

Algebraic calculi for weighted ontology alignments

Armen Inants, Manuel Atencia, Jérôme Euzenat

► **To cite this version:**

Armen Inants, Manuel Atencia, Jérôme Euzenat. Algebraic calculi for weighted ontology alignments. Paul Groth; Elena Simperl; Alasdair Gray; Marta Sabou; Markus Krötzsch; Freddy Lécué; Fabian Flöck; Yolanda Gil. 15th International semantic web conference (ISWC), Oct 2016, Kobe, Japan. Springer Verlag, Proc. 15th International semantic web conference (ISWC), pp.360-375, 2016, <10.1007/978-3-319-46523-4_22>. <hal-01382098>

HAL Id: hal-01382098

<https://hal.inria.fr/hal-01382098>

Submitted on 15 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Algebraic calculi for weighted ontology alignments

Armen Inants, Manuel Atencia, and Jérôme Euzenat

Inria & Univ. Grenoble Alpes, Grenoble, France
`{Armen.Inants,Manuel.Atencia,Jerome.Euzenat}@inria.fr`

Abstract. Alignments between ontologies usually come with numerical attributes expressing the confidence of each correspondence. Semantics supporting such confidences must generalise the semantics of alignments without confidence. There exists a semantics which satisfies this but introduces a discontinuity between weighted and non-weighted interpretations. Moreover, it does not provide a calculus for reasoning with weighted ontology alignments. This paper introduces a calculus for such alignments. It is given by an infinite relation-type algebra, the elements of which are weighted taxonomic relations. In addition, it approximates the non-weighted case in a continuous manner.

Keywords: weighted ontology alignment, algebraic reasoning, qualitative calculi

1 Introduction

Ontology alignments are used for facilitating the integration of semantically related ontologies [8]. They are sets of correspondences relating entities from two ontologies using semantic relations such as equivalence (\equiv), subsumption (\sqsubseteq , \supseteq) and disjointness (\perp). Very often, these correspondences are coupled with weights in $[0, 1]$. The intended meaning of these weights is a degree of confidence on the correspondence, i.e. a measure of how much we can trust that the correspondence is true. For example, the correspondence (`AssociateProfessor`, `SeniorLecturer`, \sqsubseteq , 0.9) states that the class `AssociateProfessor` is subsumed by the class `SeniorLecturer` with a confidence degree of 0.9, and, therefore, one should trust that this subsumption is true. The automatic treatment of ontology alignments calls for a calculus for reasoning with weighted correspondences. However, such a calculus has not been proposed yet.

In previous work, we advocated the algebraic approach to reasoning with ontology alignments [6, 13]. An algebraic calculus of alignments is given by an algebra of ontology alignment relations. In this paper, we show how to compose *weighted* ontology alignment relations, based on their algebraic semantics.

Previous work introduced a formal semantics for weighted ontology alignments [1]. A weighted correspondence between two classes C and D is written $C \ r_{[a,b]} \ D$ where $r \in \{\sqsubseteq, \equiv, \supseteq, \perp\}$ and a, b are real numbers in $[0, 1]$. The semantics is based on the classification interpretation of alignments, when a common finite set of instances is classified under classes of different ontologies. The

weighted correspondence $C \sqsubseteq_{[a,b]} D$, for example, is interpreted as “the proportion of instances classified under C that are classified under D lies in the interval $[a, b]$.” Although [1] provides some entailment rules for reasoning with weighted correspondences, none of these rules allows to compose alignment relations. In addition, the current semantics has some shortcomings, discussed below.

First, although the interval $[a, b]$ may be any subinterval of $[0, 1]$, in practice we are mostly interested in intervals of the form $[a, 1]$. Think, for example, of the previous correspondence (`AssociateProfessor`, `SeniorLecturer`, \sqsubseteq , 0.9). This should be translated into `AssociateProfessor` $\sqsubseteq_{[0.9,1]}$ `SeniorLecturer` and not `AssociateProfessor` $\sqsubseteq_{[0.9,0.9]}$ `SeniorLecturer`. Indeed, the latter is interpreted as “(exactly) 90% of the associate professors are senior lecturers” from which it follows that the crisp subsumption is not true. However, the former is interpreted as “at least 90% of the associate professors are senior lecturers” which leaves room for the possibility that the crisp subsumption is true. In general, $C \ r_{[a,b]} \ D \models \neg(C \ r \ D)$ if $b < 1$. Furthermore, from a theoretical point of view, if we restrict to $[a, 1]$ intervals, then weighted relations can be seen as relaxed crisp relations, i.e., $C \ r \ D \models C \ r_{[a,1]} \ D$, or equivalently $r \models r_{[a,1]}$. In what follows, r^a will replace $r_{[a,1]}$.

Second, one would expect that $\sqsubseteq^1 \models \neg \perp$. However, with the current semantics of the disjointness relation, this is not the case. Let us illustrate this with an example. Consider the classes `BrazilianSnakes` and `VenomousSnakes`, and imagine that 100 snakes are classified under these two classes, and that from these 100 snakes, 10 are Brazilian, and all of them are venomous. Thus, `BrazilianSnakes` $\sqsubseteq_{[1,1]}$ `VenomousSnakes` and `BrazilianSnakes` $\supseteq_{[0.1,0.1]}$ `VenomousSnakes`. The weight of the equivalence relation is the harmonic mean of 1 and 0.1, i.e. `BrazilianSnakes` $\equiv_{[0.2,0.2]}$ `VenomousSnakes`, and the weight of the disjointness relation is 1 minus the harmonic mean, i.e. `BrazilianSnakes` $\perp_{[0.8,0.8]}$ `VenomousSnakes`.

Finally, although in the crisp case equivalence entails subsumption, i.e. $\equiv \models \sqsubseteq$, this does not hold in general for weighted correspondences, that is, from equivalence with a confidence interval $[a, 1]$ one cannot entail subsumption with (at least) the same confidence: $\equiv^a \not\models \sqsubseteq^a$ and $\equiv^a \not\models \supseteq^a$ for any $a \in (0, 1)$. This becomes evident in the previous example, since, although `BrazilianSnakes` $\equiv_{[0.2,0.2]}$ `VenomousSnakes`, `BrazilianSnakes` $\not\sqsubseteq_{[0.1,0.1]}$ `VenomousSnakes`.

