RWTH
LTCS–Report

Aachen University of Technology
Research group for
Theoretical Computer Science

Query Containment Using a DLR ABox
Preliminary Version

Ian Horrocks
University of Manchester

Ulrike Sattler
RWTH Aachen

Sergio Tessaris
University of Manchester

Stephan Tobies
RWTH Aachen

LTCS-Report 99-15

RWTH Aachen
LuFg Theoretische Informatik
http://www-lti.informatik.rwth-aachen.de

Ahornstr. 55
52074 Aachen
Germany

Query Containment Using a DLR ABox
Preliminary Version

Ian Horrocks
University of Manchester

Ulrike Sattler
RWTH Aachen

Sergio Tessaris
University of Manchester

Stephan Tobies
RWTH Aachen

November 16, 1999

Abstract

Query containment under constraints is the problem of determining whether the result of one query is contained in the result of another query for every database satisfying a given set of constraints. This problem is of particular importance in information integration and warehousing where, in addition to the constraints derived from the source schemas and the global schema, inter-schema constraints can be used to specify relationships between objects in different schemas. A theoretical framework for tackling this problem using the $\mathcal{DLR}$ logic has been established, and in this paper we show how the framework can be extended to a practical decision procedure. The proposed technique is to extend $\mathcal{DLR}$ with an Abox (a set of assertions about named individuals and tuples), and to transform query subsumption problems into $\mathcal{DLR}$ Abox satisfiability problems. We then show how such problems can be decided, via a reification transformation, using a highly optimised reasoner for the $\mathcal{SHIQ}$ description logic.

1 Introduction

Query containment under constraints is the problem of determining whether the result of one query is contained in the result of another query for every database satisfying a given set of constraints (derived, for example, from a schema). This problem is of particular importance in information integration (see [9]) and data warehousing where, in addition to the constraints derived from the source schemas and the global schema, inter-schema constraints can be used to specify relationships between objects in different schemas (see [6]).

This problem has been studied by Calvanese et al. [4]; they have established a theoretical framework using the logic $\mathcal{DLR}$,[1] presented several (un)decidability results, and described a method for solving the decidable cases using an embedding in the propositional dynamic logic CPDL$_g$ [12, 11]. However, this method does not lead directly

[1] Set semantics is assumed in this framework.

to a practical decision procedure as there is no (known) implementation of a CPDL$_g$ reasoner. Moreover, even if such an implementation were to exist, similar embedding techniques [10] have resulted in severe tractability problems when used, for example, to embed the $\mathcal{SHIF}$ description logic in $\mathcal{SHF}$ by eliminating inverse roles [13].

In this paper we present a practical decision procedure for the case where neither the queries nor the constraints contain regular expressions. This represents a restriction with respect to the framework described in Calvanese et al., where it was shown that the problem is still decidable if regular expressions are allowed in the schema and the (possibly) containing query, but this seems to be acceptable when modelling classical relational information systems, where regular expressions are seldom used [7, 6]. Moreover, the use of $\mathcal{DLR}$ in both schema and queries still allows for relatively expressive queries, and by staying within a strictly first order setting we are able to use a decision procedure that has demonstrated good empirical tractability.

The procedure is based on the method described by Calvanese et al., but extends $\mathcal{DLR}$ by defining an *ABox*, a set of axioms that assert facts about named *individuals* and tuples of named individuals (see [5]). This leads to a much more natural encoding of queries (there is a direct correspondence between variables and individuals), and allows the problem to be reduced to that of determining the satisfiability of a $\mathcal{DLR}$ *knowledge base* (KB), i.e., a combined schema and ABox. This problem can in turn be reduced to a KB satisfiability problem in the $\mathcal{SHIQ}$ description logic, with n-ary relations reduced to binary ones by reification. In [16], a similar approach is presented. However, the underlying description logic ($\mathcal{ALCNR}$) is less expressive than $\mathcal{DLR}$ and $\mathcal{SHIQ}$ (for example, it is not able to capture Entity-Relationship schemas).

We have good reasons to believe that this approach represents a practical solution. In the FaCT system [13] we already have an (optimised) implementation of the decision procedure for $\mathcal{SHIQ}$ schema satisfiability described in [15], and using FaCT we have been able to reason very efficiently with a realistic schema derived from the integration of several Entity-Relationship schemas using $\mathcal{DLR}$ inter-schema constraints.[2] In Section 4 we show how this algorithm can be straightforwardly extended to deal with ABox axioms. As the number of individuals generated by the encoding of realistic problems will be relatively small, this extension should not compromise the empirical tractability.

2 Preliminaries

In this section we will (briefly) define the key components of our framework, namely the logic $\mathcal{DLR}$, (conjunctive) queries, and the logic $\mathcal{SHIQ}$.

2.1 The Logic $\mathcal{DLR}$

We will begin with $\mathcal{DLR}$ as it is used in the definition of both schemas and queries. $\mathcal{DLR}$ is a description logic (DL) extended with the ability to describe relations of any arity.

[2]The schemas and constraints were taken from a case study undertaken as part of the Esprit DWQ project [7, 6].

Definition 2.1 Given a set of atomic concept names NC and a set of atomic relation names NR, every $C \in \mathsf{NC}$ is a concept and every $\mathbf{R} \in \mathsf{NR}$ is a relation, with every $\mathbf{R}$ having an associated arity. If C, D are concepts, $\mathbf{R}, \mathbf{S}$ are relations of arity n, i is an integer $1 \leqslant i \leqslant n$, and k is a non-negative integer, then

$$\begin{array}{rl} \top, \neg C, C \sqcap D, \exists[\$i]\mathbf{R}, (\leq k[\$i]\mathbf{R}) & \text{are } \mathcal{DLR} \textit{ concepts}, \text{ and} \\ \top_n, \neg\mathbf{R}, \mathbf{R} \sqcap \mathbf{S}, (\$i/n : C) & \text{are } \mathcal{DLR} \textit{ relations} \text{ with arity } n. \end{array}$$

Relation expressions must be well typed in the sense that only relations with the same arity can be conjoined, and in constructs like $\exists[\$i]\mathbf{R}$ the value of i must be less than or equal to the arity of $\mathbf{R}$.

The semantics of $\mathcal{DLR}$ is given in terms of *interpretations* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is the domain (a non-empty set), and $\cdot^{\mathcal{I}}$ is an interpretation function that maps every concept to a subset of $\Delta^{\mathcal{I}}$ and every n-ary relation to a subset of $(\Delta^{\mathcal{I}})^n$ such that the following equations are satisfied ("$\sharp$" denotes set cardinality).

$$\begin{array}{rr} \top^{\mathcal{I}} = \Delta^{\mathcal{I}} & (C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}} \\ \neg C^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}} & (\exists[\$i]\mathbf{R})^{\mathcal{I}} = \left\{ d \in \Delta^{\mathcal{I}} \mid \exists (d_1, \ldots, d_n) \in \mathbf{R}^{\mathcal{I}}.d_i = d \right\} \\ & (\leq k[\$i]\mathbf{R})^{\mathcal{I}} = \left\{ d \in \Delta^{\mathcal{I}} \mid \sharp \left\{ (d_1, \ldots, d_n) \in \mathbf{R}^{\mathcal{I}}.d_i = d \right\} \leq k \right\} \\ {\top_n}^{\mathcal{I}} \subseteq (\Delta^{\mathcal{I}})^n & \mathbf{R}^{\mathcal{I}} \subseteq {\top_n}^{\mathcal{I}} \\ (\neg\mathbf{R})^{\mathcal{I}} = {\top_n}^{\mathcal{I}} \setminus \mathbf{R}^{\mathcal{I}} & (\mathbf{R} \sqcap \mathbf{S})^{\mathcal{I}} = \mathbf{R}^{\mathcal{I}} \cap \mathbf{S}^{\mathcal{I}} \\ & (\$i/n : C)^{\mathcal{I}} = \{ (d_1, \ldots, d_n) \in {\top_n}^{\mathcal{I}} \mid d_i \in C^{\mathcal{I}} \} \end{array}$$

Note that $\top_n$ does not need to be interpreted as the set of all tuples of arity n, but only as a subset of them, and that the negation of a relation $\mathbf{R}$ with arity n is relative to $\top_n$.

In our framework, a schema consists of a set of logical inclusion axioms expressed in $\mathcal{DLR}$. These axioms could be derived from the translation into $\mathcal{DLR}$ of schemas expressed in some other data modelling formalism (such as Entity-Relationship modelling [3, 8]), or could directly stem from the use of $\mathcal{DLR}$ to express, for example, inter-schema constraints to be used in data warehousing, (see [6]).

Definition 2.2 A $\mathcal{DLR}$ *schema* $\mathcal{S}$ is a set of *axioms* of the form $C \sqsubseteq D$ and $\mathbf{R} \sqsubseteq \mathbf{S}$, where C, D are $\mathcal{DLR}$ concepts and $\mathbf{R}, \mathbf{S}$ are $\mathcal{DLR}$ relations of the same arity; an interpretation $\mathcal{I}$ *satisfies* $C \sqsubseteq D$ (written $\mathcal{I} \models C \sqsubseteq D$) iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, and it satisfies $\mathbf{R} \sqsubseteq \mathbf{S}$ (written $\mathcal{I} \models \mathbf{R} \sqsubseteq \mathbf{S}$) iff $\mathbf{R}^{\mathcal{I}} \subseteq \mathbf{S}^{\mathcal{I}}$. An interpretation $\mathcal{I}$ *satisfies* a schema $\mathcal{S}$ iff $\mathcal{I}$ satisfies every axiom in $\mathcal{S}$.

Crucially, we extend $\mathcal{DLR}$ to assert properties of *individuals*, names representing single elements of the domain. An *ABox* is a set of axioms asserting facts about individuals and tuples of individuals.

Definition 2.3 Given a set of individuals NI, a $\mathcal{DLR}$ *ABox* $\mathcal{A}$ is a set of *axioms* of the form $w{:}C$ and $\vec{w}{:}\mathbf{R}$, where C is a concept, $\mathbf{R}$ is a relation of arity n, w is an individual and $\vec{w}$ is an n-tuple $\langle w_1, \ldots, w_n \rangle$ such that $w_1, \ldots, w_n$ are individuals. We will often write w_i to refer to the ith element of an n-tuple $\vec{w}$, where $1 \leqslant i \leqslant n$.

Additionally, the interpretation function $\cdot^{\mathcal{I}}$ maps every individual to an element of $\Delta^{\mathcal{I}}$. An interpretation $\mathcal{I}$ *satisfies* an axiom $w{:}C$ (written $\mathcal{I} \models w{:}C$) iff $w^{\mathcal{I}} \in C^{\mathcal{I}}$, and

it *satisfies* an axiom $\vec{w}$:**R** (written $\mathcal{I} \models \vec{w}$:**R**) iff $\vec{w}^{\mathcal{I}} \in \mathbf{R}^{\mathcal{I}}$. An interpretation $\mathcal{I}$ *satisfies* an ABox $\mathcal{A}$ iff $\mathcal{I}$ satisfies every axiom in $\mathcal{A}$.

A *knowledge base* (KB) $\mathcal{K}$ is a pair $\langle \mathcal{S}, \mathcal{A} \rangle$, where $\mathcal{S}$ is a schema and $\mathcal{A}$ is an ABox. An interpretation $\mathcal{I}$ *satisfies* a KB $\mathcal{K}$ iff it satisfies both $\mathcal{S}$ and $\mathcal{A}$.

If an interpretation $\mathcal{I}$ satisfies a concept, schema, or ABox X, then we say that $\mathcal{I}$ is a *model* of X, call X *satisfiable*, and write $\mathcal{I} \models X$.

Note that it is not assumed that individuals with different names are mapped to different elements in the domain (the so-called unique name assumption).

Definition 2.4 If $\mathcal{K}$ is a KB, $\mathcal{I}$ is a model of $\mathcal{K}$, and $\mathcal{A}$ is an ABox, then $\mathcal{I}'$ is called an *extension* of $\mathcal{I}$ to $\mathcal{A}$ iff $\mathcal{I}'$ satisfies $\mathcal{A}$, $\Delta^{\mathcal{I}} = \Delta^{\mathcal{I}'}$, and all concepts, roles and individuals occuring in $\mathcal{K}$ are interpreted identically by $\mathcal{I}$ and $\mathcal{I}'$.

Given two ABoxes $\mathcal{A}, \mathcal{A}'$ and a schema $\mathcal{S}$, $\mathcal{A}$ is *included* in $\mathcal{A}'$ w.r.t. $\mathcal{S}$ (written $\langle \mathcal{S}, \mathcal{A} \rangle | \approx \mathcal{A}'$) iff every model $\mathcal{I}$ of $\langle \mathcal{S}, \mathcal{A} \rangle$ can be extended to $\mathcal{A}'$.

2.2 Queries

In this extended abstract we will consider only conjunctive queries (see [1, chap. 4]). A *conjunctive query* q is an expression

$$q(\vec{x}) \leftarrow term_1(\vec{x}, \vec{y}, \vec{c}) \wedge \ldots \wedge term_n(\vec{x}, \vec{y}, \vec{c})$$

where $\vec{x}$, $\vec{y}$, and $\vec{c}$ are tuples of *distinguished* variables, variables, and constants, respectively (distinguished variables appear in the answer, "ordinary" variables are used only in the query expression, and constants are fixed values). Each term $term_i(\vec{x}, \vec{y}, \vec{c})$ is called an atom in q and is in one of the forms $C(w)$ or $\mathbf{R}(\vec{w})$, where w (resp. $\vec{w}$) is a variable or constant (resp. tuple of variables and constants) in $\vec{x}, \vec{y}, \vec{c}$, C is a $\mathcal{DLR}$ concept and $\mathbf{R}$ is a $\mathcal{DLR}$ relation.[3]

For example, a query designed to return the bus number of the city buses travelling in both directions between two stops is:

$$\text{BUS}(nr) \leftarrow \text{bus_route}(nr, stop_1, stop_2) \wedge \text{bus_route}(nr, stop_2, stop_1) \wedge \text{city_bus}(nr)$$

where nr is a distinguished variable (it appears in the answer), $stop_1$ and $stop_2$ are non-distinguished variables, city_bus is a $\mathcal{DLR}$ concept and bus_route is a $\mathcal{DLR}$ relation.

