Aspects of Distributed and Modular Ontology Reasoning

Luciano Serafini
ITC-IRST
Via Sommarive 18
38050 Trento, Italy
serafini@itc.it

Alex Borgida
Department of Computer Science
Rutgers University
Piscataway, NJ 08855, USA
borgida@cs.rutgers.edu

Andrei Tamilin
DIT - University of Trento
Via Sommarive 16
38050 Trento, Italy
tamilin@dit.unitn.it

Abstract

We investigate a formalism for reasoning with multiple local ontologies, connected by directional semantic mappings. We propose: (1) a relatively small change of semantics which localizes inconsistency (thereby making unnecessary global satisfiability checks), and preserves directionality of "knowledge import"; (2) a characterization of inferences using a fixed-point operator, which can form the basis of a cache-based implementation for local reasoners; (3) a truly distributed tableaux algorithm for cases when the local reasoners use subsets of $\mathcal{SHIQ}$. Throughout, we indicate the applicability of the results to several recent proposals for knowledge representation and reasoning that support modularity, scalability and distributed reasoning.

1 Introduction

In applications ranging from information and service integration to the semantic web, it is expected that ontologies will be fragmented and distributed, rather than monolithic. Moreover, there are well-known advantages to breaking up a large ontology into modules (e.g., see [9]). Recently, there have been a number of proposals for systems supporting modularized ontologies, where one can have a localized view either by "importing" specific information [9, 3] or by distributing reasoning [8, 4].

Distributed Description Logics (DDLs) [2] are a particular formalization of the notion of distributed ontology and reasoning, which starts from local T-boxes, and expresses semantic connections between them as "bridge rules" between pairs of connects. This paper contributes to the study of DDLs, with the hope of shedding light both on the semantic and the implementation aspects of *all* similar mechanisms.

In particular, the claimed contributions of the paper are three-fold. First, on the semantic front, we propose a new semantics for DDLs, which better captures the properties of *localized inconsistency* and *directionality*. According to the directionality property, semantic mappings have a direction from a source ontology to a target ontology, and support knowledge propagation only in such a direction. According to localized inconsistency, the inconsistency in one component T-box, or in some subgroup of connected components, should not automatically render the *entire* system inconsistent. These properties are desirable for all multi-module ontologies, and we discuss how our results apply to other proposals, such as [9, 3, 8].

Second, we examine an approach which views the bridge rules connecting two local ontologies as describing an *operator* that propagates knowledge in the form of DL subsumption axioms. This is used as the basis of a characterization of distributed DL reasoning using a fixed point operator, which does forward-propagation of axioms. In addition to its intrinsic interest, as part of a syntactic characterization of the DDL consequence relation, this result can be used to extend the caching technique described in [9] to a sound and complete one.

The third contribution is a sound and complete distributed tableaux algorithm that determines the satisfiability of a $\mathcal{SHIQ}$ concept in the context of the local axioms of an ontology and the extra knowledge imparted by the bridge rules. This algorithm is the basis of a peer-to-peer implementation of a proof mechanism for checking concept subsumption in a distributed manner.

The rest of the paper is organized as follows. In Section 2 we review the formal framework of DDLs. In Section 3 we introduce the notion of inconsistent local interpretation that allows us to model partially inconsistent distributed ontologies, and show exactly how the new and old entailments are related. The new semantics is also used in Section 4 to characterize the knowledge flow induced by bridge rules, and in Section 5 to present a sound and complete distributed tableaux algorithm that computes satisfiability/subsumption.

2 Distributed Description Logics

We briefly recall the definition of DDL as given by Borgida and Serafini [2].

Syntax Given a non-empty set I of indexes, used to enumerate local ontologies, let $\{\mathcal{D}L_i\}_{i\in I}$ be a collection of description logics[1]. For each $i \in I$, let us denote a T-box of $\mathcal{D}L_i$ as $\mathcal{T}_i$.[2] To make every description D distinct, we will prefix it with the index of ontology it belongs to, as in $i : C$. We use $i : C \sqsubseteq D$ to say that $C \sqsubseteq D$ is being considered in the i-th ontology.

[1]We assume the reader is familiar with description logics and related reasoning systems, as described in [1].

[2]We assume that a T-box will contain all the information necessary to define the terminology of a domain, including not just concept and role definitions, but also general axioms relating descriptions, as well as declarations such as the transitivity of certain roles. This is in keeping with the intent of the original paper introducing the terms T-Box and A-box.

Semantic mappings between different ontologies are expressed via *bridge rules*. A bridge rule from i to j is an expression, which in this paper is restricted to being one of the following two forms:

$i : x \xrightarrow{\sqsubseteq} j : y$ — an *into-bridge rule*

$i : x \xrightarrow{\sqsupseteq} j : y$ — an *onto-bridge rule*

where x and y are concepts. The derived bridge rule $i : x \xrightarrow{\equiv} j : y$ can be defined as the conjunction the corresponding into and onto bridge rule.

Bridge rules from i to j express relations between i and j viewed from the *subjective* point of view of the j-th ontology. For example, the into-bridge rule $i : C \xrightarrow{\sqsubseteq} j : D$ intuitively says that, from the j-th point of view, the individuals in concept C in i correspond (via an approximation introduced by an implicit semantic domain relation) to a subset of the individuals in its local concept D. Therefore, bridge rules from i to j provide the possibility of translating *into* j's ontology some of the concepts of a foreign ontology i.

A *distributed T-box (DTBox)* $\mathfrak{T} = \langle \{\mathcal{T}_i\}_{i \in I}, \mathfrak{B} \rangle$ therefore consists of a collection of T-boxes $\{\mathcal{T}_i\}_{i \in I}$, and a collection of bridge rules $\mathfrak{B} = \{\mathfrak{B}_{ij}\}_{i \neq j \in I}$ between them.

