Formalizing Ontology Alignment and its Operations with Category Theory[1]

Antoine ZIMMERMANN [a], Markus KRÖTZSCH [b], Jérôme EUZENAT [a] and Pascal HITZLER [b]

[a] *INRIA Rhône-Alpes, 655 avenue de l'Europe, 38330 Montbonnot (France)*
[b] *AIFB, Universität Karlsruhe, D-76128 Karlsruhe (Germany)*

Abstract.
An ontology alignment is the expression of relations between different ontologies. In order to view alignments independently from the language expressing ontologies and from the techniques used for finding the alignments, we use a category-theoretical model in which ontologies are the objects. We introduce a categorical structure, called V-alignment, made of a pair of morphisms with a common domain having the ontologies as codomain. This structure serves to design an algebra that describes formally what are ontology merging, alignment composition, union and intersection using categorical constructions. This enables combining alignments of various provenance. Although the desirable properties of this algebra make such abstract manipulation of V-alignments very simple, it is practically not well fitted for expressing complex alignments: expressing subsumption between entities of two different ontologies demands the definition of non-standard categories of ontologies. We consider two approaches to solve this problem. The first one extends the notion of V-alignments to a more complex structure called W-alignments: a formalization of alignments relying on "bridge axioms." The second one relies on an elaborate concrete category of ontologies that offers high expressive power. We show that these two extensions have different advantages that may be exploited in different contexts (*viz.*, merging, composing, joining or meeting): the first one efficiently processes ontology merging thanks to the possible use of categorical institution theory, while the second one benefits from the simplicity of the algebra of V-alignments.

Keywords. Ontology alignment, category theory

Introduction

In its most general form, the term "ontology alignment" can refer to almost any formal description of the (semantic) relationship between ontologies. A more restricted conception of the term is used in surveys [6,23] or alignment API [7], that conceive alignments as pairs of elements of the ontologies[2], together with information on the type of the rela-

[1] This work has been partly supported by the European Knowledge Web network of excellence (IST-2004-507482), by the German Ministry for Education and research (BMBF) under the SmartWeb project, and by the European Commission under contract IST-2003-506826 SEKT.

[2] We speak of "elements of an ontology" to refer to arbitrary semantic entities of a given ontology language, *e.g.*, concepts, relations, or instances.

tion and the confidence in its correctness. An alignment is thus given a set-theoretic definition as a set of correspondences. Though perfectly acceptable in applications, it has the disadvantage of using entities—something local—as the basis of ontology alignment—something global.

Here we present a complementary approach that treats alignments as first class citizens where locally defined elements are not needed. Our main goal is to build a formal theory of ontology alignments and their associated operations that is independent of the internal representation language. To achieve this objective, we start from a category-theoretic definition of ontology alignment that was already sketched in [24,4,15,12,14]: a pair of *morphisms** with a common domain. However, none of these works provided in-depth investigations of this abstract formulation of the alignment problem.

Based on this related work, we build up a framework around the notion of *V-alignments*, with an abstract definition of the merge of two ontologies and an algebra allowing composition, intersection and union of alignments (§2). Then, disposing of a well-behaved, language-independent infrastructure, concrete alignments must be embedded into it as an example of the generality of our approach. A critical issue is the ability to express non-symmetrical relations (*e.g.*, a class in one ontology being subsumed by another class in another ontology) within such theoretical framework. No solution is given to this problem in all cited papers. The difficulty does not reside in the theoretical ability to express such relations, but in the manner we should instantiate the abstract formulation. In §3, we propose two solutions to this problem:

- define a more complex structure for the definition of category theoretic alignments, while reusing already existing categories of ontologies;
- design a category—or rather a class of categories—with elaborate morphisms enabling the expression of complex, non-symmetrical relations.

The former approach, presented in §4, is a re-construction of the framework with the new notion of *W-alignments*. The latter solution presents a home-made category of ontologies that increases the expressivity of morphisms, in comparison to previously published categories of ontologies (§5). Both approaches have interesting advantages and constraining drawbacks that we consider in §6.

1. Related work

In [24], a similar categorical approach is mentioned but not rigorously formalized. [4] uses morphisms of algebraic specifications[3] to define morphisms between ontologies and say a relation (an alignment in their sense) between ontologies O_1 and O_2 consists of an ontology O and a pair of morphisms $\chi_1 : O \to O_1$ and $\chi_2 : O \to O_2$. This is precisely the definition of V-alignment given below, but it does not provide any means of representing complex alignments as we do. In [15], a category-theoretic approach using the information flow theory of Barwise and Seligman [3] is given, with no concrete representation of alignments. Kent also gives an intuition of the bridge ontology idea in [16] but he does not formalize it within category theory and only describes the merge of

*Words marked with a * are category-theoretic terms defined in the Appendix.