In this framework, the *evaluation* of a query q with n distinguished variables w.r.t. a $\mathcal{DLR}$ interpretation[4] $\mathcal{I}$ (written $q(\mathcal{I})$) is the set of n-tuples $\vec{d} \in (\Delta^{\mathcal{I}})^n$ such that

$$\mathcal{I} \models \exists \vec{y}.term_1(\vec{d}, \vec{y}, \vec{c}) \wedge \ldots \wedge term_n(\vec{d}, \vec{y}, \vec{c}).$$

A query $q_1(\vec{x})$ is *contained* in a query $q_2(\vec{x})$ w.r.t. a schema $\mathcal{S}$ (written $\mathcal{S} \models q_1 \sqsubseteq q_2$), iff, for every possible model $\mathcal{I}$ of $\mathcal{S}$, $q_1(\mathcal{I}) \subseteq q_2(\mathcal{I})$.

For example, the schema containing the axioms

$$(\text{bus_route} \sqcap (\$1/3 : \text{city_bus})) \sqsubseteq \text{city_bus_route} \quad \text{and}$$
$$\text{city_bus_route} \sqsubseteq (\text{bus_route} \sqcap (\$1/3 : \text{city_bus})),$$

[3] The fact that these concepts and relations can also appear in the schema is one of the distinguishing features of this approach.

[4] Here perceived as standard FO interpretation.

states that the relation city_bus_route contains exactly the bus_route information that concerns city buses. It is easy to see that the following CITY_BUS query

$$\text{CITY_BUS}(nr) \leftarrow \text{city_bus_route}(nr, stop_1, stop_2) \wedge \text{city_bus_route}(nr, stop_2, stop_1)$$

is equivalent to the previous BUS query w.r.t. the given schema. In an information integration scenario, for example, this could be exploited by reformulating the BUS query as a CITY_BUS query ranging over a smaller database without any loss of information.

2.3 The Logic $\mathcal{SHIQ}$

$\mathcal{SHIQ}$ is a standard DL, in the sense that it deals with concepts and (only) binary relations (called *roles*), but it is unusually expressive in that it supports reasoning with inverse roles, qualifying number restrictions on roles, transitive roles, and role inclusion axioms.

Definition 2.5 Given a set of atomic concept names NC and a set of atomic role names NR with transitive role names $\mathsf{NR}_+ \subseteq \mathsf{NR}$, every $C \in \mathsf{NC}$ is a concept, every $R \in \mathsf{NR}$ is a role, and every $R \in \mathsf{NR}_+$ is a transitive role. If R is a role, then R^- is also a role (and if $R \in \mathsf{NR}_+$ then R^- is also a transitive role). If S is a (possibly inverse) role, C, D are concepts, and k is a non-negative integer, then

$$\top, \neg C, C \sqcap D, \exists S.C, \leqslant k S.C \text{ are also } \mathcal{SHIQ} \text{ concepts.}$$

The semantics of $\mathcal{SHIQ}$ is given in terms of *interpretations* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is the domain (a non-empty set), and $\cdot^{\mathcal{I}}$ is an interpretation function that maps every concept to a subset of $\Delta^{\mathcal{I}}$ and every role to a subset of $(\Delta^{\mathcal{I}})^2$ such that the following equations are satisfied.

$$\begin{array}{rr}
\top^{\mathcal{I}} = \Delta^{\mathcal{I}} & (\exists S.C)^{\mathcal{I}} = \{d \mid \exists d'.(d,d') \in S^{\mathcal{I}} \text{ and } d' \in C^{\mathcal{I}}\} \\
\neg C^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}} & (\leqslant k S.C)^{\mathcal{I}} = \{d \mid \sharp\{d'.(d,d') \in S^{\mathcal{I}} \text{ and } d' \in C^{\mathcal{I}}\} \leqslant k\} \\
(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}} & R^{\mathcal{I}} = (R^{\mathcal{I}})^+ \text{ for all } R \in \mathsf{NR}_+ \\
 & (R^-)^{\mathcal{I}} = \{(d',d) \mid (d,d') \in R^{\mathcal{I}}\}
\end{array}$$

$\mathcal{SHIQ}$ schemas, ABoxes, and KBs are defined similarly to those for $\mathcal{DLR}$: if C, D are concepts, R, S are roles, and v, w are individuals, then a schema $\mathcal{S}$ consists of axioms of the form $C \sqsubseteq D$ and $R \sqsubseteq S$, and an ABox $\mathcal{A}$ consists of axioms of the form $w{:}C$ and $\langle v, w\rangle{:}R$. Again, a KB $\mathcal{K}$ is a pair $\langle \mathcal{S}, \mathcal{A}\rangle$, where $\mathcal{S}$ is a schema and $\mathcal{A}$ is an ABox.

The definitions of interpretations, satisfiability, and models also parallel those for $\mathcal{DLR}$, and there is again no unique name assumption.

Note that, in order to maintain decidability, the roles that can appear in number restrictions are restricted [15]: if a role R occurs in a number restriction $\leqslant k S.C$, then neither S nor any of its sub roles may be transitive (i.e., if the schema contains a $\sqsubseteq$-path from S' to S, then S' is not transitive).

3 Determining Query Containment

In this section we will describe how the problem of deciding whether one query is contained in another one w.r.t. a schema can be reduced to the problem of deciding KB satisfiability in the $\mathcal{SHIQ}$ description logic. There are three steps to this reduction. Firstly, the queries are transformed into $\mathcal{DLR}$ ABoxes $\mathcal{A}_1$ and $\mathcal{A}_2$ such that $\mathcal{S} \models q_1 \sqsubseteq q_2$ iff $\langle \mathcal{S}, \mathcal{A}_1 \rangle |\approx \mathcal{A}_2$ (see Definition 2.4). Secondly, the ABox inclusion problem is transformed into one or more KB satisfiability problems. Finally, we show how a $\mathcal{DLR}$ KB can be transformed into an equisatisfiable $\mathcal{SHIQ}$ KB.

3.1 Transforming Query Containment into ABox Inclusion

We will first show how a query can be transformed into a *canonical* $\mathcal{DLR}$ ABox. Such an ABox represents a generic pattern that must be matched by all tuples in the evaluation of the query.

Definition 3.1 Let q be a conjunctive query. The *canonical ABox* for q is defined by

$$\mathcal{A}_q = \{\vec{w}{:}\mathbf{R} \mid \mathbf{R}(\vec{w}) \text{ is an atom in } q\} \cup \{w{:}C \mid C(w) \text{ is an atom in } q\}.$$

We introduce a new atomic concept P_w for every individual w in $\mathcal{A}$ and define the *completed* canonical ABox for q by

$$\widehat{\mathcal{A}}_q = \mathcal{A}_q \cup \{w{:}P_w \mid w \text{ occurs in } \mathcal{A}_q\} \cup \{w_i{:}\neg P_{w_j} \mid w_i, w_j \text{ are constants in } q \text{ and } i \neq j\}.$$

The axioms $w{:}P_w$ in $\widehat{\mathcal{A}}_q$ introduce *representative concepts* for each individual w in $\mathcal{A}_q$. They are used (in the axioms $w_i{:}\neg P_{w_j}$) to ensure that individuals corresponding to different constants in q cannot have the same interpretation,[5] and will also be useful in the transformation to KB satisfiability.

By abuse of notation we will say that an interpretation $\mathcal{I}$, and an assignment ρ of distinguished variables, non-distinguished variables and constants to elements in the domain of $\mathcal{I}$ such that $\mathcal{I} \models \rho(q)$, define a model for $\mathcal{A}_q$ with the interpretation of the individuals corresponding with ρ and the interpretation $P_w^{\mathcal{I}} = \{w^{\mathcal{I}}\}$.

We can use this definition to transform the query containment problem into a (very similar) problem involving $\mathcal{DLR}$ ABoxes. We can assume that the names of the non-distinguished variables in q_2 differ from those in q_1 (arbitrary names can be chosen without affecting the evaluation of the query), and that the names of distinguished variables and constants appear in both queries (if a name is missing in one of the queries, it can be simply added using a term like $\top(v)$).

The following Theorem shows that a canonical ABox really captures the structure of a query, allowing the query containment problem to be restated as an ABox inclusion problem.

Theorem 1 *Given a schema $\mathcal{S}$ and two queries q_1 and q_2, $\mathcal{S} \models q_1 \sqsubseteq q_2$ iff $\langle \mathcal{S}, \widehat{\mathcal{A}}_{q_1} \rangle |\approx \mathcal{A}_{q_2}$.*

[5] A standard assumption in the database setting.

PROOF: For the if direction, assume $\mathcal{S} \not\models q_1 \sqsubseteq q_2$. Then there exists a model $\mathcal{I}$ of $\mathcal{S}$ and a tuple $(d_1, \ldots, d_n) \in (\Delta^{\mathcal{I}})^n$ such that $(d_1, \ldots, d_n) \in q_1(\mathcal{I})$ and $(d_1, \ldots, d_n) \notin q_2(\mathcal{I})$. $\mathcal{I}$ and the assignment of variables leading to $(d_1, \ldots, d_n)$ define a model for $\widehat{\mathcal{A}}_{q_1}$. If $\cdot^{\mathcal{I}}$ could be extended to satisfy $\mathcal{A}_{q_2}$, then the extension would correspond with an assignment of the non-distinguished variables in q_2 such that $(d_1, \ldots, d_n) \in q_2(\mathcal{I})$, thus contradicting the assumption.

For the only if direction, assume there is a model $\mathcal{I}$ of both $\mathcal{S}$ and $\widehat{\mathcal{A}}_{q_1}$ that cannot be extended to a model of $\mathcal{A}_{q_2}$. Hence there is a tuple $(d_1, \ldots, d_n) \in q_1(\mathcal{I})$ and a corresponding assignment of variables that define $\mathcal{I}$. If there is an assignment of the non-distinguished variables in q_2 such that $(d_1, \ldots, d_n) \in q_2(\mathcal{I})$, then this assignment would define the extension of $\mathcal{I}$ such that $\mathcal{A}_{q_2}$ is also satisfied. ∎

The representative concepts P_w in a completed canonical ABox $\widehat{\mathcal{A}}$ have the useful property that, without loss of generality, we can often restrict our attention to interpretations in which, for every w occurring in $\widehat{\mathcal{A}}$, $P_w^{\mathcal{I}} = \{w^{\mathcal{I}}\}$.

Lemma 1 *Let $\mathcal{S}$ be a schema, $\mathcal{A}$ a canonical ABox and $\widehat{\mathcal{A}}$ the completed version of $\mathcal{A}$. If $\mathcal{I}$ is an interpretation such that $\mathcal{I} \models \langle \mathcal{S}, \widehat{\mathcal{A}} \rangle$, then there exists an interpretation $\mathcal{I}'$ where $\Delta^{\mathcal{I}} = \Delta^{\mathcal{I}'}$, $P_w^{\mathcal{I}'} = \{w^{\mathcal{I}}\}$ for all individuals w occuring in $\mathcal{A}$ and their corresponding representative concepts P_w, $\cdot^{\mathcal{I}'}$ is the same as $\cdot^{\mathcal{I}}$ in every other respect, and $\mathcal{I}' \models \langle \mathcal{S}, \widehat{\mathcal{A}} \rangle$.*

PROOF: From the semantics it is clear that the interpretation of a concept C depends only on the interpretations of the atomic concepts and roles that appear syntactically in C, and from Definition 3.1, P_{w_i} appears only in axioms of the form $w_i{:}P_{w_i}$ and $w_j{:}\neg P_{w_i}$ in $\widehat{\mathcal{A}} \setminus \mathcal{A}$. Therefore $\mathcal{I}'$ satisfies all the axioms $C \sqsubseteq D$ and $\mathbf{R} \sqsubseteq \mathbf{S}$ in $\mathcal{S}$ and all the axioms $w{:}C$ in $\mathcal{A}$, because $\mathcal{I} \models \langle \mathcal{S}, \widehat{\mathcal{A}} \rangle$ and all the C, D, $\mathbf{R}$, $\mathbf{S}$ and w are identically interpreted by $\mathcal{I}$ and $\mathcal{I}'$.