Example 2.1 Figure 1 shows fragments of class hierarchies from two ontologies, SWRC[3] and SHOE[4], available from the DAML on-line library. These can be viewed as local T-boxes. The following are examples of bridge rules from SWRC to SHOE.

$$\text{SW} : \text{Publication} \xrightarrow{\equiv} \text{SH} : \text{Publication} \quad (1)$$

$$\text{SW} : \text{InProceedings} \xrightarrow{\sqsubseteq} \text{SW} : \text{ConferencePaper} \sqcup \text{WorkshopPaper} \quad (2)$$

$$\text{SW} : \text{InBook} \xrightarrow{\sqsupseteq} \text{SH} : \text{BookArticle} \quad (3)$$

Semantics DDL semantics is a customization of the Local Models Semantics for Multi Context Systems [5, 6]. Each ontology $\mathcal{T}_i$ is *locally interpreted* by a standard DL interpretation $\mathcal{I}_i = \langle \Delta^{\mathcal{I}_i}, \cdot^{\mathcal{I}} \rangle$. Since local domains may be heterogeneous (e.g., time may be represented by Rationals and Integers in two ontologies), we need relations that model semantic correspondences between heterogeneous domains. A *domain relation* r_{ij} from $\Delta^{\mathcal{I}_i}$ to $\Delta^{\mathcal{I}_j}$ is a subset of $\Delta^{\mathcal{I}_i} \times \Delta^{\mathcal{I}_j}$. For example, if $\Delta^{\mathcal{I}_1}$ and $\Delta^{\mathcal{I}_2}$ are the Rationals and the Naturals, then r_{ij} could be the round-off function. We use $r_{ij}(d)$ to denote $\{d' \in \Delta^{\mathcal{I}_j} \mid \langle d, d' \rangle \in r_{ij}\}$; for $D \subseteq \Delta^{\mathcal{I}_i}$, we use $r_{ij}(D)$ for $\bigcup_{d \in D} r_{ij}(d)$.

A *distributed interpretation* $\mathfrak{I} = \langle \{\mathcal{I}_i\}_{i \in I}, \{r_{ij}\}_{i \neq j \in I} \rangle$ of a DTBox $\mathfrak{T}$ therefore combines the above two notions. and is said to satisfy (written $\mathfrak{I} \vDash_d$) the elements of $\mathfrak{T}$ if

1. $\mathcal{I}_i \vDash_d A \sqsubseteq B$ for all $A \sqsubseteq B$ in $\mathcal{T}_i$
2. $\mathfrak{I} \vDash_d i : x \xrightarrow{\sqsubseteq} j : y$, if $r_{ij}(x^{\mathcal{I}_i}) \subseteq y^{\mathcal{I}_j}$
3. $\mathfrak{I} \vDash_d i : x \xrightarrow{\sqsupseteq} j : y$, if $r_{ij}(x^{\mathcal{I}_i}) \supseteq y^{\mathcal{I}_j}$
4. $\mathfrak{I} \vDash_d \mathfrak{T}$, if for every $i, j \in I$, $\mathfrak{I} \vDash_d \mathcal{T}_i$ and $\mathfrak{I} \vDash_d \mathfrak{B}_{ij}$.

Finally, $\mathfrak{T} \vDash_d i : C \sqsubseteq D$ (read as "$\mathfrak{T}$ d-entails $i : C \sqsubseteq D$") if for every $\mathfrak{I}$, $\mathfrak{I} \vDash_d \mathfrak{T}$ implies $\mathfrak{I} \vDash_d i : C \sqsubseteq D$. We say $\mathfrak{T}$ is *satisfiable* if there exists a $\mathfrak{I}$ such that $\mathfrak{I} \vDash_d \mathfrak{T}$.

[3] www.semanticweb.org/ontologies/swrc-onto-2000-09-10.daml

[4] www.cs.umd.edu/projects/plus/DAML/onts/univ1.0.daml

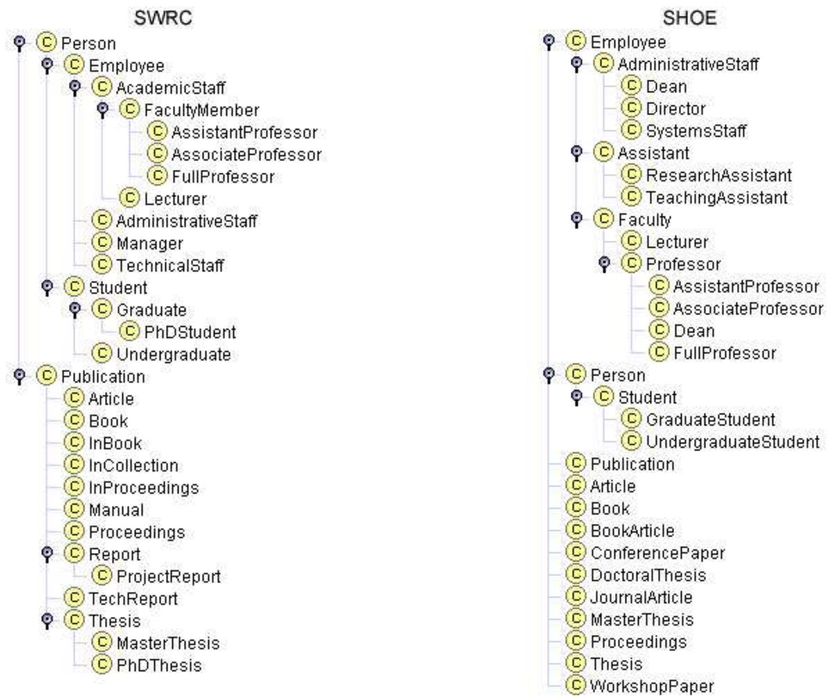

Figure 1: Extracts of the class hierarchies

Concept $i : C$ is *satisfiable* with respect to $\mathfrak{T}$ if there is a $\mathfrak{I}$ such that $\mathfrak{I} \models_d \mathfrak{T}$ and $C^{\mathcal{I}_i} \neq \emptyset$.