[3]Algebraic specifications are closely related to specifications in institution theory, which we often refer to, without using it explicitly.

two ontologies. Information Flow is also the basis of an implemented system called IF-Map [13] designed for automated ontology mapping, but they do not have a categorical representation for rich alignments. [12] gives a concrete example of a representation of an alignment in category theory, but since it is so simplistic, it is hard to see the generality of the approach. Joseph Goguen's work on *institution** theory [10], especially [9,8], advertises the use of *colimits** for ontology integration. This theory, though not sufficient to model complex alignments with a V-alignment structure, can be used as a grounding for our so-called W-alignments. More details on the categorical approach are given in a survey on ontology mapping [14]. Finally, ontology alignment and schema matching are closely related. In particular, [5] describes a schema matching algebra that we partially generalize with our approach. See [21,14,23] for surveys on ontology and schema mappings in general.

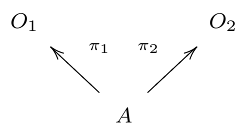
2. Simple alignments

This part presents a categorical formulation of various operations employed when manipulating different ontologies and alignments. These operations are ontology merging, alignment composition, alignment union and intersection. They are presented in more details in [11]. They are but mere application of common categorical constructions to the abstract alignment structure and will only be sketched here.

Remark: In the remainder, some familiarity with category theory would be useful. However, definitions are given in Appendix[4]. For more details on the basics of category theory, see [20] for an easy yet good introduction. [1,19] give something more elaborated.

2.1. Category-theoretic alignments

As said before, an ontology alignment is a description of the relationship between two ontologies. Category theory generalizes the set-theoretic notion of relation, and offers a definition for a (generalized) relation between two arbitrary objects in a *category**. An alignment thus corresponds to the diagram given below, where *objects** O_1, O_2 and A are ontologies and π_1 and π_2 are ontology *morphisms**. In categorical terminology, such diagram is called a *span*.



When the objects O_1, O_2, A are ontologies and π_1, π_2 are ontology morphisms, we call this structure a V-alignment due to the shape of the associated diagram and in order not to confuse it with the informal notion of alignment. The very same definition is used in [4,12,14] with different names (ontology relation, ontology alignment, ontology articulation).

[4]See footnote on page 2.

2.2. Merging with V-alignments

Once a V-alignment between two ontologies is known, it is desirable to integrate the aligned ontologies into a single ontology. This operation, called *ontology merging*, aims at uniting heterogeneous specifications into a larger, more precise one which allows more information sharing. The categorical formalization of V-alignments allows for a simple description of the merge, and this can be described in terms of the category-theoretic *pushout** construction. [15,13,12] give more details about this construction, and the present paper also discusses this point in §3.

2.3. Algebra for V-alignments

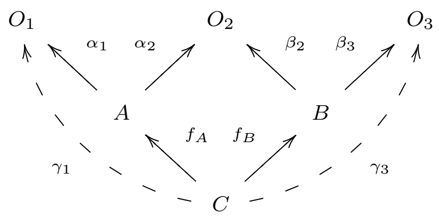
The need for ontology alignment naturally arises when information from many ontologies is relevant to a given task. However, since the task of constructing alignments is not an easy one and can hardly be accomplished in a fully automatic fashion, it is reasonable to store and reuse known alignments. The purpose of this section is to introduce a sound algebra of V-alignments that allows for essential operations that enable us to compose, join, and intersect alignments.

These operations are application of general operations over categorical *spans*, so interested readers shall refer to [19] for their general properties. They are also more detailed in [11]. This is why here we present them briefly.

2.3.1. Composing alignments

Composition is a central operation for the reuse of alignments: if we have alignments between ontologies O_1 and O_2, and between O_2 and O_3, then it should be possible to obtain an alignment of O_1 and O_3. The definition is the same as the composition of spans in category theory (see [19]). So it is obtained by the use of the categorical construction called *pullback**.

The following *commutative diagram** shows two V-alignments $\langle A, \alpha_1, \alpha_2 \rangle$ and $\langle B, \beta_2, \beta_3 \rangle$. The composition is the alignment $\langle C, \alpha_1 \circ f_A, \beta_3 \circ f_B \rangle$, where $\langle C, f_A, f_B \rangle$ is the pullback of α_2 and β_2.



Composition is associative and identity exists, which confirms that the proposed operation is well-behaved as a composition of alignments.

2.3.2. Intersection and union of alignments

Intersection gives the consensual correspondences in two alignments. Union gathers all asserted relations specified in two alignments. These operations are indeed very useful in the context of the Semantic Web since they allow a modularization of alignments. In this respect, one can give a partial alignment with only part of the relevant correspondences and expect to retrieve more on the Web when needed.

Figure 1 **a.** gives the diagram of intersected alignments $\langle A, f_1, f_2 \rangle$ and $\langle B, g_1, g_2 \rangle$. Object C together with morphisms k_A, k_B, h_1 and h_2 make the *limit** of the diagram composed of the two alignments. The resulting alignment is $\langle C, h_1, h_2 \rangle$.