This weighted semantics is a generalization of the crisp or Boolean semantics: if all weights are 1, then the semantics is exactly the crisp semantics. However, the way it approaches the crisp semantics is, in a sense which will be explained in this paper, discontinuous: as close as the weighted semantics approaches the crisp one, these two properties ($\sqsubseteq^1 \models \neg \perp$ and $\equiv \models \sqsubseteq$) do not hold, but as soon as all weights are 1, they do.

In this paper, we propose a calculus for reasoning with weighted alignments based on the semantics that overcomes the shortcomings explained above.

The paper is structured as follows. Section 2 introduces the state of the art and other related work. Section 3 contains some mathematical notions and preliminary results, upon which the developments of this paper are based. The key notion that we employ is that of a (relational) constraint language. Section 4 introduces the constraint language QTAX of *quantified taxonomic relations*. We

show that both crisp and weighted taxonomic relations can be expressed in QTAX. In Section 5, we specify a sublanguage of QTAX consisting of the relaxed taxonomic relations. We compare the revisited semantics of \equiv^a and \perp^a with the old one and discuss its advantages. In Section 6, we develop the calculus of relaxed taxonomic relations. Section 7 discusses how this calculus can be used to reason with weighted ontology alignments. Finally, Section 8 summarizes the results and provides some concluding remarks.

2 Related work

Different semantics to weighted ontology alignments have been proposed [16, 1]. [16] relies on tightly integrated description logics programs, i.e., pairs of description logic T-boxes and answer set programs. In that work, weights are interpreted as probabilistic distributions over *models*. We here concentrate on extensional interpretations.

The semantics proposed in [1] is based on a classificational interpretation of alignments: if O_1 and O_2 are two ontologies used to classify a common set X , then correspondences between O_1 and O_2 are interpreted as encoding how elements of X classified in the concepts of O_1 are re-classified in the concepts of O_2 , and weights are interpreted to measure how precise and complete re-classifications are. Syntactically, a weighted correspondence between ontologies O_1 and O_2 , expressed in a description logic [2], is an expression of the form:

$$1:C \ r_{[a,b]} \ 2:D,$$

such that C and D are concepts in O_1 and O_2 respectively, $r \in \{\sqsubseteq, \supseteq, \equiv, \perp\}$ and $a, b \in [0, 1]$ ($a \leq b$).

The semantics of such correspondences is based on pairs of description logic interpretations $\mathcal{I}_1 = (U_1, \cdot^{\mathcal{I}_1})$ and $\mathcal{I}_2 = (U_2, \cdot^{\mathcal{I}_2})$ of O_1 and O_2 respectively. A pair of interpretations is a model of a weighted correspondence if the degree ds_X that can be computed from the interpretations lies within the interval $[a, b]$ assigned to the correspondence. The degrees are defined as follows:

$$\begin{aligned} ds_X(\mathcal{I}_1, \mathcal{I}_2, C \sqsubseteq D) &= \frac{|C_X^{\mathcal{I}_1} \cap D_X^{\mathcal{I}_2}|}{|C_X^{\mathcal{I}_1}|} \\ ds_X(\mathcal{I}_1, \mathcal{I}_2, C \supseteq D) &= \frac{|C_X^{\mathcal{I}_1} \cap D_X^{\mathcal{I}_2}|}{|D_X^{\mathcal{I}_2}|} \\ ds_X(\mathcal{I}_1, \mathcal{I}_2, C \equiv D) &= \frac{2 \times ds_X(\mathcal{I}_1, \mathcal{I}_2, C \sqsubseteq D) \times ds_X(\mathcal{I}_1, \mathcal{I}_2, C \supseteq D)}{ds_X(\mathcal{I}_1, \mathcal{I}_2, C \sqsubseteq D) + ds_X(\mathcal{I}_1, \mathcal{I}_2, C \supseteq D)} \\ ds_X(\mathcal{I}_1, \mathcal{I}_2, C \perp D) &= 1 - ds_X(\mathcal{I}_1, \mathcal{I}_2, C \equiv D) \end{aligned}$$

The interpretation of \sqsubseteq and \supseteq are expressed as the proportion of common individuals in the class interpretations, which can also be interpreted as the probability of reclassification of individuals [1]. This is justifiable by extensional practice of ontology matching. The interpretation of \equiv mitigates the impact of these two

through the use of the F-measure between them, while \perp is interpreted as the complement of equivalence. This semantics of weighted alignments approximates classical crisp semantics [4, 18] in the sense that if all weights are assigned 1 or 0, i.e., $[1, 1]$ or $[0, 0]$, then the models are those of the crisp semantics.

However, as mentioned in the introduction, this semantics has some undesirable consequences and we will show how they can be addressed. For that purpose, we will reconsider it in the framework of algebras of relations.

The algebraic approach to reasoning with relational assertions, which we adopt in this paper, comes from the domain of qualitative spatial and temporal reasoning. This approach may also be applied to reasoning with aligned ontologies [6, 15, 13] and was extended to support relations between different kinds of entities [12]. The central notion is that of a *qualitative calculus* [14, 5], which is a finite symbolic algebra used for constraint-based reasoning based on the path-consistency method. There exist reasoning toolboxes which support qualitative calculi [17, 9]. The only principal difference of \mathbb{A}_{QTAX} from qualitative calculi is that it contains infinitely many relations. This may call for adjustments to existing reasoning algorithms.

3 Preliminaries

Here we introduce the notion of constraint languages for relations (Section 3.1) and the algebras of relations that they generate (Section 3.2).

3.1 Constraint languages

Constraint languages are a mathematical framework for defining semantics of relational assertions. A (relational) constraint language is given by a collection of relation symbols and their interpretations. We use the formal definition of a constraint language as a relational structure in the model-theoretic sense [11].

