



Region-Based Merging of Open-Domain Terminological Knowledge

Zied Bouraoui, Sébastien Konieczny, Thanh Ma, Nicolas Schwind, Ivan Varzinczak

► To cite this version:

Zied Bouraoui, Sébastien Konieczny, Thanh Ma, Nicolas Schwind, Ivan Varzinczak. Region-Based Merging of Open-Domain Terminological Knowledge. 19th International Conference on Principles of Knowledge Representation and Reasoning KR-2022, Jul 2022, Haifa, Israel. pp.81-90, 10.24963/kr.2022/9 . hal-03855336

HAL Id: hal-03855336

<https://hal.science/hal-03855336>

Submitted on 16 Nov 2022

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.

Region-Based Merging of Open-Domain Terminological Knowledge

Zied Bouraoui¹, Sébastien Konieczny¹, Thanh Ma¹, Nicolas Schwind², Ivan Varzinczak¹

¹CRIL, Univ. Artois & CNRS, France

²National Institute of Advanced Industrial Science and Technology, Tokyo, Japan

nicolas-schwind@aist.go.jp, {bouraoui,konieczny,ma,varzinczak}@cril.fr

Abstract

This paper introduces a novel method for merging open-domain terminological knowledge. It takes advantage of the Region Connection Calculus (RCC5), a formalism used to represent regions in a topological space and to reason about their set-theoretic relationships. To this end, we first propose a faithful translation of terminological knowledge provided by several and potentially conflicting sources into region spaces. The merging is then performed on these spaces, and the result is translated back into the underlying language of the input sources. Our approach allows us to benefit from the expressivity and the flexibility of RCC5 while dealing with conflicting knowledge in a principled way.

1 Introduction

Commonsense knowledge is playing an increasingly important role in the development of AI systems. Such knowledge is available, for example, in large open-domain terminological knowledge bases Cyc or SUMO as ontological knowledge, in knowledge graphs (KGs) such as DBpedia and WikiData, and as semantic markup (e.g., RDFa). Ontologies, as frameworks for expressing terminological knowledge, encode structured relations about the concepts and properties of a given domain. In this paper, we focus on terminological knowledge about concepts that can be expressed using ontologies as these latter are playing an important role in areas such as semantic web (Homburg, Staab, and Janke 2020), information retrieval (Chen et al. 2019), natural language processing (Rospoche and Corcoglioniti 2018), and machine learning (Hohenecker and Lukasiewicz 2020), among others. For instance, Bouraoui and Schockaert (2018) have shown that knowledge encoded in ontologies, as prior conceptual knowledge, is useful for learning concept representations from few examples.

However, the available ontologies (and KGs, as simple ontologies) are inevitably incomplete, where several rules and facts are missing. Several methods have been proposed for automated ontology (KG) completion (Beltagy et al. 2013; Rocktäschel and Riedel 2017; Li, Bouraoui, and Schockaert 2019) that exploit statistical regularities in a given ontology to predict plausible missing rules or facts. Unfortunately, meaningful knowledge is difficult to predict, especially when only a few examples of facts or rules are available. Moreover, as most of the existing approaches are

mainly based on statistical regularities, the resulting predictions might be conflicting with each other. A repair-based mechanism is then required to maintain the consistency of the set of terminological statements, i.e., ensure that there are no conflicting (or contradictory) statements. In the same perspective, to widen the coverage of terminological knowledge to several domains and to deal with incompleteness and conflicting statements, one may combine knowledge from several sources. However, it turns out that merging open-domain knowledge bases is a particularly challenging task as pointed out, for example, in (Tanon et al. 2016) reporting the different problems and difficulties encountered when merging Freebase with WikiData. Conflicting information may occur when the statements of several sources are simply gathered together. Ontology merging and alignment / matching has also attracted much attention in the literature (Thiéblin, Haemmerlé, and Trojahn 2018; Chang, wei Chen, and Zhang 2019; Benferhat et al. 2019; Zhao and Zhang 2016; Laadhar et al. 2017; Thiéblin, Haemmerlé, and Trojahn 2018). The approach generally consists in studying an equivalence matching between a source and target taxonomy. In this work, we assume that all knowledge encoded by the different sources are already aligned and mapped to each other. Namely, we suppose that they share the same terminology. Hence, while ontology alignment (or matching) is the process of determining correspondences between terminologies of ontologies, ontology *merging* aims to combine two (or more) ontologies having the same terminology while handling conflicting statements.

Let us consider an example to illustrate the merging problem. Assume that a first source says that the concept *Paper* is disjoint with the concept *Document*, while another source says that every *Paper* is a *Document*. Obviously enough, these two statements are conflicting. To be faithful to both sources while resolving conflicts, a sensible choice would be to assume that *Paper* and *Document* are not disjoint concepts, but every *Paper* is not necessarily a *Document*, that is, the two concepts *partially overlap*. This kind of result is clearly consistent and can be seen as a good compromise between both sources. Finding a meaningful and relevant compromise between sources during the merging process is difficult to obtain. This is mainly due the fact that ontology languages (Description Logics for example) are not expressive enough to capture salient knowledge that might emerge

during the merging process. The simple example pointed out above clearly shows some pieces of knowledge that should be taken into account during the merging process, but that cannot be captured in the ontology language. The problem of ontology (or DL) merging is close to the problem of belief merging in a propositional setting (Benferhat et al. 2019; Benferhat et al. 2014; Kumar and Harding 2016; Wang et al. 2012; Patricia, Konieczny, and Marquis 2008). For instance, Benferhat et al. (2019) studied merging assertional bases in the *DL-Lite* fragment. They have determined the minimal subsets of assertions to resolve conflicts based on the inconsistency minimization principle. Bouraoui et al. (2020a) proposed a model-based merging operator for merging $\mathcal{EL}$ ontologies which solves semantic conflicts that arise during the merging process.

In this paper, to handle the merging problem, we take inspiration from conceptual spaces, which are geometric representation frameworks, in which the objects are represented as points in a topological space, and concepts are modelled as regions (Haarslev and Möller 1997; Gärdenfors 2000; Douven et al. 2022). Motivated by the fact that conceptual knowledge in an ontology can be to some extent modelled as geometric objects and constraints on metric spaces (Bouraoui et al. 2020b), this paper proposes a method for ontology merging that takes advantage of *qualitative spatial reasoning* to find out a relevant compromise between sources while resolving conflicts. Qualitative spatial reasoning is a suitable paradigm for efficiently reasoning about spatial entities and their relationships, where knowledge is represented as a so-called *qualitative constraint network* (QCN). Spatial information is usually represented in terms of basic or non-basic relations in a qualitative calculus, where reasoning tasks are then formulated as solving a set of qualitative constraints (Randell, Cui, and Cohn 1992; Cohn et al. 1997; Bhatt et al. 2011; Sioutis, Long, and Janhunen 2020). In particular, the Region Connection Calculus (RCC) is a well-studied formalism for qualitative topological representation and reasoning, including its subsets *RCC-5* (Schockaert and Li 2013) and *RCC-8* (Randell, Cui, and Cohn 1992). Two significant advantages of the RCC framework are its ability to reason efficiently about the relationships between spatial entities, and its ability to deal with conflicts in qualitative constraint merging as shown in (Condotta et al. 2010; Thau, Bowers, and Ludäscher 2009). In short, the representation of region constraints into QCNs allows for more expressivity than when using DL rules (or constraints). In particular, QCNs are expressive enough to allow for disjunctions in the constraints.

Several QCN merging operators have been introduced in the literature (e.g., (Condotta et al. 2009; Condotta et al. 2010; Thau, Bowers, and Ludäscher 2009; Julien and Westphal 2012). Roughly speaking, these operators compute a distance between QCN scenarios and the input QCNs. Then the scenarios with a minimal distance are selected as the best candidates for the merged result. Taking inspiration from these works, in this paper we take advantage of the *RCC-5* formalism and propose a method for merging open-domain terminological knowledge (simply called ontologies) using QCNs. We first show how to translate such knowledge into

qualitative spaces while preserving its semantics and properties. Second, we propose a merging operator that produces a single and consistent region space representing a compromise between sources. Lastly, we show how to express the region space in the input ontology language while maintaining all relevant information.

The proofs of propositions are available online¹.

