Towards Richer Rule Languages with PTime Data Complexity for the Semantic Web

Linh Anh Nguyen\textsuperscript{1,2}, Thi-Bich-Loc Nguyen\textsuperscript{3}, and Andrzej Szalas\textsuperscript{2,4}

\textsuperscript{1} Division of Knowledge and System Engineering for ICT, Ton Duc Thang University, No. 19, Nguyen Huu Tho Street, Tan Phong Ward, District 7, Ho Chi Minh City, Vietnam
\texttt{nguyenanhlinh@tdt.edu.vn}

\textsuperscript{2} Institute of Informatics, University of Warsaw
Banacha 2, 02-097 Warsaw, Poland
\{nguyen,andsz\}@mimuw.edu.pl

\textsuperscript{3} Department of Information Technology, Hue University of Sciences
77 Nguyen Hue, Hue city, Vietnam
\texttt{ntbichloc@hueuni.edu.vn}

\textsuperscript{4} Dept. of Computer and Information Science, Linköping University
SE-581 83 Linköping, Sweden

Abstract. We introduce a Horn description logic called Horn-DL, which is strictly and essentially richer than Horn-\textit{Reg}, Horn-SHIQ and Horn-SROIQ, while still has \textit{PTime} data complexity. In comparison with Horn-SROIQ, Horn-DL additionally allows the universal role and assertions of the form \textit{irreflexive}(a), \texttt{¬s}(a,b), \texttt{a ¬= b}. More importantly, in contrast to all the well-known Horn fragments \textit{EL}, DL-Lite, DLP, Horn-SHIQ, Horn-SROIQ of description logics, Horn-DL allows a form of the concept constructor “universal restriction” to appear at the left hand side of terminological inclusion axioms. Namely, a universal restriction can be used in such places in conjunction with the corresponding existential restriction. We develop the first algorithm with \textit{PTime} data complexity for checking satisfiability of Horn-DL knowledge bases.

1 Introduction

The Semantic Web is a rapidly growing research area that has received lots of attention in the last decade. One of the layers of the Semantic Web is OWL (Web Ontology Language). The first version of OWL is based on the description logic \textit{SHIQ}, the second version OWL 2, recommended by W3C in 2009, is based on the description logic \textit{SROIQ} \cite{14}. These logics are highly expressive but have intractable combined complexity (\textit{ExpTime}-complete and \textit{N2ExpTime}-complete, respectively) and data complexity (\textit{NP}-hard) for basic reasoning problems. Thus, W3C also recommended profiles OWL 2 EL, OWL 2 QL and OWL 2 RL, which are restricted sublanguages of OWL 2 Full with \textit{PTime} data complexity.\footnote{The full names of the complexity classes mentioned in this paper are as follows: \textit{PTime} means \textit{deterministic polynomial time}, \textit{NP} means \textit{nondeterministic polynomial time}, \textit{ExpTime} means \textit{deterministic exponential time}, \textit{N2ExpTime} means \textit{nondeterministic double exponential time}. The data complexity is measured in the size of the ABox, which is in the “reduced” form, while assuming that the RBox, the TBox and the query are fixed.}

Description logics (DLs) are formal languages suitable for representing terminological knowledge. They are of particular importance in providing a logical formalism for ontologies and the Semantic Web. DLs represent the domain of interest in terms of concepts, individuals, and roles. A concept is interpreted as a set of individuals, while a role is interpreted as a binary relation among individuals. A knowledge base in a DL consists of axioms about roles (grouped into an RBox), terminology axioms (grouped into a TBox), and assertions about individuals (grouped into an ABox). A DL is usually specified by: i) a set of constructors that allow building complex concepts and complex roles from concept names, role names and individual names, ii) allowed forms of axioms and assertions. The basic DL \textit{ALC} allows basic concept constructors listed in Table 1, but does not allow role constructors nor role axioms. The most common additional features for extending \textit{ALC} are also listed in Table 1: \textit{I} is a role constructor, \textit{Q} and \textit{O} are concept constructors, while \textit{H} and \textit{S} are allowed forms of role axioms.
## 1.1 Rule Languages in Description Logics

Rule languages in DLs have attracted lots of attention due to their applications to the Semantic Web. The profiles OWL 2 EL, OWL 2 QL and OWL 2 RL of OWL 2 are based on the families of DLs $\mathcal{E}L$ [1, 2], DL-Lite [4] and DLP [13], which are monotonic rule languages with PTIME data complexity. Such monotonic rule languages are Horn fragments of the corresponding full languages with appropriate restrictions adopted to eliminate nondeterminism.

A number of Horn fragments of DLs with PTIME data complexity have also been investigated in [15, 18, 20, 32, 27, 31]. The fragments Horn-SHIFQ [15] and Horn-SROIQ [31] are notable, with considerable rich sets of allowed constructors and features, where Horn-SROIQ is richer than Horn-SHIFQ. The combined complexities of Horn fragments of DLs were considered, amongst others, in [19]. Some tractable Horn fragments of DLs without ABoxes have also been isolated in [1, 3]. Combinations of rule languages like Datalog or its extensions with DLs have also been studied in a considerable number of works [7, 21, 23, 33, 17, 22, 8, 11, 5].

To eliminate nondeterminism, all $\mathcal{E}L$ [1, 2], DL-Lite [4], DLP [13], Horn-SHIFQ [15] and Horn-SROIQ [31] disallow (any form of) the universal restriction at the left hand side of $\sqsubseteq$ in terminological axioms. The problem is that roles are not required to be serial (i.e., satisfying the condition $\forall x \exists y R(x, y)$) and the general Horn fragment of the basic DL $\mathcal{ALC}$ that allows $\forall R.C$ at the left hand side of $\sqsubseteq$ has NP-complete data complexity [27].

In [25] Nguyen introduced the deterministic Horn fragment of $\mathcal{ALC}$, in which the constructor $\forall R.C$ is allowed at the left hand side of $\sqsubseteq$ in the combination with $\exists R.C$ (in the form $\forall R.C \sqsubseteq \exists R.C$, denoted by $\forall 3 R.C$ [3]). He proved that that fragment has PTIME data complexity by providing a bottom-up method for constructing a (logically) least model for a given deterministic positive knowledge base in the restricted language. In [27] Nguyen applied the method of [25] to regular DL $\mathcal{Reg}$, which extends $\mathcal{ALC}$ with regular role inclusion axioms characterized by finite automata. Let us denote the Horn fragment of $\mathcal{Reg}$ that allows the constructor $\forall 3 R.C$ at the left hand side of $\sqsubseteq$ by Horn-$\mathcal{Reg}$. As not every positive Horn-$\mathcal{Reg}$ knowledge base has a (logically) least model, Nguyen [27] proposed to approximate the instance checking problem in Horn-$\mathcal{Reg}$ by using weakenings with PTIME data complexity.

In [29] we studied a Horn fragment called Horn-$\mathcal{Reg}^I$ of the regular DL with inverse $\mathcal{Reg}^I$ and provided an algorithm with PTIME data complexity for checking satisfiability of Horn-$\mathcal{Reg}^I$ knowledge bases. This fragment extends Horn-$\mathcal{Reg}$ with inverse roles. Our work [29] overcomes the difficulties encountered in [25, 27] by using the top-down rather than bottom-up approach, and thus enables to show that both Horn-$\mathcal{Reg}$ and Horn-$\mathcal{Reg}^I$ have PTIME data complexity, solving an open problem of [27].

### Table 1. Concept constructors for $\mathcal{ALC}$ and some additional constructors/features of other DLs.

<table>
<thead>
<tr>
<th>Constructor</th>
<th>Syntax</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>complement</td>
<td>$\neg C$</td>
<td>$\neg \text{Male}$</td>
</tr>
<tr>
<td>intersection</td>
<td>$C \cap D$</td>
<td>$\text{Human} \cap \text{Male}$</td>
</tr>
<tr>
<td>union</td>
<td>$C \cup D$</td>
<td>$\text{Doctor} \cup \text{Lawyer}$</td>
</tr>
<tr>
<td>existential restriction</td>
<td>$\exists r.C$</td>
<td>$\exists \text{hasChild} \text{Male}$</td>
</tr>
<tr>
<td>universal restriction</td>
<td>$\forall r.C$</td>
<td>$\forall \text{hasChild} \text{Female}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Constructor/Feature</th>
<th>Syntax</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>inverse roles ($I$)</td>
<td>$r^-$</td>
<td>$\text{hasChild}^-$ (i.e., $\text{hasParent}$)</td>
</tr>
<tr>
<td>qualified number</td>
<td>$\geq n R.C$</td>
<td>$\geq 3 \text{hasChild} \text{Male}$</td>
</tr>
<tr>
<td>restrictions ($Q$)</td>
<td>$\leq n R.C$</td>
<td>$\leq 2 \text{hasParent}$.†</td>
</tr>
<tr>
<td>nominals ($O$)</td>
<td>${a}$</td>
<td>${\text{John}}$</td>
</tr>
<tr>
<td>hierarchies of roles ($H$)</td>
<td>$R \sqsubseteq S$</td>
<td>$\text{hasChild} \sqsubseteq \text{hasDescendant}$</td>
</tr>
<tr>
<td>transitive roles ($S$)</td>
<td>$R \circ R \sqsubseteq R$</td>
<td>$\text{hasDescendant} \circ \text{hasDescendant} \sqsubseteq \text{hasDescendant}$</td>
</tr>
</tbody>
</table>
1.2 Constructivism in Defining Rules

We now justify the usefulness of the concept constructor $\forall \exists R.C$. It is related to constructivism in defining rules [27]. Consider the TBox $\mathcal{T}$ consisting of the following rules:

$$\exists \text{hasChild}. \top \sqsubseteq \text{Parent} \tag{1}$$
$$\forall \text{hasChild}. \text{Happy} \sqsubseteq \text{HappyParent} \tag{2}$$
$$\text{HappyParent} \sqsubseteq \text{Happy}. \tag{3}$$

The symbol $\top$ stands for truth and the first rule states that if someone has a child then he or she is a parent. The second rule states that if all children of someone are happy then that person is a happy parent. The meaning of third rule is clear.

Let $\mathcal{R}$ be the empty RBox and $\mathcal{A}$ be the ABox specified by:

$$\mathcal{A} = \{ \text{hasChild} (\text{Jane}, \text{Peter}), \text{hasChild} (\text{Peter}, \text{Christ}),$$
$$\langle \leq 1 \text{hasChild}. \top \rangle (\text{Peter}), \text{Happy} (\text{Christ}) \}. $$

The first assertion states that Peter is a child of Jane, the second assertion states that Christ is a child of Peter, the third assertion states that Peter has no more than one child, and the fourth assertion states that Christ is happy.