Definition 1 (Constraint language). A *relational signature* is a set σ of relation symbols (also called predicate symbols), each with an associated finite arity. A (*relational*) *constraint language* over σ , or shortly a σ -*language*, is a tuple $\Gamma = (\sigma, U, \cdot^{\Gamma})$, where σ is a relational signature, U is a set called the *universe* and \cdot^{Γ} is the *interpretation* function defined on σ , which maps each relation symbol with arity n to an n -ary relation over U .

In this paper we confine ourselves to binary constraint languages, i.e., those that consist of binary relations.

Given a constraint language $\Gamma = (\sigma, U, \cdot^{\Gamma})$, we say that R is a Γ -*relation*, if R is equal to r^{Γ} for some relation symbol $r \in \sigma$. We may write $R \in \Gamma$, meaning that R is a Γ -relation. When the interpretation of relation symbols in σ is clear from the context, we will specify a constraint language over a finite signature as $\Gamma = (U; r_1, r_2, \dots, r_n)$, where U is the universe and r_1, r_2, \dots, r_n are the relation symbols.

Example 1. The constraint language of base taxonomic relations between sets. $\text{baseTAX5} = (U; \equiv, \sqsubset, \sqsupset, \overline{\cap}, \perp)$, where U is some powerset and $\overline{\cap}$ the partial overlap relation symbol.

Constraint languages can be compared in terms of granularity. We start with a general definition of granularity relations [7].

Definition 2 (Granularity). Let \mathcal{X} and \mathcal{Y} be two collections of sets. \mathcal{X} is said to be

- *finer* than \mathcal{Y} if, for every $X \in \mathcal{X}$, there exists $Y \in \mathcal{Y}$ such that $X \subseteq Y$;
- *coarser* than \mathcal{Y} if, for every $X \in \mathcal{X}$, there exists $\mathcal{Y}_0 \subseteq \mathcal{Y}$ such that $X = \cup \mathcal{Y}_0$;
- a *refinement* of \mathcal{Y} , if \mathcal{X} is finer than \mathcal{Y} and \mathcal{Y} is coarser than \mathcal{X} .

The relations “finer than”, “coarser than” and “refinement of” are transitive. A σ -language Γ is said to be *finer* than, *coarser* than, or a *refinement* of a σ' -language Γ' , if so is the set of Γ -relations w.r.t. the set of Γ' -relations.

Definition 3 (Disjunctive expansion). Let $\Gamma = (\sigma, U, \cdot^\Gamma)$ be a constraint language. The *disjunctive expansion* of Γ is the constraint language $\Gamma_\vee = (\widehat{\sigma}, U, \cdot^{\Gamma_\vee})$, where $\widehat{\sigma}$ consists of all subsets of σ ($\widehat{\sigma} = \wp(\sigma)$) and, for every $r \in \widehat{\sigma}$, $r^{\Gamma_\vee} = \cup \{r_0^\Gamma : r_0 \in r\}$.

The signature of Γ_\vee can be also defined, following the logical notation, as the set of all disjunctions of relation symbols from σ . For the signature of Γ_\vee we will use the set-theoretic notation with one reservation: we will identify a singleton set $\{r\} \in \wp(\sigma)$ with the element $r \in \sigma$. Thus, for $r \in \sigma$ we may also write that $r \in \wp(\sigma)$. If $r \in \sigma$, then the relation r^{Γ_\vee} is called a *base Γ_\vee -relation*. If $r \subseteq \sigma$, then r^{Γ_\vee} is said to be a *disjunctive Γ_\vee -relation*.

Example 2. The disjunctive expansion of baseTAX5 (Example 1) is called the *constraint language of taxonomic relations between sets*, denoted as TAX5 . Among the disjunctive TAX5 -relations is subsumption and its converse: $\sqsubseteq = \{\sqsubset, \equiv\}$ and $\sqsupseteq = \{\sqsupset, \equiv\}$.

We will usually assume that different relation symbols correspond to different relations. In these cases, for a binary relation $R \in \Gamma$, by R^σ we will denote the relation symbol $r \in \sigma$, for which $r^\Gamma = R$. If $R \in \Gamma_\vee$, then $R^\sigma := \{r \in \sigma : r^\Gamma \subseteq R\}$.

3.2 Algebras generated by constraint languages

If a constraint language Γ is closed under all intersections (finite or infinite) and contains the universal relation, then we can define weak composition of Γ -relations as follows: for $R, S \in \Gamma$, their *weak composition* is defined as $R \diamond_\Gamma S = \cap \{T \in \Gamma : R \circ S \subseteq T\}$. (When it causes no ambiguity, we will write \diamond instead of \diamond_Γ .) Likewise, *weak converse* is defined as $R^\vee = \cap \{T \in \Gamma : R^{-1} \subseteq T\}$. The operations of weak composition and weak converse are naturally induced on the relation symbols: $r \diamond s = (r^\Gamma \diamond s^\Gamma)^\sigma$ and $r^\vee = ((r^\Gamma)^\vee)^\sigma$.

A more specific and well-studied case is when a constraint language is obtained by the disjunctive expansion of a *partition scheme*. The notion of a partition scheme was introduced in [14] and then extended in [5]. We refer to the former definition as strong partition schemes and to the latter as abstract partition schemes.

Definition 4 (Partition scheme). Let X be some nonempty set and \mathcal{P} a set of its subsets. \mathcal{P} is said to be a *partition* of X if each element of X belongs to one and only one element of \mathcal{P} . A constraint language $\Gamma = (\sigma, U, \cdot^\Gamma)$ is said to be an *(abstract) partition scheme*, if Γ -relations make up a partition of $U \times U$. In this case Γ -relations are also said to be *jointly exhaustive and pairwise disjoint* (JEPD) on U . An abstract partition scheme Γ is said to be *strong*, if it is closed under converse and contains the identity relation over U .