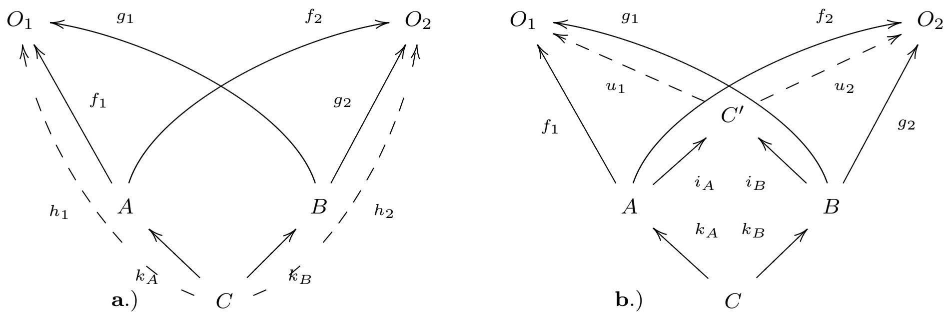


Figure 1. Intersection and union of alignments.

Union is defined via intersection. In order to unify two alignments, one has to know what is common to both of them. Then the union is the disjoint union of this common part and the non-common parts. In Figure 1 **b.**, this is done by way of a categorical *pushout** of $\langle k_A, k_B \rangle$. Morphisms u_1 (resp. u_2) is obtained by *factorizing** f_1 (resp. f_2) through i_A (resp. i_B). So informally, we say that union is the pushout of intersection.

These operations are, as expected, commutative and associative.

This algebra concisely formalizes operations combining two or more alignments provided by different alignment algorithms or experts, either by composing (when there is no alignment between two related ontologies), joining (if both algorithms take into account different aspects of ontologies) or meeting them (if on the contrary they should agree for considering correspondences to be correct).

3. Concretizing V-alignments

In order to apply the previous framework to real cases, it is necessary to instantiate it with concrete categories. For instance, one can consider the most basic way to describe relationships between two ontologies: identifying those elements which represent the same semantic entities. This can be adequately described by a binary relation between the sets of elements, that is, consider morphisms as functions.

In the literature, the most adapted categories of ontologies are found in institution theory [10], where *specifications** (*i.e.*, ontologies in our terms) are mapped with truth-preserving functions. Notably the language OWL can be described as an institution [17]. Unfortunately, a pair of functions (even structure-preserving), is only adequate to express equivalence of entities. In many cases, though, the two ontologies to align were designed in such way that some concepts do not have their equivalent in the other ontology, although several concepts are closely related. For instance, one may find that concept $Woman$ in ontology O_1 is a subclass of $Person$ in ontology O_2. In this case, the merge should contain concepts $Person$ and $Woman$ with a subsumption relation between them (see Figure 2). However, assuming this is the result of a pushout operation, it is not clear what the alignment should be.

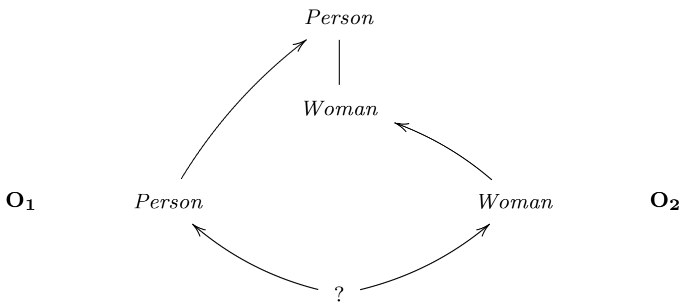


Figure 2. Expressing non-symmetrical relations with V-alignments.

A pair of functions cannot lead to such a pushout. So the problem of expressing complex alignments requires investigation, and we therefore propose the following solutions to work out this issue:

1. Find more complex categories, where objects still are ontologies, but with morphisms able to express other relations;
2. Keep the category simple, and complexify the definition of an alignment using more elaborate structure;
3. Change the definition of the merge, for example by using different type of *colimit**.

The last item implies that the operations defined in §2 are to be abandoned. Since they are built on well established work, we will not challenge this idea. We first discuss item (2) in §4, where a new alignment structure is defined on top of the previous one and leads to an upgrade of the associated infrastructure. Item (1) is considered in §5, which defines a new category of ontologies where morphisms are family of relations instead of functions.

4. W-alignments

In this section, we combine existing material into an extended formulation of alignments that we will suggestively name W-alignments. It corresponds to a categorization of the notion of bridge axioms. Merging and composing are defined in this framework, but the algebra suffers from defects.

4.1. Categorical formulation of bridge axioms

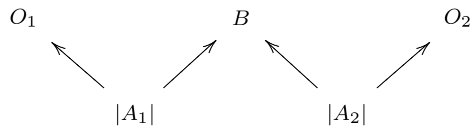
Let us start with the example from §3: consider two OWL-ontologies O_1 and O_2 that contain the atomic concepts $Woman$ and $Person$, respectively. Assuming that none of the ontologies contains both concepts, it is not possible to express the intended subsumption of $Woman$ and $Person$ with V-alignments in a standard category of ontologies.