Consider the knowledge base $\langle \mathcal{R}, \mathcal{T}, \mathcal{A} \rangle$. As expected, we have

$$\langle \mathcal{R}, \mathcal{T}, \mathcal{A} \rangle \models (\text{Parent} \sqcup \text{HappyParent})(\text{Peter}),$$

which means that Peter is a happy parent is a logical consequence of the knowledge base; but unexpectedly, we also have

$$\langle \mathcal{R}, \mathcal{T}, \mathcal{A} \rangle \models (\text{Parent} \sqcup \text{HappyParent})(\text{Christ}).$$

The reason is that the premise of the rule (2) is not “constructive”. In other words, the instance $\exists \text{hasChild}. \top \sqcup \forall \text{hasChild}. \lnot \top$ of the “law of excluded middle” holds for every individual $a$. If $a$ satisfies $\exists \text{hasChild}. \top$ then it is an instance of $\text{Parent}$; else it satisfies $\forall \text{hasChild}. \lnot \top$, and therefore satisfies any formula $\forall \text{hasChild}. C$, and hence is an instance of $\text{HappyParent}$ (by (2)).

The rule (2) should be changed to

$$\text{Parent} \sqcap \forall \text{hasChild}. \text{Happy} \sqsubseteq \text{HappyParent}$$

or to

$$\forall \exists \text{hasChild}. \text{Happy} \sqsubseteq \text{HappyParent}. $$

This shows that, in premises of rules, the constructor $\forall \exists$ is more “constructive” than $\forall$.

1.3 Our Contributions

This work is an extension of our conference papers [28, 29]. In this work, we introduce a Horn description logic called Horn-DL, which is strictly and essentially richer than Horn-$\text{Reg}^I$, Horn-$\text{SHIQ}$ and Horn-$\text{SROIQ}$, while still has $\text{PTime}$ data complexity.

In comparison with Horn-$\text{Reg}^I$ [29], Horn-DL additionally allows nominals, qualified number restrictions, the $\exists r. \text{Self}$ constructor, the universal role as well as assertions of the form $\text{Disj}(s, s')$, $\text{irreflexive}(s)$, $\lnot s(a, b)$, $a \neq b$.

In comparison with Horn-$\text{SROIQ}$, Horn-DL additionally allows the universal role and assertions of the form $\text{irreflexive}(s)$, $\lnot s(a, b)$, $a \neq b$. More importantly, like Horn-$\text{Reg}^I$, but in contrast to all the well-known Horn fragments $\mathcal{EL}$ [1, 2], DL-Lite [4], DLP [13], Horn-$\text{SHIQ}$ [15], Horn-$\text{SROIQ}$ [31] of DLs, Horn-DL allows the concept constructor $\forall \exists R.C$ to appear at the left hand side of terminological inclusion axioms.
We develop a direct algorithm with PTIME data complexity for checking satisfiability of
Horn-DL knowledge bases. It is based on constructing a graph using a global caching technique
similar to the one of [25–27, 30, 10, 29]. The keys for our algorithm are to follow the top-down
approach of [29] (instead of the bottom-up approach of [27]) and use special techniques to deal
with non-seriality, inverse roles and other concept constructors of Horn-DL.

1.4 The Structure of This Paper

The rest of this paper is structured as follows. In Section 2 we present syntax and semantics
of Horn-DL. In Section 3 we give a detailed comparison between Horn-DL and Horn-SROIQ.
In Section 4 we present our algorithm of checking satisfiability of Horn-DL knowledge bases,
starting with a recall of automaton-modal operators. Section 5 contains examples illustrating
the algorithm and Section 6 concludes this work.

2 Syntax and Semantics of Horn-DL

Our language uses a finite set $C$ of concept names, a finite set $R_+$ of role names including a
subset of simple role names, and a finite set $I$ of individual names. We use letters like $a, b$
to denote individual names, letters like $A, B$ to denote concept names, and letters like $r, s$
to denote role names.

For $r \in R_+$, we call the expression $\overline{r}$ the inverse of $r$. Let $R_-=\{\overline{r} \mid r \in R_+\}$ and
$\mathcal{R} = R_+ \cup R_-$. For $\mathcal{R} = \tau$, let $\overline{\mathcal{R}}$ stand for $r$. We call elements of $R$
basic roles and use letters like $R, S$ to denote them. We define a simple role to be either a simple role name or the inverse
of a simple role name.\footnote{As will be seen later, only the universal role is not basic and only simple roles are allowed in number restrictions.}

2.1 Syntax of RBoxes

A role inclusion axiom (RIA for short) is an expression of the form $\mathcal{S}_1 \circ \cdots \circ \mathcal{S}_k \subseteq \mathcal{R}$, where
$k \geq 0$. In the case $k = 0$, the left hand side of the inclusion axiom stands for the identity binary
relation $\varepsilon$. A role assertion is an expression of the form $\text{Irr}(S)$ or $\text{Disj}(S, S')$, where $S$ and $S'$
are simple roles, $\text{Irr}$ stands for “irreflexive”, and $\text{Disj}$ stands for “disjoint”.

A context-free semi-Thue system $\mathcal{S}$ over $R$ is a finite set of context-free production rules
over alphabet $R$. It is symmetric if, for every rule $R \rightarrow \mathcal{S}_1 \cdots \mathcal{S}_k$ of $\mathcal{S}$, the rule $\overline{\mathcal{R}} \rightarrow \overline{\mathcal{S}_k} \cdots \overline{\mathcal{S}_1}$ is
also in $\mathcal{S}$.\footnote{In the case $k = 0$, the right hand side of the production rules stand for the empty word, also denoted by $\varepsilon$.} It is regular if, for every $R \in R$, the set of words derivable from $R$ using the system
is a regular language over $R$.

A context-free semi-Thue system is like a context-free grammar, but it has no designated
start symbol and there is no distinction between terminal and non-terminal symbols. We assume
that, for $R \in R$, the word $R$ is derivable from $R$ using such a system.

A regular box of RIAs is a finite set $\mathcal{R}_h$ of RIAs such that
\[
\{ R \rightarrow \mathcal{S}_1 \cdots \mathcal{S}_k \mid (\mathcal{S}_1 \circ \cdots \circ \mathcal{S}_k \subseteq \mathcal{R}) \in \mathcal{R}_h \}
\]
is a symmetric regular semi-Thue system $\mathcal{S}$ over $R$ such that if $R \in R$ is a simple role then only
words with length 1 or 0 are derivable from $R$ using $\mathcal{S}$. We assume that $\mathcal{R}_h$ is given together
with a mapping $\mathbf{A}$ that associates every $R \in R$ with a finite automaton $\mathbf{A}_R$ recognizing the
words derivable from $R$ using $\mathcal{S}$. We call that $\mathbf{A}$ the RIA-automaton-specification of $\mathcal{R}_h$.

A regular RBox is a set $\mathcal{R} = \mathcal{R}_a \cup \mathcal{R}_h$, where $\mathcal{R}_a$ is a finite set of role assertions and $\mathcal{R}_h$ is
a regular box of RIAs.

Let $\mathcal{R}$ be a regular RBox and $\mathbf{A}$ be the RIA-automaton-specification of $\mathcal{R}_h$. For $R, S \in R$, we say that $R$ is a subrole of $S$ w.r.t. $\mathcal{R}$, denoted by $R \subseteq_R S$, if the word $R$ is accepted by $\mathbf{A}_S$.\footnote{As will be seen later, only the universal role is not basic and only simple roles are allowed in number restrictions.}
Recall that a finite automaton $A$ over alphabet $R$ is a tuple $(R, Q, q_0, \delta, F)$, where $Q$ is a finite set of states, $q_0 \in Q$ is the initial state, $\delta \subseteq Q \times R \times Q$ is the transition relation, and $F \subseteq Q$ is the set of accepting states. A run of $A$ on a word $R_1 \ldots R_k$ over alphabet $R$ is a finite sequence of states $q_0, q_1, \ldots, q_k$ such that $\delta(q_{i-1}, R_i, q_i)$ holds for every $1 \leq i \leq k$. It is an accepting run if $q_k \in F$. We say that $A$ accepts a word $w$ if there exists an accepting run of $A$ on $w$.

Example 2.1. Let $R_+ = \{ \text{link}, \text{path} \}$ and let $\mathcal{R}_h$ consist of the following RIAs:

\[
\begin{align*}
\text{link} \subseteq & \text{path}, \\
\text{link} \circ \text{path} \subseteq & \text{path}, \\
\text{path} \circ \text{link} \subseteq & \text{path}.
\end{align*}
\]

The set $\mathcal{R}_h$ is a regular box of RIAs. Its RIA-automaton-specification $A$ is specified by:

\[
\begin{align*}
A_{\text{link}} &= \langle R, \{0, 1\}, 0, \{\{0, \text{link}, 1\}\}, \{1\} \\
A_{\text{link}}^{-} &= \langle R, \{0, 1\}, 1, \{\{1, \text{link}, 0\}\}, \{0\} \\
A_{\text{path}} &= \langle R, \{0, 1\}, 0, \{\{0, \text{link}, 0\}, \{0, \text{link}, 1\}, \{\text{path}, 1\}\}, \{1\} \\
A_{\text{path}}^{-} &= \langle R, \{0, 1\}, 1, \{\{0, \text{link}, 0\}, \{1, \text{link}, 0\}, \{\text{path}, 0\}\}, \{0\}.
\end{align*}
\]

Example 2.2. DLs are variants of modal logics. Individuals in DLs are like states in Kripke models, and roles in DLs are like accessibility relations in Kripke models. Let $R_+ = \{B, G, I\}$, where $B$, $G$ and $I$ stand for the accessibility relations of belief, goal and intention, respectively. Positive and negative introspection of these modalities w.r.t. belief as well as the axiom “intention implies goal” correspond to the following RIAs [9]:

\[
\begin{align*}
B \circ B & \subseteq B, \\
\overline{B} \circ B & \subseteq \overline{B}, \\
B \circ G & \subseteq G, \\
\overline{G} \circ B & \subseteq \overline{G}, \\
B \circ I & \subseteq I, \\
\overline{I} \circ B & \subseteq \overline{I}, \\
G \subseteq I, \\
\overline{C} & \subseteq \overline{I}.
\end{align*}
\]

The set consisting of the above RIAs is a regular box of RIAs, because the sets of words derivable from the symbols $B$, $G$, $I$ using the corresponding semi-Thue system are characterized, respectively, by the following regular expressions:

\[
\begin{align*}
&- (B \cup \overline{B})^* B \\
&- (B \cup \overline{B})^* G \\
&- (B \cup \overline{B})^* (I \cup G).
\end{align*}
\]

2.2 Syntax of Concepts and TBoxes

Concepts are defined by the following BNF grammar, where $A \in C$, $R \in R$, $a \in I$, $n$ is a natural number and $S$ is a simple role:

\[
C ::= \top \mid \bot \mid A \mid \neg C \mid C \cap C \mid C \cup C \mid \forall R.C \mid \exists R.C \mid \{a\} \mid \forall U.C \mid \exists U.C \mid \exists R.\text{Self} \mid \geq n S.C \mid \leq n S.C
\]

The symbol $U \notin R$ is called the universal role.