The signature of the disjunctive expansion Γ_\vee of a constraint language $\Gamma = (\sigma, U, \cdot^\Gamma)$ is a powerset algebra, hence a complete atomic Boolean algebra [10]. If Γ is an abstract partition scheme, then Γ_\vee is closed under intersection and contains the universal relation $U \times U$. Thus, there are two additional operations on $\wp(\sigma)$: namely, weak composition and weak converse. The algebra $\mathbb{A}_\Gamma = (\wp(\sigma), \cup, \cap, -, \emptyset, \sigma, \diamond, \smile)$ is said to be generated by the abstract partition scheme Γ . The algebra \mathbb{A}_Γ provides a symbolic calculus of Γ -relations.

Example 3. The constraint language **baseTAX5** is a partition scheme only if its universe U does not contain the empty set. Then it generates an algebra \mathbb{A}_5 , which is specified in [6]. If the universe U contains the empty set, then the relations of **baseTAX5** are not pairwise disjoint any more. In that case it takes 8 base relations to refine **baseTAX5** into a partition scheme (for more details see [13, 12]).

Proposition 1 establishes an important property of algebras generated by partition schemes, which says that it is enough to define weak composition and weak converse on atoms.

Proposition 1 ([12]). *Let Γ be an arbitrary (finite or infinite) abstract partition scheme over a set U . Then weak composition and weak converse operations of \mathbb{A}_Γ are completely additive, i.e., they completely distribute over the union.*

In addition to the algebraic method for reasoning with constraint languages, there are other approaches coming from recent research in CSP [3]. The main advantage of the algebraic approach is that it is polynomial (cubic) time. The disadvantage is that its reasoning capabilities vary from one constraint language to another and in many cases are rather limited.

4 The constraint language of quantified taxonomic relations

In this section, we consider the universe of all finite sets and define a constraint language, called QTAX, of cardinality-based binary relation over this universe.

We show that QTAX contains the crisp taxonomic relations (TAX5) and also the weighted taxonomic relations introduced in [1].

Let D be some countably infinite set, we consider the set of nonempty finite subsets of D as the *universe* and denote it as \mathcal{U}_D , or simply \mathcal{U} :

$$\mathcal{U}_D = \{X : X \subseteq D \text{ and } 0 < |X| < \omega\},$$

where ω is the first uncountable ordinal number. The set of all rational numbers from 0 to 1 will be denoted as $[0, 1]_{\mathbb{Q}}$. We define a binary relational signature σ as a set of ordered pairs (α, β) , where $\alpha, \beta \in [0, 1]_{\mathbb{Q}}$. Further, we define a σ -language Δ on the universe \mathcal{U} as follows:

$$(\alpha, \beta)^{\Delta} = \left\{ (X, Y) \in \mathcal{U} \times \mathcal{U} : \frac{|X \cap Y|}{|X|} = \alpha \text{ and } \frac{|X \cap Y|}{|Y|} = \beta \right\}.$$

Clearly, if $\alpha = 0$ and $\beta \neq 0$, or $\alpha \neq 0$ and $\beta = 0$, then $(\alpha, \beta)^{\Delta} = \emptyset$. This means that the relation symbols $(0, \beta)$ or $(\alpha, 0)$, in which $\alpha, \beta \neq 0$, are synonyms and all denote the empty relation; we will exclude such relation symbols from consideration. For the rest of σ -symbols we will say (α, β) is equal to (α', β') iff $\alpha = \alpha'$ and $\beta = \beta'$.

We denote the *disjunctive expansion* of Δ as QTAX and call it *the constraint language of quantified taxonomic relations*. A base QTAX-relation can be visually represented as a point on the unit square of α, β parameters (Figure 1a), which we will call the (α, β) -space. A disjunctive QTAX-relation correspond then to a regions of the (α, β) -space, as shown in Figure 1b.

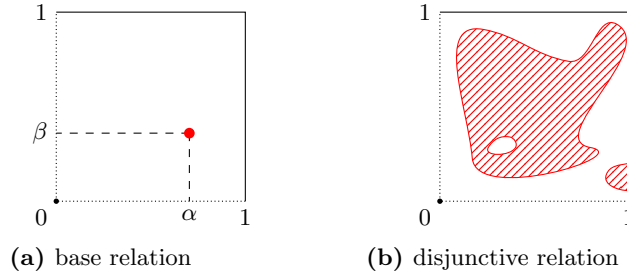


Fig. 1: Visual representation of QTAX-relations on the (α, β) -space.

Recall the constraint language TAX5 of taxonomic relations considered in Example 2. Proposition 2 shows that, if defined on the same universe, QTAX is a refinement of TAX5.

Proposition 2. *QTAX is a refinement of TAX5.*

Proof (Sketch). Figure 2a shows that each taxonomic relation can be presented as a disjunction of quantified taxonomic relations. This means that TAX5 is coarser than QTAX. QTAX is also finer than TAX5, because the latter contains the universal relation of the former. Hence, QTAX is a refinement of TAX5.

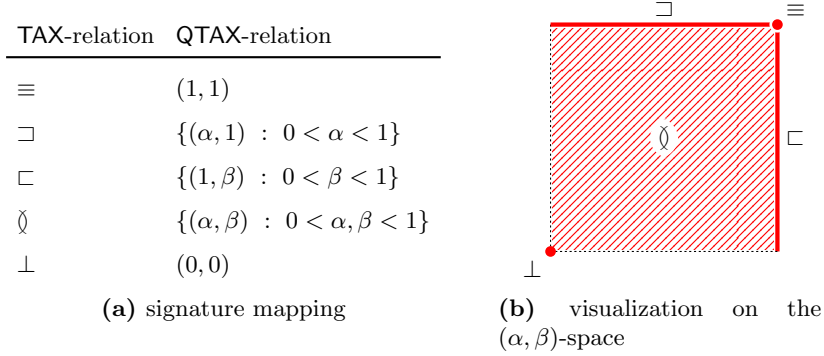


Fig. 2: The constraint language TAX5 of taxonomic relations is a sublanguage of the constraint language of quantified taxonomic relations QTAX.

The base taxonomic relations are visualized on the (α, β) -space in Figure 2b. The weighted taxonomic relations $r_{[a,b]}$ (Section 2) can also be expressed in QTAX, as shown in Table 1. Figure 3 visualizes these relations on the (α, β) -space.