Moreover, $\mathcal{I}'$ also satisfies both kinds of axiom in $\widehat{\mathcal{A}} \setminus \mathcal{A}$. It obviously satisfies the axioms of the form $w{:}P_w$ because $w^{\mathcal{I}'} = w^{\mathcal{I}}$ and $P_w^{\mathcal{I}'} = \{w^{\mathcal{I}}\}$, so $w^{\mathcal{I}'} \in P_w^{\mathcal{I}'}$. It also satisfies the axioms of the form $w_i{:}\neg P_{w_j}$, where $w_i \neq w_j$, because from $w_j^{\mathcal{I}} \in P_{w_j}^{\mathcal{I}}$ and $w_i^{\mathcal{I}} \notin P_{w_j}^{\mathcal{I}}$ we have $w_i^{\mathcal{I}} \neq w_j^{\mathcal{I}}$, and $P_{w_j}^{\mathcal{I}'} = \{w_j^{\mathcal{I}}\}$, so $w_i^{\mathcal{I}'} \notin P_{w_j}^{\mathcal{I}'}$. ∎

3.2 Transforming ABox Inclusion into ABox Satisfiability

Next, we will show how to transform the ABox inclusion problem into one or more KB satisfiability problems. In order to do this, there are two main difficulties that must be overcome. The first is that, in order to transform inclusion into satisfiability, we would like to be able to "negate" axioms. This is easy for axioms of the form $w{:}C$, because an interpretation satisfies $w{:}\neg C$ iff it does not satisfy $w{:}C$. However, we cannot deal with axioms of the form $\vec{w}{:}\mathbf{R}$ in this way, because $\mathcal{DLR}$ only has a weak form of negation for relations relative to $\top_n$. Our solution is to transform all axioms in $\mathcal{A}_{q_2}$ into the form $w{:}C$.

The second difficulty is that $\mathcal{A}_{q_2}$ may contain individuals corresponding to non-distinguished variables in q_2 (given the symmetry between queries and ABoxes, we will refer to them from now on as non-distinguished individuals). These individuals

introduce an extra level of quantification that we cannot deal with using our standard reasoning procedures: $\langle \mathcal{S}, \widehat{\mathcal{A}}_{q_1} \rangle \approx \mathcal{A}_{q_2}$ iff for all models $\mathcal{I}$ of $\langle \mathcal{S}, \widehat{\mathcal{A}}_{q_1} \rangle$ *there exists* some extension of $\mathcal{I}$ to $\mathcal{A}_{q_2}$. We deal with this problem by eliminating the non-distinguished individuals from $\mathcal{A}_{q_2}$.

We will begin by exploiting some general properties of ABoxes that allow us to *compact* $\mathcal{A}_{q_2}$ so that it contains only one axiom $\vec{w}{:}\mathbf{R}$ for each tuple $\vec{w}$, and one axiom $w{:}C$ for each individual w that is not an element in any tuple. It is obvious from the semantics that we can combine all ABox axioms relating to the same individual or tuple: $\mathcal{I} \models \{w{:}C, w{:}D\}$ (resp. $\{\vec{w}{:}\mathbf{R}, \vec{w}{:}\mathbf{S}\}$) iff $\mathcal{I} \models \{w{:}(C \sqcap D)\}$ (resp. $\{\vec{w}{:}(\mathbf{R} \sqcap \mathbf{S})\}$). The following lemma shows that we can also absorb $w_i{:}C$ into $\vec{w}{:}\mathbf{R}$ when w_i is an element of $\vec{w}$.

Lemma 2 *Let $\mathcal{A}$ be a $\mathcal{DLR}$ ABox with $\{w_i{:}C, \vec{w}{:}\boldsymbol{R}\} \subseteq \mathcal{A}$, where w_i is the ith element in $\vec{w}$. Then $\mathcal{I} \models \mathcal{A}$ iff $\mathcal{I} \models \{\vec{w}{:}(\boldsymbol{R} \sqcap \$i : C)\} \cup \mathcal{A} \setminus \{w_i{:}C, \vec{w}{:}\boldsymbol{R}\}$.*

PROOF: From the semantics, if $\vec{w}^{\mathcal{I}} \in (\mathbf{R} \sqcap \$i : C)^{\mathcal{I}}$, then $\vec{w}^{\mathcal{I}} \in \mathbf{R}^{\mathcal{I}}$ and $w_i^{\mathcal{I}} \in C^{\mathcal{I}}$, and if $w_i^{\mathcal{I}} \in C^{\mathcal{I}}$ and $\vec{w}^{\mathcal{I}} \in \mathbf{R}^{\mathcal{I}}$, then $\vec{w}^{\mathcal{I}} \in (\mathbf{R} \sqcap \$i : C)^{\mathcal{I}}$. ∎

The ABox resulting from exhaustive application of Lemma 2 can be represented as a graph, with a node for each tuple, a node for each individual, and edges connecting tuples with the individuals that compose them. The graph will consist of one or more connected components, where each component is either a single individual (representing an axiom $w{:}C$, where w is not an element in any tuple) or a set of tuples linked by common elements (representing axioms of the form $\vec{w}{:}\mathbf{R}$). As they do not have any individuals in common, we can deal independently with the inclusion problem for each connected set of axioms: $\langle \mathcal{S}, \mathcal{A} \rangle \approx \mathcal{A}'$ iff $\langle \mathcal{S}, \mathcal{A} \rangle \approx \mathcal{G}$ for every connected set of axioms $\mathcal{G} \subseteq \mathcal{A}'$.

Returning to our original problem, we will now show how we can *collapse* a connected component of $\mathcal{A}_{q_2}$

$$\mathcal{G} = \{\vec{w}_1{:}\mathbf{R}_1, \ldots, \vec{w}_n{:}\mathbf{R}_n\}$$

into a single axiom of the form $w{:}C$, where w (the "root" individual) is an element of one of the tuples $\vec{w}_1 \ldots \vec{w}_n$ occurring in $\mathcal{G}$, C is a concept that describes $\mathcal{G}$ from the point of view of w, and $\langle \mathcal{S}, \widehat{\mathcal{A}}_{q_1} \rangle \approx \mathcal{G}$ iff $\langle \mathcal{S}, \widehat{\mathcal{A}}_{q_1} \rangle \approx \{w{:}C\}$. The collapsing procedure works by replacing each axiom $\vec{w}_i{:}\mathbf{R}_i$ with an axiom of the form $w{:}C$ (where w is an element of $\vec{w}_i$), which can then be absorbed into another axiom $\vec{w}_j{:}\mathbf{R}_j$ (where w is an element of $\vec{w}_j$) using Lemma 2. A recursive traversal of the graphical representation of $\mathcal{G}$ is used to choose the order in which to apply the replacements and absorptions so that $\mathcal{G}$ is collapsed into a single axiom (a similar technique is used in [4] to transform queries into concepts). During the collapsing procedure, new concepts Q_{w_i} may be introduced to represent non-distinguished individuals w_i that occur in $\mathcal{G}$. These concepts serve only as "place-holders", and will be replaced when $\mathcal{G}$ is completely collapsed.

A traversal starts at an (arbitrary) individual node w (the "root") and proceeds as follows.

- At an individual node x, the node is first marked as visited. Then, while there remains an unmarked tuple node connected to x, one of these, $\vec{w}$, is selected,

visited, and the axiom $\vec{w}$:**R** is replaced with the axiom

$$x{:}\exists[\$i](\mathbf{R} \sqcap \prod_{1 \leqslant j \leqslant n . j \neq i} (\$j/n : C_{w_j})),$$

where $\vec{w} = \langle w_1, \ldots, w_n \rangle$, x is the ith element of $\vec{w}$, w_j is the jth element of $\vec{w}$, and C_{w_j} is either the representative concept P_{w_j}, if w_j is an individual occurring in $\widehat{\mathcal{A}}_{q_1}$, or a concept Q_{w_j} otherwise. Finally, any axioms $x{:}C_1, \ldots, x{:}C_n$ resulting from visiting the unmarked tuples connected to x are merged into a single axiom $x{:}(C_1 \sqcap \ldots \sqcap C_n)$.

- At a tuple node $\vec{w}$, the node is first marked as visited. Then, while there remains an unmarked individual node connected to $\vec{w}$, one of these, x, is selected, visited, and any axiom $x{:}C$ that results from the visit is merged into the axiom $\vec{w}$:**R** using Lemma 2.

After the traversal, $\mathcal{G}$ has been reduced to a single axiom $w{:}C$, but it may contain concepts Q_{w_i} that were introduced during the collapsing procedure as representatives for non-distinguished individuals. As these concepts do not occur in $\langle \mathcal{S}, \widehat{\mathcal{A}}_{q_1} \rangle$, they must be eliminated if the inclusion relationship is to be preserved. This is easy for concepts Q_{w_i} that occur only once in C, and where w_i is not the root individual (i.e., $w \neq w_i$): as w_i is "referred to" only once in the collapsed axiom, and can be freely interpreted when a model $\mathcal{I}$ of $\langle \mathcal{S}, \widehat{\mathcal{A}}_{q_1} \rangle$ is extended to $\mathcal{G}$, Q_{w_i} can simply be replaced with $\top$ (this will be shown more formally in Lemma 3).

This solution cannot be adopted for a concept Q_{w_i} that occurs more than once in C, or that occurs at least once in C when $w = w_i$, because w_i must have the same interpretation everywhere it is "referred to" in the collapsed axiom. However, in this case we can deal with Q_{w_i} by exploiting the fact that the individual w_i must occur in a cycle in the graph representing $\mathcal{G}$. An individual w is in a cycle in the graph if there is a path leading from the node representing w back to itself in which the same edge is never traversed (in either direction) more than once. As the marking of nodes during the traversal ensures that the same edge is never traversed more than once, w_i must have been in such a cycle.

Given the correspondence between the graph and the axioms in $\mathcal{G}$, it is obvious that $\mathcal{G}$ can only be satisfied by an interpretation $\mathcal{I}$ in which $w_i^{\mathcal{I}}$ is also in a relational cycle (the cycle is explicitly asserted by the axioms in $\mathcal{G}$). Moreover, given that $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle \not\approx \mathcal{G}$, and that extending an interpretation of $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle$ to $\mathcal{G}$ cannot extend the interpretation of any relation, then such a cycle must already exist in every interpretation of $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle$. Finally, the properties of $\mathcal{DLR}$ mean that an interpretation $\mathcal{I}$ of $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle$ can only be guaranteed to contain a relational cycle if the cycle is explicitly asserted in axioms of the form $\vec{w}$:**R** in $\widehat{\mathcal{A}}$, so that each element in the cycle must be the interpretation of one of the individuals forming the tuples in these axioms. We can therefore conclude that the individual w_i must have the same interpretation as some individual w_j occurring in $\widehat{\mathcal{A}}$, and that Q_{w_i} can be replaced with the representative concept P_{w_j} (and that if w_i is the root individual, the axiom $w_i : C$ can be replaced by $w_j : C$).

Of course we do not know which individual occurring in $\widehat{\mathcal{A}}$ corresponds to a given Q_{w_i}, but we can simply try all possible replacements (of which there can only be

finitely many), so that $\langle \mathcal{S}, \widehat{\mathcal{A}}_{q_1} \rangle |\!\approx \mathcal{G}$ iff, for one of these replacements, $\langle \mathcal{S}, \widehat{\mathcal{A}}_{q_1} \rangle |\!\approx \{w{:}C\}$. An extra level of non-determinism is thus added to the procedure, but this should be manageable as the numbers of such Q_{w_i} will typically be very small.[6] These replacements can obviously be performed either before or after the collapsing procedure without affecting the the result. In practice, it will be more efficient to delay the replacement as long as possible, but in the following Lemma (Lemma 3) we will assume that the replacements have been performed before the collapsing procedure.

The correctness of the collapsing procedure does not depend on the traversal (whose purpose is simply to choose a suitable ordering), but only on the correctness of the individual transformations. We have already shown that the compacting and absorbing transformations preserve (un)satisfiability, and so obviously preserve the implication relationship; it only remains to show that the implication relationship is also preserved by each replacement of an axiom of the form $\vec{w}{:}\mathbf{R}$ with one of the form $w{:}C$.

Lemma 3 *Let $\mathcal{S}$ be a schema, $\widehat{\mathcal{A}}$ a completed canonical ABox and $\mathcal{A}_1$ an ABox where $\vec{w}{:}\boldsymbol{R} \in \mathcal{A}_1$, $\vec{w} = \langle w_1, \ldots, w_n \rangle$, w_i is the ith element of $\vec{w}$, and every other element of $\vec{w}$ is either an individual that occurs in $\widehat{\mathcal{A}}$ or an individual that occurs nowhere else in either $\widehat{\mathcal{A}}$ or $\mathcal{A}_1$. Let C be the concept*

$$\exists[\$i](\boldsymbol{R} \sqcap \bigsqcap_{1 \leqslant j \leqslant n . j \neq i} (\$j/n : C_{w_j})),$$

where C_{w_j} is the representative concept P_{w_j} when w_j is individual that occurs in $\widehat{\mathcal{A}}$, and $\top$ otherwise. If $\mathcal{A}_2$ is the ABox that results from the replacement of $\vec{w}{:}\boldsymbol{R} \in \mathcal{A}_1$ with the axiom $w_i{:}C$, then $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle |\!\approx \mathcal{A}_1$ iff $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle |\!\approx \mathcal{A}_2$.

PROOF: It is only necessary to show that $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle |\!\approx \{\vec{w}{:}\mathbf{R}\}$ iff $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle |\!\approx \{w_i{:}C\}$: if $\mathcal{A}_1$ contains other axioms, then any interpretation that satisfies these axioms will still satisfy them after the replacement. For the only if direction, it is easy to show that if $\mathcal{I} \models \langle \mathcal{S}, \widehat{\mathcal{A}} \rangle$, and $\mathcal{I}'$ is an extension of $\mathcal{I}$ that satisfies $\vec{w}{:}\mathbf{R}$, then $\mathcal{I}'$ also satisfies $w_i{:}C$.

Obviously, $w_i^{\mathcal{I}'}$ is the ith element of $\vec{w}^{\mathcal{I}'}$, and $\vec{w}^{\mathcal{I}'} \in \mathbf{R}^{\mathcal{I}'}$. For each component $(\$j/n : C_{w_j})$ in C there are two possible cases

1. When w_j is an individual occurring in $\widehat{\mathcal{A}}$, C_{w_j} is P_{w_j}, the representative concept for w_j. In this case, $w_j{:}P_{w_j}$ is an axiom in $\widehat{\mathcal{A}}$, so $w_j^{\mathcal{I}'} \in P_{w_j}^{\mathcal{I}'}$ and $\vec{w}^{\mathcal{I}'} \in (\$j/n : P_{w_j})^{\mathcal{I}'}$.