We use letters like $C$, $D$ to denote concepts.

Let $\forall R.C$ stand for $\forall R.C \cap \exists R.C$. Left-hand-side (Horn-DL) concepts, called LHS concepts for short, are defined by the following BNF grammar, where $A \in C$, $R \in R$, $a \in I$, and $S$ is a simple role:

\[
C ::= \top \mid A \mid C \cap C \mid C \cup C \mid \forall R.C \mid \exists R.C \mid \{a\} \mid \forall U.C \mid \exists U.C \mid \exists S.\text{Self}
\]
2.3 Syntax of ABoxes and Knowledge Bases

A (Horn-DL) ABox is a finite set of assertions of the form $C(a)$, $r(a, b)$, $\neg s(a, b)$, $a = b$ or $a \neq b$, where $C$ is an RHS concept, $r \in \mathbf{R}$, and $s$ is a simple role name.

A reduced (Horn-DL) ABox is a finite set of assertions of the form $A(a)$, $\neg A(a)$, $r(a, b)$, $\neg s(a, b)$ or $a \neq b$, where $r \in \mathbf{R}$, and $s$ is a simple role name.

A (Horn-DL) knowledge base is a tuple $\langle \mathbf{R}, \mathbf{T}, \mathbf{A} \rangle$, where $\mathbf{R}$ is a regular RBox, $\mathbf{T}$ is a Horn-DL TBox and $\mathbf{A}$ is a Horn-DL ABox. When $\mathbf{T}$ is a clausal Horn-DL TBox and $\mathbf{A}$ is a reduced Horn-DL ABox, we call such a knowledge base a clausal (Horn-DL) knowledge base.

Example 2.3. Let

$$\mathbf{C} = \{ \text{Male}, \text{Female}, \text{Mother}, \text{Father}, \text{Parent}, \text{Grandparent}, \text{OnlyChild}, \text{ParentWithOnlyChild}, \text{ParentWithOnlySons}, \text{ParentWithOnlyDaughters} \}$$

$$\mathbf{R}_+ = \{ \text{hasChild}, \text{hasParent}, \text{hasDescendant} \}$$

$$\mathbf{I} = \{ \text{Alice}, \text{Bob}, \text{Confucius}, \text{Dave}, \text{Ellie}, \text{Frank}, \text{George} \},$$

where hasChild is a simple role name, and let

$$\mathbf{A} = \{ \text{Male}(\text{Bob}), \text{Male}(\text{Confucius}), \text{Male}(\text{Dave}), \text{Male}(\text{Frank}), \text{Male}(\text{George}), \text{Female}(\text{Alice}), \text{Female}(\text{Ellie}), \text{hasChild}(\text{Alice}, \text{Dave}), \text{hasChild}(\text{Alice}, \text{Ellie}), \text{hasChild}(\text{Bob}, \text{Dave}), \text{hasChild}(\text{Bob}, \text{Ellie}), \text{hasChild}(\text{Dave}, \text{Frank}), \text{hasChild}(\text{Ellie}, \text{George}), \text{OnlyChild}(\text{George}), \text{ParentWithOnlyChild}(\text{Dave}) \} \cup \{ x \neq y \mid x, y \in \mathbf{I}, x \neq y \}$$

$$\mathbf{R} = \{ \text{hasChild} \sqsubseteq \text{hasParent}, \text{hasChild} \sqsubseteq \overline{\text{hasParent}}, \text{hasChild} \sqsubseteq \text{hasDescendant}, \text{hasChild} \sqsubseteq \overline{\text{hasDescendant}}, \text{hasDescendant} \circ \text{hasChild} \sqsubseteq \text{hasDescendant}, \text{hasDescendant} \circ \overline{\text{hasChild}} \sqsubseteq \overline{\text{hasDescendant}} \}$$

and let $\mathbf{T}$ be the TBox consisting of the following axioms:

$$\exists \text{hasChild}. \top \sqsubseteq \text{Parent} \quad (4)$$

$$\exists \text{hasChild}. \text{Parent} \sqsubseteq \text{Grandparent} \quad (5)$$
Fig. 1. An illustration for the ABox given in Example 2.3.

\[
\begin{align*}
\text{Confucius} & : \text{Male} \\
\text{Alice} & : \text{Female} \\
\text{Dave} & : \text{Male, ParentWithOnlyChild} \\
\text{Ellie} & : \text{Female} \\
\text{Frank} & : \text{Male} \\
\text{George} & : \text{Male, OnlyChild} \\
\text{Bob} & : \text{Male} \\
\text{Dave} & : \text{Male, ParentWithOnlyChild} \\
\text{Ellie} & : \text{Female} \\
\text{Frank} & : \text{Male} \\
\text{George} & : \text{Male, OnlyChild}
\end{align*}
\]

The ABox \( \mathcal{A} \) is illustrated in Figure 1, where the edges denote instances of the role \( \text{hasChild} \). Observe that \( KB = \langle \mathcal{R}, \mathcal{T}, \mathcal{A} \rangle \) is a Horn-DL knowledge base. Treating \( \text{hasChild} \) as a “primitive” role and \( \text{Male, Female} \) as “primitive” concepts, using the approach of logic programming, the RBox \( \mathcal{R} \) “defines” the roles \( \text{hasParent} \) and \( \text{hasDescendant} \), while the TBox axioms (4)–(10) “define” the concepts \( \text{Parent}, \text{Grandparent}, \text{Father}, \text{Mother}, \text{ParentWithOnlySons}, \text{ParentWithOnlyDaughters} \) and \( \text{DescendantOfConfucius} \). That is, \( \text{hasParent} \) is “defined” to be the inverse of \( \text{hasChild} \), \( \text{hasDescendant} \) is “defined” to be the transitive closure of \( \text{hasChild} \), and \( \subseteq \) in the axioms (4)–(10) can be treated as equivalence. The axiom (11) is not a definition of \( \text{OnlyChild} \). It states only a necessary condition but not a sufficient condition for being an instance of \( \text{OnlyChild} \). The problem is that the converse of (11) is not allowed in Horn-DL. The axioms (12) and (13) define \( \text{ParentWithOnlyChild} \) to be equivalent to \( \exists \text{hasChild}. \text{OnlyChild} \). We use (12) as a constraint about \( \text{OnlyChild} \). The axiom (14) is a constraint stating that \( \text{Male} \) and \( \text{Female} \) are “disjoint” concepts.

\[\text{Fig. 2. An illustration for the TBox axioms.}\]

2.4 Semantics

An interpretation is a pair \( \mathcal{I} = \langle \Delta^I, \mathcal{I}^I \rangle \), where \( \Delta^I \) is a non-empty set called the domain of \( \mathcal{I} \) and \( \mathcal{I}^I \) is a mapping called the interpretation function of \( \mathcal{I} \) that associates each individual \( a \in \mathbf{I} \) with an element \( a^I \in \Delta^I \), each concept name \( A \in \mathbf{C} \) with a set \( A^I \subseteq \Delta^I \), and each role name \( r \in \mathbf{R}_+ \) with a binary relation \( r^I \subseteq \Delta^I \times \Delta^I \). Define \( e^I = \{ \langle x, x \rangle \mid x \in \Delta^I \} \), \( U^I = \Delta^I \times \Delta^I \), and \( (r^I)^{-1} = \{ \langle y, x \rangle \mid \langle x, y \rangle \in r^I \} \) for \( r \in \mathbf{R}_+ \). The interpretation function \( \mathcal{I}^I \) is extended to complex concepts as shown in Figure 2.

Note that, \( (\forall U.C)^I = \Delta^I \) or \( (\forall U.C)^I = \emptyset \), and \( (\exists U.C)^I = \Delta^I \) or \( (\exists U.C)^I = \emptyset \).
Given an interpretation $\mathcal{I}$ and an axiom/assertion $\varphi$, the satisfaction relation $\mathcal{I} \models \varphi$ is defined as follows:

- $\mathcal{I} \models \text{irr}(S)$ if $S^\mathcal{I}$ is irreflexive
- $\mathcal{I} \models \text{Disj}(S, S')$ if $S^\mathcal{I}$ and $S'^\mathcal{I}$ are disjoint
- $\mathcal{I} \models S_1 \circ \cdots \circ S_k \subseteq R$ if $S_1^\mathcal{I} \circ \cdots \circ S_k^\mathcal{I} \subseteq R^\mathcal{I}$
- $\mathcal{I} \models x \subseteq R$ if $x^\mathcal{I} \subseteq R^\mathcal{I}$
- $\mathcal{I} \models C \subseteq D$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$
- $\mathcal{I} \models C(a)$ if $a^\mathcal{I} \in C^\mathcal{I}$
- $\mathcal{I} \models r(a, b)$ if $\langle a^\mathcal{I}, b^\mathcal{I} \rangle \in r^\mathcal{I}$
- $\mathcal{I} \models \neg s(a, b)$ if $\langle a^\mathcal{I}, b^\mathcal{I} \rangle \not\in s^\mathcal{I}$
- $\mathcal{I} \models a = b$ if $a^\mathcal{I} = b^\mathcal{I}$
- $\mathcal{I} \models a \neq b$ if $a^\mathcal{I} \neq b^\mathcal{I}$.

If $\mathcal{I} \models \varphi$ then we say that $\mathcal{I}$ validates $\varphi$.

An interpretation $\mathcal{I}$ is a model of an RBox $\mathcal{R}$, a TBox $\mathcal{T}$ or an ABox $\mathcal{A}$ if it validates all the axioms/assertions of that “box”. It is a model of a knowledge base $\langle \mathcal{R}, \mathcal{T}, \mathcal{A} \rangle$ if it is a model of all $\mathcal{R}$, $\mathcal{T}$ and $\mathcal{A}$.

A knowledge base is satisfiable if it has a model. For a knowledge base $KB$, we write $KB \models \varphi$ to mean that every model of $KB$ validates $\varphi$. If $KB \models C(a)$ then we say that $a$ is an instance of $C$ w.r.t. $KB$.

Example 2.4. Let us continue Example 2.3. One can ask, for example:

- whether Alice is an instance of the concept DescendantOfConfucius w.r.t. $KB$,
- whether Bob is an instance of the concept Grandparent w.r.t. $KB$,
- whether Ellie is an instance of the concept ParentWithOnlySons w.r.t. $KB$.