Proposition 3 says that base QTAX-relations make up a strong partition scheme, thus they generate an algebra \mathbb{A}_{QTAX} .

Proposition 3. Δ is an infinite strong partition scheme.

Proof (Sketch). First, any $\alpha, \beta \in [0, 1]_{\mathbb{Q}}$, such that α and β are either both zero or both nonzero, the relation $(\alpha, \beta)^{\Delta}$ is not empty. Further, it is easy to check that Δ -relations are jointly exhaustive and pairwise disjoint. Finally, it remains to check that Δ is closed under converse and contains the identity relation. Indeed, $((\alpha, \beta)^{\Delta})^{-1} = (\beta, \alpha)^{\Delta}$ and $(1, 1)^{\Delta} = Id_{\mathcal{U}}$.

5 The relaxed taxonomic relations

In this section, we discuss the shortcomings of weighted equivalence and disjointness and propose different semantics for these relations. The revisited set of weighted relations constitutes a sublanguage of QTAX, called the constraint language of *relaxed taxonomic relations*. We compare the relaxed semantics of equivalence and disjointness with the former one and discuss its advantages.

As mentioned in the introduction, in a weighted relation $r_{[a,b]}$, if the upper bound b of the confidence interval $[a, b]$ is less than 1, then $r_{[a,b]}$ negates the crisp relation r (in symbols, $r_{[a,b]} \models \neg r$), which is counter-intuitive. This issue can be solved by confining to confidence intervals $[a, 1]$, in which the upper bound is always 1, as shown in Figure 4.

We denote the relations $r_{[a,1]}$ as r^a and call them *relaxed taxonomic relations*, since they are weaker than r , i.e., $r \models r^a$ for any $r \in \{\equiv, \sqsubset, \sqsupset, \perp\}$ and any $a \in [0, 1]$. The semantics of relaxed equivalence \equiv^a and relaxed disjointness

| Weighted taxonomic relation | QTAX-relation |
|-----------------------------|--|
| $\sqsubseteq_{[a,b]}$ | $\{(\alpha, \beta) \in \sigma : a \leq \alpha \leq b\}$ |
| $\sqsupseteq_{[a,b]}$ | $\{(\alpha, \beta) \in \sigma : a \leq \beta \leq b\}$ |
| $\equiv_{[a,b]}$ | $\{(\alpha, \beta) \in \sigma : a \leq \frac{2\alpha\beta}{\alpha+\beta} \leq b\}$ |
| $\perp_{[a,b]}$ | $\{(\alpha, \beta) \in \sigma : a \leq 1 - \frac{2\alpha\beta}{\alpha+\beta} \leq b\}$ |

Table 1: Weighted taxonomic relations $r_{[a,b]}$ expressed in the constraint language QTAX.

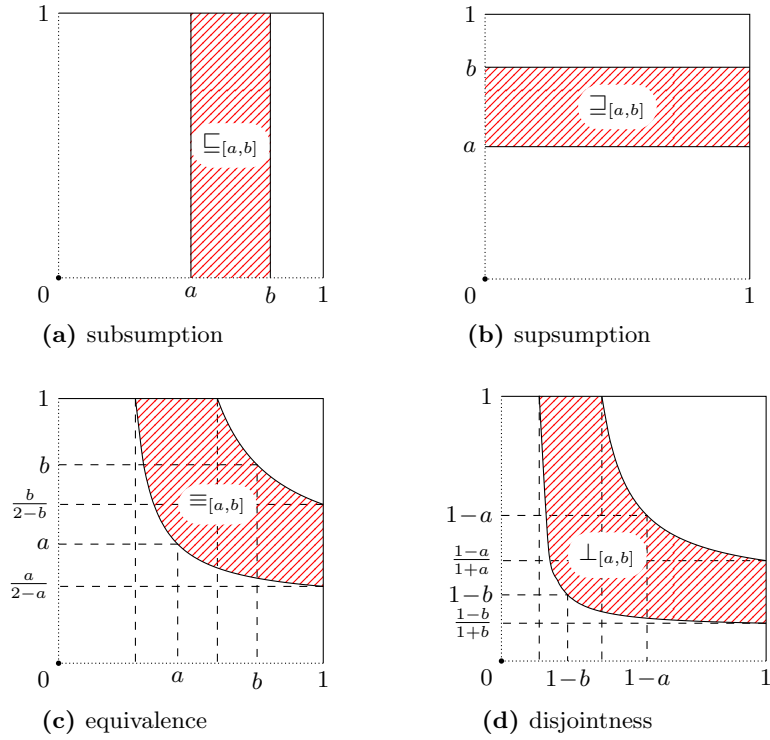


Fig. 3: Visualization of weighted relations $r_{[a,b]}$ (in the sense of [1]) on the (α, β) -space.

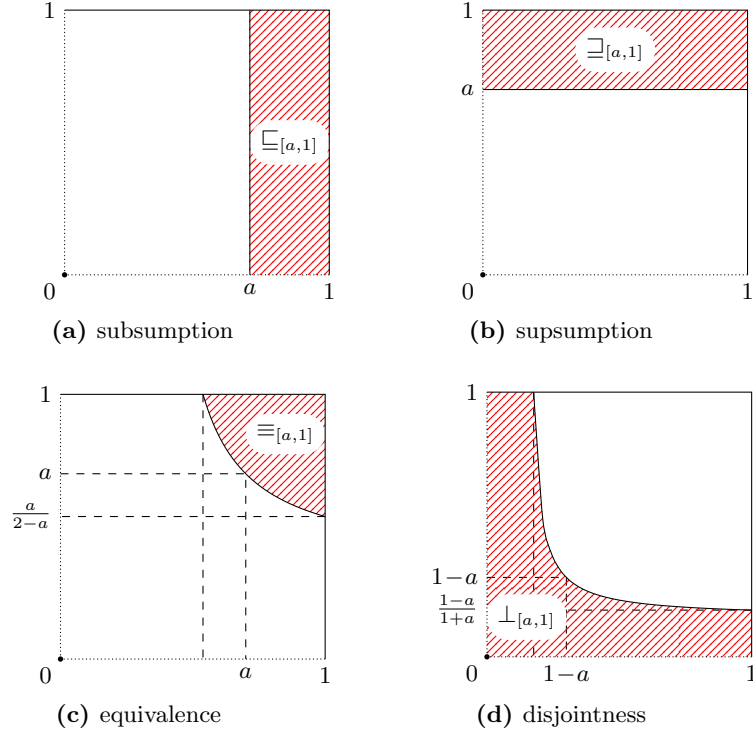


Fig. 4: Visualization of weighted relations $r_{[a,1]}$ (in the sense of [1]) on the (α, β) -space.