In such cases, the relation is nonetheless expressible in the ontology language but cannot be represented with the vocabulary of any of the two ontologies. So the idea is to externalize the assertion "$Woman \sqsubseteq Person$" in another ontology. These external assertions are called *bridge axioms* (described from a logical point of view in *e.g.*, [18]). As observed in the introduction, alignments described as sets of bridge axioms give a local description of the correspondences between two ontologies. In order to conform to the categorical paradigm, we must first give a globalized definition of these axioms.

We do this by representing bridge axioms in form of an additional *bridge ontology*. The fact that certain concepts of the aligned ontologies occur within the bridge ontology is captured by V-alignments between the bridge and each of the aligned ontologies. We thus arrive at the following definition:

Definition 4.1 *A* W-alignment *between two ontologies O_1 and O_2 is a triple $\langle B, A_1, A_2 \rangle$ where B is a* bridge ontology *and A_1 and A_2 are two V-alignments between O_1 and B and between O_2 and B, respectively.*

The following diagram depicts the situation, which also serves to illustrate why the above terminology was chosen. Note also that we do not impose any restrictions on the bridge ontology B. In particular B could contain axioms that are related to neither O_1 nor O_2.



Based on this categorical formulation, here we give a suitable definition for merging of ontologies that are aligned with a W-alignment.

Definition 4.2 *Given two ontologies O_1 and O_2 and a W-alignment between them, the* merge *of O_1 and O_2 is defined to be the* colimit* *of the alignment diagram. More explicitly, this colimit M is computed by successive pushouts as in Figure 3.*

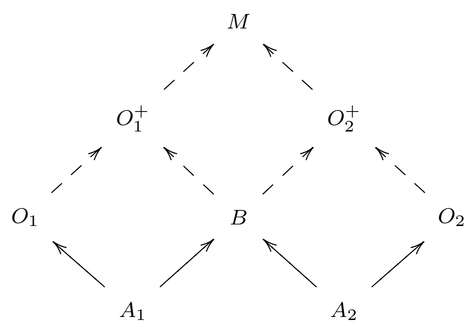


Figure 3. Merging with W-alignments.

Intuitively, O_1^+ and O_2^+ represent the original ontologies O_1 and O_2 extended with axioms and elements that enable us to express their alignment as a simple V-alignment. This idea is not entirely new, and in [16] O_1^+ and O_2^+ have been called *portal ontolo-*

gies, referring to their specific role in making the knowledge of each of the ontologies accessible to the other one.

Since this merge is obtained by successive pushouts, this operation for W-alignments is in the same class of complexity as merging with V-alignments.

Example 4.3 *A more demonstrative example consists in expressing the semantic connection between a n-ary relation and its reification with only binary relations. For instance, property "sells" relates a seller to a buyer and to an object (3-ary relation) in the first ontology. The second ontology has a class "Sale" that has three properties "hasSeller", "hasBuyer" and "hasObject". The bridge ontology will contain the axiom* $R(x, y, z) \Leftrightarrow \exists t S(t) \wedge r_1(t, x) \wedge r_2(t, y) \wedge r_3(t, z)$. *The first V-alignment matches* R *with "sells" and the second matches* S, r_1, r_2, r_3 *with "Sale", "hasSeller", "hasBuyer", "hasObject", respectively. The merge will contain both the relation and its reification, together with the axiom.*

4.2. Composing W-alignments

A full-featured algebra for W-alignments, along the lines of §2.3, would be complicated and unintuitive. However, we can easily describe a useful operation for composing W-alignments.

Definition 4.4 *Consider ontologies* O_1, O_2, *and* O_3 *with W-alignments as in Figure 4. The* composition *of the W-alignments of Figure 4 is described as follows:*

- *The bridge ontology* B *is obtained as the merge of the bridge ontologies* B_1 *and* B_2, *according to the W-alignment* $\langle O_2, A_2, A_3 \rangle$,
- *the V-alignment of* O_1 *and* B *is* $\langle A_1, f_1, b_1 \circ g_1 \rangle$, *and*
- *the V-alignment of* O_3 *and* B *is* $\langle A_4, g_4, b_3 \circ f_4 \rangle$.

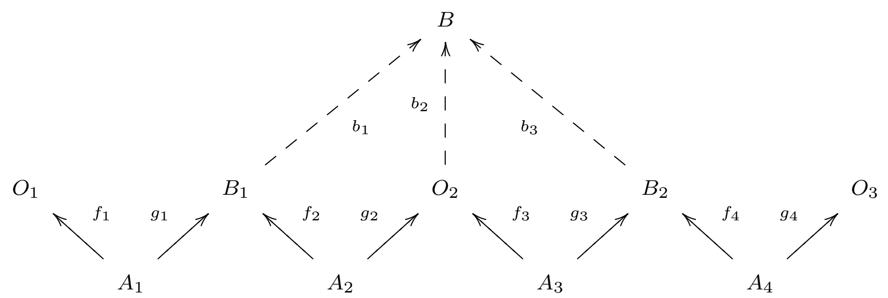


Figure 4. Composing W-alignments.