2 Background

Our method is based on two complementary frameworks for merging open-domain terminological knowledge: we rely on a lightweight Description Logic (DL) framework to encode knowledge and use *RCC-5* and qualitative constraints for performing the merging. This section briefly recalls the technical background required on these two topics.

Description Logics. $\mathcal{EL}$ is a family of lightweight DLs, which underlies the Ontology Web Language profile OWL2-EL, that is considered as one of the main representation formalisms to express terminological knowledge (Baader, Brandt, and Lutz 2005). The main ingredients of DLs are individuals, concepts, and roles, which correspond at the semantic level to objects, sets of objects, and binary relations between objects. More formally, let N_C , N_R , N_I be three pairwise disjoint sets where N_C denotes a set of atomic concepts, N_R denotes a set of atomic relations (roles), and N_I denotes a set of individuals. In this paper, we consider $\mathcal{EL}_\perp$ concept expressions (Kriegel 2020) which are built according to the following grammar:

$$C ::= \top \mid \perp \mid N_C \mid C \sqcap C \mid \exists r.C \text{ where } r \in N_R.$$

Let $C, D \in N_C$, $a, b \in N_I$, and $r \in N_R$. An $\mathcal{EL}$ ontology $O = \langle \mathcal{T}, \mathcal{A} \rangle$ (a.k.a. knowledge base) comprises two components, the TBox (Terminological Box denoted by $\mathcal{T}$) and ABox (denoted by $\mathcal{A}$). The TBox consists of a set of General Concept Inclusion (GCI) axioms of the form $C \sqsubseteq D$, meaning that C is more specific than D or simply C is subsumed by D , and axioms of the form $C \sqcap D \sqsubseteq \perp$, meaning that C and D are disjoint concepts. The ABox is a finite set of assertions on individual objects of the form $C(a)$ or $r(a, b)$.

The semantics is given in terms of interpretations $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, which consist of a non-empty interpretation domain $\Delta^{\mathcal{I}}$ and an interpretation function $\cdot^{\mathcal{I}}$ that maps each individual $a \in N_I$ into an element $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$, each concept $A \in N_C$ into a subset $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, and each role $r \in N_R$ into a subset $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$.

A summary of the syntax and semantics of $\mathcal{EL}_\perp$ is shown in Table 1. An interpretation $\mathcal{I}$ is said to be a model of (or satisfies) an axiom Φ in the form of the left column in the table, denoted by $\mathcal{I} \models \Phi$, when the corresponding condition in the right column is satisfied. For instance, $\mathcal{I} \models C \sqsubseteq D$ if and only if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. Similarly, $\mathcal{I}$ satisfies a concept (resp. role) assertion, denoted by $\mathcal{I} \models C(a)$ (resp. $\mathcal{I} \models r(a, b)$), if $a^{\mathcal{I}} \in C^{\mathcal{I}}$ (resp. $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$). An interpretation $\mathcal{I}$ is a model of an ontology O if it satisfies all the axioms and assertions in O . An ontology is said to be consistent if it has

¹<https://arxiv.org/abs/2205.02660>

Syntax	Semantics
$C \sqsubseteq D$	$C^I \subseteq D^I$
r	$r^I \subseteq \Delta^I \times \Delta^I$
a	$a^I \in \Delta^I$
$C \sqcap D$	$C^I \cap D^I$
$\top$	Δ^I
$\perp$	$\emptyset$
$\exists r.C$	$\{x \in \Delta^I \mid \exists y \in \Delta^I s.t. (x, y) \in r^I, y \in C^I\}$

Table 1: Syntax and semantics of $\mathcal{EL}_\perp$

a model. Otherwise, it is inconsistent. An axiom Φ is entailed by an ontology, denoted by $O \models \Phi$, if Φ is satisfied by every model of O . We say that C is subsumed by D w.r.t. an ontology O iff $O \models C \sqsubseteq D$. Similarly, we say that a is an instance of C w.r.t. O iff $O \models C(a)$. An interpretation $\mathcal{I} = (\Delta^I, \cdot^I)$ is said to be *fulfilling* when each concept name in the ontology is non-empty in $\mathcal{I}$, i.e., for each concept $C_i \in N_C$, $\cdot^I(C_i) \neq \emptyset$.

The main reasoning task that is considered in terminological ontologies is classification. It consists in computing all the entailed subsumptions ($C \sqsubseteq D$) (and equivalences ($C \equiv D$)) that hold between atomic concepts of an ontology, or the concepts $\top$ or $\perp$. Such a procedure is described in (Baader, Brandt, and Lutz 2005), which first consists in transforming the ontology into a normal form using a set of rules, and then performing a classification reasoning process using the set of inference (completion) rules (see (Baader, Brandt, and Lutz 2005) for more details). In this paper, we assume that the input ontologies are provided in a specific normal form, to which we apply completion rules for classification. This classification step is to normalize the input ontologies. The reason of conducting normalization is to handle and transform the complex axioms into the axioms of all atomic concepts to be simpler for the translation process. That is, before translating the input ontologies into a region-based representation, we assume that each of them is in the *strict normal form*, i.e., if its TBox $\mathcal{T}$ only consists of inclusions of the fundamental form: $A \sqsubseteq B$, $A \sqcap B \sqsubseteq \perp$, $A \sqsubseteq \exists r.B$ and $\exists r.A \sqsubseteq B$ where $A, B \in N_C$. Such an assumption is made without loss of generality, since for each ontology O , one can compute an ontology O' in the strict normal form in polynomial time (Baader, Brandt, and Lutz 2005).

Region Connection and Qualitative Constraints. The RCC (Region Connection Calculus) formalism allows one to represent and reason about the relationships between spatial entities (Randell, Cui, and Cohn 1992). Among the fragments of the RCC theory, RCC-5 fragment is expressive enough to reason about set-theoretic relations between regions (Bennett 1994; Peter and Thomas 1997). In RCC-5, regions can simply be interpreted as non-empty subsets of a given set and the focus is given on a set $\mathcal{B} = \{DR, PO, EQ, PP, PPI\}$ of five binary relations between regions called *basic relations*. The set $\mathcal{B}$ forms a jointly exhaustive and pairwise disjoint set of relations, that is, each

Name (Symbol)	Syntax	Semantics
Proper Part of (PP)	$v_C\{PP\}v_D$	$v_C^S \subset v_D^S$
Inverse PP of (PPI)	$v_C\{PPI\}v_D$	$v_D^S \subset v_C^S$
Equals (EQ)	$v_C\{EQ\}v_D$	$v_C^S = v_D^S$
Disjoint From (DR)	$v_C\{DR\}v_D$	$v_C^S \cap v_D^S = \emptyset$
Partially Overlaps (PO)	$v_C\{PO\}v_D$	$v_C^S \cap v_D^S \neq \emptyset$ $v_C^S \not\subseteq v_D^S, v_D^S \not\subseteq v_C^S$

Table 2: Syntax and semantics of $RCC\text{-}5$, $v_C, v_D \in V$.

pair of regions satisfies exactly one relation from $\mathcal{B}$: the relation DR (resp. PO, EQ, PP) holds between two regions whenever the two regions are disjoint (resp. when they partially overlap, are equal, when the first is a strict subset of the second), and PPI is the converse of PP . Based on $\mathcal{B}$, more complex pieces of information about the relative positions of a set of regions can be represented by means of qualitative constraint networks (QCNs). Formally, a QCN is a pair $\mathcal{N} = \langle V, \Psi \rangle$, where $V = \{v_C, v_D, \dots\}$ is a set of region variables representing the spatial entities and Ψ is a set of binary constraints between these entities. Each constraint $\Psi_{CD} \in \Psi$ is a mapping from $V \times V$ to $2^{\mathcal{B}}$, and is simply denoted by $\Psi_{CD} = v_C \varphi v_D$, where $\varphi \subseteq 2^{\mathcal{B}}$; and Ψ_{CD} is said to be a singleton constraint whenever φ is a singleton.