To answer any of these questions, one can extend $KB$ with the negation of the question and check whether the extended knowledge base is unsatisfiable. The answer is “no” for the first question and “yes” for the two remaining questions.

We justify the answer for the first question. Let $\mathcal{A}' = \mathcal{A} \cup \{ \neg \text{DescendantOfConfucius(Alice)} \}$ and $KB' = \langle \mathcal{R}, \mathcal{T}, \mathcal{A}' \rangle$. This extended knowledge base has a model $\mathcal{I} = \langle \Delta^\mathcal{I}, \cdot^\mathcal{I} \rangle$ specified by:

- $\Delta^\mathcal{I} = \mathcal{I}$, $\text{Male}^\mathcal{I} = \{ x \mid \text{Male}(x) \in \mathcal{A} \}$, $\text{Female}^\mathcal{I} = \{ x \mid \text{Female}(x) \in \mathcal{A} \}$,
- $\text{hasChild}^\mathcal{I} = \{ \langle x, y \rangle \mid \text{hasChild}(x, y) \in \mathcal{A} \}$, $\text{hasParent}^\mathcal{I} = (\text{hasChild}^\mathcal{I})^{-1}$,
- $\text{hasDescendant}^\mathcal{I} = (\text{hasChild}^\mathcal{I})^+$ (the transitive closure of $\text{hasChild}^\mathcal{I}$),
- $\text{Parent}^\mathcal{I} = \{ \text{Alice}, \text{Bob}, \text{Dave}, \text{Ellie} \}$, $\text{Father}^\mathcal{I} = \{ \text{Bob}, \text{Dave} \}$, $\text{Mother}^\mathcal{I} = \{ \text{Alice}, \text{Ellie} \}$,
- $\text{OnlyChild}^\mathcal{I} = \{ \text{Frank}, \text{George} \}$, $\text{ParentWithOnlyChild}^\mathcal{I} = \{ \text{Dave}, \text{Ellie} \}$,
- $\text{ParentWithOnlySons}^\mathcal{I} = \{ \text{Dave}, \text{Ellie} \}$, $\text{ParentWithOnlyDaughters}^\mathcal{I} = \emptyset$,
- $\text{Grandparent}^\mathcal{I} = \{ \text{Alice}, \text{Bob} \}$, $\text{DescendantOfConfucius}^\mathcal{I} = \emptyset$.

To make things more interesting consider the Horn-DL knowledge base $KB_2 = \langle \mathcal{R}, \mathcal{T}_2, \mathcal{A} \rangle$, where

$$\mathcal{T}_2 = \mathcal{T} \cup \{ \top \subseteq (\exists \text{hasParent.Father} \land \exists \text{hasParent.Mother}) \}.$$
The additional axiom states that everyone has a father and a mother. To answer the question whether \( KB_2 \models \text{DescendantOfConfucius}(Alice) \) one can proceed in a similar way as before. Let \( \mathcal{A}_2 = A \cup \{ \neg \text{DescendantOfConfucius}(Alice) \} \) and \( KB_2 = (\mathcal{R}, T_2, \mathcal{A}_2) \). This extended knowledge base is satisfiable. Every “natural” model of \( KB_2 \) should be infinite (unless one assumes that there were the first people in the world). We give here a finite model \( \mathcal{J} = (\Delta^\mathcal{J}, -^\mathcal{J}) \) of \( KB'_2 \), which is specified by:

- \( \Delta^\mathcal{J} = I \cup \{ Adam, Eve \} \),
- \( \text{Male}^\mathcal{J} = \{ Adam \} \cup \{ x \mid \text{Male}(x) \in A \} \), \( \text{Female}^\mathcal{J} = \{ Eve \} \cup \{ x \mid \text{Female}(x) \in A \} \),
- \( \text{hasChild}^\mathcal{J} = \{ (x,y) \mid \text{hasChild}(x,y) \in A \} \cup \{ (Adam,x),(Eve,x) \mid x \in \{ Adam, Eve, Confucius, Alice, Ellie \} \} \),
- \( \text{hasParent}^\mathcal{J} = (\text{hasChild}^\mathcal{J})^{-1} \), \( \text{hasDescendant}^\mathcal{J} = (\text{hasChild}^\mathcal{J})^+ \),
- \( \text{Parent}^\mathcal{J} = \{ Adam, Eve, Alice, Bob, Dave, Ellie \} \),
- \( \text{Father}^\mathcal{J} = \{ Adam, Bob, Dave \} \), \( \text{Mother}^\mathcal{J} = \{ Eve, Alice, Ellie \} \),
- \( \text{OnlyChild}^\mathcal{J} = \{ Frank, George \} \), \( \text{ParentWithOnlyChild}^\mathcal{J} = \{ Dave, Ellie \} \),
- \( \text{ParentWithOnlySons}^\mathcal{J} = \{ Dave, Ellie \} \), \( \text{ParentWithOnlyDaughters}^\mathcal{J} = \emptyset \),
- \( \text{Grandparent}^\mathcal{J} = \{ Adam, Eve, Alice, Bob \} \), \( \text{DescendantOfConfucius}^\mathcal{J} = \emptyset \). <

### 2.5 Data Complexity

The size of a concept, an assertion or an axiom \( \varphi \) is defined to be the number of bits needed to encode \( \varphi \) in the usual way. The size of an ABox is the sum of the sizes of its assertions. The size of a TBox is the sum of the sizes of its axioms.

The data complexity class of Horn-DL is defined to be the complexity class of the problem of checking satisfiability of a Horn-DL knowledge base \( (\mathcal{R}, \mathcal{T}, \mathcal{A}) \), measured in the size of \( \mathcal{A} \) when assuming that \( \mathcal{R} \) and \( \mathcal{T} \) are fixed and \( \mathcal{A} \) is a reduced Horn-DL ABox.

**Proposition 2.5.** Let \( KB = (\mathcal{R}, \mathcal{T}, \mathcal{A}) \) be a Horn-DL knowledge base.

1. If \( C \) is an LHS concept then \( KB \models C(a) \) iff the Horn-DL knowledge base \( (\mathcal{R}, \mathcal{T} \cup \{ C \subseteq A \}, \mathcal{A} \cup \{ \neg A(a) \} ) \) is unsatisfiable, where \( A \) is a fresh concept name.
2. \( KB \) can be converted in polynomial time in the sizes of \( \mathcal{T} \) and \( \mathcal{A} \) to a Horn-DL knowledge base \( KB' = (\mathcal{R}, \mathcal{T}', \mathcal{A}') \) with \( \mathcal{A}' \) being a reduced Horn-DL ABox such that \( KB \) is satisfiable iff \( KB' \) is satisfiable.
3. \( KB \) can be converted in polynomial time in the size of \( \mathcal{T} \) to a Horn-DL knowledge base \( KB' = (\mathcal{R}, \mathcal{T}', \mathcal{A}) \) with \( \mathcal{T}' \) being a clausal Horn-DL TBox such that \( KB \) is satisfiable iff \( KB' \) is satisfiable.

**Proof.** The first assertion is clear. For the second assertion, we start with \( \mathcal{T}' := \mathcal{T} \) and \( \mathcal{A}' := \mathcal{A} \) and then modify them as follows:

- for each \( C(a) \in \mathcal{A} \) where \( C \) is not a concept name, replace \( C(a) \) in \( \mathcal{A}' \) by \( A(a) \), where \( A \) is a fresh concept name, and add to \( \mathcal{T}' \) the axiom \( A \subseteq C \),
- for each \( (a \# b) \in \mathcal{A}' \), delete that assertion from \( \mathcal{A}' \) and then replace every occurrence of \( b \) in \( \mathcal{T}' \) and \( \mathcal{A}' \) by \( a \).

It is easy to check that the resulting Horn-DL knowledge base \( KB' = (\mathcal{R}, \mathcal{T}', \mathcal{A}') \) is satisfiable iff \( KB \) is satisfiable.

For the third assertion, we apply the technique that replaces complex concepts by fresh concept names. For example, if \( \forall \exists R.C \subseteq \exists S.D \) is an axiom of \( \mathcal{T} \), where \( C \) and \( D \) are complex concepts, then we replace it by axioms \( C \subseteq A_C, \forall \exists R.A_C \subseteq \exists S.A_D \) and \( A_D \subseteq D \), where \( A_C \) and \( A_D \) are fresh concept names. Furthermore, every \( \exists r.Self \) is replaced by \( \exists r.Self \).

**Corollary 2.6.** Every Horn-DL knowledge base \( KB = (\mathcal{R}, \mathcal{T}, \mathcal{A}) \) can be converted in polynomial time in the sizes of \( \mathcal{T} \) and \( \mathcal{A} \) to a clausal Horn-DL knowledge base \( KB' = (\mathcal{R}, \mathcal{T}', \mathcal{A}') \) such that \( KB \) is satisfiable iff \( KB' \) is satisfiable.
3 Comparing Horn-DL with Horn-\textit{SROIQ}

We first recall the syntax of Horn-\textit{SROIQ} [31]:

- An RBox in Horn-\textit{SROIQ} consists of role assertions $\text{Disj}(S,S')$, where $S$ and $S'$ are simple roles, and RIAs of the following forms, where $k \geq 1$ and $R, R_1, \ldots, R_k$ are roles strictly less than $r$ in a fixed partial order:

  $r \sqsubseteq r$, $R_1 \circ \cdots \circ R_k \sqsubseteq r$
  $R \sqsubseteq r$, $r \circ R_1 \circ \cdots \circ R_k \sqsubseteq r$
  $r \circ r \sqsubseteq r$, $R_1 \circ \cdots \circ R_k \circ r \sqsubseteq r$.

- A TBox in Horn-\textit{SROIQ} consists of axioms of the following forms, where $C, C', C''$ are concepts of the form $A, \{a\}$ or $\exists S.\text{Self}$, and $S$ is a simple role:

  $C \sqcap C' \sqsubseteq C''$, $\exists R.C \sqsubseteq C'$
  $C \sqsubseteq \forall R.C'$, $C \sqsubseteq 1.S.C'$
  $C \sqsubseteq \exists R.C'$, $C \sqsubseteq \geq n.S.C'$.

- An ABox in Horn-\textit{SROIQ} consists of assertions of the form $A(a)$ or $R(a,b)$.

Comparing Horn-\textit{SROIQ} with Horn-DL, observe that:

- Every set of RIAs in Horn-\textit{SROIQ} is a regular box of RIAs, but the converse does not hold. For example, the regular box of RIAs given in Example 2.2 is not a set of RIAs in Horn-\textit{SROIQ} (because $B$ and $\overline{B}$ are “dependent” on each other). Roughly speaking using the notion of regular expressions, “regularity” of RIAs in Horn-\textit{SROIQ} allows only a bounded nesting depth of the star operator $^*$, while “regularity” of RIAs in Horn-DL is not so restricted. Furthermore, reflexivity of a role is expressible in our case, but not in the case of Horn-\textit{SROIQ}.