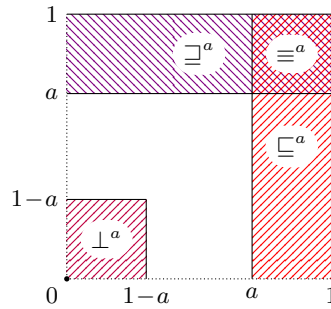


Fig. 5: Visualization of base relaxed taxonomic relations on the (α, β) -space.

\perp^a , proposed in [1], has some shortcomings. First, the “equivalence entails subsumption” property, which holds for crisp equivalence and crisp subsumption (in symbols, $\equiv \models \sqsubseteq$), is not preserved by their relaxed counterparts. That is, from equivalence with a confidence interval $[a, 1]$ one cannot entail subsumption with (at least) the same confidence: $\equiv^a \not\models \sqsubseteq^a$, for any $0 < a < 1$. Second, one would intuitively expect the relaxed disjointness and subsumption to be mutually exclusive, as it is the case with the crisp relations. However, this property does not hold either: for any $0 < a < 1$, the assertions $A \perp^a B$ and $A \sqsubseteq^a B$ do not contradict each other.

We overcome these drawbacks by refining the semantics of relaxed equivalence and disjointness as follows:

$$\equiv^a = \{(\alpha, \beta) \in \sigma : \alpha, \beta \geq a\}, \quad \perp^a = \{(\alpha, \beta) \in \sigma : \alpha, \beta \leq 1 - a\}.$$

These relations are visualized in Figure 5. From this definition it follows that \equiv^a is the intersection of \sqsubseteq^a and \supseteq^a . Moreover, relaxed disjointness \perp^a does not overlap with relaxed subsumption \sqsubseteq^a for any $a > 0.5$.

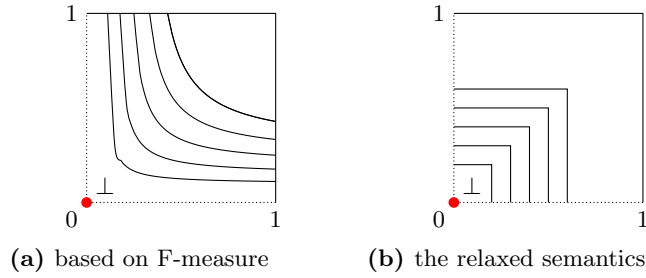


Fig. 6: Comparison of semantics for weighted disjointness.

It is now time to justify the discontinuity observed in the weighted semantics of [1] with the help of QTAX. The crisp semantics of \perp is the $(0, 0)$ point. The weighted semantics approaches it, but because it is the result of using the F-measure it always preserves the possibility that the segments $(0,1)$ and $(1,0)$ denote \perp because $\text{F-measure}(1, 0) = \text{F-measure}(0, 1) = 0$. This is what is shown in Figure 6a. Hence, the discontinuity comes from preserving these segments — and the points $(0,1)$ and $(1,0)$ which are in the interpretation of \sqsubseteq and \supseteq — whatever closed the weights are from crisp.

This is different when relations are approached by reducing a distance. This is illustrated in Figure 6b where \perp is continuously approximated through α with the Manhattan distance.

6 The calculus of relaxed taxonomic relations

In this section, we define the algebraic calculus of QTAX, which allows for composing the relaxed taxonomic relations (Section 6.1) and introduce two algebras which can be used for reasoning with alignments (Section 6.2).

6.1 Composition of relaxed taxonomic relations

Composition in QTAX distributes over union (Proposition 1). Thus, to compose two relaxed taxonomic relations, one has to compose pairwise all constituent base relations.

$$r^a \diamond s^b = \bigcup_{\substack{(\alpha, \beta) \in r \\ (\alpha', \beta') \in s}} (\alpha, \beta) \diamond (\alpha', \beta').$$

Before providing the formula for composing base QTAX-relations, let us introduce abbreviations for some relation symbols in $\wp(\sigma)$.

$$\begin{aligned} \text{INT}(\alpha_0, \alpha_1, k) &:= \{(\alpha, k\alpha) : \alpha_0 \leq \alpha \leq \alpha_1\} \\ \text{REC}(\alpha_0, \beta_0, \alpha_1, \beta_1) &:= \{(\alpha, \beta) : \alpha_0 \leq \alpha \leq \alpha_1 \text{ and } \beta_0 \leq \beta \leq \beta_1\} \end{aligned}$$

The relation symbols $\text{INT}(\alpha_0, \alpha_1, k)$, where $\alpha_0 \leq \alpha_1 \in [0, 1]_{\mathbb{Q}}$ and $0 < k\alpha_1 \leq 1$, correspond to intervals on the (α, β) -space, as shown in Figure 7a. We call them *interval relations* (not to confuse with Allen's temporal intervals). On the (α, β) -space these relations lie on a line which passes through the point $(0, 0)$. The relation symbols $\text{REC}(\alpha_0, \beta_0, \alpha_1, \beta_1)$, where $\alpha_0, \beta_0, \alpha_1, \beta_1 \in [0, 1]_{\mathbb{Q}}$, correspond to rectangles on the (α, β) -space, the edges of which are parallel to those of the unit square (Figure 7b). We call them *rectangle relations*.