An interpretation of a QCN $\mathcal{N}$ is defined as $\mathcal{S} = (\mathcal{D}^S, \cdot^S)$, where $\mathcal{D}^S$ is a non-empty set (the domain of the regions), and $\cdot^S$ is an interpretation function which maps each variable v_C to a non-empty subset v_C^S of $\mathcal{D}^S$. Table 2 precises how singleton constraints from Ψ are interpreted in $RCC\text{-}5$, i.e., an interpretation $\mathcal{S}$ of $\mathcal{N}$ satisfies a singleton constraint Ψ_{CD} , denoted by $\mathcal{S} \models \mathcal{N}$, if the relation between v_C^S and v_D^S according to the table is satisfied (e.g., $\mathcal{S} \models v_C\{PP\}v_D$ whenever $v_C^S \subset v_D^S$). The satisfaction relation is extended to any (non-singleton) constraint from $2^{\mathcal{B}}$ as follows: for each $\varphi \in 2^{\mathcal{B}}$, $\mathcal{S} \models v_C \varphi v_D$ iff $\mathcal{S} \models v_C \{\varphi_i\} v_D$ for some $\varphi_i \in \varphi$ (e.g., $\mathcal{S} \models v_C\{PP, EQ\}v_D$ iff $\mathcal{S} \models v_C\{PP\}v_D$ or $\mathcal{S} \models v_C\{EQ\}v_D$). An interpretation $\mathcal{S}$ of a QCN $\mathcal{N} = \langle V, \Psi \rangle$ is said to be a solution of $\mathcal{N}$, denoted by $\mathcal{S} \models \mathcal{N}$, iff $\mathcal{S} \models \Psi_{CD}$ for each $\Psi_{CD} \in \Psi$. A QCN is consistent iff it admits a solution. A sub-network of $\mathcal{N}$ is a QCN $\mathcal{N}' = (V, \Psi')$ such that $\Psi' \subseteq \Psi$. A quasi-atomic QCN $\langle V, \Psi \rangle$ is a QCN where for each $v_C, v_D \in V$, there is a unique constraint $\Psi_{CD} \in \Psi$, and where Ψ_{CD} is either a singleton or $\Psi_{CD} \in \{\{PP, EQ\}, \{PPI, EQ\}\}$. A scenario of a QCN is a quasi-atomic sub-network of $\mathcal{N}$. Noteworthy, a QCN is consistent if it admits a consistent scenario.

3 Merging Framework Description

This short section summarizes our method for merging open-domain terminological knowledge (i.e., ontologies) using qualitative constraint networks (QCNs). As highlighted in the introduction, QCNs are expressive enough to capture some relevant information that might emerge during the merging process, which in turn allows one to select a consistent compromise between sources when expressing

the merging result.

The first task is to find a “faithful translation” from an ontology to a QCN that preserves the semantics and maintains the initial knowledge encoded in the ontology. Then, it is assumed that every input ontology is translated into a QCN, so we are given a multiset of QCNs, also called *profile*. The second task is then to define a merging operator that associates the QCN profile into a single QCN, representing the QCN profile in a global and consistent way. As constraints of the merged QCN are *sets* of basic relations, this QCN cannot be translated back into the target ontology language in the general case. So we select one of its “best” consistent scenarios which, in contrast, can be expressed as an ontology.

More precisely, our approach involves the following main steps:

1. The translation of terminological knowledge sources (i.e., the input ontologies given in the strict normal form) into QCNs (Section 4). In this step, we present a translation function that ensures the faithfulness of the translation from an ontology into a QCN.
2. The definition of a QCN merging operator (Section 5). Exploiting the notion of “distance” between basic relations and constraints, this step associates the input QCN profile (the translated ontologies) with a single merged, consistent QCN.
3. The selection of a “best” consistent scenario of the merged QCN as a candidate of the merged result (Section 6). This selection process takes advantage of the notion of distance between scenarios and some plausible instantiation of the input ontologies (i.e., some interpretations).
4. The translation of the selected consistent scenario back into an ontology, i.e., the underlying language of the input sources (Section 7).

4 Translating Terminological Knowledge into QCNs

In this section, we present a translation function from any ontology to a QCN. More precisely, we (1) map atomic concepts names into QCN variables and axioms into constraints, and (2) show that the translation is faithful to the TBox of the initial ontology.

Definition 1 (Forward translation $\tau_{\triangleright}$). *A forward translation is a function $\tau_{\triangleright} : N_C \rightarrow V$ s.t. $\tau_{\triangleright}(C) \stackrel{\text{def}}{=} v_C$. $\tau_{\triangleright}$ is extended to map axioms in the (strict) normal form into constraints as follows:*

- $\tau_{\triangleright}(C \sqsubseteq D) \stackrel{\text{def}}{=} \tau_{\triangleright}(C)\{PP, EQ\}\tau_{\triangleright}(D)$, and
- $\tau_{\triangleright}(C \sqcap D \sqsubseteq \perp) \stackrel{\text{def}}{=} \tau_{\triangleright}(C)\{DR\}\tau_{\triangleright}(D)$.

Moreover, $\tau_{\triangleright}$ is extended to translate ontologies in the (strict) normal form into a set of constraints in the expected way: $\tau_{\triangleright}(O) \stackrel{\text{def}}{=} \{\tau_{\triangleright}(\Phi) \mid \Phi \in O\}$.

To show that the translation is faithful, we provide a “semantic” mapping from O to $\tau_{\triangleright}(O)$, and conversely. Let us first show how models of O correspond to solutions of $\tau_{\triangleright}(O)$:

Definition 2 (Flattening of an interpretation). *Let $I = (\Delta^I, \cdot^I)$ be a fulfilling interpretation. With $S_I \stackrel{\text{def}}{=} (\Delta^I, \cdot^{S_I})$ we denote the **flattening** of I , where $\cdot^{S_I} : V \rightarrow 2^{\Delta^I}$ is such that $(v_C)^{S_I} = C^I$.*

Notably, non-fulfilling interpretations are not considered (are irrelevant) in the paper because the corresponding translation regions cannot be empty (are non-empty subsets) in RCC-5.

Theorem 1. *Let O be an ontology, and let I be a fulfilling interpretation of O such that $I \models O$. Then $S_I \models \tau_{\triangleright}(O)$.*

The other way around, let us show how solutions of $\tau_{\triangleright}(O)$ correspond to interpretations satisfying all axioms from O .

Definition 3 (Inflation of a solution). *Let $S = (\mathcal{D}^S, \cdot^S)$ be a semantic solution to a QCN N over V . With $I_S \stackrel{\text{def}}{=} (\Delta^{I_S}, \cdot^{I_S})$, where $\Delta^{I_S} = \mathcal{D}^S$ and, for every $A \in N_C$, $A^{I_S} = (v_A)^S$, we call I_S an **inflation** of S .*

Intuitively, the inflation of S corresponds to an interpretation blown up from S by interpreting atomic concept names in the same way their corresponding variable names are “populated” by the solution. Notice that there are as many possible inflations of S as there are ways of interpreting N_R and N_I over Δ^{I_S} . An immediate consequence of Definition 3 is that every inflation of a solution S is fulfilling.

Theorem 2. *Let O be an ontology and let S be a solution of $\tau_{\triangleright}(O)$. Then there is an inflation I_S of S s.t. $I_S \models \Phi$ for each axiom Φ of O .*

Theorems 1 and 2 establish that our translation is faithful, i.e., that the set of all fulfilling models of an ontology O are captured precisely in its translated QCN $\tau_{\triangleright}(O)$.

5 QCN Merging

We reduce the merging of a profile of ontologies $\mathcal{P} = \langle O^1, \dots, O^n \rangle$ to the merging of a profile of QCNs $\mathcal{N} = \langle N^1, \dots, N^n \rangle$, where for each $i \in \{1, \dots, n\}$, $N^i = (V, \Psi^i) = \tau_{\triangleright}(O^i)$, based on the faithful translation given the previous section. Inspired by works on syntactical QCN merging (Condotta et al. 2010), this QCN merging process is summarized as follows. We associate with the profile $\mathcal{N}$ a single merged and consistent QCN $N = (V, \Psi)$ representing $\mathcal{N}$ in a “global” way. This is performed in a constraint-wise fashion: for each pair of variables $v_C, v_D \in V$, we associate each basic relation $b \in \mathcal{B}$ with a value representing its *distance* to the profile of constraints $\mathcal{E}_{CD} = \langle \Psi_{CD}^1, \dots, \Psi_{CD}^n \rangle$. This distance is the key tool to form the constraint Ψ_{CD} of the merged QCN N . Intuitively, each constraint Ψ_{CD} corresponds to the set of basic relations with the lowest distances to the profile $\mathcal{E}_{CD}$, while ensuring that the resulting QCN is consistent.