On injective domain correspondences A key novelty of the semantic mappings in DDL is support for multiple levels of granularity and perspective: allowing individuals to be related using arbitrary binary relations rather than just bijection. For example, while it is traditional to state correspondences such as "Wife in Ontology 1 corresponds to Moglie in Ontology 2", DDLs support domain correspondences that are general binary relations, so that one can say that "Husband and Wife in ontology 1 correspond to Couple in Ontology 2", which can be formalized by using onto-bridge rules $\{1 : \mathsf{Wife} \xrightarrow{\sqsupseteq} 2 : \mathsf{Couple}, 1 : \mathsf{Husband} \xrightarrow{\sqsupseteq} 2 : \mathsf{Couple}\}$. In [4], DDLs are faulted because the collection of bridge rules $\{1 : \mathsf{Bird} \xrightarrow{\sqsupseteq} 2 : \mathsf{Penguin},\ 1 : \neg\mathsf{Fly} \xrightarrow{\sqsupseteq} 2 : \mathsf{Penguin}\}$ do not render Penguin unsatifi able even if Bird is subsumed by Fly in Ontology 1. As the example involving Couple shows, the general formal pattern is correct in some cases, so this is actually a problem of incomplete modeling.

In the case of Penguins, the extra information is that the domain relation is one-to-one. In such cases, one should also have added bridge rules stating that *non*-birds and flying objects are *non*-penguins: $\{1 : \neg\mathsf{Bird} \xrightarrow{\sqsubseteq} 2 : \neg\mathsf{Penguin}, 1 : \mathsf{Fly} \xrightarrow{\sqsubseteq} 2 : \neg\mathsf{Penguin}\}$. This would indeed lead to the conclusion $\mathfrak{T} \models_d 2 : \mathsf{Penguin} \sqsubseteq \bot$.

Since the property that the domain relation is one-one over some concept B arises quite frequently, we might consider adding a new kind of bridge rule to express it, writing something like $\xrightarrow{\leq 1}$ Penguin. Interestingly, it can be proven that in the context of DDLs, such rules can be eliminated by syntactically manipulating the DTBox, so that whenever $\xrightarrow{\leq 1} G$ and $A \xrightarrow{\sqsupseteq} H$ are present, a new bridge rule $\neg A \xrightarrow{\sqsubseteq} \neg(H \sqcap G)$ is added. The tableaux technique in Section 5 could however use such rules more efficiently without the encoding.

Properties and Desiderata for DDL We first give some basic ways in which subsumption and a combination of onto- and into-bridge rules allows to propagate subsumptions across ontologies.

Lemma 2.1 *If* $\mathfrak{B}_{ij}$ *contains* $i : A \xrightarrow{\sqsupseteq} j : G$ *and* $i : B \xrightarrow{\sqsubseteq} j : H$, *then* $\mathfrak{T} \models_d i : A \sqsubseteq B \Longrightarrow \mathfrak{T} \models_d j : G \sqsubseteq H$

Thus, in Example 2.1, the subsumption BookArticle $\sqsubseteq$ Publication : $SHOE$ can be inferred in DDL through bridge rules (1) and (3), and the subsumption InBook $\sqsubseteq$ Publication contained in $\mathcal{T}_{SWRC}$.

If the local languages support disjunction as a concept constructor then a more general form of propagation can occur:

Lemma 2.2 *If* $\mathfrak{B}_{ij}$ *contains* $i : A \xrightarrow{\sqsupseteq} j : G$ *and* $i : B_k \xrightarrow{\sqsubseteq} j : H_k$ *for* $1 \leq k \leq n$ *(with* $n \geq 0$*), then*
$\mathfrak{T} \models_d i : A \sqsubseteq \bigsqcup_{k=1}^{n} B_k \Longrightarrow \mathfrak{T} \models_d j : G \sqsubseteq \bigsqcup_{k=1}^{n} H_k$

Additional properties would be desirable for DDL entailment. In particular, since the intended meaning is that bridge rules $\mathfrak{B}_{ij}$ constitute a semantic channel which allows ontology j to access and import knowledge *from* ontology i, we want information flow to be "directional" in some sense. To express this, we first introduce the notion of bridge graph.

Definition 2.1 *The* bridge graph $G_{\mathfrak{T}}$ *of a DTBox* $\mathfrak{T}$ *is a directed graph with an arc from* i *to* j *exactly when the set of bridge rules* $\mathfrak{B}_{ij}$ *is non-empty.*

We can then state the main property we are looking for as:

Directionality desideratum If in $G_{\mathfrak{T}}$ there is no path from i to j, then $\mathfrak{T} \models_d j : A \sqsubseteq B$ if and only if $\mathfrak{T}' \models_d j : A \sqsubseteq B$, where $\mathfrak{T}'$ is obtained by removing $\mathcal{T}_i$, $\mathfrak{B}_{ki}$, and $\mathfrak{B}_{ik}$ from $\mathfrak{T}$.

This says that knowledge is propagated *only* through bridge rules, so that if there are no bridge rules that go from i towards j, then j is not affected by i. The following two isolation properties are special cases of this:

Isolation 1 A T-box without incoming bridge rules is not affected by other T-boxes. (Formally, if $\mathfrak{B}_{ki} = \emptyset$ for all $k \neq i \in I$, then $\mathfrak{T} \models_d i : A \sqsubseteq B \Longrightarrow \mathcal{T}_i \vDash A \sqsubseteq B$)

Isolation 2 A T-box without outgoing bridge rules does not affect the other T-boxes.