2. Otherwise, C_{w_j} is $\top$, and as $(\$j/n : \top)$ is equivalent to $\top_n$, obviously $\vec{w}^{\mathcal{I}'} \in (\$j/n : \top)^{\mathcal{I}'}$.

Therefore, we also have $\vec{w}^{\mathcal{I}'} \in \bigcap_{1 \leqslant j \leqslant n . j \neq i} (\$j/n : C_{w_j})^{\mathcal{I}'}$, and so $w_i^{\mathcal{I}'} \in C^{\mathcal{I}'}$.

The converse direction is more complicated. Let $\mathcal{I}$ be an interpretation such that $\mathcal{I} \models \langle \mathcal{S}, \widehat{\mathcal{A}}_{q_1} \rangle$, and $\mathcal{I}$ cannot be extended to satisfy $\vec{w}{:}\mathbf{R}$. From Lemma 1 we can assume, without loss of generality, that $P_{w_i}^{\mathcal{I}} = \{w_i^{\mathcal{I}}\}$ for every representative concept

[6] This represents a useful refinement over the procedure described in [4], where all z_i that occur in cycles are non-deterministically replaced with one of the w_i, regardless of whether or not they are used to enforce a co-reference.

P_{w_i} occurring in $\widehat{\mathcal{A}}$. Assume that $\mathcal{I}$ can be extended to an interpretation $\mathcal{I}'$ that satisfies $w_i{:}C$. Then there must be some $(d_1, \ldots, d_n) \in \mathbf{R}^{\mathcal{I}'}$ such that $d_i = w_i^{\mathcal{I}'}$, and for each d_j with $j \neq i$, $d_j \in (\$j/n : C_{w_j})^{\mathcal{I}'}$. Again, for each component $(\$j/n : C_{w_j})$ in C there are two possible cases.

1. When w_j is an individual occurring in $\widehat{\mathcal{A}}$, C_{w_j} is P_{w_j}, the representative concept for w_j. In this case, $P_{w_j}^{\mathcal{I}'} = \{w_j^{\mathcal{I}'}\}$, so $d_j = w_j^{\mathcal{I}'}$.

2. Otherwise, C_{w_j} is $\top$. In this case, w_j occurs nowhere else in either $\widehat{\mathcal{A}}$ or $\mathcal{A}_1$, so when $\mathcal{I}$ was extended to $\mathcal{I}'$, w_j could have been interpreted as any element in $\Delta^{\mathcal{I}}$ without affecting the satisfiability of any other axiom. We can therefore assume, without loss of generality, that in this particular interpretation $w_j^{\mathcal{I}'} = d_j$ (obviously, $d_j \in \top^{\mathcal{I}'}$).

We therefore have $\vec{w}^{\mathcal{I}'} = (d_1, \ldots, d_n)$ and $(d_1, \ldots, d_n) \in \mathbf{R}^{\mathcal{I}'}$, so $\mathcal{I}' \models \vec{w}{:}\mathbf{R}$, in contradiction of the assumption. ∎

Having collapsed $\mathcal{G}$, and (non-deterministically) replaced the Q_{w_i}, we finally have a problem that we can decide using KB satisfiability.

Lemma 4 *If $\mathcal{S}$ is a schema, $\widehat{\mathcal{A}}$ is a completed canonical ABox and C is a concept composed only of relations and concepts occurring in $\mathcal{S}$ or $\widehat{\mathcal{A}}$, then $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle |\!\approx \{w{:}C\}$ iff w is an individual in $\widehat{\mathcal{A}}$ and $\langle \mathcal{S}, (\widehat{\mathcal{A}} \cup \{w{:}\neg C\}) \rangle$ is not satisfiable, or w is not an individual in $\widehat{\mathcal{A}}$ and $\langle (\mathcal{S} \cup \{\top \sqsubseteq \neg C\}), \widehat{\mathcal{A}} \rangle$ is not satisfiable.*

PROOF: In the case where w is an individual in $\widehat{\mathcal{A}}$, there are no longer any non-distinguished individuals in $w{:}C$, so $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle |\!\approx \{w{:}C\}$ iff every model of $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle$ is also a model of $w{:}C$. This is obviously true iff there are no models of $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle |\!\approx \{w{:}C\}$ that are also models of $w{:}\neg C$, i.e., iff $\langle \mathcal{S}, (\widehat{\mathcal{A}} \cup \{w{:}\neg C\}) \rangle$ is not satisfiable.

In the case where w is not an individual in $\widehat{\mathcal{A}}$, $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle |\!\approx \{w{:}C\}$ iff for every model $\mathcal{I}$ of $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle$, $\mathcal{I}$ can be extended to $\{w{:}C\}$. As w is the only remaining non-distinguished individual in $\{w{:}C\}$, $\mathcal{I}$ can be extended to $\{w{:}C\}$ iff $C^{\mathcal{I}} \neq \emptyset$ (equivalently, $(\neg C)^{\mathcal{I}} \neq \Delta^{\mathcal{I}}$), i.e., iff $\langle (\mathcal{S} \cup \{\top \sqsubseteq \neg C\}), \widehat{\mathcal{A}} \rangle$ is not satisfiable. ∎

To illustrate the inclusion to satisfiability transformation, we will refer to the example given in Section 2.2. The containment of BUS in CITY_BUS w.r.t. the schema is demonstrated by the inclusion $\langle \mathcal{S}, \widehat{\mathcal{A}}_1 \rangle |\!\approx \mathcal{A}_2$, where $\mathcal{S}$, $\widehat{\mathcal{A}}_1$ and $\mathcal{A}_2$ are the schema and two canonical ABoxes (completed in the case of $\widehat{\mathcal{A}}_1$) corresponding to the given queries:

$$
\begin{array}{rcl}
\mathcal{S} & = & \left\{ \begin{array}{l} (\text{bus_route} \sqcap (\$1/3 : \text{city_bus})) \sqsubseteq \text{city_bus_route}, \\ \text{city_bus_route} \sqsubseteq (\text{bus_route} \sqcap (\$1/3 : \text{city_bus})) \end{array} \right\} \\
\widehat{\mathcal{A}}_1 & = & \{\ \langle n, y_1, y_2 \rangle{:}\text{bus_route}, \langle n, y_2, y_1 \rangle{:}\text{bus_route}, n{:}\text{city_bus}, n{:}P_n, y_1{:}P_{y_1}, y_2{:}P_{y_2}\ \} \\
\mathcal{A}_2 & = & \{\ \langle n, z_1, z_2 \rangle{:}\text{city_bus_route}, \langle n, z_2, z_1 \rangle{:}\text{city_bus_route}\ \}
\end{array}
$$

The two axioms in $\mathcal{A}_2$ are connected, and can be collapsed into a single axiom using the described procedure. If z_1 is chosen as the root, and the traversal visits $\langle n, z_1, z_2 \rangle$,

$$
\begin{array}{rl}
\sigma(\top_n) = \top_n & \qquad \sigma(\top) = \sigma(\top_1) \\
\sigma(\mathbf{P}) = \mathbf{P} & \qquad \sigma(A) = A \\
\sigma(\$i/n : C) = \top_n \sqcap \exists f_i.\sigma(C) & \qquad \sigma(\neg C) = \neg\sigma(C) \\
\sigma(\neg \mathbf{R}) = \top_n \sqcap \neg\sigma(\mathbf{R}) & \qquad \sigma(C_1 \sqcap C_2) = \sigma(C_1) \sqcap \sigma(C_2) \\
\sigma(\mathbf{R}_1 \sqcap \mathbf{R}_2) = \sigma(\mathbf{R}_1) \sqcap \sigma(\mathbf{R}_2) & \qquad \sigma(\exists[\$i]\mathbf{R}) = \exists f_i^-.\sigma(\mathbf{R}) \\
 & \qquad \sigma(\leq k[\$i]\mathbf{R}) = (\leq k\ f_i^-\ \sigma(\mathbf{R}))
\end{array}
$$

Figure 1: Reification of $\mathcal{DLR}$ concepts and relations

z_2, and $\langle n, z_2, z_1\rangle$, in that order, then the resulting axiom (describing $\mathcal{A}_2$ from the point of view of z_1) is $z_1{:}C$, where C is the concept

$$\exists[\$2](\text{city_bus_route} \sqcap (\$3 : (P_{z_2} \sqcap \exists[\$2](\text{city_bus_route} \sqcap \$1 : P_n \sqcap \$3 : P_{z_1}))) \sqcap \$1 : P_n),$$

and P_{z_1}, P_{z_2} are "place-holders" for z_1, z_2.[7] As z_2 is referred to only once, P_{z_2} can be replaced with $\top$. However, as z_1 is referred to twice (as P_{z_1} and as the root), it must be replaced (non-deterministically) with one of the individuals in $\widehat{\mathcal{A}}_1$, and $\langle \mathcal{S}, \widehat{\mathcal{A}}_1\rangle \not\approx \mathcal{A}_2$ iff $\langle \mathcal{S}, \widehat{\mathcal{A}}_1\rangle \not\approx \{z_1{:}C\}$ for one of these replacements. Substituting P_{z_2} with $\top$, z_1 with y_1 and P_{z_1} with P_{y_1} results in an axiom $y_1{:}C'$, and $\langle \mathcal{S}, \widehat{\mathcal{A}}_1\rangle \not\approx \{y_1{:}C'\}$ holds because $\langle \mathcal{S}, (\widehat{\mathcal{A}}_1 \cup \{y_1{:}\neg C'\})\rangle$ is not satisfiable.

Summing up, we thus have:

Theorem 2 *For a $\mathcal{DLR}$ KB $\mathcal{K} = \langle \mathcal{S}, \mathcal{A}\rangle$ and a $\mathcal{DLR}$ ABox $\mathcal{A}'$, the problem whether $\mathcal{A}$ is included in $\mathcal{A}'$ w.r.t. $\mathcal{S}$ can be reduced to (possibly several) $\mathcal{DLR}$ ABox satisfiability problems.*

3.3 Transforming $\mathcal{DLR}$ satisfiability into $\mathcal{SHIQ}$ satisfiability

We decide satisfiability of $\mathcal{DLR}$ knowledge bases by means of a satisfiability-preserving translation $\sigma(\cdot)$ from $\mathcal{DLR}$ KBs to $\mathcal{SHIQ}$ KBs. This translation deals with the fact that $\mathcal{DLR}$ allows for arbitrary n-ary relations while $\mathcal{SHIQ}$ only allows for unary predicates and binary relations; this is achieved by a process called *reification*. The main idea behind this is easily described: each n-ary tuple in a $\mathcal{DLR}$-interpretation is represented by an individual in a $\mathcal{SHIQ}$-interpretation that is linked via the dedicated functional relations $f_1, \ldots, f_n$ to the elements of the tuple.

For $\mathcal{DLR}$ without regular expressions, the mapping $\sigma(\cdot)$ shown in Figure 1 (given by Calvanese et al. [4]) reifies $\mathcal{DLR}$ expressions into $\mathcal{SHIQ}$-concepts. This mapping can be extended to a knowledge base as follows.

Definition 3.2 Let $\mathcal{K} = (\mathcal{S}, \mathcal{A})$ be a $\mathcal{DLR}$ knowledge base. The reification of $\mathcal{S}$ is given by

$$\{(\sigma(\mathbf{R}_1) \sqsubseteq \sigma(\mathbf{R}_2)) \mid (\mathbf{R}_1 \sqsubseteq \mathbf{R}_2) \in \mathcal{S}\} \cup \{(\sigma(C_1) \sqsubseteq \sigma(C_2)) \mid (C_1 \sqsubseteq C_2) \in \mathcal{S}\}.$$

[7] The reader will recall that, in practice, we use such "place-holders" during the collapsing procedure and then make appropriate substitutions.

To reify the ABox $\mathcal{A}$, we have to reify all tuples appearing in the axioms. For each distinct tuple $\vec{w} = \langle w_1, \ldots, w_n \rangle$ occurring in $\mathcal{A}$ we chose a distinct individual $t_{\vec{w}}$ (called the "reification of $\vec{w}$") and define:

$$\sigma(\vec{w}{:}\mathbf{R}) = \{t_{\vec{w}}{:}\sigma(\mathbf{R})\} \cup \{\langle t_{\vec{w}}, w_i \rangle{:}f_i \mid 1 \leq i \leq n\} \quad \text{and}$$
$$\sigma(\mathcal{A}) = \bigcup \{\sigma(\vec{w}{:}\mathbf{R}) \mid \vec{w}{:}\mathbf{R} \in \mathcal{A}\} \cup \{w{:}\sigma(C) \mid w{:}C \in \mathcal{A}\}.$$

We need a few additional inclusion and ABox axioms to guarantee that any model of $(\sigma(\mathcal{S}), \sigma(\mathcal{A}))$ can be "un-reified" into a model of $(\mathcal{S}, \mathcal{A})$. Let $n_{\max}$ denote the maximum arity of the $\mathcal{DLR}$ relations appearing in $\mathcal{K}$. We define $f(\mathcal{S})$ to consist of the following axioms (where $x \equiv y$ is an abbreviation for $x \sqsubseteq y$ and $y \sqsubseteq x$):

$\top \equiv \top_1 \sqcup \cdots \sqcup \top_{n_{\max}}$	
$\top \sqsubseteq (\leq 1\ f_1) \sqcap \cdots \sqcap (\leq 1\ f_{n_{\max}})$	
$\forall f_i.\bot \sqsubseteq \forall f_{i+1}.\bot$	for $2 \leq i < n_{\max}$
$\top_i \equiv \exists f_1.\top_1 \sqcap \cdots \sqcap \exists f_i.\top_1 \sqcap \forall f_{i+1}.\bot$	for $2 \leq i \leq n_{\max}$
$\mathbf{P} \sqsubseteq \top_n$	for each atomic relation $\mathbf{P}$ of arity n
$A \sqsubseteq \top_1$	for each atomic concept A

These are the standard axioms needed for reification in schema reasoning, and can already be found in [4].