Example 1. *Consider the profile of ontologies $\mathcal{P} = \langle O^1, O^2, O^3, O^4 \rangle$ that encodes the following knowledge about the four concepts of Paper, Text, Document and Book, respectively denoted by P, T, D and B^2 .*

²An implementation to illustrate our method is also made available at the following link: <https://github.com/ontologymerging/MergingOntologyWithQCN>.

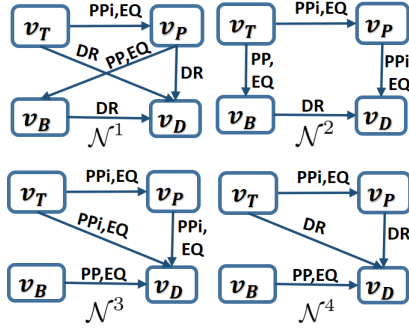


Figure 1: QCN profile of Example 1

- $O^1 = \langle \mathcal{T}^1 = \{P \sqsubseteq T, T \sqcap D \sqsubseteq \perp, P \sqsubseteq B, P \sqcap D \sqsubseteq \perp, B \sqcap D \sqsubseteq \perp\}, \mathcal{A}^1 = \{P(p_1), T(p_1), T(t_1), D(d_1), B(b_1)\} \rangle$,
- $O^2 = \langle \mathcal{T}^2 = \{P \sqsubseteq T, T \sqsubseteq B, D \sqsubseteq B, D \sqsubseteq P\}, \mathcal{A}^2 = \{P(p_2), P(d_2), T(p_2), T(t_2), D(d_2), B(p_2), B(b_2)\} \rangle$,
- $O^3 = \langle \mathcal{T}^3 = \{B \sqsubseteq D, D \sqsubseteq P, P \sqsubseteq T, D \sqsubseteq T\}, \mathcal{A}^3 = \{P(p_3), T(t_3), T(p_3), D(d_3), B(b_3)\} \rangle$,
- $O^4 = \langle \mathcal{T}^4 = \{D \sqcap P \sqsubseteq \perp, P \sqsubseteq T, B \sqsubseteq D, T \sqcap B \sqsubseteq \perp, T \sqcap D \sqsubseteq \perp\}, \mathcal{A}^4 = \{P(p_4), T(t_4), T(p_4), D(b_4), D(d_4), B(b_4)\} \rangle$.

Using the forward translation $\tau_{\triangleright}$ presented in the previous section, one associates with $\mathcal{P}$ a profile of QCNs $\langle \tau_{\triangleright}(O^i) \rangle = \langle \mathcal{N}^i \rangle$ ($i \in \{1, \dots, 4\}$). The four QCNs are depicted in Figure 1 (to alleviate the figures, we do not represent the Ψ_{CD} when $\Psi_{CD} = \mathcal{B}$, i.e., when the QCN does not provide any information between the relationship between v_C and v_D).

Although we do not require it in the general case, note that in this example each ontology O^i is consistent, i.e., the TBox of each input ontology does not contain conflicting information. As a direct consequence of Theorem 1, each QCN is consistent. However, simply combining these QCNs can easily lead to conflicts. For instance, there is no basic relation shared in the constraints Ψ_{BD}^1 and Ψ_{BD}^3 , since in $\mathcal{N}^1$ we have that $Book \{DR\} Document$ whereas in $\mathcal{N}^3$ we have that $Book \{PP, EQ\} Document$. This calls for our QCN merging procedure.

5.1 Computing a Distance between a Basic Relation and a Profile of Constraints

We start by considering a distance between basic relations. Firstly introduced in (Freksa 1992) in the context of temporal reasoning, the notion of *conceptual neighborhood* (CN) between relations was later adapted to QCN merging in (Condotta, Kaci, and Schwind 2008) to define such a distance. Intuitively, two basic relations $b, b' \in \mathcal{B}$ are CNs if a continuous transformation of two regions which satisfy the basic relation $b \in \mathcal{B}$ leads them to directly satisfy the basic relation b' without satisfying any other basic relation from $\mathcal{B}$. For instance, PP and EQ are CNs since “shrinking” a first region C initially equal to another

region D directly makes it a proper part of D . This results in the neighborhood relation $\{(DR, PO), (PO, DR), (PO, PP), (PP, PO), (PO, PPI), (PPI, PO), (PP, EQ), (EQ, PP), (PPI, EQ), (EQ, PPI)\}$. This neighborhood relation induces a *neighborhood graph* $\mathcal{G}$ whose vertices are the elements of $\mathcal{B}$, and where there is an edge between two basic relations $b, b' \in \mathcal{B}$ whenever b and b' are CNs according to the neighborhood relation. The distance $d(b, b')$ between two basic relations b and b' is defined as the length of the shortest path between b and b' in the neighborhood graph. For instance, $d(DR, PO) = 1$ since DR and PO are CNs, and $d(DR, EQ) = 3$ since DR and PO (resp. (PO, PP) and (PP, EQ)) are CNs, but DR and EQ are not. This distance is extended to a distance between a basic relation $b \in \mathcal{B}$ and a constraint $\varphi \in 2^{\mathcal{B}}$, defined as $d(b, \varphi) = \min_{b' \in \varphi} d(b, b')$. Lastly, given two variables $v_C, v_D \in V$, the distance between each $b \in \mathcal{B}$ and the profile of constraints $\mathcal{E}_{CD} = \langle \Psi_{CD}^1, \dots, \Psi_{CD}^n \rangle$ is defined by $d(b, \mathcal{E}_{CD}) = \sum_{i \in \{1, \dots, n\}} d(b, \Psi_{CD}^i)$.

Example 1 (continued). Let us focus on Text (T) and Document (D). We have $\mathcal{E}_{TD} = \langle \Psi_{TD}^1, \Psi_{TD}^2, \Psi_{TD}^3, \Psi_{TD}^4 \rangle = \langle \{DR\}, \mathcal{B}, \{PPI, EQ\}, \{DR\} \rangle$. For the distance between the basic relation PP and $\mathcal{E}_{TD}$, we have that $d(PP, \mathcal{E}_{TD}) = d(b, \Psi_{TD}^1) + d(b, \Psi_{TD}^2) + d(b, \Psi_{TD}^3) + d(b, \Psi_{TD}^4)$, where:

$$\begin{aligned} d(PP, \Psi_{TD}^1) &= d(PP, \{DR\}) = d(PP, DR) = 2, \\ d(PP, \Psi_{TD}^2) &= d(PP, \mathcal{B}) = d(PP, PP) = 0, \\ d(PP, \Psi_{TD}^3) &= d(PP, \{PPI, EQ\}) = d(PP, EQ) = 1, \\ d(PP, \Psi_{TD}^4) &= d(PP, \{DR\}) = d(PP, DR) = 2. \end{aligned}$$

We get that $d(PP, \mathcal{E}_{TD}) = 5$. The distances between each basic relation from $\mathcal{B}$ and the profile of constraints $\mathcal{E}_{CD}$ for each pair of variables $v_C, v_D \in V$ is summarized in Table 3.

$\mathcal{B}$	$\mathcal{E}_{TP}$	$\mathcal{E}_{TB}$	$\mathcal{E}_{TD}$	$\mathcal{E}_{PB}$	$\mathcal{E}_{PD}$	$\mathcal{E}_{BD}$
DR	8	2	2	2	4	6
PO	4	2	3	1	4	4
PP	4	2	5	0	6	3
PPI	0	3	4	1	4	4
EQ	0	3	6	0	6	3

Table 3: Distances between relations from $\mathcal{B}$ and the profile of constraints $\mathcal{E}_{CD}$, for each pair of variables $v_C, v_D \in V$.