This definition formalizes the fact that we know there is a relation of O_1 and O_3, given by means of an intermediate ontology O_2. In order to describe this with a single bridge ontology, we integrate both of the involved bridges with O_2. This construction has the advantage that it faithfully captures all information that is available about the composed alignment.

However, there is a major problem with the above definition: by deriving bridge axioms from the ontologies B_1, B_2, and O_2, we incorporate all their embedded information into the new bridge ontology. But this set of bridge axioms might be highly redundant for the given purpose: it may involve axioms of O_2 that are neither related to O_1 nor to O_3. Another pathological case is when O_2 is the disjoint union of O_1 and O_3, while O_1 and O_3 are not related at all. In this case, we would rather wish the composed bridge ontology to be empty, instead of containing the whole information of all involved ontologies.

Overcoming this difficulty at the concrete level relates to the problem of finding a minimal non-redundant set of axioms that yields a given set of desired (or relevant) conclusions. Unfortunately, logical languages tend to be highly non-local in this respect.

Other operations like intersection and union suffer from the same kind of deficiency: there is neither canonical nor intuitive definition that satisfies the notion they are supposed to cover. We therefore omit their mentioning in this paper, and prefer to focus on a different approach that relies on a newly proposed category of ontology.

5. Improved category of ontologies

The other possible solution consists in using elaborate morphisms capable of expressing complex alignments with V-alignments alone. We describe here an enhanced category of ontologies, that we name $\mathfrak{Ont}^+$, which has ontologies as objects and particularly elaborate morphisms.

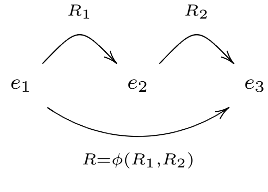
Definition 5.1 (Morphisms) *A morphism $f : O_1 \rightarrow O_2$ in $\mathfrak{Ont}^+$ is a set of triples $\langle e_1, e_2, R\rangle$ such that:*

- *e_1 and e_2 are syntactic entities (concepts, relations, individuals, etc.) from ontologies O_1 and O_2 respectively,*
- *R denotes a relationship that holds between e_1 and e_2 (e.g., subsumption, equivalence, temporal relations, etc.). The set of available relations will be denoted $\mathcal{R}$.*

The categorical composition operation associated to these morphisms is thus defined:

Definition 5.2 (Composition) *Let $f : O_1 \rightarrow O_2$ and $g : O_2 \rightarrow O_3$ be two morphisms in $\mathfrak{Ont}^+$. The composition of f and g, noted $g \circ f$ is the set of triples $\langle e_1, e_3, R\rangle$ such that there exist e_2, R_1, R_2 such that $\langle e_1, e_2, R_1\rangle \in f$, $\langle e_2, e_3, R_2\rangle \in g$ and $R = \phi(R_1, R_2)$ with $\phi : \mathcal{R}\times\mathcal{R} \rightarrow \mathcal{R}$ given by a composition table, such as the one given in Example 5.3.*

The small figure below gives the intuition of this definition: if an entity e_2 is related to entities in ontologies O_1 and O_3, then there should be some kind of relation between e_1 and e_3. This relation depends on the relationship between e_2 and the other entities, and is expressed by the function ϕ. It corresponds to the composition of relational constraints, as found in temporal [2] or spatial algebras [22].



Example 5.3 *In the following composition table, $=$ is equality, $\subset$ is strict inclusion, $\supset$ is strict containment, $\perp$ is disjointness and $\between$ is overlapping with partial disjointness.*

$\mathbf{R_1}$ \ $\mathbf{R_2}$	$=$	$\subset$	$\supset$	$\perp$	$\between$
$=$	$\{=\}$	$\{\subset\}$	$\{\supset\}$	$\{\perp\}$	$\{\between\}$
$\subset$	$\{\subset\}$	$\{\subset\}$	$\{=,\subset,\supset,\perp,\between\}$	$\{\perp\}$	$\{\subset,\perp,\between\}$
$\supset$	$\{\supset\}$	$\{=,\subset,\supset,\between\}$	$\{\supset\}$	$\{\supset,\perp,\between\}$	$\{\supset,\between\}$
$\perp$	$\{\perp\}$	$\{\subset,\perp,\between\}$	$\{\perp\}$	$\{=,\subset,\supset,\perp,\between\}$	$\{\subset,\perp,\between\}$
$\between$	$\{\between\}$	$\{\subset,\between\}$	$\{\supset,\perp,\between\}$	$\{\supset,\perp,\between\}$	$\{=,\subset,\supset,\perp,\between\}$

Property 5.4 *The composition is associative iff ϕ is associative.*

The associativity of ϕ is not a severe constraint because all usual relations in description logics, temporal and spatial reasoning have associative composition tables. Moreover, in order to have the identity morphism, $\mathcal{R}$ must contain equality. Given these somewhat reasonable constraints, $\mathfrak{Ont}^+$-morphisms together with ontologies as objects form a category. Relations in $\mathcal{R}$ are not restricted to the ontology language. So, for example, two OWL[5] ontologies can be related with temporal or spatial relations, as well as fuzzy ones.

