4.3. Role Design Pattern

This pattern consists of one rule:

Rule 4: The alignment (equivalence) of a <<role>> class C1 of an ontology A to a <<role>> class C2 of an ontology B is only possible if the <<kind>> rigid class that the <<role>> classes specialize, and the relation its instantiation depends on, are aligned to each other.

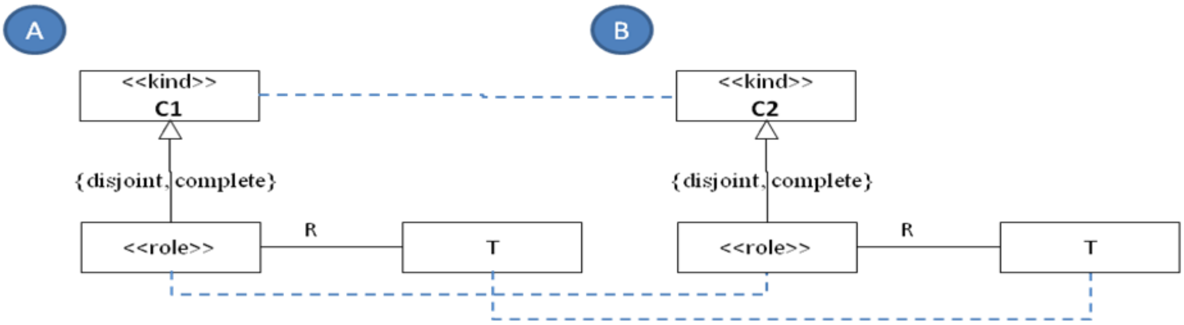

Figure 10. Rule 4 of Role Design Pattern

5. Examples of use

Is this section we will exemplify the application of the alignment patterns presented in section 4. The ontologies describe the domain of organizing conferences, which corresponds to one track of the Ontology Alignment Evaluation Initiative (OAEI) 2011[1]. The ontologies and the reference alignments indicated by the initiative are available on the Conference Track[2].

A prerequisite for application of the alignment patterns is that the ontologies to be aligned must comply with the UFO constraints set out by Guizzardi (2005) and it is necessary the identification of the OntoUML stereotype applicable to each class, considering the design patterns presented in section 2. Because the conference domain is well understandable for every researcher, this task was executed by the authors for the fragments discussed in this paper.

We have analyzed the alignment of two ontologies identified in Table 1 by considering the submitted alignments results of tools evaluated in group 1, which consists of best evaluated matchers of the track. We will explore two common errors committed by the four matchers of this group, one affecting the precision and other the recall measure.

Table 1. Ontologies Iasted and SigKdd

Name	Type	Number of classes
Iasted	Web*	140
SigKdd	Web*	49
* Ontologies have been based upon actual conference (series) and its web pages		

All four matchers have identified a correspondence between the classes Iasted::Document and SigKdd::Document that is not indicated by the reference alignment (which harms the precision). The fragments are illustrated in figure11.

In this case, both Document classes are of the type kind and correspond to a generalization set of subkind classes, disjoint and complete.