- In contrast to Horn-DL, Horn-\textit{SROIQ} does not allow the universal role and assertions of the form $\text{Irr}(S)$, $\neg s(a,b)$, $a \neq b$. Also, in Horn-\textit{SROIQ} a concept $\exists R.\text{Self}$ can occur at the right hand side of $\sqsubseteq$ only when $R$ is a simple role, while in Horn-DL such $R$ can be an arbitrary role.

- The most remarkable difference is that, in contrast to Horn-DL, Horn-\textit{SROIQ} does not allow the concept constructor $\forall \exists R.C$ in the left hand side of TBox inclusion axioms.

Therefore, every RBox (resp. TBox or ABox) in Horn-\textit{SROIQ} is also a (regular) RBox (resp. TBox or ABox) in Horn-DL, but not vice versa. Thus, Horn-DL is strictly and essentially richer than Horn-\textit{SROIQ}.

4 Checking Satisfiability of Horn-DL Knowledge Bases

In this section we present an algorithm that, given a clausal Horn-DL knowledge base $\langle R, T, A \rangle$ together with the RIA-automaton-specification $A$ of $R_h$, checks whether the knowledge base is satisfiable.

\textsuperscript{8} It can be proved that extending Horn-\textit{SROIQ} only by allowing the concept constructor $\forall \exists R.C$ in the left hand side of TBox inclusion axioms already increases its expressiveness. This can be done by using bisimulation as in [6].
Given an interpretation \( \mathcal{I} \) and a finite automaton \( A \) over alphabet \( R \), define \( A^\mathcal{I} = \{(x, y) \in \Delta^\mathcal{I} \times \Delta^\mathcal{I} \mid \text{there exist a word } R_1 \ldots R_k \text{ accepted by } A \text{ and elements } x_0 = x, x_1, \ldots, x_k = y \text{ of } \Delta^\mathcal{I} \text{ such that } (x_{i-1}, x_i) \in R_i^\mathcal{I} \text{ for all } 1 \leq i \leq k \} \).

We will use auxiliary modal operators \( [A] \) and \( \langle A \rangle \), where \( A \) is a finite automaton over alphabet \( R \). We call \( [A] \) (resp. \( \langle A \rangle \)) a universal (resp. existential) automaton-modal operator. Such operators were used earlier, among others, in \([12, 26, 10, 27]\).

In the extended language, if \( C \) is a concept then \( [A]C \) and \( \langle A \rangle C \) are also concepts. The semantics of \( [A]C \) and \( \langle A \rangle C \) in an interpretation \( \mathcal{I} \) are defined as follows:

\[
([A]C)^\mathcal{I} = \{ x \in \Delta^\mathcal{I} \mid \forall y((x, y) \in A^\mathcal{I} \implies y \in C^\mathcal{I}) \}
\]

\[
(\langle A \rangle C)^\mathcal{I} = \{ x \in \Delta^\mathcal{I} \mid \exists y((x, y) \in A^\mathcal{I} \text{ and } y \in C^\mathcal{I}) \}.
\]

For a finite automaton \( A \) over \( R \), assume that \( A = (R, Q_A, q_A, \delta_A, F_A) \).

If \( q \) is a state of a finite automaton \( A \) then by \( A_q \) we denote the finite automaton obtained from \( A \) by replacing the initial state by \( q \).

**Lemma 4.1.** Let \( \mathcal{I} \) be a model of a regular box \( R_h \) of RIAs, \( A \) be the RIA-automaton-specification of \( R_h \), \( C \) be a concept, and \( R \in R \). Then \( (\forall R.C)^\mathcal{I} = ([A]_R(C))^\mathcal{I} \), \( (\exists R.C)^\mathcal{I} = ([A]_R(C))^\mathcal{I} \), \( C^\mathcal{I} \subseteq ([A]_R(C))^\mathcal{I} \) and \( C^\mathcal{I} \subseteq ([A]_R(C))^\mathcal{I} \).

The proof of this lemma is straightforward.

### 4.2 The Data Structure and Auxiliary Notions

For convenience, we also refer to TBox axioms as concepts and treat a TBox axiom \( C \subseteq D \) as the concept \( \neg C \cup D \) (which stands for a global requirement for all elements in the domain). Observe that, for any interpretation \( \mathcal{I} \), we have \( \mathcal{I} \models C \subseteq D \iff (\neg C \cup D)^\mathcal{I} = \Delta^\mathcal{I} \).

Let \( \text{EdgeLabels} \) be the power set of \( \{ \exists R.A, \exists R.T \mid R \in R, A \in C \} \cup R \). For \( E \in \text{EdgeLabels} \), we define \( \text{Roles}(E) = \{ R \in R \mid E \text{ contains } R \text{ or some } \exists R.C \} \) and call \( E = \{ R \mid R \in \text{Roles}(E) \} \) the inverse of \( E \).

Our algorithm for checking satisfiability of \( (R, T, A) \) uses a data structure

\[
G = (R', T', A', \Delta_0, \Delta_1, \Delta_2, \Delta, Label, Succ, Next_U, Replacement, Status),
\]

called a **Horn-DL graph**, where:

- \( R', T' \) and \( A' \) are modified versions of \( R, T \) and \( A \), respectively,
- \( \Delta \) is a non-empty set of nodes, \( \Delta_0 \subseteq \Delta_1 \subseteq \Delta, \Delta_2 \subseteq \Delta \) and \( \Delta_2 \cap \Delta_0 = \emptyset \),
- \( Label \) is a mapping that associates each \( x \in \Delta \) with a set of concepts,
- \( Succ : \Delta \times \text{EdgeLabels} \rightarrow \Delta \) is a partial mapping,
- \( Next_U : \{ \exists U.T, \exists U.A \mid A \in C \} \rightarrow \Delta \) is a partial mapping,
- \( Replacement : \mathcal{I} \rightarrow \mathcal{I} \) is a mapping,
- \( Status \in \{ \text{unknown, unsat, sat} \} \).

Informally, \( \Delta_0 \) consists of all individual names occurring in \( A' \) and \( T' \), \( \Delta_1 \setminus \Delta_0 \) consists of the nodes created for realizing “requirements” of the form \( \exists U.C \), and \( \Delta_2 \) consists of so called unnamed noncloneable nodes. For a node \( x \in \Delta \), \( Label(x) \) is called the label of \( x \). If \( Succ(x, E) = y \) then \( y \) is called a successor of \( x \), and \( x \) is called a predecessor of \( y \). If \( Next_U(\exists U.A) = y \) then \( A \in Label(y) \) and the “requirement” \( \exists U.A \) at any \( x \) is “realized” by using \( y \). When defined, \( Next_U(\exists U.T) \) stands for the “logically smallest object”. If \( Replacement(a) = b \) then \( a \equiv b \) follows from the knowledge base and \( b \) is the representative of its abstract class of the equivalence relation \( \equiv \). A fact \( Status = \text{unsat} \) (resp. \( Status = \text{sat} \)) means the knowledge base is unsatisfiable (resp. satisfiable).
The partial mapping \( \text{Succ} \) has the property that if
\[
\text{Succ}(x, \{ \exists R_1 C_1, \ldots, \exists R_k C_k, S_1, \ldots, S_h \}) = y
\]
then \( \{ \exists R_1 C_1, \ldots, \exists R_k C_k \} \subseteq \text{Label}(x), \{ C_1, \ldots, C_k \} \subseteq \text{Label}(y), \) and \( x \) is “connected” to \( y \) by all the roles \( R_1, \ldots, R_k, S_1, \ldots, S_h. \) Furthermore,

if both \( \text{Succ}(x, E) \) and \( \text{Succ}(x, E') \) are defined then \( E \cap E' \subseteq R. \quad (15) \)

That is, for each \( \exists R.C \) with \( C \in C \cup \{ \top \}, \) there exists at most one \( E \in \text{EdgeLabels} \) such that \( \exists R.C \in E \) and \( \text{Succ}(x, E) \) is defined.

We say that \( x \in \Delta \) is reachable from \( \Delta_0 \) if there exist \( x_0, \ldots, x_k \in \Delta \) and \( E_1, \ldots, E_k \) such that \( x_0 \in \Delta_1, k \geq 0, x_k = x \) and \( \text{Succ}(x_{i-1}, E_i) = x_i \) for all \( 1 \leq i \leq k. \)

Let \( \text{Next} : \Delta \times \{ \exists R.A, \exists R.\top \mid R \in \mathbf{R}, A \in C \} \to \Delta \) denote the partial mapping defined by: \( \text{Next}(x, \exists R.C) = y \) if there exists \( E \in \text{EdgeLabels} \) with \( \exists R.C \in E \) and \( \text{Succ}(x, E) = y. \)

By the assumption (15), \( \text{Next} \) is well defined. The meaning of \( \text{Next}(x, \exists R.C) = y \) is that \( \exists R.C \in \text{Label}(x), C \in \text{Label}(y) \) and the “requirement” \( \exists R.C \) is realized at \( x \) by going to \( y. \)

If \( \text{Next}(x, \exists R.\top) = y \) then \( y \) stands for the “logically smallest \( R \)-neighbor” of \( x. \)

Define \( \text{NamedEdges} \) to be the smallest subset of \( \Delta_0 \times \mathbf{R} \times \Delta_0 \) such that:

- if \( r(a, b) \in \mathbf{A}' \) then \( \langle a, r, b \rangle \in \text{NamedEdges}, \)
- if \( \text{Succ}(a, E) = b \) and \( R \in \text{Roles}(E) \) then \( \langle a, R, b \rangle \in \text{NamedEdges}, \)
- if \( \exists R.\text{Self} \in \text{Label}(a) \) then \( \langle a, R, a \rangle \in \text{NamedEdges}, \)
- if \( \langle a, R, b \rangle \in \text{NamedEdges} \) and \( R \sqsubseteq_R S \) then \( \langle a, S, b \rangle \in \text{NamedEdges}, \)
- if \( \langle a, R, b \rangle \in \text{NamedEdges} \) then \( \langle b, R, a \rangle \in \text{NamedEdges}. \)

Let \( X \) be a set of concepts. The saturation of \( X \) (w.r.t. \( \mathbf{A} \) and \( \mathbf{T} \)), denoted by \( \text{Satr}(X), \) is defined to be the least extension of \( X \) such that:

- for every \( R \in \mathbf{R}, \) if \( \mathbf{A}_R \) accepts \( \varepsilon \) then \( \exists R.\text{Self} \in \text{Satr}(X), \)
- if \( \exists R.\text{Self} \in \text{Satr}(X) \) then \( \exists R.\text{Self} \in \text{Satr}(X), \)
- if \( \exists R.\text{Self} \in \text{Satr}(X) \) and \( R \sqsubseteq_R S \) then \( \exists S.\text{Self} \in \text{Satr}(X), \)
- if \( n.S.A \in \text{Satr}(X) \) where \( n \geq 1, \) then \( \exists S.A \in \text{Satr}(X), \)
- if \( \forall R.A \in \text{Satr}(X) \) and \( R \in \mathbf{R} \) then \( [\mathbf{A}_R]A \in \text{Satr}(X), \)
- if \( U.A \in \text{Satr}(X) \) then \( A \in \text{Satr}(X), \)
- if \( [A]C \in \text{Satr}(X) \) and \( q_A \in \mathbf{F}_A \) then \( C \in \text{Satr}(X), \)
- if \( \{ [A]C, \exists R.\text{Self} \} \subseteq \text{Satr}(X) \) and \( \langle q_A, R, q \rangle \in \delta_A \) then \( [A]C \in \text{Satr}(X), \)
- for every \( R \in \mathbf{R}, \) if \( \forall R.A \) occurs in \( \mathbf{T} \) for some \( A \) then \( [\mathbf{A}_R] \exists R.\top \in \text{Satr}(X), \)
- for every \( R \in \mathbf{R}, \) if \( A \in \text{Satr}(X) \) and \( \exists R.A \) occurs in the left hand side of \( \subseteq \) in some clause of \( \mathbf{T} \) then \( [\mathbf{A}_R] \langle A.R \rangle A \in \text{Satr}(X). \)

For \( E \in \text{EdgeLabels}, \) the transfer of \( X \) through \( E \) is defined to be
\[
\text{Trans}(X, E) = \{ [A]C \mid [A]C \in X, \langle q_A, R, q \rangle \in \delta_A, R \in \text{Roles}(E) \}. \]

4.3 A Natural Approach

How can we try to construct a model for \( \langle \mathbf{R}, \mathbf{T}, \mathbf{A} \rangle? \) Let us first describe a natural approach without the termination property. (In the next subsection we will revise the approach to form an algorithm with the termination property.) The intended model is based on a Horn-DL graph, which is initialized by Procedure \text{InitializeGraph} \) (on page 14), with \( \mathbf{A}' := \mathbf{A} \) and \( \mathbf{T}' := \mathbf{T}. \)

Each individual \( a \) occurring in \( \mathbf{A} \) or \( \mathbf{T} \) is a node of the initial graph, with the label consisting of \( \{ a \}, \) all the concepts \( A \) such that \( A(a) \in \mathbf{A}, \) and all the TBox axioms of \( \mathbf{T}. \) During the construction, for each node \( x \) of the graph, we treat \( \text{Label}(x) \) as a set of requirements to be realized at \( x. \) We realize a requirement \( \varphi \in \text{Label}(x) \) as follows:
- Case $\varphi = \exists R.C$: We connect $x$ to a new node $y$ with $\text{Label}(y) = \{C\} \cup T'$ using an edge labeled by $\{3R.C\}$ by setting $\text{Succ}(x, \{3R.C\}) := y$.
- Case $\varphi = \exists U.C$: If $\text{Next}_U(\exists U.C)$ was not defined then we create a new node $y$ with $\text{Label}(y) = \{C\} \cup T'$ and set $\text{Next}_U(\exists U.C) := y$.
- Case $\varphi = \forall R.A$: We include $[A_R]A$ in $\text{Label}(x)$ by saturating $\text{Label}(x)$. Since $\forall R.A$ is semantically equivalent to $[A_R]A$, we realize the requirement $[A_R]A$ (instead of $\forall R.A$) at $x$ as in the following case.
- Case $\varphi = [A]C$:
  - If $\text{Succ}(x, E) = y$ then we extend $\text{Label}(y)$ with $\text{Trans}(\text{Label}(x), E)$.
  - If $\text{Succ}(y, E) = x$ then we extend $\text{Label}(y)$ with $\text{Trans}(\text{Label}(x), E)$.
  - If $\langle x, R, y \rangle$ or $\langle y, \overline{R}, x \rangle$ belongs to $\text{NamedEdges}$ then we extend $\text{Label}(y)$ with $\text{Trans}(\text{Label}(x), \{R\})$.
  - If $\exists R.\text{Self} \in \text{Label}(x)$ then we extend $\text{Label}(x)$ with $\text{Trans}(\text{Label}(x), \{R\})$.
  - If $q_A \in F_A$ then we extend $\text{Label}(x)$ with $C$.

The last two operations are done by saturating $\text{Label}(x)$. Notice that some changes are done set-at-a-time, not depending on a concrete concept $[A]C$.
- Case $\varphi = \forall U.A$: We extend the label of every node of the graph with $A$.
- Case $\varphi = \{a\}$: If $a \neq x$ then we merge $a$ and $x$ in an appropriate way (if $x$ is also a named individual then we make appropriate changes for $\Delta_0$, $\top$, $A'$, $\text{Label}$, $\text{Succ}$, $\text{Next}_U$ and $\text{Replacement}$).
- Case $\varphi = \exists r.\text{Self}$: The requirement is implicitly realized when dealing with requirements of the form $[A]C$ and by using saturation.
- Case $\varphi = (\geq n S.A)$: Dismissing the constraint (15) just for this discussion, similarly to the case of $\exists R.C$, we connect $x$ to new nodes $y_1, \ldots, y_n$ with label $\{A\} \cup T'$ using edges labeled by $\{3S.A\}$.
- Case $\varphi = (\leq 1 S.A)$: If $y$ and $y'$ are different “$S$-neighbors” of $x$ whose labels contain $A$ then we “merge” them together. This will be discussed in more detail later. Note that the complex form of edge labels is due to realization of requirements of the form $\leq 1 S.A$, when edge labels are merged.
- Case $\varphi = (C \subseteq D)$: If $C$ “holds” at $x$ then we extend $\text{Label}(x)$ with $D$. Suppose $C = C_1 \cap \ldots \cap C_k$. This premise “holds” at $x$ if $C_i$ “holds” at $x$ for each $1 \leq i \leq k$. There are the following cases:
  - Case $C_i = A$; $C_i$ holds at $x$ if $A \in \text{Label}(x)$.
  - Case $C_i = \{a\}$; $C_i$ holds at $x$ if $x = a$.
  - Case $C_i = \exists s.\text{Self}$; $C_i$ holds at $x$ if either $x \notin \Delta_0$ and there exists $\exists R.\text{Self} \in \text{Label}(x)$ with $R \subseteq S$, or $x \in \Delta_0$ and $\langle x, s, x \rangle \in \text{NamedEdges}$.
  - Case $C_i = \forall \exists R.A$: $C_i$ “holds” at $x$ if both $\forall R.A$ and $\exists R.\top$ “hold” at $x$. If $\exists R.\top$ “holds” at $x$ by the evidence of a path connecting $x$ to a node $z$ with (forward or backward\(^9\)) edges labeled by $E_1, \ldots, E_k$ such that $S_j \in \text{Roles}(E_j)$ for each $1 \leq j \leq k$ and the word $S_1 \ldots S_k$ is accepted by $A_R$ then, since $[A_R]3R.\top$ is included in $\text{Label}(z)$ by saturation, we can expect that $\exists R.\top \in \text{Label}(x)$ and $\text{Next}(x, \exists R.\top)$ is defined. To check whether $C_i$ “holds” at $x$ we just check whether $\exists R.\top \in \text{Label}(x)$, $\text{Next}(x, \exists R.\top)$ is defined and $A \in \text{Label}(\text{Next}(x, \exists R.\top))$. The intuition is that, $y = \text{Next}(x, \exists R.\top)$ is the “least $R$-neighbor” of $x$, and if $A \in \text{Label}(y)$ then $A$ will occur in all “$R$-neighbors” of $x$. The technique of using $\text{Next}(x, \exists R.\top)$ comes from [24].
  - Case $C_i = \exists R.A$: If $\exists R.A$ “holds” at $x$ by the evidence of a path connecting $x$ to a node $z$ with (forward or backward) edges labeled by $E_1, \ldots, E_k$ such that $S_j \in \text{Roles}(E_j)$ for each $1 \leq j \leq k$ and the word $S_1 \ldots S_k$ is accepted by $A_R$ and $A \in \text{Label}(z)$ then, since $[A_R]3R.A$ is included in $\text{Label}(z)$ by saturation, we can expect that $(A_R)A \in \text{Label}(x)$. To check whether $C_i = \exists R.A$ “holds” at $x$, we just check whether $(A_R)A \in \text{Label}(x)$. (Semantically, $(A_R)A$ is equivalent to $\exists R.A$.)

---

\(^9\) If an edge is labeled by $E$ then its inverse is an edge labeled by $\overline{E}$. 
4.4 Global Caching

- Case $C_i = \forall U.A$: $C_i$ “holds” at $x$ if $A \in \text{Label}(\text{Next}_U(\exists U.T))$. The intuition is that $\text{Next}_U(\exists U.T)$ stands for the “logically smallest node”.
- Case $C_i = \exists U.A$: $C_i$ “holds” at $x$ if there exists a node whose label contains $A$.
  - Case $\varphi = \bot$: We set $\text{Status} := \text{unsat}$ and terminate the process with the conclusion that the considered knowledge base is unsatisfiable.
  - Case $\varphi = \neg A$: If $A \in \text{Label}(x)$ then we also set $\text{Status} := \text{unsat}$ and terminate the process.