We introduce a new atomic concept Q_w for every individual w in $\mathcal{A}$ and define $f(\mathcal{A})$ to consist of the following axioms:

$$f(\mathcal{A}) = \{w{:}Q_w \mid w \text{ occurs in } \mathcal{A}\} \cup$$
$$\{w_1{:}\leqslant 1\ f_1^-.(\top_n \sqcap \exists f_2.Q_{w_2} \sqcap \ldots \sqcap \exists f_n.Q_{w_n}) \mid \langle w_1, \ldots, w_n \rangle \text{ occurs in } \mathcal{A}\}$$

These axioms are crucial when dealing with the problem of tuple-admissibility (see below) in the presence of ABoxes.

Finally, we define $\sigma(\mathcal{K}) = \langle (\sigma(\mathcal{S}) \cup f(\mathcal{S})), (\sigma(\mathcal{A}) \cup f(\mathcal{A})) \rangle$.

Theorem 3 *Let $\mathcal{K} = \langle \mathcal{S}, \mathcal{A} \rangle$ be a $\mathcal{DLR}$ knowledge-base. $\mathcal{K}$ is satisfiable iff the $\mathcal{SHIQ}$-KB $\sigma(\mathcal{K})$ is satisfiable.*

The proof of Theorem 3 is rather involved and technical. We first give a sketch of the proof.

PROOF (sketch): The same techniques that were used in [2] can be adapted to the DL $\mathcal{SHIQ}$, and extended to deal with ABox axioms. The only-if direction is straightforward. A model $\mathcal{I}$ of $\mathcal{K}$ can be transformed into a model of $\sigma(\mathcal{K})$ by introducing, for every arity n with $2 \leq n \leq n_{\max}$ and every n-tuple of elements $\vec{d} \in (\Delta^{\mathcal{I}})^n$, a new element $t_{\vec{d}}$ that is linked to the elements of $\vec{d}$ by the functional relations $f_1, \ldots, f_n$. If we interpret $\top_1$ by $\Delta^{\mathcal{I}}$, $\top_n$ by the reifications of all elements in $\top_n^{\mathcal{I}}$, and, for every w that occurs in $\mathcal{A}$, Q_w by $w^{\mathcal{I}}$, then it is easy to show that we have constructed a model of $\sigma(\mathcal{K})$.

The converse direction is more complicated since a model of $\sigma(\mathcal{K})$ is not necessarily *tuple-admissible*, i.e., in general there may be distinct elements t, t' that are reifications of the same tuple $\vec{d}$. In the "un-reification" of such a model, $\vec{d}$ would only appear

once which may conflict with assertions in the $\mathcal{DLR}$ KB about the number of tuples in certain relations. However, it can be shown that every satisfiable KB $\sigma(\mathcal{K})$ also has a tuple-admissible model. It is easy to show that such a model, by "un-reification", induces a model for the original KB $\mathcal{K}$. ∎

Theorem 3 will be an immediate consequence of the following Lemmata 5 and 7.

Lemma 5 *Let $\mathcal{K} = (\mathcal{S}, \mathcal{A})$ be a $\mathcal{DLR}$ knowledge-base. If $\mathcal{K}$ is satisfiable, then the $\mathcal{SHIQ}$-KB $\sigma(\mathcal{K})$ is satisfiable.*

PROOF: Let $\mathcal{I}$ be a model of $(\mathcal{S}, \mathcal{A})$. We will reify it into a model $\hat{\mathcal{I}}$ for $(\sigma(\mathcal{S}) \cup f(\mathcal{K}), \sigma(\mathcal{A}))$.

Let $n_{\max}$ denote the maximum arity of relations in $\mathcal{S}$ and $\mathcal{A}$. The set of individuals of $\hat{\mathcal{I}}$ is the set of individuals of $\mathcal{I}$ plus a distinct individual for each possible n-tuple with $2 \leq n \leq n_{\max}$:

$$\Delta^{\hat{\mathcal{I}}} := \Delta^{\mathcal{I}} \cup \{t_{\vec{d}} \mid \vec{d} = \langle d_1, \ldots, d_n \rangle \in (\Delta^{\mathcal{I}})^n, 2 \leq n \leq n_{\max}\}$$

We have to fix the interpretation of the atomic $\mathcal{SHIQ}$-concepts and roles. The only roles that occur in $(\sigma(\mathcal{S}) \cup f(\mathcal{K}, \sigma(\mathcal{A}))$ are the f_n with $1 \leq n \leq n_{\max}$. For each role f_n we set

$$f_n^{\hat{\mathcal{I}}} = \{\langle t_{\vec{d}}, d_n \rangle \mid \vec{d} \in (\Delta^{\mathcal{I}})^n \text{ and } d_n \text{ is the } n\text{-th component of } \vec{d}\}$$

For every atomic $\mathcal{DLR}$-concept A, we set

$$A^{\hat{\mathcal{I}}} = A^{\mathcal{I}}$$

For every atomic $\mathcal{SHIQ}$-concept $\mathbf{P}$ that corresponds to an n-ary atomic $\mathcal{DLR}$-relation with $n \geq 2$ we define

$$\mathbf{P}^{\hat{\mathcal{I}}} = \{t_{\vec{d}} \mid \vec{d} \in (\Delta^{\mathcal{I}})^n \text{ and } \vec{d} \in \mathbf{P}^{\mathcal{I}}\}$$

Finally, we have to define the interpretation of the newly-introduced atomic concepts $\top_n$ for $1 \leq n \leq n_{\max}$. This is done as follows:

$$\top_1^{\hat{\mathcal{I}}} = \Delta^{\mathcal{I}}$$
$$\top_n^{\hat{\mathcal{I}}} = \{t_{\vec{d}} \mid \vec{d} \in \top_n^{\mathcal{I}}\} \quad \text{for } 2 \leq n \leq n_{\max}$$

It is easy to see that $\hat{\mathcal{I}} \models f(\mathcal{S})$.

By induction of the structure of $\mathcal{DLR}$-concepts and relations one can show, for every $\mathcal{DLR}$-concept C, every $\mathcal{DLR}$-relation $\mathbf{R}$, every $d \in \Delta^{\mathcal{I}}$, and every $\vec{d} \in (\Delta^{\mathcal{I}})^n$ for $2 \leq n \leq n_{\max}$, that

$$d \in C^{\mathcal{I}} \text{ implies } d \in \sigma(C)^{\hat{\mathcal{I}}} \text{ and}$$
$$\vec{d} \in \mathbf{R}^{\mathcal{I}} \text{ implies } t_{\vec{d}} \in \sigma(\mathbf{R})^{\hat{\mathcal{I}}}.$$

From this it immediately follows that $\mathcal{I} \models \mathcal{S}$ implies $\hat{\mathcal{I}} \models \sigma(\mathcal{S})$ and hence $\hat{\mathcal{I}} \models \sigma(\mathcal{S}) \cup f(\mathcal{K})$. It remains to show that also $\mathcal{I} \models \sigma(\mathcal{A})$.

We fix the interpretation of the auxiliary concepts Q_w that have been introduced in $f(\mathcal{A})$ by

$$Q_w^{\hat{\mathcal{I}}} = w^{\mathcal{I}}$$

At first, we have to define the interpretation of the individuals in $\sigma(\mathcal{S})$. For any individual w that appears also in $\mathcal{A}$ we set $w^{\hat{\mathcal{I}}} = w^{\mathcal{I}}$. For each newly introduced individual $t_{\vec{w}}$ with $\vec{w} = \langle w_1, \ldots, w_n \rangle$ we set $t_{\vec{w}}^{\hat{\mathcal{I}}} = t_{\vec{d}}$ where $\vec{d} = \langle w_1^{\mathcal{I}}, \ldots, w_n^{\mathcal{I}} \rangle$. With this definition it is easy to see that $\hat{\mathcal{I}} \models \sigma(\mathcal{A})$.

It remains to show that $\hat{\mathcal{I}} \models f(\mathcal{A})$. $\hat{\mathcal{I}} \models w{:}Q_w$ follows by construction of $Q_w^{\hat{\mathcal{I}}}$ for every individual w that occurs in $\mathcal{A}$. Let $\langle w_1, \ldots, w_n \rangle$ be a tuple that occurs in $\mathcal{A}$. We have to show that $w^{\hat{\mathcal{I}}} \in (\leqslant 1\, f_1^-.(\top_n \sqcap \exists f_2.Q_{w_2} \sqcap \ldots \sqcap \exists f_n.Q_{w_n}))^{\hat{\mathcal{I}}}$. By construction we have that

$$(\top_n \sqcap \exists f_2.Q_{w_2} \sqcap \ldots \sqcap \exists f_n.Q_{w_n})^{\hat{\mathcal{I}}} = \{t_{\vec{d}} \mid \vec{d} \in \top_n^{\mathcal{I}} \text{ and } \vec{d} = \langle x, w_2^{\mathcal{I}}, \ldots, w_n^{\mathcal{I}} \rangle \text{ for some } x \in \Delta^{\mathcal{I}}\}$$

and hence

$$(\leqslant 1\, f_1^-.(\top_n \sqcap \exists f_2.Q_{w_2} \sqcap \ldots \sqcap \exists f_n.Q_{w_n}))^{\hat{\mathcal{I}}} = \{x \in \Delta^{\mathcal{I}} \mid \exists^{\leq 1} t_{\vec{d}} \in \top_n^{\mathcal{I}}\,.\, \vec{d} = \langle x, w_2^{\mathcal{I}}, \ldots, w_n^{\mathcal{I}} \rangle\}$$

Trivially, for every $x \in \Delta^{\mathcal{I}}$, there is at most one n-tuple that starts with x and continues with $w_2^{\mathcal{I}}, \ldots, w_n^{\mathcal{I}}$. Hence, we get, for every tuple $\langle w_1, \ldots, w_n \rangle$ that occurs in $\mathcal{A}$, that $\hat{\mathcal{I}} \models w_1{:}\leqslant 1\, f_1^-.(\top_n \sqcap \exists f_2.Q_{w_2} \sqcap \ldots \sqcap \exists f_n.Q_{w_n})$. ∎

The proof of the converse direction of Lemma 5 is more involved. The problem arises from the fact that a model $\mathcal{I}$ of $\sigma(\mathcal{K})$ may not be *tuple-admissible*, i.e., there may be two distinct elements $t, t' \in \Delta^{\mathcal{I}}$ that are reifications of the same tuple $\vec{d} = \langle d_1, \ldots, d_n \rangle$. This means that both $t, t' \in \top_n^{\mathcal{I}}$ and $\langle t, d_i \rangle \in f_i^{\mathcal{I}}$ as well as $\langle t', d_i \rangle \in f_i^{\mathcal{I}}$ for $1 \leq i \leq n$. The next lemma shows that any consistent $\mathcal{SHIQ}$ knowledge base always has a tuple-admissible model.

Lemma 6 *Let $\mathcal{K} = (\mathcal{S}, \mathcal{A})$ be $\mathcal{DLR}$-KB and $\sigma(\mathcal{K}) = (\sigma(\mathcal{S}) \cup f(\mathcal{K}), \sigma(\mathcal{A}))$ its reified $\mathcal{SHIQ}$-counterpart. If $\sigma(\mathcal{K})$ is consistent, then there exists a* tuple-admissible *model $\hat{\mathcal{I}}$ for $\sigma(\mathcal{K})$, i.e., a model where, for every $2 \leq n \leq n_{max}$ and $t, t' \in \top_n^{\hat{\mathcal{I}}}$ it holds that*

$$\left(\bigwedge_{1 \leq i \leq n} \exists d.(\langle t, d \rangle \in f_i^{\mathcal{I}} \wedge \langle t', d \rangle \in f_i^{\mathcal{I}}) \right) \Rightarrow t = t' \qquad (*)$$

PROOF: Let $\mathcal{I}$ be a model of $\sigma(\mathcal{K})$. We will transform $\mathcal{I}$ into a tuple-admissible model $\hat{\mathcal{I}}$ for $\sigma(\mathcal{K})$. Since $\mathcal{I} \models \sigma(\mathcal{K})$, we have that $\mathcal{I} \models f(\mathcal{K})$ and hence $f_n^{\mathcal{I}}$ is the graph of a partial function. To this function we will refer by $f_n^{\mathcal{I}}(\cdot)$.

For $2 \leq n \leq n_{\max}$ and n-tuple $\vec{d} = \langle d_1, \ldots, d_n \rangle \in (\top_1^{\mathcal{I}})^n$, we define the set of all reifications of this tuple by

$$T_{\vec{d}} = \{t \in \Delta^{\mathcal{I}} \mid \langle t, d_1 \rangle \in f_1^{\mathcal{I}} \wedge \cdots \wedge \langle t, d_n \rangle \in f_n^{\mathcal{I}}\}$$

Each set $T_{\vec{d}}$ which contains more than one element violates $(*)$. For any such set we pick an arbitrary element $t_{\vec{d}} \in T_{\vec{d}}$ and say that the other elements are conflicting with

$t_{\vec{d}}$. With *Conf* we denote the set of all elements that are conflicting with other elements. We will now transform $\mathcal{I}$ into an interpretation $\hat{\mathcal{I}}$ that contains no conflicts.