[1] http://oaei.ontologymatching.org/

[2] http://oaei.ontologymatching.org/2011/conference/index.html

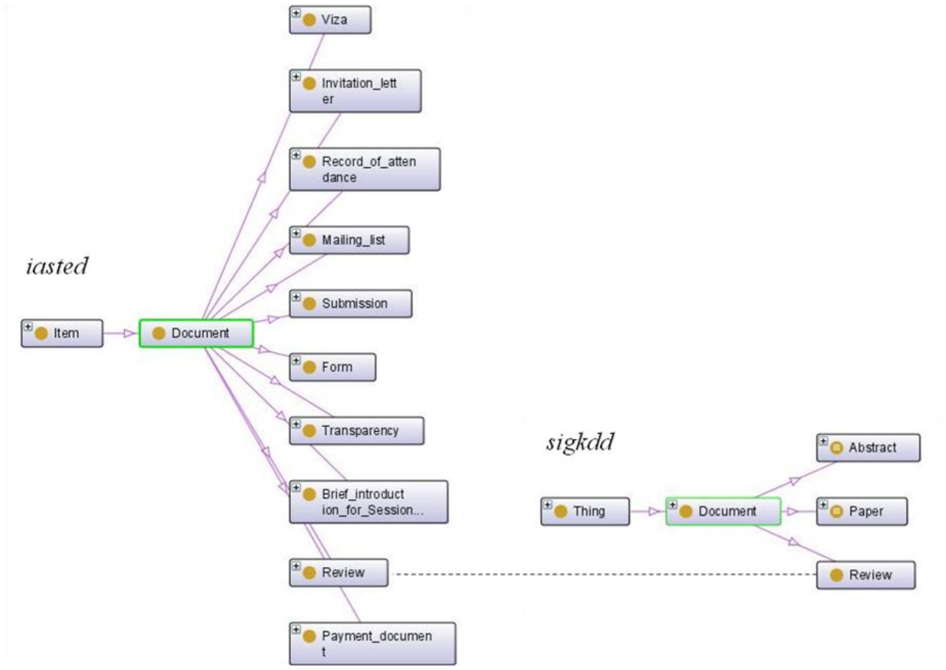

Figure 11. Alignment between Iasted::Document and SigKdd::Document

Dotted lines indicate the expected alignment between the classes of these ontologies fragments. The set of matchers considered have identified the correspondence between Iasted::Review and SigKdd::Review, besides the wrong correspondence between the Document classes. In this context, the application of Rule 3 would reject the correspondence between Iasted::Document and SigKdd::Document since some of their subkind classes are not aligned.

The other common error is that all four matchers could not identify the reference alignment between Iasted::Sponsor and SigKdd::Sponzor indicated by the reference alignment (thus reducing recall). The fragments are illustrated in figure 12.