Unfortunately, property *Isolation 1* does not always hold, because of onto-rules. Indeed, in the presence of onto-rule $1 : A \xrightarrow{\sqsupseteq} 2 : G$ from T-box $\mathcal{T}_1$ to $\mathcal{T}_2$, if $\mathcal{T}_2$ entails $\top \sqsubseteq G$, then $1 : A$ cannot be empty according to DDL semantics, and so, for example, an inconsistency would be generated if $\mathcal{T}_1$ entails $A \sqsubseteq \bot$. This is despite the fact that the bridge rules are toward $\mathcal{T}_2$.

Property *Isolation 2* may also not hold. Indeed, if $\mathcal{T}_1$ is unsatisfiable, then $\mathcal{T}_2 \models_d 2 : X \sqsubseteq Y$ for every X, Y, even if there are no bridge rules connecting $\mathcal{T}_1$ with $\mathcal{T}_2$, because there are no satisfying distributed interpretations at all. Note that in a DDL, inconsistency may arise in a connected group of T-boxes even if each T-box is locally consistent; e.g., consider the case in the hypothesis of Lemma 2.1, when $\mathcal{T}_j \models \top \sqsubseteq G$ and $\mathcal{T}_j \models H \sqsubseteq \bot$.

This is a significant problem, because a localized inconsistency spreads and contaminates reasoning in *all* other local ontologies, even in the absence of connections to them, because there will be no satisfying distributed interpretation, and hence every statement about them is true, as usual in logic. This problem plagues all modular and distributed representation systems.

In the following section we propose an extension of the initial semantics in order to fix this problem.

3 Inconsistency in DDL

There are a number of possible approaches to handle the problem of inconsistency propagation.

(1) Define d-entailment in a 2-step manner, first eliminating local T-boxes that are inconsistent, and then using the standard definition. The problem with this approach is that it is non-monotonic, and it does not deal with cases where the inconsistency arises due to several connected local sources.

(2) Use some variant of a multi-modal epistemic semantics, which allows for models of even inconsistent knowledge in the case when the set of accessible worlds is empty. Such an approach was used in [6] for Distributed First Order Logics, but its computational complexity/decidability aspects are quite worrisome, and the precise impact of such non-standard semantics on logical consequences is hard to explain in an intuitive manner to users.

(3) Introduce some special interpretation, called a "hole" in [2], whose role is to interpret even inconsistent local T-boxes. We pursue this latter option.

Definition 3.1 *A* hole *for a T-box $\mathcal{T}$ is an interpretation $\mathcal{I}^\epsilon = \langle \emptyset, \cdot^\epsilon \rangle$, where the domain is empty.*

Of course, the important property of holes is that $\mathcal{I}^\epsilon \models X \sqsubseteq Y$ for every X and Y, since both sides are interpreted as the empty set. [5] We will however continue to refer to T-boxes as *"inconsistent/unsatisfiable"* in case there are no interpretations *other* than $\mathcal{I}^\epsilon$ which satisfy all the axioms in it.

[5] We are indebted to XXX for pointing out an earlier error, and to a YYY referee for suggesting the empty domain.

Let us extend the notion of d-entailment $\models_d$, obtaining the $\models_\epsilon$ relation, by also allowing holes as interpretations for local T-boxes. Note that now even if some local T-box $\mathcal{T}_i$ is inconsistent, we still have an interpretation for the whole DTBox: one that uses $\mathcal{I}^\epsilon$ to satisfy $\mathcal{T}_i$.

Properties of the semantics with holes First, the new semantics does the intended job:

Theorem 3.1 *The earlier-stated "directionality desideratum" holds for $\models_\epsilon$.*

Non-standard semantics (such as multivalued logics) can however distort the meaning of the original semantics in unpredictable ways. The following results should be reassuring in this respect.

For any $\mathfrak{T}$ and any $i \in I$, let $\mathfrak{T}(\epsilon_i)$ (the distributed T-box with the i-th local T-box viewed as inconsistent) be obtained by removing $\mathcal{T}_i$, $\mathfrak{B}_{ij}$ and $\mathfrak{B}_{ji}$ form $\mathfrak{T}$, and by extending each $\mathcal{T}_j$ with the set of axioms $\{G \sqsubseteq \bot | i : A \xrightarrow{\sqsupseteq} j : G \in \mathfrak{B}_{ij}\}$. For any finite set $J = \{i_1, \ldots, i_n\}$, such that $J \subset I$, let $\mathfrak{T}(\epsilon_J)$ be $\mathfrak{T}(\epsilon_{i_1}) \ldots (\epsilon_{i_n})$. (If J is empty $\mathfrak{T}(\epsilon_J) = \mathfrak{T}$.) The following then precisely characterizes the relationship of $\models_d$ and $\models_\epsilon$:

Proposition 3.1 $\mathfrak{T} \models_\epsilon i : X \sqsubseteq Y$ *if and only if for every subset $J \subseteq I$ not containing i, $\mathfrak{T}(\epsilon_J) \models_d i : X \sqsubseteq Y$.*

Moreover, in acyclic cases the relationship is even clearer:

Proposition 3.2 *Let $\mathfrak{T} = \langle \mathcal{T}_1, \mathcal{T}_2, \mathfrak{B}_{12} \rangle$ be a DTBox. Then*

(i) if $\mathcal{T}_1$ is consistent, then $\mathfrak{T} \models_\epsilon j : X \sqsubseteq Y$ if and only if $\mathfrak{T} \models_d j : X \sqsubseteq Y$.