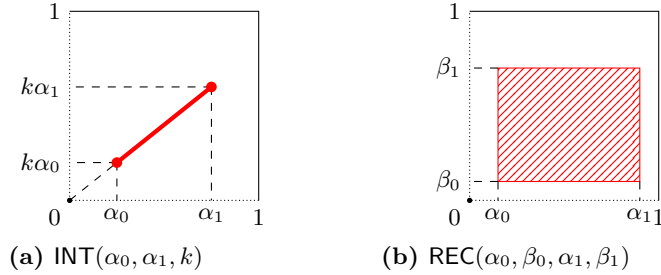


Fig. 7: Visual representation of interval and rectangle QTAX-relations.

Now we can formulate the main result. The composition of two base QTAX-relations is either a rectangle relation, if one of the base relations is the disjointness (Theorem 1). Otherwise, the composition is an interval relation.

Theorem 1.

$$(\alpha, \beta) \diamond (\alpha', \beta') = \begin{cases} \text{REC}(0, 0, 1, 1 - \beta'), & \text{if } \alpha, \beta = 0, \\ \text{REC}(0, 0, 1 - \alpha, 1), & \text{if } \alpha', \beta' = 0, \\ \text{INT}(\alpha''_0, \alpha''_1, \frac{\beta\beta'}{\alpha\alpha'}), & \text{if } \alpha, \beta, \alpha', \beta' \neq 0, \end{cases}$$

where

$$\alpha''_0 = \frac{\alpha}{\beta} \max(\alpha' + \beta - 1, 0),$$

$$\alpha''_1 = \min \left[1, \frac{\alpha\alpha'}{\beta\beta'}, \alpha \left(\min(1, \frac{\alpha'}{\beta}) + \min(\frac{\alpha'}{\beta} \frac{1 - \beta'}{\beta}, \frac{1 - \alpha}{\alpha}) \right) \right]$$

Proof. The proof can be found in [12].

6.2 Approximation and parametrization of QTAX relations

\mathbb{A}_{QTAX} is an algebra of relation symbols and not of actual binary relations. A relation symbol is a set of pairs (α, β) , where $\alpha, \beta \in [0, 1]_{\mathbb{Q}}$.

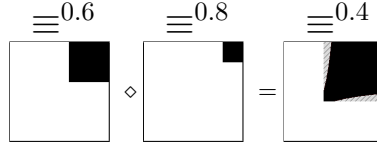


Fig. 8: REC-approximation of composition.

Composition of some relaxed taxonomic relations is visually represented in Figure 8. In general, the composition of such relations is not a relaxed taxonomic relation, but some “irregular” QTAX-relation represented by the black area. However, it can be always approximated by a rectangle relation, and in some cases even by another relaxed taxonomic relation, as shown in Figure 8 by the grayed area. The composition of relaxed equivalences $\equiv^{0.6}$ and $\equiv^{0.8}$ has a shape close to a rectangle. The REC-approximation of composition is $\equiv^{0.4}$.

All rectangle relations plus the empty relation are closed under intersection and contain the universal relation. Thus, weak composition \diamond_{REC} is a valid operation on the REC sublanguage of QTAX. The operation \diamond_{REC} can be specified based on numeric evaluation of a set of (α, β) relation symbols which constitute the composition in QTAX. A union of rectangle relations may not be a rectangle relation, but can always be approximated by one. This defines the operation of *weak union* on REC, denoted as \cup_w . The rectangle relations, together with operations of weak composition, converse intersection and weak union, form an algebra \mathbb{A}_{REC} :

$$\mathbb{A}_{\text{REC}} = (R, \cup_w, \cap, \emptyset, \text{REC}(0, 0, 1, 1), \diamond_{\text{REC}}, \smile), \quad (6.1)$$

| i | j | $(\equiv^i \diamond \equiv^j)$ | $(\equiv^i \diamond \sqsubseteq^j)$ | $(\equiv^i \diamond \supseteq^j)$ | $(\equiv^i \diamond \perp^j)$ | $(\sqsubseteq^i \diamond \sqsubseteq^j)$ | $(\sqsubseteq^i \diamond \perp^j)$ |
|-----|-----|--------------------------------|-------------------------------------|-----------------------------------|-------------------------------|--|------------------------------------|
| 0.5 | 0.5 | | | | | | |
| 0.5 | 0.6 | | | | | | |
| 0.5 | 0.7 | | | | | | |
| 0.5 | 0.8 | | | | | | |
| 0.5 | 0.9 | | | | | | |
| 0.5 | 1 | | | | | | |
| 0.6 | 0.5 | | | | | | |
| 0.6 | 0.6 | | | | | | |
| 0.6 | 0.7 | | | | | | |
| 0.6 | 0.8 | | | | | | |
| 0.6 | 0.9 | | | | | | |
| 0.6 | 1 | | | | | | |
| 0.7 | 0.5 | | | | | | |
| 0.7 | 0.6 | | | | | | |
| 0.7 | 0.7 | | | | | | |
| 0.7 | 0.8 | | | | | | |
| 0.7 | 0.9 | | | | | | |
| 0.7 | 1 | | | | | | |
| 0.8 | 0.5 | | | | | | |
| 0.8 | 0.6 | | | | | | |
| 0.8 | 0.7 | | | | | | |
| 0.8 | 0.8 | | | | | | |
| 0.8 | 0.9 | | | | | | |
| 0.8 | 1 | | | | | | |
| 0.9 | 0.5 | | | | | | |
| 0.9 | 0.6 | | | | | | |
| 0.9 | 0.7 | | | | | | |
| 0.9 | 0.8 | | | | | | |
| 0.9 | 0.9 | | | | | | |
| 0.9 | 1 | | | | | | |
| 1 | 0.5 | | | | | | |
| 1 | 0.6 | | | | | | |
| 1 | 0.7 | | | | | | |
| 1 | 0.8 | | | | | | |
| 1 | 0.9 | | | | | | |
| 1 | 1 | | | | | | |

Table 2: Composition of relaxed taxonomic relations visualized on the (α, β) -space.

where $R = \{\text{REC}(\alpha_0, \beta_0, \alpha_1, \beta_1) : \alpha_0 \leq \alpha_1, \beta_0 \leq \beta_1 \in [0, 1]_{\mathbb{Q}}\} \cup \{\emptyset\}$. A general formula for composing relaxed equivalence relations is the following:

$$\equiv^x \diamond_{\text{REC}} \equiv^y = \equiv^{\max(0, x+y-1)} \quad (6.2)$$

Similar formulas can be obtained for other pairs of relaxed taxonomic relations.