5.2 Using the Distance to Build a Merged Consistent QCN

We now describe our procedure which associates a profile of QCNs with a merged, consistent QCN. This takes advantage of the distance between basic relations and a profile of constraints $\mathcal{E}_{CD}$. Let us first formally define two intermediate functions $relax_{CD}$ and val_{CD} which are used in our procedure. Given a total preorder³ $\leq$ over a finite set E , let us denote by $\min(E, \leq)$ the set of minimal elements of E w.r.t. $\leq$,

³A total preorder over a set E is a total, symmetric and transitive relation.

i.e., $\min(E, \leq) = \{e \in E \mid \forall e' \in E, e \leq e'\}$. Each pair of variables v_C, v_D is associated with a total preorder $\leq_{CD}$ over the basic relations from $\mathcal{B}$ defined for all $b, b' \in \mathcal{B}$ as $b \leq_{CD} b'$ iff $d(b, \mathcal{E}_{CD}) \leq d(b', \mathcal{E}_{CD})$. Then the function $relax_{CD}$ is a mapping $relax_{CD} : 2^{\mathcal{B}} \mapsto 2^{\mathcal{B}}$ defined for each $\varphi \in 2^{\mathcal{B}}$ as $relax_{CD}(\varphi) = \varphi \cup \min(\mathcal{B} \setminus \varphi, \leq_{CD})$. It corresponds to the *relaxation* of a constraint φ w.r.t.. the total preordering $\leq_{CD}$. Noteworthy, $relax_{CD}(\emptyset)$ corresponds to the set of basic relations with a minimal distance to the profile of constraints $\mathcal{E}_{CD}$. The function val_{CD} is a mapping $val_{CD} : 2^{\mathcal{B}} \mapsto \mathbb{N}$ defined for each $\varphi \in 2^{\mathcal{B}}$ as $val_{CD}(\varphi) = \max_{b \in \varphi} d(b, \mathcal{E}_{CD})$. For instance, according to Table 3 and focusing on *Book* and *Document* (cf. $\mathcal{E}_{BD}$), we get that $PP, EQ \leq_{BD} PO, PPI \leq_{BD} DR$, that $relax_{BD}(\emptyset) = \{PP, EQ\}$, that $relax_{BD}(\{PP, EQ\}) = \{PP, EQ, PO, PPI\}$, and that $val_{BD}(\{PP, EQ\}) = 3$.

We are ready to introduce our main procedure, whose outline is given in Algorithm 1 that defines an initial QCN $\mathcal{N}$ by setting each one of its constraints Ψ_{CD} to the set of basic relations from $\mathcal{B}$ having a distance to the profile of constraints $\mathcal{E}_{CD}$ that is minimal (lines 2 to 5). If this QCN is consistent, then it is returned as the merged QCN (line 12). If not, some of the constraints of $\mathcal{N}$ are relaxed in line 7, in the sense that some basic relations are added to these constraints. Such a set of constraints S is selected as follows. In line 7, S is first restricted to those constraints from $\mathcal{N}$ which *can* be relaxed, i.e., those constraints not equal to $\mathcal{B}$. Among those candidate constraints, one selects in line 8 the constraints Ψ_{CD} having a highest value $val_{CD}(\Psi_{CD})$. Indeed, we do not want to relax first the constraints consisting of basic relations which are “close” to the input profile, but rather would one prioritize the relaxation of more “controversial” constraints, i.e., those with a high value according to val_{CD} . For instance, let us look back at Table 3. It can be seen that $d(PPI, \mathcal{E}_{TP}) = d(EQ, \mathcal{E}_{TP}) = 0$, and thus $val_{TP}(\{PPI, EQ\}) = 0$; this low value reflects the consensus between sources on the fact that one of the basic relations PPI, EQ holds between *Text* and *Paper*, and indeed it can be verified that the axiom $P \sqsubseteq T$ is consistent with each input TBox. On the contrary, one has that $val_{PD}(\{DR, PO, PPI\}) = 4$; this higher value reflects a disagreement between the input sources about the relationship between the concepts of *Paper* and *Document*. And in the general case, whenever possible and in order to restore the consistency of the merged QCN, it is a sensible choice to keep unchanged those constraints unanimously accepted by the input sources, and rather weaken first the most disputed constraints. This “relaxation” process is repeated iteratively until the resulting QCN is consistent which, obviously enough, is guaranteed after a finite number of iterations.

Example 1 (continued). Initially, the merged QCN $\mathcal{N} = \langle V, \Psi \rangle$ is defined by the following set of constraints, which correspond the basic relations highlighted in Table 3 (this QCN is also depicted in Figure 2(a)):

$$\begin{aligned} \Psi_{TP} &= \{PPI, EQ\} & \Psi_{TB} &= \{DR, PO, PP\} \\ \Psi_{TD} &= \{DR\} & \Psi_{PB} &= \{PP, EQ\} \\ \Psi_{PD} &= \{DR, PO, PPI\} & \Psi_{BD} &= \{PP, EQ\} \end{aligned}$$

This QCN is inconsistent. One can see that the constraints Ψ_{PB} and Ψ_{BD} imply by transitivity that the relation between

Algorithm 1: Computing a merged QCN

input: A profile of QCNs $\mathcal{N} = \langle \mathcal{N}_1, \dots, \mathcal{N}_n \rangle$
output: A merged, consistent QCN $\mathcal{N}$

```

1 begin
  // Initialization of the output QCN  $\mathcal{N}$ 
2  $\Psi \leftarrow \{\Psi_{CD} \mid v_C, v_D \in V\}$ 
3 foreach  $(v_C, v_D) \in V \times V$  do
4    $\Psi_{CD} \leftarrow relax_{CD}(\emptyset)$ 
5  $\mathcal{N} \leftarrow (V, \Psi)$ 
6 while  $\mathcal{N}$  is not consistent do
  // One relaxes some constraints of  $\mathcal{N}$ 
7  $S \leftarrow \{\Psi_{CD} \mid \Psi_{CD} \in \Psi, \Psi_{CD} \neq \mathcal{B}\}$ 
8  $S \leftarrow \arg \max\{val_{CD}(\Psi_{CD}) \mid \Psi_{CD} \in S\}$ 
9 foreach  $\Psi_{CD} \in S$  do
10   $\Psi_{CD} \leftarrow relax_{CD}(\Psi_{CD})$ 
11  $\mathcal{N} \leftarrow (V, \Psi)$ 
12 return  $\mathcal{N}$ 

```

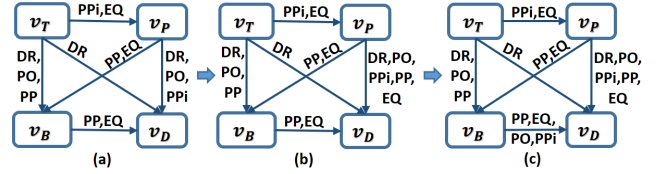


Figure 2: The QCNs iteratively generated by our algorithm. Fig. 2(c) corresponds to the final consistent merged QCN.

P and D must be PP or EQ , yet $\Psi_{PD} \cap \{PP, EQ\} = \emptyset$. Then the constraint Ψ_{PD} is selected (cf. line 8 in the algorithm) as the only candidate for relaxation at this point, since $val_{PD}(\Psi_{PD}) = 4$, which is the highest value among all constraints. And since $relax_{PD}(\Psi_{PD}) = \mathcal{B}$, one updates Ψ_{PD} to $\mathcal{B}$ which results in the QCN depicted in Figure 2(b). This QCN is, again, inconsistent (in this case, explaining its inconsistency is more complex as it involves dependencies between all four variables. We omit the details for space reasons). Then the constraint Ψ_{BD} is selected for relaxation ($val_{BD}(\Psi_{BD}) = 3$) and one updates Ψ_{BD} to $\{PP, EQ, PO, PPI\}$. The resulting QCN (cf. Figure 2(c)) is consistent and returned by the procedure.

Algorithm 1 runs in a time that is polynomial on the size of the input QCN profile, given access to an NP oracle in one step (line 6). Indeed, (i) checking the consistency of an RCC-5 QCN can be performed by taking advantage of a standard SAT solver (Condotta, Nouaouri, and Sioutis 2016), (ii) the number of possible relaxations of a given constraint is bounded by a constant (the number of RCC-5 basic relations), and as a consequence, the number of iterations performed in the loop starting from line 6 is in $O(|V \times V|)$; and (iii) the functions $relax_{CD}$ and val_{CD} are computed in $O(1)$ since the distance d is computed in $O(1)$, again because the number of RCC-5 basic relations is bounded by a constant. This makes the complexity of Algorithm 1 in FP^{NP} , that is reminiscent of the complexity of the inference problem for propositional belief merging, i.e., P^{NP} ,

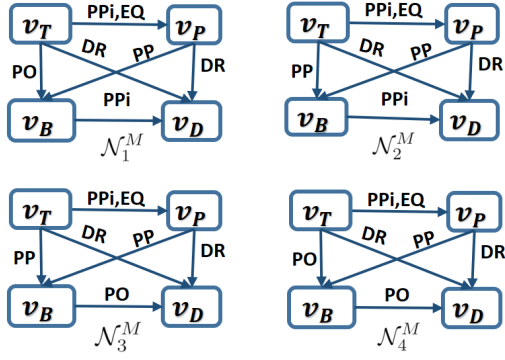


Figure 3: The consistent scenarios of the merged QCN.

for a vast majority of the existing operators (Konieczny and Marquis 2004).