4.4 Global Caching

**Procedure InitializeGraph**

1. $\text{Status} := \text{unknown}$, $\mathcal{R}' := \mathcal{R}$, $\mathcal{T}' := \mathcal{T}$, $A' := A$;
2. foreach simple role $S$ and $\text{Irr}(R) \in \mathcal{R}$ such that $S \subseteq \mathcal{R} R$ do add $\text{Irr}(S)$ into $\mathcal{R}'$;
3. foreach simple roles $S$ and $S'$ and $\text{Disj}(R, R') \in \mathcal{R}$ such that $S \subseteq \mathcal{R} R$ and $S' \subseteq \mathcal{R} R'$ do add $\text{Disj}(S, S')$ into $\mathcal{R}'$;
4. $\Delta_0 := \{ a \in I \mid a \text{ occurs in } A \text{ or } T \};$
5. if $\Delta_0 = \emptyset$ then $\Delta_0 := \{ i \}$;
6. $\Delta_1 := \Delta_0 \cup \{ u_i \}$, $\Delta_2 := \emptyset$, $\Delta := \Delta_1$;
7. foreach $a \in \Delta_0$ do $\text{Label}(a) := \text{Satr}([\{ a \}] \cup \{ A(a) \in A \} \cup T)$;
8. set $\text{Suc}$ and $\text{Next}_U$ to the empty mappings;
9. $\text{Next}_U(\exists U.T) := u_i$;
10. set Replacement to the identity function on $I$;

**Function Find($X$)**

1. if there exists $z \in \Delta \setminus \Delta_0$ with $\text{Label}(z) = X$ then return $z$
2. else add a new element $z$ to $\Delta$ with $\text{Label}(z) := X$ and return $z$;

**Procedure ExtendLabel($z, X$)**

1. if $X \subseteq \text{Label}(z)$ then return;
2. if $z \in \Delta_0$ then $\text{Label}(z) := \text{Satr}(\text{Label}(z) \cup X)$
3. else
4. $z_\ast := \text{Find}(\text{Satr}(\text{Label}(z) \cup X))$;
5. foreach $y$ and $E$ such that $\text{Suc}(y, E) = z$ do $\text{Suc}(y, E) := z_\ast$;
6. foreach $C$ such that $\text{Next}_U(\exists U.C) = z$ do $\text{Next}_U(\exists U.C) := z_\ast$;
7. if $z \in \Delta_1$ then $\Delta_1 := \Delta_1 \setminus \{ z \} \cup \{ z_\ast \}$;

**Function CheckPremise($x, C$)**

1. if $C = \top$ then return $\text{true}$ else let $C = C_1 \cap \ldots \cap C_k$
2. foreach $1 \leq i \leq k$ do
3. if $C_i = A$ and $A \notin \text{Label}(x)$ then return $\text{false}$;
4. if $C_i = \{ a \}$ and $x \neq a$ then return $\text{false}$;
5. if $C_i = \exists R.A$ and $\exists R.T \notin \text{Label}(x)$ or $\text{Next}(x, \exists R.T)$ is not defined or $A \notin \text{Label}(\text{Next}(x, \exists R.T))$ then return $\text{false}$;
6. if $C_i = \exists R.A$ and $\{ A\_i \} A \notin \text{Label}(x)$ then return $\text{false}$;
7. if $C_i = \forall U.A$ and $A \notin \text{Label}(\text{Next}_U(\exists U.T))$ then return $\text{false}$;
8. if $C_i = \exists U.A$ and there does not exist any $y \in \Delta$ reachable from $\Delta_1$ such that $A \in \text{Label}(y)$ then return $\text{false}$;
9. if $C_i = \exists s.Self$ then
10. if $x \notin \Delta_0$ and $\exists s.Self \notin \text{Label}(x)$ then return $\text{false}$;
11. if $x \in \Delta_0$ and $\{ x, s, x \} \notin \text{NamedEdges}$ then return $\text{false}$;
12. return $\text{true}$;

The graph constructed in the previous subsection may be infinite. What we want to check is whether at some step the attribute $\text{Status}$ receives value $\text{unsat}$. To guarantee the termination property we apply a global caching technique, which is similar to the one of [25–27, 30, 10].
Algorithm 1: checking satisfiability of a clausal Horn-DL knowledge base

Input: a clausal Horn-DL knowledge base \((R, T, A)\) and the RIA-automaton-specification \(A\) of \(R_0\).
Output: true if \((R, T, A)\) is satisfiable, or false otherwise.
Data structure: \((R', T', A', \Delta_0, \Delta_1, \Delta, \text{Label}, \text{Succ}_U, \text{Replacement}, \text{Status})\).

1. InitializeGraph;
2. while some rule in Tables 2 and 3 can change the graph do
   3. choose such a rule, where \((O)\) has the highest priority, and execute it; // any strategy can be used
   4. if Status = unsat then return false;
   5. return true;

(∀) \(x\) if \(r(a, b) \in A'\) then ExtendLabel(b, \(\text{Trans}(\text{Label}(a), \{r\})\), \(\text{ExtendLabel}(a, \text{Trans}(\text{Label}(b), \{r\}))\));
(∀) \(x\) if \(x\) is reachable from \(\Delta_1\) and \(\text{Succ}(x, E) = y\) then
   \(\text{Succ}(x, E) := \text{Find}(\text{Satr}(\text{Label}(y) \cup \text{Trans}(\text{Label}(x), E)))\);
(∀) \(x\) if \(x\) is reachable from \(\Delta_1\) and \(\text{Succ}(x, E) = y\) then ExtendLabel(x, \(\text{Trans}(\text{Label}(y), E)\));
(∀) \(x\) if \(x\) is reachable from \(\Delta_1\) then ExtendLabel(x, \(\{A \mid \forall U.A \in \text{Label}(y)\}\) for some \(y\) reachable from \(\Delta_1\});
(∃) \(x\) if \(x\) is reachable from \(\Delta_1\), \(\exists R.C \in \text{Label}(x)\), \(R \in R\) and \(\text{Next}(x, \exists R.C)\) is not defined
   then \(\text{Succ}(x, \{\exists R.C\}) := \text{Find}(\text{Satr}(|\{C\} \cup \text{Trans}(\text{Label}(x), \{R\})\} \cup T'))\);
(∃) \(x\) if \(x\) is reachable from \(\Delta_1\), \(\exists U.A \in \text{Label}(x)\) and \(\text{Next}_U(\exists U.A)\) is not defined
   then NextU(\(\exists U.A\)) := Find(Satr(|\{A\} \cup T'));
(0) \(x\) if \(x\) is reachable from \(\Delta_1\), \{a \in \text{Label}(x)\} and \(x \neq a\) then Merge(a, x);
(∃) \(x\) if \(x\) is reachable from \(\Delta_1\), \(C \subseteq D\) in \(\text{Label}(x)\) and \(\text{CheckPremise}(x, C)\) then ExtendLabel(x, \(\{D\}\));
(∃) \(x\) if \(x\) is reachable from \(\Delta_1\), \(a \in \text{Label}(x)\) or there exists \(A, \neg A \subseteq \text{Label}(x)\) or \(a \neq a\) in \(A'\) then Status := unsat;
(∃) \(x\) if \(\neg a, b) \in A'\) and \(a, s, b) \in \text{NamedEdges}\) then Status := unsat;
(∃) \(x\) if \(\text{Ir}(S) \in R'\) and \(\exists S.Self \in \text{Label}(x)\) then Status := unsat;
(∃) \(x\) if \(\text{Dis}(S, S') \in R'\), \{(a, S, b), (a, s', b)\} \subseteq \text{NamedEdges}\) then Status := unsat;

Table 2. Expansion rules for Horn-DL graphs - Part I

Roughly speaking, if two nodes that are not named individuals have the same label, then we merge them together. In other words, for every finite set \(X\) of concepts, the graph contains at most one node \(z\) such that \(\text{Label}(z) = X\). The function \(\text{Find}(X)\) (on page 14) returns such a node \(z\) if it exists, or creates such a node \(z\) otherwise. Global caching, however, causes some problems, which are discussed and solved below.

Suppose we want to extend the label of \(z\) in \(\Delta\) with a set \(X\) of concepts.

- Case \(z \in \Delta_0\) (i.e., \(z\) is a named individual occurring in \(A\)): as \(z\) is “fixed” by the ABox \(A\),
we have no choice but to extend \(\text{Label}(z)\) directly with \(X\) and the concepts resulting from saturation.

- Case \(z \notin \Delta_0\) and the requirements \(X\) are directly caused by \(z\) itself or its successors: if we directly extend the label of \(z\) (with \(X\)) then \(z\) may have the same label as another node not belonging to \(\Delta_0\) and global caching is not fulfilled. Hence, we simulate changing the label of \(z\) by using \(z_* = \text{Find}(\text{Satr}(\text{Label}(z) \cup X))\) for playing the role of \(z\). In particular: for each \(y\) and \(E\) such that \(\text{Succ}(y, E) = z\), we set \(\text{Succ}(y, E) := z*_4\); for each \(C\) such that \(\text{Next}_U(\exists U.C) = z\), we set \(\text{Next}_U(\exists U.C) := z*_4\); and if \(z \in \Delta_1\) then we replace \(z\) in \(\Delta_1\) by \(z_*\).

Extending the label of \(z\) for the above two cases is done by Procedure \(\text{ExtendLabel}(z, X)\) (on page 14). The third case is considered below.

Suppose \(\text{Succ}(x, E) = y\). Then, to realize the requirements at \(x\), the label of \(y\) should be extended with \(X = \text{Trans}(\text{Label}(x), E)\). How can we realize such an extension? There may exist another \(\text{Succ}(x', E') = y\) with \(x' \neq x\). That is, by global caching, we may use \(y\) as a successor for two different nodes \(x\) and \(x'\). If we directly modify the label of \(y\) to realize the requirements of \(x\), such a modification may affect \(x'\) unjustifiably. The solution is to delete the edge connecting \(x\) to \(y\) via \(E\) and reconnect \(x\) to \(y_* = \text{Find}(\text{Satr}(\text{Label}(y) \cup X))\) by setting \(\text{Succ}(x, E) := y_*\).

Another problem is that, when \(x \notin \Delta_0\), \(\text{Succ}(x, E) = x\) and \(s \in \text{Roles}(E)\) may not imply that \(\exists S.Self\) “holds” at \(x\). We may have \(\text{Succ}(x, E) = x\) just by global caching. As a solution,
for $x \not\in \Delta_0$, we explicitly keep in $\text{Label}(x)$ all concepts of the form $\exists S . \text{Self}$ that should “hold” at $x$. This is done by saturating $\text{Label}(x)$ and appropriate treatments when executing mergings forced by requirements of the form $\leq 1 \text{S.A}$.

When realizing a requirement of the form $\geq n \text{S.A}$ at a node $x$, as all the “S-successors” of $x$ created for the realization would be cached by using the same node, instead of $n$ “S-successors” we create only one “S-successor” of $x$ for that purpose. This is done by saturating $\text{Label}(x)$ with $\exists S . A$. The constraint (15) is thus resumed. For checking possible inconsistency caused by requirements of the form $\geq n \text{S.A}$, the rules $(\perp_6)$ and $(\perp_7)$ in Table 3 are used.