(ii) if $\mathcal{T}_1$ is inconsistent, then $\mathfrak{T} \models_\epsilon j : X \sqsubseteq Y$ if and only if $\mathcal{T}_j \cup \{G \sqsubseteq \bot \mid i : A \xrightarrow{\sqsupseteq} j : G \in \mathfrak{B}_{ij}\} \models X \sqsubseteq Y$

Application to Other Frameworks As noted earlier, the problem of local inconsistency polluting the inferences of all the modules in a modular representation is quite general. We examine how the approach presented here can be applied to two previously proposed schemes.

[9] proposes an elegant notion of modular ontology which starts from the semantic framework of DDLs, but restricts bridge rules to "identities" defining new local names $j : N$ using concepts $i : C$ from T-box i, modulo a semantic domain correspondence exactly like r_{ij} for DDLs.[6] This can be modeled by replacing every definition $i : C \equiv j : N$ by the composed bridge rule $i : C \xrightarrow{\equiv} j : N$, Therefore the semantics involving holes introduced in the previous section can be applied to this approach, in order to localize inconsistencies in modules.

[6]Although [9] originally defines imported names using conjunctive queries over concepts and roles in T-box j, it then says that these can be "rolled up" into descriptions. Although this may in fact not always be doable, we will deal here with exactly those definitions for which this roll-up holds.

The Somewhere Peer Data Management System [8] consists of a collection of peers which maintain local ontologies, including repositories of "extensional data". Peers are acquainted with neighbors, whose concepts they can use, and query processing involves intensional distributed reasoning for query rewriting. Since this reasoning is (semantically) based on a single global interpretation, it is subject to the above mentioned difficulties due to inconsistency. In fact, for completeness, a *global* consistency check, involving even unconnected peers, would be required. We would suggest adopting a distributed semantics with holes, such as that of DDL. In particular, current peer links in Somewhere can be reduced to subsumption expressions like $1 : C \sqcap 3 : D \sqsubseteq 2 : E$. A DDL can be constructed from this by replacing occurrences of $i : C$ in peer j by new, local symbol C_i, and adding bridge rules $i : C \xrightarrow{\equiv} j : C_i$ and $i : \neg C \xrightarrow{\equiv} j : \neg C_i$. Our semantics then provides for directionality and locality, and the next section provides a distributed satisfiability testing algorithm for the semantics with holes.

Finally, the C-OWL [3] proposal for contextualized ontologies, uses the same model theory as DDL, so the particular version of "holes" given in this paper also gives C-OWL the directionality property, in addition to the localized inconsistency it already had.

4 Fixed-point semantics of bridge rules

As we saw earlier, combinations of bridge rules allow the propagation of subsumptions across T-boxes. To better understand how this propagation happens, we will associate with a set $\mathfrak{B}_{ij}$ of bridge rules an operator of the same name, which extends the j-th T-box with a set of subsumption axioms that are the transformation via bridge rules of subsumptions in the i-th T-Box.

Before proceeding further, we need to introduce the concept of disjoint union for interpretations. To begin with, we define as usual the disjoint union of two, possibly overlapping sets S and T as $S \uplus T = (S \times \{\#\}) \cup (T \times \{@\})$, where the values are distinguished by tupling with two discriminant symbols — $\#$ and @, in this case. This is generalized to the disjoint union $\biguplus_{i \in K} S_i$ of a collection of sets $\{S_i\}_{i \in K}$ indexed with (possibly infinite) K, by using the indeces i as the discriminants.

Definition 4.1 *Given two interpretations* $\mathcal{I} = \langle \Delta_{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$ *and* $\mathcal{J} = \langle \Delta_{\mathcal{J}}, \cdot^{\mathcal{J}} \rangle$ *of the same language* $\mathcal{L}$*, the* disjoint union of $\mathcal{I}$ and $\mathcal{J}$*, denoted by* $\mathcal{I} \uplus \mathcal{J}$*, is* $\langle \Delta_{\mathcal{I} \uplus \mathcal{J}}, \cdot^{\mathcal{I} \uplus \mathcal{J}} \rangle$*, where:*

1. $\Delta_{\mathcal{I} \uplus \mathcal{J}} = \Delta_{\mathcal{I}} \times \{\#\} \cup \Delta_{\mathcal{J}} \times \{@\}$*;*
2. *for concept* A*,* $(A)^{\mathcal{I} \uplus \mathcal{J}} = A^{\mathcal{I}} \times \{\#\} \cup A^{\mathcal{J}} \times \{@\}$*;*
3. *for role* R*,* $R^{\mathcal{I} \uplus \mathcal{J}} = \{\langle (x, \#), (y, \#) \rangle \mid \langle x, y \rangle \in R^{\mathcal{I}}\} \cup \{\langle (w, @), (z, @) \rangle \mid \langle w, z \rangle \in R^{\mathcal{J}}\}$

Disjoint union for interpretations $\biguplus_{k \in K} \mathcal{I}_k$ can similarly be generalized to the case of a sets. Intuitively the interpretation $\mathcal{I} \uplus \mathcal{J}$ is interpretation that is composed of two unrelated subparts one is isomorphic to $\mathcal{I}$ and the other to $\mathcal{J}$.

Definition 4.2 *A description logic family $\mathcal{DL}$ has the* disjoint union satisfi ability property *if $E^{\mathcal{I}'} = \biguplus_{k \in K} E^{\mathcal{I}_k}$ holds for all concepts and roles E over $\mathcal{DL}$, and for all interpretations $\mathcal{I}' = \biguplus_{k \in K} \mathcal{I}_k$.*

Lemma 4.1 *$\mathcal{SHIQ}$, and its sub-languages, have the distributed union satisfiability property.*

On the other hand, languages that support nominals (such as OWL), or A-boxes do not have this property.