We start by describing this transformation for the simple case that we have only a single conflicting element t. This conflict can be resolved as follows. Let $\mathcal{I}'$ be the interpretation consisting of two disjoint copies of $\mathcal{I}$ (we will forget about the interpretation of individuals at the moment). $\mathcal{I}'$ contains the conflicting element t and a copy t' of t. We define $\hat{\mathcal{I}}$ from $\mathcal{I}'$ by setting

$$f_1^{\hat{\mathcal{I}}} = (f_1^{\mathcal{I}'} \setminus \{\langle t, f_1^{\mathcal{I}'}(t)\rangle, \langle t', f_1^{\mathcal{I}'}(t')\rangle\}) \cup \{\langle t, f_1^{\mathcal{I}'}(t')\rangle, \langle t', f_1^{\mathcal{I}'}(t)\rangle\},$$

and preserving the interpretation of all other atomic concepts and roles. The result is an interpretation that contains no more conflicting elements.

The construction in the general case is a little bit more complicated because in general *Conf* may be of arbitrary cardinality and we have to take care of the ABox axioms. To prevent interference of the later construction with the ABox axioms we will use a little bit more care when choosing *Conf*. Firstly, we show that the interpretation of two different ABox individuals may never conflict.

CLAIM 1: Let $t_{\vec{v}}, t_{\vec{w}}$ be two distinct ABox individuals. There is no conflict between $t_{\vec{v}}^{\mathcal{I}}$ and $t_{\vec{w}}^{\mathcal{I}}$.

PROOF OF CLAIM 1: If $t_{\vec{v}}^{\mathcal{I}} = t_{\vec{w}}^{\mathcal{I}}$ then there cannot be a conflict because no element conflicts with itself. Assume $t_{\vec{v}}^{\mathcal{I}} \neq t_{\vec{w}}^{\mathcal{I}}$ but, for each $1 \leq i \leq n$, $f_i^{\mathcal{I}}(t_{\vec{v}}^{\mathcal{I}}) = f_i^{\mathcal{I}}(t_{\vec{w}}^{\mathcal{I}})$ (a conflict). Since $\mathcal{I} \models \sigma(\mathcal{A})$ we have, for each $1 \leq i \leq n$, $v_i^{\mathcal{I}} = w_i^{\mathcal{I}}$ and hence $v_i^{\mathcal{I}} \in Q_{w_i}^{\mathcal{I}}$. This implies

$$\{t_{\vec{v}}^{\mathcal{I}}, t_{\vec{w}}^{\mathcal{I}}\} \subseteq (\top_n \sqcap \exists f_2.Q_{w_2} \sqcap \ldots \sqcap \exists f_n.Q_{w_n})^{\mathcal{I}},$$

which yields $w_1^{\mathcal{I}} \notin (\leqslant 1\, f_1^{-}.(\top_n \sqcap \exists f_2.Q_{w_2} \sqcap \ldots \sqcap \exists f_n.Q_{w_n}))^{\hat{\mathcal{I}}}$ because $w_1^{\mathcal{I}}$ appears as the first component of two distinct reified tuples that satisfy $\top_n \sqcap \exists f_2.Q_{w_2} \sqcap \ldots \sqcap \exists f_n.Q_{w_n}$. This is a contradiction to the assumption that $\mathcal{I} \models f(\mathcal{A})$.

Using Claim 1, we make sure that we do not have any conflicting elements that appear in the interpretation of ABox individuals. There are no two ABox individuals $t_{\vec{w}_1}, t_{\vec{w}_2}$ such that $t_{\vec{w}_1}^{\mathcal{I}}, t_{\vec{w}_2}^{\mathcal{I}}$ are conflicting. From this it follows that, in each set $T_{\vec{d}}$, there is at most one element that appears as the image of an ABox individual of the interpretation $\mathcal{I}$ (it may appear as the image of several ABox individuals). Hence, we can choose *Conf* in a way that it contains no elements that appear as images of ABox individuals of $\mathcal{I}$.

Let $\mathcal{I}'$ denote the disjoint union of $\sharp(2^{\mathit{Conf}})$ copies of $\mathcal{I}$. For a set $Z \subseteq \mathit{Conf}$ we denote the copy of $d \in \Delta^{\mathcal{I}}$ in the Z-th copy of $\mathcal{I}$ by d_Z. For two distinct sets Z, Z' and elements $d_Z, d_{Z'}$, we call *exchanging* $f_1^{\mathcal{I}'}(d_Z)$ *with* $f_1^{\mathcal{I}'}(d_{Z'})$ the operation on $\mathcal{I}'$ which changes the interpretation of f_1 under $\mathcal{I}'$ as follows:

$$f_1^{\mathcal{I}'} = (f_1^{\mathcal{I}'} \setminus \{\langle d_Z, f_1^{\mathcal{I}'}(d_Z)\rangle, \langle d_{Z'}, f_1^{\mathcal{I}'}(d_{Z'})\rangle\}) \cup \{\langle d_Z, f_1^{\mathcal{I}'}(d_{Z'})\rangle, \langle d_{Z'}, f_1^{\mathcal{I}'}(d_Z)\rangle\}$$

We define $\hat{\mathcal{I}}$ from $\mathcal{I}'$ as the result of *simultaneously* exchanging, for each $d \in \mathit{Conf}$ and each $Z \subseteq \mathit{Conf}$ with $d \in Z$, $f_1(d_Z)$ with $f_1(d_{Z \setminus \{d\}})$.

CLAIM 2: $\hat{\mathcal{I}}$ does not contain any conflicts.

PROOF OF CLAIM 2: A conflict in $\hat{\mathcal{I}}$ may either origin from $\mathcal{I}'$ or may be created during the exchange.

- A conflict in $\mathcal{I}'$ can only involve two elements in the same disjoint copy. Let d_Z, e_Z denote the conflicting elements which reside in the Z-th copy. W.o.l.g., we can assume $d \in \mathit{Conf} \cap Z$. From the fact that d_Z, e_Z are still in a conflict originating from $\mathcal{I}'$ we have $f_1^{\mathcal{I}'}(d_Z) = f_1^{\mathcal{I}'}(e_Z)$ and $f_1^{\hat{\mathcal{I}}}(d_Z) = f_1^{\hat{\mathcal{I}}}(e_Z)$.
  - If $e \in \mathit{Conf}$, then we have changed the relation f_1 for e_Z. Strictly speaking, we have to distinguish the two cases $e \in Z$ and $e \notin Z$ but these are dual. In the first case we have exchanged $f_1^{\mathcal{I}'}(e_Z)$ by $f_1^{\mathcal{I}'}(e_{Z\setminus\{e\}})$, in the latter case we have exchanged $f_1\mathcal{I}'(e_{Z\cup\{e\}})$ with $f_1^{\mathcal{I}'}(e_Z)$.
  - If $e \notin \mathit{Conf}$, then we have not changed the relation f_1 for e_Z and hence we have $f_1^{\hat{\mathcal{I}}}(e_Z) = f_1^{\mathcal{I}'}(e_Z)$. At the same time, we have exchanged $f_1(d_Z)$ by $f_1(d_{Z\setminus\{d\}})$ and hence $f_1^{\hat{\mathcal{I}}}(d_Z) \neq f_1^{\mathcal{I}'}(d_Z)$ which is a contradiction.

  In both cases, we have $f_1^{\hat{\mathcal{I}}}(d_Z) \neq f_1^{\hat{\mathcal{I}}}(e_Z)$ because these elements are in different disjoint copies. Hence, d_Z and e_Z cannot be conflicting.

- Now assume that we have created a new conflict between elements $d_Z, e_{Z'}$ in $\hat{\mathcal{I}}$. This implies that, w.l.o.g., the function f_1 has been modified for d_Z during the exchange (otherwise the conflict would already be present in $\mathcal{I}'$). Since we only change the interpretation of the role f_1, d_Z and $e_{Z'}$ must have $f_i^{\mathcal{I}'}(d_Z) = f_i^{\mathcal{I}'}(e_{Z'})$ for $2 \leq i \leq n$, and hence d_Z and $e_{Z'}$ must reside in the same disjoint copy because we do not have f_i-links between the disjoint copies in $\mathcal{I}'$ for $i \geq 2$. Hence we have $Z = Z'$. Since d_Z and e_Z do not conflict in $\mathcal{I}'$, we must have $f_1^{\mathcal{I}'}(d_Z) \neq f_1^{\mathcal{I}'}(e_Z)$.
  - If both $d, e \notin \mathit{Conf}$, then $f_1^{\hat{\mathcal{I}}}(d_Z) = f_1^{\mathcal{I}'}(d_Z) \neq f_1^{\mathcal{I}'}(e_Z) = f_1^{\hat{\mathcal{I}}}(e_Z)$, and d_Z, e_Z cannot conflict in $\hat{\mathcal{I}}$.
  - If $d \in \mathit{Conf}$ and $e \notin \mathit{Conf}$, then we have that $f_1^{\hat{\mathcal{I}}}(d_Z)$ lies in the Y-th disjoint copy for $Y \neq Z$, while $f_1^{\hat{\mathcal{I}}}(e_Z)$ lies in the Z-th disjoint copy. Thus, we cannot have a conflict between d_Z and e_Z.
  - Finally, if $d, e \in \mathit{Conf}$, then we have to distinguish between the following cases:
    - if $d, e \in Z$ then $f_1^{\hat{\mathcal{I}}}(d_Z) = f_1^{\mathcal{I}'}(d_{Z\setminus\{d\}}) = f_1^{\hat{\mathcal{I}}}(e_Z) = f_1^{\mathcal{I}'}(e_{Z\setminus\{e\}})$. Hence, $Z\setminus\{d\}$ and $Z\setminus\{e\}$ refer to the same disjoint copy and we have $d = e$ and thus d_Z and e_Z are the same element and can not conflict.
    - if $d, e \notin Z$, then $d = e$ follows analogously and hence d_Z and e_Z cannot conflict.

CLAIM 3: Let C be a $\mathcal{SHIQ}$-concept, $d \in C^{\mathcal{I}}$ and $Z \subseteq \mathit{Conf}$. Then $d_Z \in C^{\hat{\mathcal{I}}}$.

PROOF OF CLAIM 3. We use a simple induction over the stucture of $\mathcal{SHIQ}$-concepts. The claim obviously holds for all atomic concepts. Also, per induction, it immediately holds for the boolean combination of concepts. For the existential, value, and number restrictions it follows from the fact that we start with disjoint copies and only change the interpretation of roles by exchanging elements that are copies of the same element. Hence, we do not changes the number of successors for each element, and, we also exchange only links to elements which, by the induction hypothesis, cannot be distinguished by "smaller" concepts.

From Claim 3 it follows that $\hat{\mathcal{I}} \models \mathcal{S}$. It remains to show that we can fix the interpretation of the ABox individuals in $\mathcal{A}$ under $\hat{\mathcal{I}}$ such that $\hat{\mathcal{I}} \models \mathcal{A}$. This can be done by interpreting all individuals in a single copy, e.g., by setting, for every ABox individual w, $w^{\hat{\mathcal{I}}} = w_\emptyset^{\mathcal{I}}$

Again, from Claim 3, we get that, for every ABox assertion $w{:}C \in \mathcal{A}$ we have that $w^{\mathcal{I}} \in C^{\mathcal{I}}$ implies $w^{\hat{\mathcal{I}}} \in C^{\hat{\mathcal{I}}}$. Furthermore, since, for every individual w that appears in $\mathcal{A}$, we have $w^{\mathcal{I}} \notin \mathit{Conf}$ and hence the interpretation of f_i is not changed for $w_\emptyset^{\hat{\mathcal{I}}}$. For any assertion $\langle w_1, w_2\rangle{:}f_i$, we have $\langle w_1^{\mathcal{I}}, w_2^{\mathcal{I}}\rangle \in f_i^{\mathcal{I}}$ and hence $\langle w_1^{\hat{\mathcal{I}}}, w_2^{\hat{\mathcal{I}}}\rangle \in f_i^{\hat{\mathcal{I}}}$. Thus, we also have $\hat{\mathcal{I}} \models \mathcal{A}$ and thus $\hat{\mathcal{I}} \models \mathcal{K}$.

Together with Claim 2, which yields that $\hat{\mathcal{I}}$ satisfies $(*)$, we have that $\hat{\mathcal{I}}$ is a tuple-admissible interpretation with $\hat{\mathcal{I}} \models \sigma(\mathcal{K})$. ∎

Once we have solved the problem of tuple admissibility it is fairly straightforward to show the following lemma.

Lemma 7 *Let $\mathcal{K} = (\mathcal{S}, \mathcal{A})$ be a $\mathcal{DLR}$ knowledge-base. If the $\mathcal{SHIQ}$-KB $\sigma(\mathcal{K})$ is satisfiable, then $\mathcal{K}$ is satisfiable.*

PROOF: If $\sigma(\mathcal{K})$ is consistent, then, by Lemma 6 we have that there is a tuple admissible model $\hat{\mathcal{I}}$ for $\sigma(\mathcal{K})$. We will "un-reify" the reified tuples in $\hat{\mathcal{I}}$ into ordinary tuples. We use the auxiliary function ur that maps a reified tuple to its un-reified counterpart. If $t \in \top_n^{\hat{\mathcal{I}}}$ and $f_i^{\hat{\mathcal{I}}}(t) = d_i$ for $1 \leq i \leq n$, then we define $ur(t) = \langle d_1, \ldots, d_n\rangle$.