This category has strong advantages with regard to its expressivity and the elegance of the V-alignment algebra that can still be applied here. Additionally, independently from V-alignments, they are, alone, composable and tunable. However, they have a major drawback: pushouts does not generally coincide with the expected merge. In next section, we further discuss advantages, drawbacks and potential interest of both approaches.

6. Discussion

Both of the two solutions proposed have pros and cons. On the one hand, the representation of W-alignments is less intuitive as V-alignments and demands a prior understanding of V-alignments. Moreover, manipulating W-alignments necessitates a reconstruction of the infrastructure available with V-alignments. This infrastructure has a defective algebra: no identity alignment, no canonical union and intersection. Another counter-intuitive property of W-alignments is the capability to use axioms in the bridge ontology that do not relate to any of the aligned ontologies. These drawbacks make W-alignments inappropriate to build new alignments out of existing ones, so they do not fit for highly distributed semantic applications. On the other hand, W-alignments can express very rich alignments, such as relations between a n-ary property and its reification. Moreover, it is built upon the same principle as simple alignment: colimits serve for ontology integration. This is a strong advantage because, for instance, the category of OWL ontologies is cocomplete [17], *i.e.*, all pushouts exist in this category. So they are well-suited for the merging of ontologies.

Working at the concrete category level leads to different and complementary results. Certain correspondences are hard to express (*e.g.*, reification of n-ary relation), but the enhanced category has the advantage of separating the alignment language—which appears in the morphisms—from the ontology language—which appears in the objects. As long as the relations verify loose constraints, the complexity of relations can be arbi-

[5] http://www.w3.org/TR/owl/

trarily increased, offering possibilities like fuzzy relations or other uncommon relations, without interfering with the ontology language. Besides, they benefit from the algebra described in §2, which makes them easy to manipulate at an abstract level. But when the category gets more complex, allowing expression of non-symmetrical relations, the merge does not always coincide with the pushout. However, the alignment algebra is adequate for abstracting modular ontology alignment applications thanks to the operation of composition, intersection and union. Finally, complicated structure similar to W-alignments could also be constructed on top of them.

7. Conclusion

Finding suitable categorical representations of alignments is the ultimate goal of our work.

Yet, as discussed in this article, previous categorical approaches fail to represent adequately the broad variety of semantic relations found between real world ontologies.

To address this issue, the present paper (1) provides a formalization of several operations on ontologies and ontology alignments relying on simple category-theoretic constructions that is consistent with previous published category-theoretic representation of ontology alignment and integration; (2) proposes two representations of complex semantic relationships within this theory: (2.a) a categorical formulation of the notion of bridge axioms and (2.b) a proposal for a concrete category of ontologies improving the expressivity of formerly proposed categories. In both cases, we study the repercussion of each contribution to the original algebra mentioned in (1) above.

Both approaches show the lack of expressivity in existing work with respect to semantic relationship. They offer partial solutions to the problem. We presented the advantages of each solution. Though both approaches have interesting benefits, we are leaning toward the second one because we think it can lead to a more general theory of ontology alignment and coordination. Of course, the morphisms we presented have to be connected to the semantics of the ontology. Our future investigation aims at providing an abstract model theory with such complex morphisms, along the line of institution theory which encompasses both syntax and semantics. Doing this, we will be able to design a legitimate semantics for ontology alignment and distributed systems, while so far, no common agreement exists on such a semantics.

References

[1] Jiří Adámek, Horst Herrlich, and George E. Strecker. *Abstract and Concrete Categories (The Joy of Cats)*. web published, 2004.

[2] James F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.

[3] John Barwise and Jerry Seligman. *Information flow: the logic of distributed systems*. Number 44. Cambridge University Press, Cambridge, United Kingdom, 1997.

[4] Trevor J. M. Bench-Capon and Grant Malcolm. Formalising ontologies and their relations. In *Proc. 10th Database and Expert Systems Applications (DEXA'99)*, pages 250–259, 1999.

[5] Philip A. Bernstein. Applying model management to classical meta data problems. In *Proc. First Biennial Conference on Innovative Data Systems Research (CIDR'03)*, January 2003.

[6] Paolo Bouquet, Jérôme Euzenat, Enrico Franconi, Luciano Serafini, Giorgos Stamou, and Sergio Tessaris. Specification of a common framework for characterizing alignment. Deliverable 2.2.1, Knowledge Web NoE, August 2004.

[7] Jérôme Euzenat. An API for ontology alignment. In *Proc. Third International Semantic Web Conference (ISWC)*, volume 3298 of *LNCS*, pages 698–712, Hiroshima, Japan, 2004. Springer-Verlag GmbH.

[8] Joseph A. Goguen. Information integration in institutions. Paper for Jon Barwise memorial volume.

[9] Joseph A. Goguen. Data, schema and ontology integration. In *Workshop on Combination of Logics: Theory and Applications (CombLog'04)*, July 2004.

[10] Joseph A. Goguen and Rod M. Burstall. Institutions: abstract model theory for specification and programming. *Journal of the ACM*, 39(1):95–146, 1992.