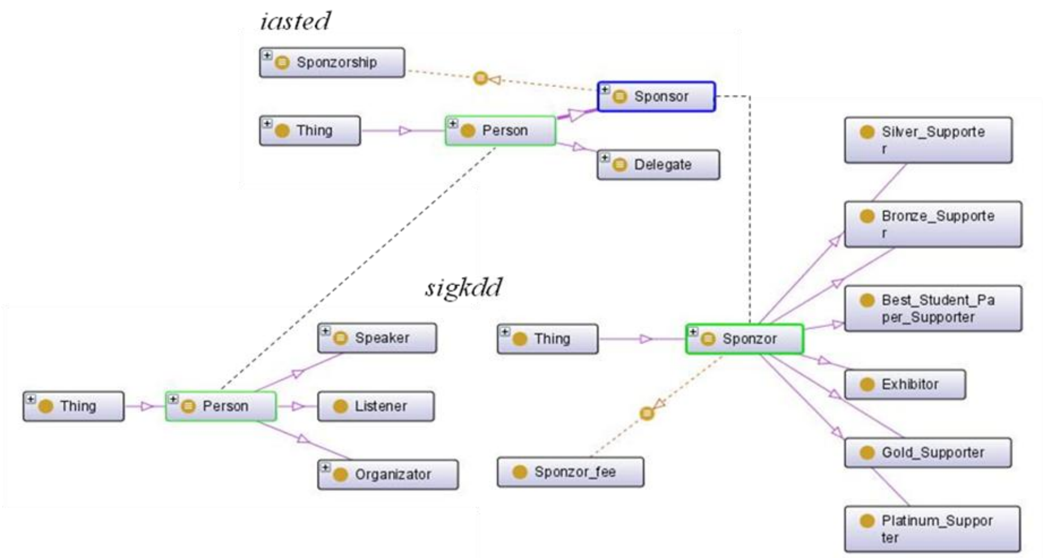

Figure 12. Alignment between Iasted::Sponsor and SigKdd::Sponzor

In this example, both Sponsor and Sponzor classes are stereotyped as roles. In Iasted ontology a Sponsor is a role played by a person that gives some Sponzorship. In SigKdd ontology a Sponzor is a role characterized by the payment of a Sponzor_fee. However, the kind class specialized by this role is not explicit in this ontology. Based on Role Design Pattern, we define this role as a specialization of the class Person, already defined in this ontology and aligned to the class Iasted::Person. With this redesign, the rigid kind classes that these roles specialize are aligned to each other, which itself brings additional information to allow the identification of the alignment between Sponsor and Sponzor classes. However, by Rule 4, to guarantee the alignment between Iasted::Sponsor and SigKdd::Sponzor, Iasted::Sponzorship and SigKdd::Sponzor_fee must be aligned, which suggests an update in the reference alignment to include this equivalence. Otherwise, this correspondence would be rejected.

In figure 13 there are fragments of other two ontologies of the same conference domain, identified in Table 2.

Table 2. Ontologies Edas and Cmt

Name	Type	Number of classes
Edas	Tool*	104
Cmt	Tool*	36
* Ontologies have been based upon actual software tool for conference organisation support		

The question identified here is not an error that affects the precision or recall, but a review of the type of the correspondence identified.

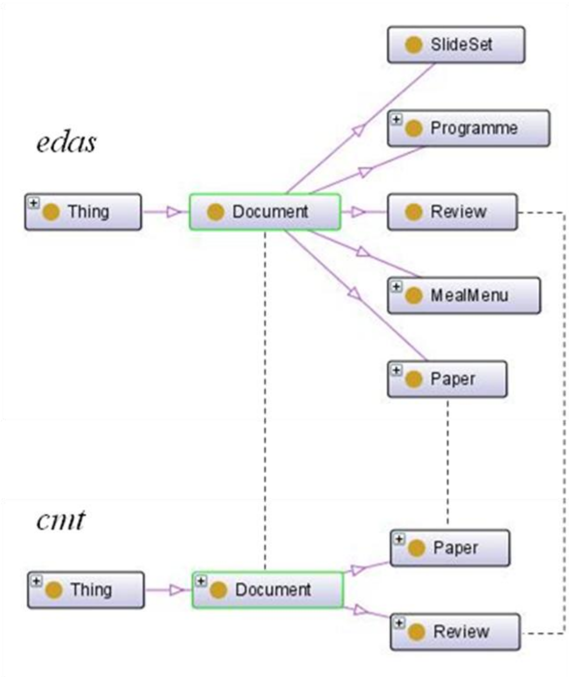

Figure 13. Alignment between Edas::Document and Cmt::Document

In this case, both Document classes are of the type kind and correspond to a generalization set of disjoint and complete subkind classes, as the first example. The dotted lines indicate the reference alignments that mean equivalence relation. Since part of <<subkind>> classes of Edas ontology have equivalent classes in Cmt ontology, and these equivalences cover all the <<subkind>> classes of Cmt ontology, the alignment between the classes Document could be refined considering that Cmt::Document is a specialization of Edas::Document.

6. Related Work

One main point that has guided the development of the approach presented in Silva et al. (2011) is the use of foundational ontologies. To establish the relationship among the foundational ontology and the domain ontologies, for each first-level concept at the domain ontology, a foundational concept was associated. Thus, the result is a unique integrated ontology, composed by the domain ontology and some of the meta-categories of a foundational ontology. Despite considering foundational ontologies, the additional information they provide was relevant for the taxonomic similarity measure (structural input) implemented by the matcher used for the tests, as it becomes possible to compare upper-level concepts in the hierarchy when a candidate pair of concepts is under analysis.

The approach presented in this paper, in turn, proposes a directly use of foundational ontologies to improve semantic integration by considering some alignment patterns based on meta-properties of the OntoUML constructs, with the determination of rules to be applied during the alignment process. Considering the suggested dissociation of the “Upper level” and “Domain specific ontologies” it is instantiated in the “Upper level ontologies” technique.