The atomic concepts and relations will be defined as follows:

$\Delta^{\mathcal{I}} = \top_1^{\hat{\mathcal{I}}}$	
$A^{\mathcal{I}} = A^{\hat{\mathcal{I}}}$	for each atomic concept A
$\mathbf{P}^{\mathcal{I}} = \{ur(t) \mid t \in \mathbf{P}^{\hat{\mathcal{I}}}\}$	for each atomic relation $\mathbf{P}$ of arity n

We also have to define the interpretation of the ABox individuals in $\mathcal{A}$. For every individual w that appears in $\mathcal{A}$ we set $w^{\mathcal{I}} = w^{\hat{\mathcal{I}}}$. Please note that, also if w appears inside a tuple of a relation assertion in $\mathcal{A}$, w will appear in $\sigma(\mathcal{A})$ and hence $w^{\hat{\mathcal{I}}}$ is defined.

Since $\hat{\mathcal{I}} \models f(\mathcal{K})$ we have that $\mathcal{I}$ is indeed a well defined interpretation. The following can easily be shown:

CLAIM: For every $\mathcal{DLR}$-concept C and $\mathcal{DLR}$-relation $\mathbf{R}$,

$$d \in \sigma(C)^{\hat{\mathcal{I}}} \text{ implies } d \in C^{\mathcal{I}}$$
$$t \in \sigma(\mathbf{R})^{\hat{\mathcal{I}}} \text{ implies } ur(t) \in \mathbf{R}^{\mathcal{I}}$$

PROOF OF THE CLAIM. The claim is obvious for atomic concepts and relations by the definition of $\mathcal{I}$. By induction it follows easily also for complex concepts and roles. We need the fact that $\hat{\mathcal{I}}$ is tuple admissible to ensure that the claim holds for concepts and relations involving counting expressions.

From this it follows that $\mathcal{I} \models \mathcal{S}$ and also $\mathcal{I} \models \mathcal{A}$, hence we have shown that $\mathcal{K}$ is consistent. ∎

3.4 Dealing with disjunctive queries

In this section we will show how the technique can be extended in order to decide the containment of disjunctive queries.

Definition 3.3 A *disjunctive query* q is an expression

$$\begin{array}{c} q(\vec{x}) \leftarrow term_{1,1}(\vec{x}, \vec{y}_1, \vec{c}) \wedge \ldots \wedge term_{1,k_1}(\vec{x}, \vec{y}_1, \vec{c}) \\ \vee \\ \vdots \\ \vee \\ term_{m,1}(\vec{x}, \vec{y}_m, \vec{c}) \wedge \ldots \wedge term_{m,k_m}(\vec{x}, \vec{y}_m, \vec{c}), \end{array}$$

where all the terms are defined exactly as in the conjunctive queries of Section 2.2. The query evaluation is defined as the union of all the evaluations for any disjunct. Given a query q with n distinguished variables, its evaluation w.r.t. the interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ is the set of n-tuples:

$$q(\mathcal{I}) = \left\{ \begin{array}{ll} \vec{d} \in (\Delta^{\mathcal{I}})^n & \mid \quad \exists \vec{y}_1 . term_{1,1}(\vec{d}, \vec{y}_1, \vec{c}) \wedge \ldots \wedge term_{1,k_1}(\vec{d}, \vec{y}_1, \vec{c}) \\ & \qquad\qquad \vee \\ & \qquad\qquad \vdots \\ & \qquad\qquad \vee \\ & \qquad\qquad \exists \vec{y}_m . term_{m,1}(\vec{d}, \vec{y}_m, \vec{c}) \wedge \ldots \wedge term_{m,k_m}(\vec{d}, \vec{y}_m, \vec{c}) \end{array} \right\}$$

Without loss of generality we can assume that all the variable names in $\vec{y}_1, \ldots, \vec{y}_m$ are distinct, and that distinguished variables and constant names appear in every disjunct (see Section 3.1). The query containment problem is defined as in the conjunctive case.

The basic idea is to consider each conjunctive subexpression as a canonical ABox, and to extend the inclusion relation of Section 2.1 to take into account the "disjunction" of ABoxes. We will first extend the definition of $\mathcal{DLR}$ ABoxes to *disjunctive* $\mathcal{DLR}$ ABoxes (in order to avoid ambiguity, we will sometimes refer to the kind of ABox defined in Section 2.1 as a conjunctive ABox).

Definition 3.4 A *disjunctive* $\mathcal{DLR}$ ABox is a finite set of conjunctive ABoxes $\{\mathcal{A}_1, \ldots, A_m\}$. The definition of interpretation and satisfiability for each conjunctive ABox A_i is the same as that given in Section 2.1. An interpretation $\mathcal{I}$ satisfies a disjunctive ABox $\mathcal{A}$ (written $\mathcal{I} \models \mathcal{A}$) iff $\mathcal{I}$ satisfies at least one of the conjunctive ABoxes in $\mathcal{A}$.

On top of the definition of a disjunctive ABox, is built the notion of a disjunctive KB and its satisfiability. All the definitions given in Section 2.1 can be naturally extended to the disjunctive case; in particular the fundamental notion of the inclusion relation between ABoxes.

To simplify the notation, we define the operator $(\cdot \times \cdot)$ which adds a set of axioms to each element of a disjunctive KB. The meaning of the operator is given by the following equations:

$$\begin{array}{rcl} \langle \mathcal{S}, \{\mathcal{A}_1, \ldots, A_m\} \rangle \times \{w{:}C\} & = & \langle \mathcal{S}, \{\mathcal{A}_1 \cup \{w{:}C\}, \ldots, A_m \cup \{w{:}C\}\} \rangle \\ \langle \mathcal{S}, \{\mathcal{A}_1, \ldots, A_m\} \rangle \times \{C \sqsubseteq D\} & = & \langle \mathcal{S} \cup \{C \sqsubseteq D\}, \{\mathcal{A}_1, \ldots, A_m\} \rangle \end{array}$$

with the natural extension to finite sets of axioms:

$$\mathcal{K} \times \{\alpha_1, \ldots, \alpha_n\} = (\ldots(\mathcal{K} \times \{\alpha_1\}) \times \ldots) \times \{\alpha_n\}\,.$$

Now we will proceed as in Sections 3.1 and 3.2 by first showing how to reduce the query containment problem to ABox inclusion, and then to ABox satisfiability.

First, we will extend the definition of canonical ABox to deal with disjunctive queries.

Definition 3.5 Let q be a disjunctive query. The *canonical disjunctive ABox* for q is defined by

$$\mathcal{A}_q = \{\mathcal{A}_1, \ldots, A_m\}$$

where each $\mathcal{A}_i$ describes a single conjunct in the query:

$$\begin{array}{rl} \mathcal{A}_i = & \{\vec{w}{:}\mathbf{R} \mid \mathbf{R}(\vec{w}) = term_{i,j}(\vec{x}, \vec{y}_i, \vec{c}) \text{ is an atom in } q \text{ for some } j\} \cup \\ & \{w{:}C \mid C(w) = term_{i,j}(\vec{x}, \vec{y}_i, \vec{c}) \text{ is an atom in } q \text{ for some } j\} \end{array}$$

The completed canonical disjunctive ABox for q is defined in a similar way to the non-disjunctive case (see Definition 3.1), the difference being that the new axioms are added to each of the conjunctive ABoxes making up the disjunction. Given the disjunctive ABox $\mathcal{A}_q = \{\mathcal{A}_1, \ldots, A_m\}$, its completed version (written as $\widehat{\mathcal{A}}_q$) is defined as:

$$\begin{array}{rl} \widehat{\mathcal{A}}_q = \mathcal{A}_q \times (& \{w{:}P_w \mid w \text{ occurs in } \mathcal{A}_q\} \cup \\ & \{w_i{:}\neg P_{w_j} \mid w_i, w_j \text{ are constants in } q \text{ and } i \neq j\}) \end{array}$$

As in the non-disjunctive case, there is a natural correspondence between database instances and interpretations of disjunctive KBs. Each element of a query evaluation corresponds to an interpretation satisfying the canonical ABox and vice versa.[8]

[8] We will consider a database as a standard $\mathcal{DLR}$ interpretation in which an individuals corresponding to a constant is taken to be interpreted as the actual constant.

Proposition 1 *Given a database $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ and a disjunctive query $q(\vec{x})$, then the tuple $\vec{d}$ is in the evaluation $q(\mathcal{I})$ iff there is an extension $\mathcal{I}'$ of $\mathcal{I}$ satisfying $\widehat{\mathcal{A}}_q$ such that ${x_i}^{\mathcal{I}'} = d_i$ for each x_i in $\vec{x}$.*

PROOF: For the "only if" direction, let $\vec{d}$ be in $q(\mathcal{I})$, then it satisfies at least one of the disjuncts in q:

$$\exists \vec{y}_i.term_{i,1}(\vec{d}, \vec{y}_i, \vec{c}) \wedge \ldots \wedge term_{i,k_i}(\vec{d}, \vec{y}_i, \vec{c})$$

for $1 \leq i \leq m$, which means that there is an assignment for the variables in $\vec{y}_i$ that makes the expression true. If $\widehat{\mathcal{A}}_i \in \mathcal{A}_q$ is the corresponding conjunctive ABox, then an extension $\mathcal{I}'$ of $\mathcal{I}$ can be defined by adding to $\cdot^{\mathcal{I}}$ a mapping from each individual in $\widehat{\mathcal{A}}_i$ to the corresponding element of $\vec{d}$. It is easy to see that $\mathcal{I}'$ satisfies $\widehat{\mathcal{A}}_i$ and thus satisfies $\widehat{\mathcal{A}}_q$.

For the "if" direction, let $\mathcal{I}'$ be an extension of $\mathcal{I}$ satisfying $\widehat{\mathcal{A}}_q$ such that ${x_i}^{\mathcal{I}'} = d_i$ for each x_i in $\vec{x}$. Then, from the definition of satisfiability of a disjunctive ABox, there is some $\widehat{\mathcal{A}}_i \in \widehat{\mathcal{A}}_q$ such that $\mathcal{I}' \models \widehat{\mathcal{A}}_i$. Note that $\mathcal{I}'$ must interpret all the individual appearing in $\mathcal{A}_i$; it therefore defines an assignment for the variables $\vec{y}_i$ in the corresponding disjunct of q. It is easy to see that this assignment satisfies the formula

$$\exists \vec{y}_i.term_{i,1}(\vec{d}, \vec{y}_i, \vec{c}) \wedge \ldots \wedge term_{i,k_i}(\vec{d}, \vec{y}_i, \vec{c}).$$

■

Given the Proposition 1 above, we can extend Theorem 1 to the disjunctive case.

Theorem 4 *Given a schema $\mathcal{S}$ and two disjunctive queries q_1 and q_2, $\mathcal{S} \models q_1 \sqsubseteq q_2$ iff $\langle \mathcal{S}, \widehat{\mathcal{A}}_{q_1} \rangle |\approx \mathcal{A}_{q_2}$.*

PROOF: The same as for Theorem 1. ■

The next step consists of reducing ABox inclusion to ABox satisfiability. As in the conjunctive case, we only consider a particular kind of ABox on the r.h.s. of the inclusion, namely those containing only axioms of the form $w{:}C$. This assumption can be made without loss of generality because the connected components of each conjunctive ABox can be collapsed into a single concept assertion, as shown in Section 3.2.

Proposition 2 *Let $\mathcal{S}$ be a schema, $\widehat{\mathcal{A}}$ a completed canonical disjunctive ABox and $\mathcal{A}'$ a disjunctive ABox. Then $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle |\approx \mathcal{A}'$ iff there is a disjunctive ABox $\mathcal{A}''$ containing only axioms of the form $w{:}C$, such that $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle |\approx \mathcal{A}''$*

PROOF: (SKETCHED) The same considerations set out in Section 3.2, which enable us to "collapse" connected components into single axioms of the form $w{:}C$, also apply in the disjunctive case, and can be used to transform each conjunctive ABox in $\mathcal{A}'$ so that it contains only such axioms. ■

In the following Lemma (Lemma 8), which provides the reduction to ABox satisfiability, we use the notation $w\not{:}C$ to describe the axiom which forces the interpretation of the individual w not to be in the extension of C. If w is a non-distinguished individual, then it is the schema axiom $\top \sqsubseteq \neg C$; otherwise it is the ABox axiom $w{:}\neg C$.

Lemma 8 *Let $\mathcal{S}$ be a schema, $\widehat{\mathcal{A}}$ a disjunctive ABox, and $\{\mathcal{A}_1, \ldots, A_m\}$ a disjunctive ABox, where each $\mathcal{A}_i$ contains only axioms of the form $w{:}C$. Then $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle \not\approx \{\mathcal{A}_1, \ldots, A_m\}$ iff for every possible KB*

$$\mathcal{K} = \langle \mathcal{S}, \widehat{\mathcal{A}} \rangle \times \bigcup_{1 \leqslant i \leqslant m} \{w \not: C\} \text{ for some } (w{:}C) \in \mathcal{A}_i,$$

$\mathcal{K}$ is unsatisfiable.