[11] Pascal Hitzler, Jérôme Euzenat, Markus Krötzsch, Luciano Serafini, Heiner Stuckenschmidt, Holger Wache, and Antoine Zimmermann. Integrated view and comparison of alignment semantics. Deliverable, Knowledge Web NoE, August 2005.

[12] Pascal Hitzler, Markus Krötzsch, Marc Ehrig, and York Sure. What is ontology merging? - a category-theoretic perspective using pushouts. In *Proc. First International Workshop on Contexts and Ontologies: Theory, Practice and Applications (C&O)*, pages 104–107. AAAI Press, July 2005.

[13] Yannis Kalfoglou and Marco Schorlemmer. Information flow based ontology mapping. In *On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE : Confederated International Conferences 2002*, volume 2519 of *LNCS*, pages 1132–1151. Springer-Verlag GmbH, 2002.

[14] Yannis Kalfoglou and Marco Schorlemmer. Ontology mapping: The state of the art. In *Semantic Interoperability and Integration*, number 04391 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2005. Internationales Begegnungs und Forschungszentrum.

[15] Robert E. Kent. The information flow foundation for conceptual knowledge organization. In *Proc. 6th International Conference of the International Society for Knowledge Organization (ISKO)*, 2000.

[16] Robert E. Kent. The IFF foundation for ontological knowledge organization. *Cataloging and Classification Quarterly*, 37(1):187–203, 2003.

[17] Dorel Lucanu. A logical foundation for the OWL languages. In *Proc. First International Symposium on Leveraging Applications of Formal Methods (ISoLA)*, 2004.

[18] Christopher Menzel. Basic semantic integration. In *Semantic Interoperability and Integration*, number 04391 in Dagstuhl Seminar Proceedings. Internationales Begegnungs und Forschungszentrum, 2005.

[19] Maria Cristina Pedicchio and Walter Tholen, editors. *Categorical Foundations: Special Topics in Order, Topology, Algebra, and Sheaf Theory*, volume 97 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2004.

[20] Benjamin C. Pierce. *Basic Category Theory for Computer Scientists*. Foundations of Computing. The MIT Press, Cambridge, MA, 1991.

[21] Erhard Rahm and Philip A. Bernstein. A survey of approaches for automatic schema matching. *The Very Large Data Base Journal*, 10(4):334–350, December 2001.

[22] David A. Randell, Zhan Cui, and Anthony Cohn. A spatial logic based on regions and connection. In *Proc. Third International Conference on Principles of Knowledge Representation and Reasoning (KR'92)*, pages 165–176. Morgan Kaufmann, 1992.

[23] Pavel Shvaiko and Jérôme Euzenat. A survey of schema-based matching approaches. *Journal on Data Semantics*, 4:146–171, 2005.

[24] Danladi Verheijen, Gio Wiederhold, Jan Jannink, and Srinivasan Pichai. Encapsulation and composition of ontologies. In *Proc. 15th National Conference on Artificial Intelligence (AAAI'98)*, July 1998.

Appendix

This appendix gives basic definitions from Category Theory and Institution Theory. Only the very basics of category theory is needed to fully understand the present paper.

Definition 7.1 (Category) *A* category $\mathfrak{C}$ *consists of:*

- *a collection of* objects $\mathrm{obj}(\mathfrak{C})$*;*
- *for all pair* $A, B \in \mathrm{obj}(\mathfrak{C})$*, a set* $\mathfrak{C}(A, B)$ *of* morphisms*;*
- *for all triple* $A, B, C \in \mathrm{obj}(\mathfrak{C})$*, a composition law*

$$\circ : \mathfrak{C}(B,C) \times \mathfrak{C}(A,B) \to \mathfrak{C}(A,C),$$

such that the following axioms be true:

- *for all objects* A, B, C, D *and all morphisms* $f \in \mathfrak{C}(A,B), g \in \mathfrak{C}(B,C), h \in \mathfrak{C}(C,D)$, $((h \circ g) \circ f) = (h \circ (g \circ f))$ *(associativity);*
- *for all* A *there exists a morphism* $\mathrm{id}_A \in \mathfrak{C}(A,A)$ *called* identity morphism *such that for all* $f \in \mathfrak{C}(A,B)$,
  $f \circ \mathrm{id}_A = f = \mathrm{id}_B$ *(idempotence of identity).*

When $f \in \mathfrak{C}(A,B)$*, one writes* $f : A \to B$*, and* A *is said to be the* domain *of* f *and* B *is the* codomain.