The bridge operator The bridge operator essentially applies generalized subsumption propagation Lemma 2.2, to fi nd new subsumptions:

Definition 4.3 *Given a set of bridge rules $\mathfrak{B}_{12}$ from $\mathcal{DL}_1$ to $\mathcal{DL}_2$, the* bridge operator *$\mathfrak{B}_{12}(\cdot)$, taking as input a T-box in $\mathcal{DL}_1$ and producing a T-box in $\mathcal{DL}_2$, is defined as follows:*

$$\mathfrak{B}_{12}(\mathcal{T}_1) = \left\{ G \sqsubseteq \bigsqcup_{k=1}^{n} H_k \;\middle|\; \begin{array}{l} \mathcal{T}_1 \models A \sqsubseteq \bigsqcup_{k=1}^{n} B_k, \\ 1 : A \xrightarrow{\sqsupseteq} 2 : G \;\; \in \mathfrak{B}_{12}, \\ 1 : B_k \xrightarrow{\sqsubseteq} 2 : H_k \in \mathfrak{B}_{12}, \\ \textit{for } 1 \le k \le n, n \ge 0 \end{array} \right\}$$

(Notationally, $\bigsqcup_{k=1}^{0} D_k$ denotes $\bot$.)

It is remarkable that these are essentially *all* the inferences that one can get, *if* we use the semantics with holes:

Theorem 4.1 *Let $\mathfrak{T}_{12} = \langle \mathcal{T}_1, \mathcal{T}_2, \mathfrak{B}_{12} \rangle$ be a distributed T-box. If $\mathcal{DL}_1$ and $\mathcal{DL}_2$ have the distributed union satisfiability property then:*
$\mathfrak{T}_{12} \models_\epsilon 2 : X \sqsubseteq Y \iff \mathcal{T}_2 \cup \mathfrak{B}_{12}(\mathcal{T}_1) \models X \sqsubseteq Y$

For any family $\mathfrak{B} = \{\mathfrak{B}_{ij}\}_{i,j \in I}$ of bridge rules, we can combine these into a new operator $\mathfrak{B}$ on a family of T-boxes as follows:

$$\mathfrak{B}(\{\mathcal{T}_i\}_{i \in I}) = \left\{ \mathcal{T}_i \cup \bigcup_{j \neq i} \mathfrak{B}_{ji}(\mathcal{T}_j) \right\}_{i \in I}$$

If I is fi nite and each $\mathfrak{B}_{ij}$ is fi nite, then there is a positive integer b such that for every family of T-boxes $\mathbf{T}$, $\mathfrak{B}^b(\mathbf{T}) = \mathfrak{B}^{b+1}(\mathbf{T})$. Let us then defi ne $\mathfrak{B}^*(\mathbf{T})$ as $\mathfrak{B}^b(\mathbf{T})$, where b is the fi rst positive integer such that $\mathfrak{B}^b(\mathbf{T}) = \mathfrak{B}^{b+1}(\mathbf{T})$. Furthermore let $\mathfrak{B}^{b+1}(\mathbf{T})_i$, be the i-th T-box in $\mathfrak{B}^{b+1}(\mathbf{T})$.

Theorem 4.2 *For every $\mathfrak{T} = \langle \mathbf{T}, \mathfrak{B} \rangle$, $\mathfrak{T} \models_\epsilon j : X \sqsubseteq Y$ if and only if the j-th T-box of $\mathfrak{B}^*(\mathbf{T})$ entails $X \sqsubseteq Y$.*

Applications to Caching A number of researchers have considered the idea of caching locally the necessary information from the imported ontology $\mathcal{T}_{other}$, since this is assumed to be both more efficient (there is no need to interrupt local reasoning, while waiting for answers from the other ontology), and more prespicuous from the point of view of the local user: in order to understand an imported concept F, it is not necessary to understand *all* of $\mathcal{T}_{other}$, only the locally cached part, which is presumed to be much smaller. (This idea is also known as "subsetting", and there is considerable research on this topic in the ontology community .)

Theorem 4.2 above indicates that it is possible to finitely pre-compile in a sound and complete manner the subsumption information imported into a local ontology T_j by the bridge rules in a DTB $\mathfrak{T}$: compute and store it.

In a similar vein, [9] takes the set of imported concept definitions $\{N_k \equiv \mathit{other} : D_k \mid k = 1, \ldots, n\}$, and then computes and caches the subsumption hierarchy of the $\{N_k\}$. Since we have explained in Section 3 that the module mechanism in [9] can be represented as a DDL, Lemma 2.2 indicates that if the language contains at least $\mathcal{ALC}$, and if if it is possible to ask subsumption queries about complex concepts composed using the imported definitions, then it is not sufficient to cache only subsumptions of the form $D_{k1} \sqsubseteq D_{k2}$, since there may be additional subsumptions entailed, involving disjunctions. On the other hand, by Theorem 4.2 it is *sufficient* to cache all subsumptions of the form $N_k \sqsubseteq N_{k_1} \sqcup \ldots \sqcup N_{k_m}$, whose definitions satisfy the condition[7] $\mathcal{T}_{other} \models D_k \sqsubseteq D_{k_1} \sqcup \ldots \sqcup D_{k_m}$.

5 A distributed tableaux algorithm for DDL

In this section we describe a tableaux-based decision procedure for $\mathfrak{T} \models_\epsilon i : X \sqsubseteq Y$, for DTBoxes whose bridge graph $G_\mathfrak{T}$ is acyclic. The cyclic case is left for future work, pending the identification of a loop blocking strategy that preserves the independence of the local proofs.