Notably, the QCN merging step we propose in our framework can be seen as an adaptation of (Condotta et al. 2010) that allows one to work with a single QCN as a merged result (the output of Algorithm 1). Nevertheless, the work of Condotta et al. (Condotta et al. 2010) does not provide such a procedure.

6 Selecting a Representative Scenario of the Merged QCN

Once we have obtained a merged, consistent QCN, our goal is to express it in our initial (target) ontology language. However, not every constraint Ψ_{CD} from the QCN (i.e., a subset of $\mathcal{B}$) can easily be translated as a set of axioms, since non-singleton constraints express some disjunctive information between two concepts/regions. Moreover, since the merged QCN is consistent, one can remark it necessarily admits at least one consistent scenario, and since a scenario involves singleton constraints (as well as the two constraints $\{PP, EQ\}$, $\{PPI, EQ\}$), it can easily be translated into a single ontology, as will be shown in the next section. Then our aim is to (i) focus on all consistent scenarios of the merged QCN, and (ii) select one representative scenario. This can be done by exploiting information provided by the input ABoxes as we intend to show in the rest of this section.

In our running example, the merged QCN admits four consistent scenarios which are depicted in Figure 3. Let us first discuss why these four scenarios seem to be reasonable candidates to the input ontologies / QCNs provided by the sources. First, note that all input ontologies state that $Paper \sqsubseteq Text$ ($v_T\{PPI, EQ\}v_P$ in the QCN profile). And it can be seen that this consensus is reflected in the four candidate scenarios, which entail that information. Second, while the two sources O_1 and O_4 state that $Text$ and $Document$ are disjoint concepts ($v_T\{DR\}v_D$ in the corresponding input QCNs), only one source says that $Document \sqsubseteq Text$ ($v_T\{PP, EQ\}v_D$). In this case, it make sense to follow the point of view of the majority of the sources. And accordingly, in all four scenarios we have that $v_T\{DR\}v_D$, thus $Text$ and $Document$ are disjoint concepts. Third, the source O_1 states that $Paper \sqsubseteq Book$ and all other sources have

no information on these concepts. It is sensible to keep this information in the merged result, and accordingly all four scenarios entail that information. More, one sees that $v_P\{PP\}v_B$ holds in all scenarios, i.e., that *Paper* is a *strict* part of *Book*, or stated otherwise, that both concepts cannot be equal while keeping the relationships between the remaining concepts consistent. This emergent property is also an interesting feature of the merging process. Last, the reason why there are four, equally reasonable, candidate scenarios is that some strong disagreements hold on the relationships between the concepts *Text* and *Book* on the one hand, and the concepts *Book* and *Document* on the other hand. Accordingly, the only differences between these two pairs of concepts hold on the constraints between these two pairs of concepts ($v_T\{PO\}v_B / v_T\{PP\}v_B$, and $v_B\{PPI\}v_D / v_B\{PO\}v_D$).

What remains to be done is to select one of these four scenarios. For this purpose, one takes advantage of the ABoxes from the input ontologies and see how these ABoxes relate to each scenario. To be as faithful as possible to what each input source says, instead of simply considering each input ABox as such, one considers the “closure” of it according to its corresponding TBox. For instance, if a given source states that $Paper \sqsubseteq Text$ in its TBox and that $Paper(p)$ in its ABox, then it makes sense to also consider that $Text(p)$ also holds in that source’s implicit knowledge. Formally, let $O = \langle \mathcal{T}, \mathcal{A} \rangle$ be an ontology. The *deductive closure* of $\mathcal{A}$ w.r.t. $\mathcal{T}$, denoted by $Cl_{\mathcal{T}}(\mathcal{A})$, is defined as $Cl_{\mathcal{T}}(\mathcal{A}) \stackrel{\text{def}}{=} \{B(a) \mid B \in N_C, a \in N_I, O \models B(a)\}$ (Baader et al. 2007; Benferhat, Bouraoui, and Tabia 2015). Accordingly, $\langle \mathcal{T}, \mathcal{A} \rangle$ is logically equivalent to $\langle \mathcal{T}, Cl_{\mathcal{T}}(\mathcal{A}) \rangle$.

To select the representative scenario, we define a distance between a scenario and the set of all input (closed) ABoxes, and then use this distance to choose the scenario. The representative scenario is the one having a minimal distance. So given a scenario, the idea is to count the number of individuals in each input ABox which raise a conflict w.r.t. the constraints of that scenario. This can naturally be done for any scenario constraint $v_C \varphi v_D$ where $\varphi \neq \{PO\}$. For instance, if $Paper(p)$ and $Document(p)$ hold in the ABox of a given source, and $v_P\{DR\}v_D$ holds in the scenario under consideration, then according to that ABox p is an individual that raises a conflict with that scenario. Another example is if $Paper(p)$ holds but not $Text(p)$, then p is not a member of the concept *Text* in the ABox (recall that ABoxes are closed w.r.t. their TBox); in that case p raises a conflict with the constraints $v_P \varphi v_T$ when $\varphi \in \{\{PP\}, \{EQ\}, \{PP, EQ\}\}$. More formally, given an ABox $\mathcal{A}$, a scenario constraint $v_C \varphi v_D$ where $\varphi \neq \{PO\}$, and an individual p , we say that p raises a conflict with φ w.r.t. $\mathcal{A}$ when:

$$\begin{aligned} C(p) \in \mathcal{A}, D(p) \notin \mathcal{A} & \quad \text{when } \varphi \subseteq \{PP, EQ\}, \\ D(p) \in \mathcal{A}, C(p) \notin \mathcal{A} & \quad \text{when } \varphi \subseteq \{PPI, EQ\}, \\ C(p), D(p) \in \mathcal{A} & \quad \text{when } \varphi = \{DR\}. \end{aligned}$$

And the number of conflicts raised by an ontology $O = \langle \mathcal{T}, \mathcal{A} \rangle$ w.r.t. a scenario constraint $v_C \varphi v_D$ where $\varphi \neq \{PO\}$, is defined as $nbConf(O, \varphi) = |\{p \in N_I \mid p \text{ raises a conflict with } \varphi \text{ w.r.t. } Cl_{\mathcal{T}}(\mathcal{A})\}|$.

The case of PO is more complex. Indeed, it can be easily seen that no individual can raise a conflict with

$v_C\{PO\}v_D$. This is because all basic relations $\mathcal{B} \setminus \{PO\}$ express explicit dependencies between concepts / regions, whereas PO is a complementary relation that (explicitly) expresses a notion of *independency* between concepts. For instance, the concepts *Smoker* and *Researcher* can naturally be thought of as independent concepts, in the sense that one can easily find in a real-world context individuals that are members of either both concepts, only one of them, and none of them (note that this should not be confused with the case of DR , which expresses an explicit dependency between concepts, e.g., the concepts *Dog* and *Cat*.) So to evaluate the number of “conflicts” raised by an ABox w.r.t. a constraint $v_C\{PO\}v_D$, we propose to count how “unbalanced” the number of conflicts are w.r.t. the remaining forms of constraints. Formally, focusing on the scenario constraint between two variables v_C and v_D , $nbConf(O, \varphi) = \max_{\varphi' \neq \{PO\}} nbConf(O, \varphi') - \min_{\varphi' \neq \{PO\}} nbConf(O, \varphi')$. For instance, when $nbConf(O, \{PP, EQ\}) = nbConf(O, \{PPI, EQ\}) = nbConf(O, \{DR\})$, then $nbConf(O, \{PO\}) = 0$: since individuals can be found equally (i) in both underlying concepts, and (ii) in one concept but not the other, O raises no conflict w.r.t. $\{PO\}$.