Definition 7.2 (Pushout) *Let* $f : A \to O, f' : A \to O'$ *be two morphisms with a common domain. The* pushout *of* $\langle f, f' \rangle$ *is a triple* $\langle g, g', P \rangle$ *such that:*

- $g : O \to P, g' : O' \to P$ *and* $g \circ f = g' \circ f'$*;*
- *for each* $\langle h, h', Q \rangle$ *satisfying the condition above, there is a unique* $k : P \to Q$ *such that* $h = k \circ g$ *and* $h' = k \circ g'$*;*

In this definition, k *is called the* factorization *of* h *(resp.* h'*) through* g *(resp.* g'*).*

For all notions defined in category theory, there exists a corresponding notion called the dual, characterized by reversing all arrows in the first definition. For instance, the dual of the pushout is the pullback:

Definition 7.3 (Pullback) *Let* $f : O \to A, f' : O' \to A$ *be two morphisms with a common codomain. The* pullback *of* $\langle f, f' \rangle$ *is a triple* $\langle g, g', P \rangle$ *such that:*

- $g : P \to O, g' : P \to O'$ *and* $f \circ g = f' \circ g'$*;*
- *for each* $\langle h, h', Q \rangle$ *satisfying the condition above, there is a unique* $k : Q \to P$ *such that* $h = g \circ k$ *and* $h' = g' \circ k$*;*

Limits generalize the notion of pullback. The dual notion of a limit is a colimit.

Definition 7.4 (Limit) *Let* D *be a collection of objects in a category, and* $(f_{ij})_{i,j \in I \times J}$ *be a collection morphisms s.t.* f_{ij} *has* $i \in D$ *as domain and* $j \in D$ *as codomain. The* limit *of* (f_{ij}) *is a family* $(g_i)_{i \in D}$ *associated to an object* L *such that:*

- *for all* $i \in D, g_i : L \to i$ *and* $f_{ij} \circ g_i = g_j$*;*
- *for each* $(h_i)_{i \in D}, L'$ *satisfying the condition above, there is a unique* $k : L' \to L$ *such that* $h_i = g_i \circ k$ *for all* $i \in D$*;*

A functor relates two categories. It is composed of a mapping over objects and a collection of mappings over morphisms that preserve identity and composition.

Definition 7.5 (Functor) *A* functor $F : \mathfrak{A} \to \mathfrak{B}$ *from a category* $\mathfrak{A}$ *to a category* $\mathfrak{B}$ *consists of*

- *a function*
  $$\begin{array}{lcl} \mathrm{obj}(\mathfrak{A}) & \to & \mathrm{obj}(\mathfrak{B}) \\ A & \mapsto & FA \end{array}$$
- *for each* $A, A' \in \mathrm{obj}(\mathfrak{A})$*, a function*
  $$\begin{array}{lcl} \mathfrak{A}(A, A') & \to & \mathfrak{A}(FA, FA') \\ f & \mapsto & Ff \end{array}$$

such that

- $F(f' \circ f) = Ff' \circ Ff$ *for all* $f : A \to A'$ *and* $f' : A' \to A''$ *in* $\mathfrak{A}$
- $F\mathrm{id}_A = \mathrm{id}_{FA}$ *for all* $A \in \mathfrak{A}$

The theory of institutions is a categorical abstract model theory which formalises the intuitive notion of logical system, including syntax, semantics, and the satisfaction between them.

Definition 7.6 (Institution) *An* institution *is a tuple* $\langle \mathfrak{Sign}, Sen, Mod, \models \rangle$ *such that:*

- $\mathfrak{Sign}$ *is a category (the category of* signatures*)*
- $Sen : \mathfrak{Sign} \rightarrow \mathfrak{Set}$ *is a functor that associates a set of* sentences *to any signature.*
- $Mod : \mathfrak{Sign}^{op} \rightarrow \mathfrak{Set}$ *is a functor that associates to any signature a* class *of* models *(of model theory).*
- $\models$ *is a family of relations parameterized by a signature* Σ*. They are called the* satisfaction relations *and for each* Σ, $\models_\Sigma$ *is a relation between models of* Σ *(i.e. elements in* $\mathrm{obj}(Mod(\Sigma))$*) and sentences of* Σ *(i.e. elements in* $\mathrm{obj}(Sen(\Sigma))$*). For all such relation, the satisfaction condition must hold, i.e.* $\forall \Sigma' \in \mathfrak{Sign}, \forall M' \in \mathrm{obj}(Mod(\Sigma')), \forall e \in \mathrm{obj}(Sen(\Sigma)), \forall f : \Sigma \rightarrow \Sigma', M' \models_{\Sigma'} f(e) \Leftrightarrow Mod(f)(M') \models_\Sigma e$*.*

We define specification and specification morphisms here because they correspond to the abstract notion of ontologies and ontology morphisms in our paper.

Definition 7.7 (Specification) *A* specification *(with regard to a given institution) is a pair* $\langle \Sigma, Ax \rangle$ *with* Σ *being a* signature*, and* Ax *a subset of* $Sen(\Sigma)$.

Definition 7.8 (Specification morphism) *A specification morphism* $f : \langle \Sigma, Ax \rangle \rightarrow \langle \Sigma', Ax' \rangle$ *is a signature morphism* $f : \Sigma \rightarrow \Sigma'$ *such that for all sentences* e *in* Ax*, all models* M' *of* $\langle \Sigma', Ax' \rangle$ *satisfy* $f(e)$*, i.e.* $\forall e \in Ax, \forall M' \in \langle \Sigma', Ax' \rangle, M' \models f(e)$*.*