To simplify the description, we suppose that local ontologies are expressed in (a subset of) the $\mathcal{SHIQ}$ language — one of the most widely known DLs. Also, we will assume that the consequences of bridge rules are atomic names. (This condition can easily be achieved by introducing, through definitions, names for the consequent concepts.). We need the usual notion of axiom internalization, as in [7]: Given T-box $\mathcal{T}_i$, the concept $C_{\mathcal{T}_i}$ is defined as $C_{\mathcal{T}_i} = \bigsqcap_{E \sqsubseteq D \in \mathcal{T}_i} \neg E \sqcup D$; also, the role hierarchy $R_{\mathcal{T}_i}$ contains the role axioms of $\mathcal{T}_i$, plus additional axioms $P \sqsubseteq U$, for each role P of $\mathcal{T}_i$, with U some fresh role.

The algorithm for testing j-satisfiability of concept expression D (i.e., checking $\mathfrak{T} \not\models_\epsilon j : D \sqsubseteq \bot$) builds, as usual, a finite representation of a distributed interpretation $\mathfrak{I}$, by running local *autonomous* $\mathcal{SHIQ}$ tableaux procedures to find each local interpretation $\mathcal{I}_i$ of $\mathfrak{I}$.

Definition 5.1 *For each $j \in I$, the function* **DTab**$_j$ *takes as input a concept X and tries to build a representation of $\mathcal{I}_j$ with $D^\mathcal{I} \neq \emptyset$ (called a* completion tree *[7]) for the*

[7] There is no need to iterate if we assume that imported names cannot be used in additional axioms of the local ontology – only for labeling information on the semantic web, for example.

concept $X \sqcap C_{\mathcal{T}_j} \sqcup \forall U.C_{\mathcal{T}_j}$*, using the* $\mathcal{SHIQ}$ *expansion rules, w.r.t. the role hierarchy* R_{T_j}*, plus the following additional "bridge" expansion rules*

Unsat-$\mathfrak{B}_{ij}$*-rule*

if 1. $G \in \mathcal{L}(x)$, $i : A \xrightarrow{\sqsupseteq} j : G \in \mathfrak{B}_{ij}$, *and*
2. *IsSat*$_i(A \sqcap \neg \bigsqcup \mathbf{B}') =$ *False, for some* $\mathbf{B}' \not\subseteq \mathcal{L}(x)$,
then $\mathcal{L}(x) \longrightarrow \mathcal{L}(x) \cup \{\bigsqcup \mathbf{B}'\}$

New-$\mathfrak{B}_{ij}$*-rule*

if 1. $G \in \mathcal{L}(x)$, $i : A \xrightarrow{\sqsupseteq} j : G \in \mathfrak{B}_{ij}$, *and*
2. $\mathbf{B} \subseteq \{B \mid i : B \xrightarrow{\sqsubseteq} j : H \in \mathfrak{B}_{ij}\}$, *and*
3. *for no* $\mathbf{B}' \subseteq \mathbf{B}$ *is IsSat*$_i(A \sqcap \neg \bigsqcup \mathbf{B}') =$ *False, and*
4. *for no* $\mathbf{B}' \supseteq \mathbf{B}$ *is IsSat*$_i(A \sqcap \neg \bigsqcup \mathbf{B}') =$ *True ,*
then if $\mathbf{DTab}_i(A \sqcap \neg \bigsqcup \mathbf{B}) =$ *Satisfi able*
then IsSat$_i(A \sqcap \neg \bigsqcup \mathbf{B}) =$ *True*
else IsSat$_i(A \sqcap \neg \bigsqcup \mathbf{B}) =$ *false*

The idea, inspired by bridge operator $\mathfrak{B}_{ij}(.)$, is that whenever $\mathbf{DTab}_j$ encounters a node x that contains a label G which is a consequence of an onto-bridge rule, then if $G \sqsubseteq \sqcup\mathbf{H}$ is entailed by the bridge rules, the label $\bigsqcup \mathbf{H}$, is added to x. To determine if $G \sqsubseteq \sqcup\mathbf{H}$ is entailed by bridge rules $\mathfrak{B}_{ij}$, $\mathbf{DTab}_j$ invokes $\mathbf{DTab}_i$ on the satisfi ability of the concept $A \sqcap \neg(\sqcup\mathbf{B})$. $\mathbf{DTab}_i$ will build (independently from $\mathbf{DTab}_j$) an interpretation $\mathcal{I}_i$, as illustrated in Figure 2. To avoid redundant calls, $\mathbf{DTab}_j$ caches the calls to $\mathbf{DTab}_i$ in a data structure *IsSat*$_i$, which caches the subsumption propagations that have been computed so far. Specifi cally, *IsSat*$_j(C)$ is initialized by $\mathbf{DTab}_j$ to *Undefined* for every C, and then *IsSat*$_i(C)$ will be set to *True/False* whenever $\mathfrak{T} \not\models_\epsilon i : C \sqsubseteq \bot$ is determined.

Theorem 5.1 (Termination) *For any acyclic DTBox* $\mathfrak{T}$ *and for any* $\mathcal{SHIQ}$ *concept* X, $\mathbf{DTab}_j(X)$ *terminates.*

Theorem 5.2 (Soundness and completeness) $j : X$ *is satisfiable in* $\mathfrak{T}$ *iff* $\mathbf{DTab}_j(X)$ *can generate a complete and clash-free completion tree.*

Note that the construction of the distributed interpretation can be parallelized, as each local tableaux procedure can run independently from the others, without checking for blocking conditions with nodes generated by the other local tableaux. The distributed tableaux proposed above has been implemented in a peer-to-peer setting, and is available for download. Some preliminary experiments have been done to see how distributed algorithm performs w.r.t. a global tableaux algorithm based on the encoding described in [2]. The results are reported in the technical report.