We have now a way to select the representative scenario from a given set of candidates. The distance between a scenario N^M and the input profile of ontologies $\mathcal{P} = \langle O^1, \dots, O^n \rangle$ is simply defined as the overall number of conflicts raised by all input ABoxes w.r.t. all constraints of N^M , i.e., $d(N^M, \mathcal{P}) = \sum_{i \in \{1, \dots, n\}, \varphi \in N^M} nbConf(O^i, \varphi)$. Given a set of candidate scenarios, the representative scenario is then the one having a minimal distance.

Example 1 (continued). Let us go back to our running example. We have that $Cl_{\mathcal{T}^3}(\mathcal{A}^3) = \{P(p_3), P(b_3), P(d_3), T(t_3), T(d_3), T(b_3), T(p_3), D(d_3), D(b_3), B(b_3)\}$. So focusing on *Text* and *Book* (i.e., on the scenario constraints between the variables v_T and v_B), we have that $nbConf(O^3, \{PP\}) = |\{t_3, d_3, p_3\}| = 3$, and one can easily verify that $nbConf(O^3, \{PO\}) = |\{t_3, d_3, p_3\}| = 3 - 0 = 3$. Summing up all conflicts, we get that $d(N_1^M, \mathcal{P}) = 20$, $d(N_2^M, \mathcal{P}) = 18$, $d(N_3^M, \mathcal{P}) = 22$, and $d(N_4^M, \mathcal{P}) = 24$. Hence, the scenario N_2^M is selected as a representative scenario of the merged QCN.

7 Translating the Representative Scenario into a Terminological Knowledge

The last step of our framework is to translate back the result- ing selected representative scenario into an ontology. The translation is defined as follows.

Definition 4 (Backward translation $\tau_{\triangleleft}$). A *backward translation* is a function $\tau_{\triangleleft} : V \rightarrow C$ s.t. $\tau_{\triangleleft}(v_C) \stackrel{\text{def}}{=} C$. $\tau_{\triangleleft}$ is extended to map constraints into an ontology as follows, where A' , C' , and D' are new concept names and a , b , and c new individual names:⁴

- $\tau_{\triangleleft}(v_C\{EQ\}v_D) \stackrel{\text{def}}{=} \langle \{C \sqsubseteq D\}, \emptyset \rangle$;

⁴Notice that the constraint $\{PO\}$ cannot be translated into a set of GCIs only, whence the use of ABox assertions in the translation.

- $\tau_{\triangleleft}(v_C\{DR\}v_D) \stackrel{\text{def}}{=} \langle \{C \sqcap D \sqsubseteq \perp\}, \emptyset \rangle$;
- $\tau_{\triangleleft}(v_C\{PO\}v_D) \stackrel{\text{def}}{=} \langle \{A' \sqsubseteq C \sqcap D, C' \sqsubseteq C, C' \sqcap D \sqsubseteq \perp, D' \sqsubseteq D, D' \sqcap C \sqsubseteq \perp\}, \{A'(a), C(c), C(a), D(d), D(a), C'(c), D'(d)\} \rangle$;
- $\tau_{\triangleleft}(v_C\{PP, EQ\}v_D) \stackrel{\text{def}}{=} \langle \{C \sqsubseteq D\}, \emptyset \rangle$;
- $\tau_{\triangleleft}(v_C\{PPI, EQ\}v_D) \stackrel{\text{def}}{=} \langle \{D \sqsubseteq C\}, \emptyset \rangle$;
- $\tau_{\triangleleft}(v_C\{PP\}v_D) \stackrel{\text{def}}{=} \langle \{C \sqsubseteq D, D' \sqsubseteq D, C \sqcap D' \sqsubseteq \perp\}, \{D'(d), C(c), D(d), D(c)\} \rangle$, and
- $\tau_{\triangleleft}(v_C\{PPI\}v_D) \stackrel{\text{def}}{=} \langle \{D \sqsubseteq C, C' \sqsubseteq C, D \sqcap C' \sqsubseteq \perp\}, \{C'(c), D(d), C(d), D(c)\} \rangle$.

Moreover, $\tau_{\triangleleft}$ is extended to translate a set of constraints into an ontology in the (strict) normal form in the expected way: $\tau_{\triangleleft}(N) \stackrel{\text{def}}{=} \langle \mathcal{T}, \mathcal{A} \rangle$, where $\mathcal{T} \stackrel{\text{def}}{=} \bigcup_{\tau_{\triangleleft}(\Psi) = \langle \mathcal{T}', \mathcal{A}' \rangle, \Psi \in N} \mathcal{T}'$ and $\mathcal{A} \stackrel{\text{def}}{=} \bigcup_{\tau_{\triangleleft}(\Psi) = \langle \mathcal{T}', \mathcal{A}' \rangle, \Psi \in N} \mathcal{A}'$.

Accordingly, our back translation is faithful. Using again the notions of inflation and flattening (cf. Definitions 2 and 3), we show that the set of solutions of a scenario N are captured precisely in its translated ontology $\tau_{\triangleleft}(N)$.

Theorem 3. Let N be a scenario and S be solution of N . Then there is an inflation I_S of S s.t. I_S is a model of $\tau_{\triangleleft}(N)$.

Theorem 4. Let N be a scenario and let I be a fulfilling interpretation of $\tau_{\triangleleft}(N)$ such that I is a model of $\tau_{\triangleleft}(N)$. Then $S_I \models N$.

Example 1 (continued). Let us translate the selected scenario N_2^M into an ontology. From Definition 4, we get that: $\tau_{\triangleleft}(N_2^M) = \langle \{(\Psi_{TP}) P \sqsubseteq T, (\Psi_{TD}) T \sqcap D \sqsubseteq \perp, (\Psi_{PD}) P \sqcap D \sqsubseteq \perp, (\Psi_{TB}) T \sqsubseteq B, SubBo1 \sqsubseteq B, SubBo1 \sqcap T \sqsubseteq \perp, (\Psi_{PB}) P \sqsubseteq B, SubBo2 \sqsubseteq B, SubBo2 \sqcap P \sqsubseteq \perp, (\Psi_{BD}) D \sqsubseteq B, SubBo3 \sqsubseteq B, SubBo3 \sqcap D \sqsubseteq \perp\}, \{T(t_1), SubBo1(s_1), B(t_1), B(s_1), P(p_1), SubBo2(s_2), B(p_1), B(s_2), D(d_1), SubBo3(s_3), B(d_1), B(s_3)\} \rangle$.

8 Conclusion

In this paper, we introduced a terminological knowledge merging procedure based on qualitative reasoning over region spaces, by providing a two-way translation between terminological knowledge and RCC-5 based qualitative constraint networks. Accordingly, even if the input sources are inconsistent when simply combined together, our approach returns a consistent result. It also remains as close as possible to the input sources, i.e., by preserving as much information as possible.

We have assumed that all input ontologies were provided in a strict normal form. Notably, the process of translating an ontology into another equivalent ontology in strict normal form can be done in polynomial time, but this process also allows one to obtain a full classification of all initial atomic concepts, for instance by taking advantage of axioms involving constraints between complex concepts of the type $\exists r.C$ (Baader, Brandt, and Lutz 2005). For the scope of our paper, we have focused on the merging of all atomic concepts from the input ontologies, since considering more complex concepts may lead to consider constraints of higher arity in the translated qualitative constraint networks (e.g., for each role

r in the signature, the assertion $\exists r.A \sqsubseteq \exists r.B$ holds when $A \sqsubseteq B$ holds, which would correspond to a constraint of arity four in the translation), and our merging procedure does not allow for these types of constraints. These kind of considerations will be investigated in a future work. Further work will also include the empirical evaluation of our method. Moreover, an approach for representing regions and points in vector spaces (combining natural language processing) is another next interesting direction for merging ontologies.

Acknowledgments

This work has benefited from the support of the AI Chair BE4musIA of the French National Research Agency (ANR-20-CHIA-0028) and FEI INS2I 2022-EMILIE.