PROOF:

- For the "only if" direction, assume that $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle \not\approx \{\mathcal{A}_1, \ldots, A_m\}$ and that there is some $A_i \in \{\mathcal{A}_1, \ldots, A_m\}$ such that the KB

  $$\mathcal{K}' = \langle \mathcal{S}, \widehat{\mathcal{A}} \rangle \times \bigcup_{1 \leqslant i \leqslant m} \{w \not: C\} \text{ for some } (w{:}C) \in \mathcal{A}_i$$

  is satisfiable. Let $\mathcal{I}$ be an interpretation satisfying $\mathcal{K}'$, and $\mathcal{I}'$ the restriction of this interpretation to exclude the non-distinguished individuals in $\{\mathcal{A}_1, \ldots, A_m\}$. Obviously $\mathcal{I}'$ satisfies $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle$. Therefore there is an extension $\mathcal{I}''$ of $\mathcal{I}'$ satisfying $\{\mathcal{A}_1, \ldots, A_m\}$. Let $\mathcal{A}_\ell \in \{\mathcal{A}_1, \ldots, A_m\}$ be a conjunctive ABox satisfied by $\mathcal{I}''$. By construction of $\mathcal{K}'$, there must be an assertion $w{:}C \in \mathcal{A}_\ell$ such that if $\mathcal{K}' = \langle \mathcal{S}', \mathcal{A}' \rangle$, then either $(w \not: C)$ is in $\mathcal{S}'$ or $(w \not: C)$ is in every conjunctive ABox in $\mathcal{A}'$. Therefore $\mathcal{I} \models w \not: C$, and $\mathcal{I}'' \models w{:}C$. Moreover, as both $\mathcal{I}$ and $\mathcal{I}''$ are extensions of $\mathcal{I}'$ (see 2.4), they differ only in the interpretation of non-distinguished variables. There are two cases, depending on whether or not w is a non-distinguished individual.

  - If w is a non-distinguished individual, then $w^{\mathcal{I}''} \in C^{\mathcal{I}''}$ and $\Delta^{\mathcal{I}} \subseteq (\neg C)^{\mathcal{I}}$. As $C^{\mathcal{I}''} = C^{\mathcal{I}}$, this implies that $w^{\mathcal{I}''} \in \emptyset$, an obvious contradiction.
  - Otherwise, $w^{\mathcal{I}} = w^{\mathcal{I}''}$, $w^{\mathcal{I}} \in (\neg C)^{\mathcal{I}}$ and $w^{\mathcal{I}''} \in C^{\mathcal{I}''}$. As $C^{\mathcal{I}''} = C^{\mathcal{I}}$, this implies that $(\neg C)^{\mathcal{I}} \cap C^{\mathcal{I}} \neq \emptyset$, again an obvious contradiction.

- For the "if" direction, assume that there is an interpretation $\mathcal{I}$ satisfying $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle$ which cannot be extended to one satisfying $\{\mathcal{A}_1, \ldots, A_m\}$. For each $\mathcal{A}_i \in \{\mathcal{A}_1, \ldots, A_m\}$ there must be at least one axiom $(w_i{:}C_i) \in \mathcal{A}_i$ that $\mathcal{I}$ cannot be extended to satisfy. Therefore, there is a KB

  $$\mathcal{K}' = \langle \mathcal{S}, \widehat{\mathcal{A}} \rangle \times \bigcup_{1 \leqslant i \leqslant m} \{w \not: C\} \text{ for some } (w{:}C) \in \mathcal{A}_i$$

  such that the interpretation $\mathcal{I}$ cannot be extended to satisfy any of the selected axioms $w{:}C \in \mathcal{A}_i$. The interpretation $\mathcal{I}$ satisfies $\langle \mathcal{S}, \widehat{\mathcal{A}} \rangle$, and it also satisfies all the axioms $\{w \not: C\}$ added in $\mathcal{K}'$. Again, there are two cases, depending on whether or not w is a non-distinguished individual.

  - If w is a distinguished variable, then $w^{\mathcal{I}} \notin C^{\mathcal{I}}$ otherwise any extension will satisfy $w{:}C$. Therefore $\mathcal{I} \models w{:}\neg C$.
  - If w is a non-distinguished variable, then $C^{\mathcal{I}}$ must be empty, otherwise an extension satisfying $w{:}C$ can be defined by mapping w to one of the element of $C^{\mathcal{I}}$. Therefore $\mathcal{I} \models \top \sqsubseteq \neg C$. ∎

Lemma 8 shows how the problem can be reduced from ABox inclusion to ABox satisfiability. Unfortunately, the resulting KB is still disjunctive, and Section 3.2 only shows how to solve satisfiability problems for conjunctive KBs. However, we can reduce the problem to conjunctive KB satisfiability by introducing one more non-deterministic step.

Lemma 9 *A disjunctive KB* $\langle \mathcal{S}, \{\mathcal{A}_1, \ldots, A_m\}\rangle$ *is satisfiable iff for some* $\mathcal{A} \in \{\mathcal{A}_1, \ldots, A_m\}$, $\langle \mathcal{S}, \mathcal{A}\rangle$ *is satisfiable.*

PROOF: Obviously, if $\mathcal{I} \models \langle \mathcal{S}, \mathcal{A}\rangle$, then $\mathcal{I} \models \langle \mathcal{S}, \{\mathcal{A}_1, \ldots, A_m\}\rangle$, and if $\mathcal{I} \models \langle \mathcal{S}, \{\mathcal{A}_1, \ldots, A_m\}\rangle$, then from Definition 3.4, $\mathcal{I}$ must satisfy at least on of the conjunctive ABoxes in $\{\mathcal{A}_1, \ldots, A_m\}$. ∎

We now have the machinery to transform a query containment problem into one or more $\mathcal{SHIQ}$ schema and ABox satisfiability problems. In the next section we will present a decision procedure that will enable us to solve such problems.

4 Deciding Satisfiability of $\mathcal{SHIQ}$ Knowledge Bases

To test satisfiability of a knowledge base $\mathcal{K} = \langle \mathcal{S}, \mathcal{A}\rangle$, we first *internalise* the schema $\mathcal{S}$ into the ABox $\mathcal{A}$, i.e, we add, for each individual w that occurs in $\mathcal{A}$, an axiom $w{:}C_{\mathcal{S}}$, where

$$C_{\mathcal{S}} := \bigsqcap_{C \sqsubseteq D \in \mathcal{S}} \neg(C \sqcap \neg D) \sqcap \forall U. \bigsqcap_{C \sqsubseteq D \in \mathcal{S}} \neg(C \sqcap \neg D),$$

for $U \in \mathsf{NR}_+$ a new transitive role with $R \sqsubseteq U$ for all roles R occurring in $\mathcal{K}$. Since U functions as a universal role, the ABox resulting from this internalisation is satisfiable iff $\mathcal{K}$ is satisfiable. Thus it only remains to decide satisfiability of $\mathcal{SHIQ}$-ABoxes.

Satisfiability of $\mathcal{SHIQ}$-ABoxes can be decided by a tableaux algorithm that tries to construct a model for the input ABox $\mathcal{A}$ by breaking down concepts occurring in $\mathcal{A}$ into sub-concepts, possibly introducing new individual variables, and thus making explicit the constraints imposed on individuals in models of $\mathcal{A}$. To this purpose, it works on a completion forest (i.e., a collection of trees whose root nodes are possibly connected to each other) some of whose nodes correspond to individuals in a model. The forest's edges denote role-successorships, and each node is labelled with concepts it must be an instance of. This algorithm is similar to the one that decides satisfiability of $\mathcal{SHIQ}$-concepts presented in Horrocks et al. [15]. Due to lack of space, we can neither describe the algorithm in detail nor prove its soundness and completeness, and refer the reader to [14], pages 38–49. Instead, we will simply point out the differences between the concept- and the ABox-satisfiability algorithm.

Firstly, instead of working on a completion *tree*, it works on a completion *forest*, that is, a collection of completion trees whose nodes correspond to individuals of a model of the input ABox and whose root nodes correspond to those individuals that occur explicitly in the ABox. Secondly, the rules of the algorithm had to be modified to correctly handle completion forests. This mainly involves the rule that identifies some of the neighbours of a node x whenever it has $\geqslant n$ neighbour nodes with respect to a role R, and we learn that, due to an at-most number restriction, x must only have at

most $n-1$ of these "R-successors". Here, we must take special care when root nodes are involved in this identification. Thirdly, the *blocking condition* which guarantees termination had to be modified in order to deal properly with root nodes. Basically, this means that root nodes can never be blocked.

5 Discussion

In this paper we have shown how the problem of query containment under constraints can be decided using a KB (schema plus ABox) satisfiability tester for the $\mathcal{SHIQ}$ description logic, and we have indicated how a $\mathcal{SHIQ}$ schema satisfiability testing algorithm can be extended to deal with an ABox. We have only talked about conjunctive queries, but extending the procedure to deal with disjunctions of conjunctive queries should be straightforward. Although there is some loss of expressive power with respect to the framework presented in [4], this seems to be acceptable when modelling classical relational information systems, where regular expressions are seldom used.

Given that the FaCT implementation of the $\mathcal{SHIQ}$ schema satisfiability algorithm has been shown to work well with realistic problems, and that the number of individuals generated by query containment problems will be relatively small, there is good reason to believe that a combination of the ABox encoding and the extended algorithm will lead to a practical decision procedure for query containment problems. Work is underway to test this hypothesis by extending the FaCT system to deal with $\mathcal{SHIQ}$ ABoxes.

Acknowledgements. We would like to thank Diego Calvanese and Maurizio Lenzerini for helpful discussions.

References

[1] S. Abiteboul, R. Hull, and V. Vianu. *Foundations of databases*. Addison-Wesley, 1995.

[2] D. Calvanese. *Unrestricted and Finite Model Reasoning in Class-Based Representation Formalisms*. PhD thesis, Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", 1996. see http://www.dis.uniroma1.it/pub/calvanese/thesis.ps.gz.

[3] D. Calvanese, G. De Giacomo, and M. Lenzerini. Structured objects: modeling and reasoning. In *Proceedings of the 4th International Conference on Deductive and Object-Oriented Databases (DOOD'95)*, number 1013 in Lecture Notes in Computer Science, pages 229–246. Springer-Verlag, 1995. see http://www.dis.uniroma1.it/ calvanes/papers-html/DOOD-95.html.

[4] D. Calvanese, G. De Giacomo, and M. Lenzerini. On the decidability of query containment under constraints. In *Proceedings of the Seventeenth ACM SIGACT SIGMOD SIGART Symposium on Principles of Database Systems (PODS-98)*, 1998. see http://www.dis.uniroma1.it/ calvanes/papers-html/PODS-98.html.

[5] D. Calvanese, G. De Giacomo, and M. Lenzerini. Answering queries using views in description logics. In E. Franconi and M. Kifer, editors, *Proceedings of the 6th International Workshop KRDB'96*, pages 6–10. CEUR, July 1999. see ftp://SunSITE.Informatik.RWTH-Aachen.DE/pub/publications/CEUR-WS/Vol-21.tar.gz.

[6] D. Calvanese, G. De Giacomo, M. Lenzerini, D. Nardi, and R. Rosati. Source integration in data warehousing. In *Proceedings of the Ninth International Workshop on Database and Expert Systems Applications (DEXA-98)*, pages 192–197. IEEE Computer Society Press, 1998.

[7] D. Calvanese, G. De Giacomo, M. Lenzerini, D. Nardi, and R. Rosati. Use of the data reconciliation tool at telecom italia. DWQ deliverable D.4.3, Foundations of Data Warehouse Quality (DWQ), 1999.

[8] D. Calvanese, M. Lenzerini, and D. Nardi. Description logics for conceptual data modeling. In Jan Chomicki and Günter Saake, editors, *Logics for Databases and Information Systems*, pages 229–263. Kluwer Academic Publisher, 1998. see http://www.dis.uniroma1.it/ calvanes/papers-html/DBIS-Chom-Saak-98.html.

[9] T. Catarci and M. Lenzerini. Representing and using interschema knowledge in cooperative information systems. *Journal for Intelligent and Cooperative Information Systems*, 2(4):375–399, 1993.

[10] G. De Giacomo. *Decidability of Class-Based Knowledge Representation Formalisms*. PhD thesis, Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", 1995.

[11] G. De Giacomo and M. Lenzerini. TBox and ABox reasoning in expressive description logics. In L. C. Aiello, J. Doyle, and S. C. Shapiro, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifth International Conference (KR'96)*, pages 316–327. Morgan Kaufmann Publishers, San Francisco, California, November 1996.

[12] M. J. Fischer and R. E. Ladner. Propositional dynamic logic of regular programs. *Journal of Computer and System Sciences*, 18:194–211, 1979.

[13] I. Horrocks. FaCT and iFaCT. In P. Lambrix, A. Borgida, M. Lenzerini, R. Möller, and P. Patel-Schneider, editors, *Proceedings of the International Workshop on Description Logics (DL'99)*, pages 133–135, 1999. see http://SunSITE.Informatik.RWTH-Aachen.DE/pub/publications/CEUR-WS/Vol-22/.

[14] I. Horrocks, U. Sattler, and S. Tobies. A description logic with transitive and converse roles, role hierarchies and qualifying number restrictions. LTCS-Report 99-08, LuFg Theoretical Computer Science, RWTH Aachen, Germany, 1999. Revised version. See http://www-lti.informatik.rwth-aachen.de/Forschung/Reports.html.

[15] I. Horrocks, U. Sattler, and S. Tobies. Practical reasoning for expressive description logics. In *Proceedings of the 6th International Conference on Logic for Programming and Automated Reasoning (LPAR'99)*, pages 161–180, 1999.

[16] A. Y. Levy and M.-C. Rousset. CARIN: A representation language combining horn rules and description logics. In Wolfgang Wahlster, editor, *Proceedings of the 12th European Conference on Artificial Intelligence (ECAI'96)*, pages 323–327. John Wiley & Sons Ltd., 1996.

[17] A. Schaerf. Reasoning with individuals in concept languages. *Data and Knowledge Engineering*, 13(2):141–176, 1994.