Another approach to make \mathbb{A}_{QTAX} computationally feasible is to discretize the (α, β) -space as an $n \times n$ matrix and thus obtain a finite algebra $\mathbb{A}_{\text{QTAX}}^n$. This approach was used for computing the composition of relaxed taxonomic relations in Table 2.

7 Application to reasoning with ontology alignments

The relaxed semantics of taxonomic relations can be used by ontology matchers. Some matchers induce relations between classes based on the instance-level data. Since the semantic web is an open environment with potentially invalid data, many instance-based matchers induce a relation between two concepts, if it holds for *most* instances of these concepts. The level of fault-tolerance is usually set by a threshold. This threshold may be expressed as the weight of an ontology alignment relation, in compliance with the relaxed semantics.

To reason with weighted ontology alignments, both algebras \mathbb{A}_{REC} or $\mathbb{A}_{\text{QTAX}}^n$ can be used. The algebra \mathbb{A}_{REC} contains infinitely many relations, but is computationally feasible, since REC-relations are finitely parametrized. However, using \mathbb{A}_{REC} for automated reasoning requires adjustments to the existing reasoning algorithms, which are designed for finite algebras. The algebras $\mathbb{A}_{\text{QTAX}}^n$ are finite, thus can be used with existing reasoning tools that support qualitative calculi.

8 Summary and conclusion

Weights in ontology alignments have been widely adopted. This paper shows how to define algebraic calculi which can be used for expressing both the relation and the weight of correspondences. Its goal is to be able to provide sound compositional reasoning for alignments.

We introduced the \mathbb{A}_{QTAX} calculus of relaxed taxonomic relations generalising the previous weighted semantics as well as the semantics of crisp relations. We provided a semantics that overcomes the problems identified and, in particular, discontinuity. \mathbb{A}_{QTAX} composition is not computationally feasible, however we discussed two different ways to make it computationally feasible: \mathbb{A}_{REC} based on rectangular approximation of these relations and $\mathbb{A}_{\text{QTAX}}^n$ based on a discretization of the (α, β) -space.

On the one hand, this proposal provides a way to reason by composition with weighted alignment that is well grounded and can compose any relation. On the other hand, [1] gave rules for reasoning with concept constructors which are absent here. It would be worth studying if such rules still holds and can be generalised to the new context.

Acknowledgement: This research has been partially supported by the joint NSFC-ANR Lindicle project (12-IS01-0002) with Tsinghua university.

Bibliography

1. M. Atencia, A. Borgida, J. Euzenat, C. Ghidini, and L. Serafini. “A formal semantics for weighted ontology mappings”. In: *Proc. of ISWC-12*. Springer, 2012, pp. 17–33.
2. F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, eds. *The description logic handbook: theory, implementation, and applications*. Cambridge University Press, 2003.
3. M. Bodirsky and V. Dalmau. “Datalog and Constraint Satisfaction with Infinite Templates”. *J. Comput. Syst. Sci.* 79.1 (2013), pp. 79–100.
4. A. Borgida and L. Serafini. “Distributed description logics: assimilating information from peer sources”. *J. Data Semant.* 1 (2003), pp. 153–184.
5. F. Dylla, T. Mossakowski, T. Schneider, and D. Wolter. “Algebraic properties of qualitative spatio-temporal calculi”. In: *Proc. of COSIT-13*. Springer, 2013, pp. 516–536.
6. J. Euzenat. “Algebras of Ontology Alignment Relations”. In: *Proc. of ISWC-08*. 2008, pp. 387–402.
7. J. Euzenat. “An Algebraic Approach to Granularity in Qualitative Time and Space Representation”. In: *Proc. of IJCAI-95*. 1995, pp. 894–900.
8. J. Euzenat and P. Shvaiko. *Ontology Matching*. Second edition. Heidelberg: Springer, 2013.
9. Z. Gantner, M. Westphal, and S. Wöfl. “GQR-A fast reasoner for binary qualitative constraint calculi”. In: *Proc. of AAAI*. Vol. 8. 2008.
10. S. Givant and P. Halmos. *Introduction to Boolean Algebras*. Springer Undergraduate Texts in Mathematics and Technology. Springer, 2009.
11. W. Hodges. *Model Theory. Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 1993.
12. A. Inants. “Qualitative calculi with heterogeneous universes”. PhD thesis. University of Grenoble Alpes, 2016.
13. A. Inants and J. Euzenat. “An algebra of qualitative taxonomical relations for ontology alignments”. In: *Proc. of ISWC-15*. 2015, pp. 253–268.
14. G. Ligozat and J. Renz. “What Is a Qualitative Calculus? A General Framework”. In: *Proc. of PRICAI-04*. 2004, pp. 53–64.
15. A. Locoro, J. David, and J. Euzenat. “Context-Based Matching: Design of a Flexible Framework and Experiment”. *J. Data Semant.* 3.1 (2014), pp. 25–46.
16. T. Lukasiewicz, L. Predoiu, and H. Stuckenschmidt. “Tightly integrated probabilistic description logic programs for representing ontology mappings”. *Ann. Math. Artif. Intell.* 63.3-4 (2011), pp. 385–425.
17. D. Wolter. “SparQ - A Spatial Reasoning Toolbox”. In: *AAAI Spring Symposium: Benchmarking of Qualitative Spatial and Temporal Reasoning Systems*. 2009, p. 53.
18. A. Zimmermann and J. Euzenat. “Three Semantics for Distributed Systems and Their Relations with Alignment Composition”. In: *Proc. of ISWC-06*. 2006, pp. 16–29.