References

- Baader, F.; Calvanese, D.; McGuinness, D.; Nardi, D.; and Patel-Schneider, P. 2007. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.
- Baader, F.; Brandt, S.; and Lutz, C. 2005. Pushing the $\mathcal{EL}$ Envelope. In *Proceedings of the 19th International Joint Conferences on Artificial Intelligence (IJCAI'05)*, 364–369.
- Beltagy, I.; Chau, C.; Boleda, G.; Garrette, D.; Erk, K.; and Mooney, R. 2013. Montague meets Markov: Deep semantics with probabilistic logical form. In *Second Joint Conference on Lexical and Computational Semantics (*SEM)*, 11–21.
- Benferhat, S.; Bouraoui, Z.; Lagrue, S.; and Rossit, J. 2014. Min-based Assertional Merging Approach for Prioritized DL-Lite Knowledge Bases. In *Proceedings of the 8th International Conference on Scalable Uncertainty Management (SUM'14)*, 8–21.
- Benferhat, S.; Bouraoui, Z.; Papini, O.; and Würbel, E. 2019. Assertional removed sets merging of dl-lite knowledge bases. In *Proceedings of the 13th International Conference of Scalable Uncertainty Management (SUM'19)*, 207–220.
- Benferhat, S.; Bouraoui, Z.; and Tabia, K. 2015. How to select one preferred assertional-based repair from inconsistent and prioritized dl-lite knowledge bases? In *Proceedings of the 24th International Conference on Artificial Intelligence (IJCAI'15)*, 1450–1456.
- Bennett, B. 1994. Spatial Reasoning with Propositional Logics. In *Proceedings of the 4th International Conference on Principles of Knowledge Representation and Reasoning (KR'94)*, 51–62.
- Bhatt, M.; Guesgen, H.; Wölfl, S.; and Hazarika, S. 2011. Qualitative Spatial and Temporal Reasoning: Emerging Applications, Trends, and Directions. *Spatial Cognition & Computation* 11:1–14.
- Bouraoui, Z., and Schockaert, S. 2018. Learning Conceptual Space Representations of Interrelated Concepts. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI'18)*, 1760–1766.
- Bouraoui, Z.; Konieczny, S.; Ma, T. T.; and Varzinczak, I. 2020a. Model-based Merging of Open-Domain Ontologies. In *Proceedings of the 33rd IEEE International Conference on Tools with Artificial Intelligence (ICTAI'20)*, 29–34.
- Bouraoui, Z.; Camacho-Collados, J.; Anke, L. E.; and Schockaert, S. 2020b. Modelling Semantic Categories using Conceptual Neighborhood. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI'20)*, 7448–7455.
- Chang, F.; wei Chen, G.; and Zhang, S. 2019. FCAMap-KG Results for OAEI 2019. In *Proceedings of the 18th International Semantic Web Conference (OM@ISWC'19)*, 138–145.
- Chen, Z.; Yu, S.; Shengxian, W.; and Dianhai, Y. 2019. RLTM: An Efficient Neural IR Framework for Long Documents. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI'19)*, 5457–5463.
- Cohn, A.; Bennett, B.; Gooday, J.; and Gotts, M. 1997. Qualitative Spatial Representation and Reasoning with the Region Connection Calculus. *GeoInformatica* 275–316.
- Condotta, J.-F.; Kaci, S.; Marquis, P.; and Schwind, N. 2009. Merging Qualitative Constraint Networks in a Piecewise Fashion. In *Proceedings of the 21st IEEE International Conference on Tools with Artificial Intelligence (ICTAI'09)*, 605–608.
- Condotta, J.-F.; Kaci, S.; Marquis, P.; and Schwind, N. 2010. A Syntactical Approach to Qualitative Constraint Networks Merging. In *Proceedings of the 17th International Conference on Logic Programming and Automated Reasoning (LPAR'10)*, 233–247.
- Condotta, J.-F.; Kaci, S.; and Schwind, N. 2008. A Framework for Merging Qualitative Constraints Networks. In *Proceedings of the 21st International Florida Artificial Intelligence Research Society Conference (FLAIRS'08)*, 586–591.
- Condotta, J.-F.; Nouaouri, I.; and Sioutis, M. 2016. A SAT Approach for Maximizing Satisfiability in Qualitative Spatial and Temporal Constraint Networks. In *Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning (KR'16)*, 342–442.
- Douven, I.; Elqayam, S.; Gärdenfors, P.; and Mirabile, P. 2022. Conceptual spaces and the strength of similarity-based arguments. *Cognition* 218:104951.
- Freksa, C. 1992. Temporal reasoning based on semi-intervals. *Artificial Intelligence* 54(1):199–227.
- Gärdenfors, P. 2000. *Conceptual Spaces: The Geometry of Thought*. US: MIT Press.
- Haarslev, V., and Möller, R. 1997. SBox: A Qualitative Spatial Reasoner. In *Proceedings of the 11th International Workshop on Qualitative Reasoning (QR'97)*, 105–113.
- Hohenecker, P., and Lukasiewicz, T. 2020. Ontology Reasoning with Deep Neural Networks. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI'20)*, 5060–5064.
- Homburg, T.; Staab, S.; and Janke, D. 2020. Geosparql+: Syntax, semantics and system for integrated querying of

- graph, raster and vector data. In *Proceedings of the 19th International Semantic Web Conference (ISWC'20)*, 258–275.
- Julien, H., and Westphal, M. 2012. Revising Qualitative Constraint Networks: Definition and Implementation. In *Proceedings of the 24th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'12)*, 548–555.
- Konieczny, S., and Marquis, P. 2004. DA2 merging operators. *Artificial Intelligence* 157:49–79.
- Kriegel, F. 2020. Most specific consequences in the description logic $\mathcal{EL}$. *Discrete Applied Mathematics* 172–204.
- Kumar, S. K., and Harding, J. A. 2016. Description logic-based knowledge merging for concrete- and fuzzy-domain ontologies. *Journal of Engineering Manufacture* 954–971.
- Laadhar, A.; Ghazzi, F.; Megdiche, I.; Ravat, F.; Teste, O.; and Gargouri, F. 2017. POMap results for OAEI 2017. In *Proceedings of the 16th International Semantic Web Conference (ISWC'17)*, 1–7.
- Li, N.; Bouraoui, Z.; and Schockaert, S. 2019. Ontology Completion Using Graph Convolutional Networks. In *Proceedings of the 18th International Semantic Web Conference (ISWC'19)*, 435–452.
- Patricia, E.; Konieczny, S.; and Marquis, P. 2008. Conflict-based merging operators. In *Proceedings of the 11st International Conference on Principles of Knowledge Representation and Reasoning (KR'08)*, 348–357.
- Peter, J., and Thomas, D. 1997. A Complete Classification of Tractability in RCC-5. *Journal of Artificial Intelligence Research* 211–221.
- Randell, D. A.; Cui, Z.; and Cohn, A. G. 1992. A spatial logic based on regions and connection. In *Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning (KR'92)*, 165–176.
- Rocktäschel, T., and Riedel, S. 2017. End-to-end differentiable proving. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*, 3791–3803.
- Rospocher, M., and Corcoglioniti, F. 2018. Joint Posterior Revision of NLP Annotations via Ontological Knowledge. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI'18)*, 4316–4322.
- Schockaert, S., and Li, S. 2013. Combining RCC5 relations with betweenness information. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13)*, 1083–1089.
- Sioutis, M.; Long, Z.; and Janhun, T. 2020. On Robustness in Qualitative Constraint Networks. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI'20)*, 1813–1819.
- Tanon, T. P.; Vrandečić, D.; Schaffert, S.; Steiner, T.; and Pintscher, L. 2016. From Freebase to Wikidata: The Great Migration. In *Proceedings of the 25th International World Wide Web Conference (WWW'16)*, 1419–1428.
- Thau, D.; Bowers, S.; and Ludäscher, B. 2009. Merging Taxonomies under RCC-5 Algebraic Articulations. *Journal of Computing Science and Engineering* 109–126.
- Thiéblin, É.; Haemmerlé, O.; and Trojahn, C. 2018. CANARD complex matching system. In *Proceedings of the 17th International Semantic Web Conference (OM@ISWC'18)*, 138–143.
- Wang, Z.; Wang, K.; Jin, Y.; and Qi, G. 2012. Ontomerge: A system for merging DL-Lite ontologies. *CEUR Workshop Proceedings* 969:16–27.
- Zhao, M., and Zhang, S. 2016. FCA-Map results for OAEI 2016. In *Proceedings of the 15th International Semantic Web Conference (OM@ISWC'16)*, 1–7.