Other works address foundational ontologies in the context of ontology alignment but they are more directly related to the use of domain ontologies to support the alignment of other ontologies on the same domain. In Mascardi et al. (2010) the techniques applied to associate the classes of the domain ontologies to the classes of the foundational ontologies are typically used to associate concepts of domain ontologies. A higher precision was only obtained with foundational ontologies that include many domain-specific concepts in addition to the upper-level ones. In Gonçalves et al. (2011) the hypothesis is that a domain reference ontology that considers the ontological distinctions of OntoUML can be employed to achieve semantic integration between data standards. The hypothesis is tested by means of an experiment that uses an electrocardiogram (ECG) ontology and conceptual models of the ECG standards. Considering the suggested dissociation of the “Upper level” and “Domain specific ontologies”, these approaches would be instantiated in the “Domain specific ontologies”.

7. Conclusion and Future Work

Ontology alignment is an active research area and some challenges consider semantic issues. In this paper we discussed how the use of OntoUML oriented by some design patterns improves the ontological adequacy of the ontologies being aligned and allows the application of some rules based on alignment patterns. We have used some ontologies from the main initiative for evaluation of ontology alignment to demonstrate how the design patterns and the alignment patterns may improve precision, recall and refine the type of the alignment, with a manually performed example. However, the process of annotation the classes with the correct OntoUML stereotypes can be assisted by a software tool (Benevides and Guizzardi 2009).

Another contribution of the paper is a review in the classification of Euzenat (2007) concerning the technique called "Upper level, domain specific ontologies" to better organize the works that address foundational ontologies in the process of ontology alignment.

Future work includes formalization of indicative and restrictive rules based on the meta-properties of a larger set of constructs of OntoUML to be applied during the alignment process. Moreover, the automatization of the proposal and its application in complete scenarios will also be considered.

References

Benevides, A. B., Guizzardi, G. (2009) "A Model-Based Tool for Conceptual Modeling and Domain Ontology Engineering in OntoUML", In: ICEIS 2009, pp. 528-538

Ehrig, M. (2007), Ontology Alignment: Bridging the Semantic Gap, Springer

Euzenat, J. and Shvaiko, P. (2007), Ontology Matching, Springer

Gonçalves, B., Guizzardi, G. and Pereira Filho, J. G. (2011) "Using an ECG reference ontology for semantic interoperability of ECG data", In: Journal of Biomedical Informatics, vol. 44 , pp 126–136

Gruber, T. R. (1995) "Toward Principles for the Design of Ontologies Used for Knowledge Sharing", In: International Journal of Human and Computer Studies, vol. 43, issues 5/6. pp. 907–928

Guarino, N. (1998), Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction and Integration

Guizzardi, G., Graças, A. P and Guizzardi, R. S. S. (2011) "Design Patterns and Inductive Modelling Rules to Support the Construction of Ontologically Well-Founded Conceptual Models in OntoUML", In: 3rd International Workshop on Ontology-Driven Information Systems (ODISE 2011), London, UK

Guizzardi, G. (2005) Ontological Foundations for Structural Conceptual Models, Ph.D. Thesis, University of Twente, The Netherlands

Mascardi, V., Locoro, A. and Rosso, P. (2010) "Automatic Ontology Matching Via Upper Ontologies: A Systematic Evaluation" In: IEEE Transactions on Knowledge and Data Engineering, vol. 22, n. 5, pp. 609–623

Silva, V.S., Campos, M. L. M, Silva, J. C. P. and Cavalcanti, M. C. (2011) "An Approach for the Alignment of Biomedical Ontologies based on Foundational Ontologies", In: Journal of Information and Data Management, vol. 2, n. 3, pp. 557–572

Ziegler, P. and Dittrich, K. R. (2007) "Data Integration - Problems, Approaches, and Perspectives", In: Krogstie, J., Opdahl, A. L. and Brinkkemper, S. (eds.) Conceptual Modelling in Information Systems Engineering, pp. 39–58. Springer, Heidelberg