Example 5.1 *To clarify how* $\mathbf{DTab}_j$ *works, Table 1 reports the trace the call of*

$$\mathbf{DTab}_{\text{shoe}}(\textsf{BookAarticle} \sqcap \neg\textsf{Publication})$$

in Example 2.1. Notice that at step (3) $\mathbf{DTab}_{\text{shoe}}$ *applies the new-*$\mathfrak{B}_{ij}$*-rule, using the bridge rules (1) and (3) on page 3.*

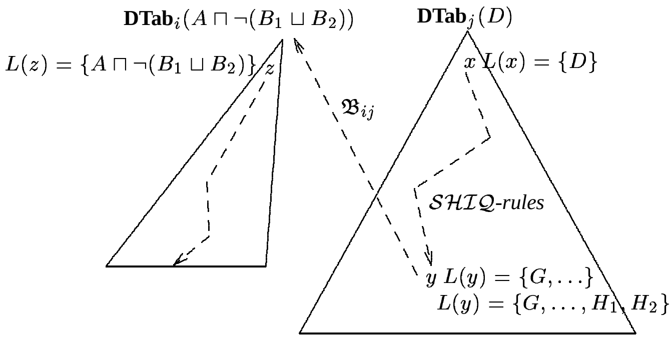

Figure 2: Subsumption propagation forced by $i : A \xrightarrow{\sqsupseteq} j : G$, $i : B_1 \xrightarrow{\sqsubseteq} j : H_1$ and $i : B_2 \xrightarrow{\sqsubseteq} j : H_2$

6 Conclusions

This paper has focused on the properties of a KR&R formalism that supports multiple, local knowledge bases connected by semantic mappings. It has proposed a *minor* modifi cation to the semantics of DDL, which has the salutary effect of localizing inconsistency to the connected component of a distributed reasoner in which it occurs, and of making "knowledge import" be directional. This resolves a problem left open in [2]. Not only is the new semantics easy to explain, but so are its effect on the inferences that continue to be valid. We believe that these features are desirable for *all* distributed and modular KR&R formalisms, and we provide several such applications [9, 3, 8]. We note here that although the e-connections formalism, and its perspectival restriction [4], can simulate DDLs, their fundamental goal is to obtain hybrid logics, and in such situations there is no motivation for localizing inconsistency.

The same semantics forms the foundation of a fi xed-point characterization of the formulas entailed by a DDL, which can be used as the basis of a sound and complete cache-based solution to the problem of distributed reasoning. It is also the standard with respect to which one can prove the completeness (as well as, of coursded, soudness) of a distributed tableaux reasoner which we have implemented in a peer-to-peer setting.

References

[1] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, P.F. Patel-Schneider, editors. *The Description Logic Handbook* Cambridge University Press, 2003.

[2] A. Borgida and L. Serafi ni. "Distributed description logics: Assimilating information from peer sources", *J. of Data Semantics 1(1)*, pp.153–184, 2003.

[3] P. Bouquet, F. Giunchiglia, F. van Harmelen, L. Serafi ni, and H. Stuckenschmidt. "Contextualizing ontologies", *J. of Web Semantics 1(4)*, 2004.

DTab$_{\mathtt{shoe}}$(BookAarticle ⊓ ¬Publication)

Step	**tree**	**labeling** and ***IsSat***	**rule**
(1)	x	$L(x) = \{$BookAarticle ⊓ ¬Publication$\}$ *IsSat* $= \{\}$	initial node
(2)	x	$L(x) = \{$BookAarticle, ¬Publication$\}$ *IsSat* $= \{\}$	⊓-rule
(3*)	x	$L(x) = \{$BookAarticle, ¬Publication$\}$ *IsSat* $= \{$InBook ⊓ ¬Publication $= False\}$	new-$\mathfrak{B}_{ij}$-rule
(4)	x	$L(x) = \{$ BookAarticle, ¬Publication, Publication $\}$ *IsSat* $= \{$InBook ⊓ ¬Publication $= False\}$	Unsat-$\mathfrak{B}_{ij}$-rule
(5)	x	$L(x) = \{\mathrm{Clash}\}$ *IsSat* $= \{$InBook ⊓ ¬Publication $= False\}$	Clash-rule

DTab$_{\mathtt{swrc}}$(InBook ⊓ ¬Publication)

Step	**tree**	**labeling** and ***IsSat***	**rule**
(1)	x	$L(x) = \{$ InBook ⊓ ¬Publication, ¬InBook ⊔ Publication $\}$ *IsSat* $= \{\}$	initial node
⋮			$\mathcal{SHIQ}$-rules
(n)	x	clash	

Table 1:

[4] B. Cuenca Grau, B. Parsia, E. Sirin. "Working with Multiple Ontologies on the Semantic Web", *Proc. ISWC'04*, LNCS vol.3298, pp.620–634. 2004.

[5] C. Ghidini and F. Giunchiglia. "Local model semantics, or contextual reasoning = locality + compatibility", *Artificial Intelligence*, 127(2):221–259, 2001.

[6] C. Ghidini and L. Serafini. "Distributed first order logics", *Proc. Frontiers of Combining Systems*, pp.121–139, 2000.

[7] I. Horrocks, U. Sattler, and S. Tobies. A description logic with transitive and converse roles, role hierarchies and qualifying number restriction. Technical Report 99-08, Technische Universität Dresden, LTCS, 1999.

[8] M-C. Rousset. "Small can be beautiful in the Semantic Web", *Proc. ISWC'04*, LNCS vol.3298, pp.6–16, 2004.

[9] H. Stuckenschmidt and M. Klein. "Integrity and change in modular ontologies", *IJCAI'03*, pp.900-908, 2003.