Table 3. Expansion rules for Horn-DL graphs - Part II

---

(Q1) if $(\leq 1 \text{S.A}) \in \text{Label}(x)$, $(\{x, R, y\}, \{x, R', y'\}) \subseteq \text{NamedEdges}, y \neq y'$, $R \subseteq R$, $R' \subseteq R$, $A \in \text{Label}(y)$ and $A \notin \text{Label}(y')$
then $\text{ExtendLabel}(y, \{y\})$; // this will merge $y'$ to $y$

(Q2) if $(\leq 1 \text{S.A}) \in \text{Label}(x), \{x, R, y\} \in \text{NamedEdges}$, $y' = \text{Succ}(x, E') \not\in \Delta_0$, $R' \in \text{Roles}(E')$, $R \subseteq R$, $R' \subseteq R$, $A \in \text{Label}(y)$ and $A \in \text{Label}(y')$
then $\text{Succ}(x, E') := y$, $\text{ExtendLabel}(y, \text{Label}(y'))$;

(Q3) if $(\leq 1 \text{S.A}) \in \text{Label}(x), \{x, R, y\} \in \text{NamedEdges}, x = \text{Succ}(y, E')$, $y' \notin \Delta_0$, $y'$ is reachable from $\Delta_1$, $R' \in \text{Roles}(E')$, $R \subseteq R$, $R' \subseteq R$, $A \in \text{Label}(y)$ and $A \in \text{Label}(y')$
then $\text{ExtendLabel}(y', \{y\})$; // this will simulate merging $y'$ to $y$

(Q4) if $x \notin \Delta_0$ is reachable from $\Delta_1$, $(\leq 1 \text{S.A}, \exists R' . \text{Self}) \subseteq \text{Label}(x)$, $y = \text{Succ}(x, E)$, $y \neq x$, $R \in \text{Roles}(E)$, $R \subseteq R$, $R' \subseteq R$, $A \in \text{Label}(x)$ and $A \notin \text{Label}(y)$
then let $X = \{\exists R'' . \text{Self} \mid R'' \in \text{Roles}(E)\}$,
if $\text{Label}(y) \cup X \subseteq \text{Label}(x)$ then $\text{Succ}(x, E') := x$
else $\text{ExtendLabel}(x, \text{Label}(y) \cup X)$;

(Q5) if $x \notin \Delta_0$, $(\leq 1 \text{S.A}, \exists R' . \text{Self}) \subseteq \text{Label}(x), y$ is reachable from $\Delta_1$, $x = \text{Succ}(y, E)$, $x \neq y$, $R \in \text{Roles}(E)$, $R \subseteq R$, $R' \subseteq R$, $A \in \text{Label}(x)$ and $A \notin \text{Label}(y)$
then let $X = \{\exists R'' . \text{Self} \mid R'' \in \text{Roles}(E)\}$,
if $\text{Label}(x) \cup X \subseteq \text{Label}(y)$ then $\text{Succ}(y, E') := y$
else $\text{ExtendLabel}(y, \text{Label}(x) \cup X)$;

(Q6) if $x$ is reachable from $\Delta_1$, $(\leq 1 \text{S.A}) \in \text{Label}(x), E \neq E'$, $\text{Succ}(x, E') = y$, $\text{Succ}(x, E')} = y'$, $R \in \text{Roles}(E)$, $R' \in \text{Roles}(E')$, $R \subseteq R$, $R' \subseteq R$, $A \in \text{Label}(y)$ and $A \in \text{Label}(y')$
then $\text{Succ}(x, E') := \text{Find}(\text{Satr}(\text{Label}(y) \cup \text{Label}(y')))$;

(Q7) if $x \notin \Delta_0$, $(\leq 1 \text{S.A}) \in \text{Label}(x), y$ is reachable from $\Delta_1$, $\text{Succ}(y, E) = x$, $\text{Succ}(x, E') = y'$, $R \in \text{Roles}(E)$, $R' \in \text{Roles}(E')$, $R \subseteq R$, $R' \subseteq R$, $A \in \text{Label}(x)$ and $A \notin \text{Label}(y)$
then if $\text{Label}(y') \not\subseteq \text{Label}(y)$ then $\text{ExtendLabel}(y, \text{Label}(y'))$
else let $E_2 = E \cup \{\text{Roles}(E'') | R'' \in \text{Roles}(E')\}$,
let $X = \text{Trans}(\text{Label}(y), E_2) \cap \{\exists S .B \mid B \in \text{Label}(y), B \text{ occurs in } T\}$,
if $E \not\subseteq E_2$ or $X \not\subseteq \text{Label}(x)$ then $\text{Succ}(y, E), \text{Succ}(y, E_2) := \text{Find}(\text{Satr}(\text{Label}(x) \cup X))$;

(Q8) if $x \in \Delta_0 \cup \Delta_2$, $(\leq 1 \text{S.A}) \in \text{Label}(x), x = \text{Succ}(y, E) = \text{Succ}(y', E')$, $y$ and $y'$ are reachable from $\Delta_1$, $R \in \text{Roles}(E)$, $R' \in \text{Roles}(E')$, $R \subseteq R$, $R' \subseteq R$, $A \in \text{Label}(y)$ and $A \in \text{Label}(y')$
then $\text{ExtendLabel}(y, \text{Label}(y'))$, $\text{ExtendLabel}(y', \text{Label}(y))$;

(Q9) if $x \in \Delta_0 \cup \Delta_2$, $(\leq 1 \text{S.A}) \in \text{Label}(x), x = \text{Succ}(y, E), R \in \text{Roles}(E)$, $R \subseteq R$, $y \notin \Delta_0 \cup \Delta_2$ and $A \in \text{Label}(y)$ then $\text{add } y \text{ into } \Delta_2$;

(\perp_6) if $(\geq n \text{S.A}) \in \text{Label}(x), n \geq 2, \text{Succ}(x, E) = y, \exists S . A \in E$ and $y \in \Delta_0 \cup \Delta_2$
then $\text{Status} := \text{unsat}$;

(\perp_7) if $(\geq n \text{S.A}, \leq 1 \text{S}' . A') \subseteq \text{Label}(x), n \geq 2, \text{Succ}(x, E) = y, A' \in \text{Label}(y), E$ contains $\exists S . A$ and some $R$ or $\exists R . B$ with $R \subseteq R$, $S'$ then $\text{Status} := \text{unsat}$.
Procedure \texttt{Merge}(x, a)

\textbf{Input:} \(x \in \Delta\) and \(a \in \Delta_0\) s.t. \(x\) is reachable from \(\Delta_1\), \(\{a\} \in \text{Label}(x)\) and \(x \neq a\).

1. \textbf{if} \(x \in \Delta_0\) \textbf{then} \(a_0 := x\) \textbf{else} \(a_0 := a\);
2. let \(a_1, \ldots, a_k\) be all the individuals \(\neq a_0\) s.t. \(\{a_i\} \in \text{Label}(x)\) for \(1 \leq i \leq k\);
3. \textbf{foreach} \(b\) such that \(\text{Replacement}(b) \in \{a_1, \ldots, a_k\}\) \textbf{do} \(\text{Replacement}(b) := a\);
4. \textbf{foreach} \(1 \leq i \leq k\) \textbf{do}

\begin{itemize}
  \item \(\text{Replacement}(a_i) := a_0\), \(\text{Label}(a_0) := \text{Satr}(\text{Label}(a_0) \cup \text{Label}(a_i))\);
  \item \(\text{foreach} y \text{ and } E \text{ such that } \text{Succ}(y, E) = a \text{ do } \text{Succ}(y, E) := a_0\);
  \item \(\text{foreach} C \text{ such that } \text{Next}_U(\exists U.C) = a_1 \text{ do } \text{Next}_U(\exists U.C) := a_0\);
\end{itemize}
5. 
6. \textbf{replace} all occurrences of \(a_1, \ldots, a_k\) in \(A'\) and \(T'\) by \(a_0\);
7. \textbf{foreach} \(x \in \Delta\) \textbf{do} \(\text{all occurrences of } a_1, \ldots, a_k \text{ in } \text{Label}(x) \text{ by } a_0\);
8. \textbf{CompactGraph};
9. delete the nodes \(a_1, \ldots, a_k\) from \(\Delta_0\), \(\Delta_1\), \(\Delta\) and delete the tuples of \(\text{Label}, \text{Succ}, \text{Next}_U\) that are directly involved with those nodes;
10. \textbf{if} \(x \notin \Delta_0\) \textbf{then} \(/ / \text{merge } x \text{ to } a_0\)
11. \(\text{Label}(a_0) := \text{Satr}(\text{Label}(a_0) \cup \text{Label}(x))\);
12. \textbf{foreach} \(y \text{ and } E \text{ such that } \text{Succ}(y, E) = x \text{ do } \text{Succ}(y, E) := a_0\);
13. \textbf{foreach} \(C \text{ such that } \text{Next}_U(\exists U.C) = x \text{ do } \text{Next}_U(\exists U.C) := a_0\);
14. \textbf{if} \(x \in \Delta_1\) \textbf{then} \(\Delta_1 := \Delta_1 \setminus \{x\}\);

Procedure \texttt{CompactGraph}

1. \textbf{foreach} \(x, x' \in \Delta \setminus \Delta_0\) reachable from \(\Delta_1\) such that \(\text{Label}(x) = \text{Label}(x')\) \textbf{do}

\begin{itemize}
  \item \(/ / \text{merge } x' \text{ to } x\)
  \item \(\text{foreach} y \text{ and } E \text{ such that } \text{Succ}(y, E) = x' \text{ do } \text{Succ}(y, E) := x\);
  \item \(\text{foreach} C \text{ such that } \text{Next}_U(\exists U.C) = x' \text{ do } \text{Next}_U(\exists U.C) := x\);
\end{itemize}
2. \textbf{if} \(x' \in \Delta_1\) \textbf{then} \textbf{add} \(x\) into \(\Delta_1\);
3. \textbf{delete} \(x'\) from \(\Delta\) and \(\Delta_1\) (if \(x' \in \Delta_1\)) and delete the tuples of \(\text{Label}, \text{Succ}, \text{Next}_U\) that are directly involved with \(x'\).

Reconsider realization of a requirement \(\leq 1.S.A\) at a node \(x\). As mentioned earlier, the approach is: if \(y\) and \(y'\) are different “S-neighbors” of \(x\) whose labels contain \(A\) then we “merge” them together. A simple case when \(\text{Succ}(x, E) = y, \text{Succ}(x, E') = y', E \neq E'\), \(R \in \text{Roles}(E), R' \in \text{Roles}(E'), R \subseteq R, R' \subseteq R, S, A \in \text{Label}(y)\) and \(A \in \text{Label}(y')\) is solved by undefining \(\text{Succ}(x, E)\) and \(\text{Succ}(x, E')\) and setting \(\text{Succ}(x, E \cup E') := \text{Find}(\text{Satr}(\text{Label}(y) \cup \text{Label}(y')))).\) This is done by the rule \((Q_6)\) in Table 3. The rules \((Q_1)-(Q_5), (Q_7), (Q_8)\) in Table 3 deal with the other cases. To see that the rules \((Q_1)-(Q_8)\) cover all of possible cases, observe that:

- that \(y\) and \(y'\) are “S-neighbors” of \(x\) is directly related to \texttt{NamedEdges}, \texttt{Self}, \texttt{Succ} or “inverse of \texttt{Succ}”;
- by definition, a condition \(\langle x, R, y \rangle \in \texttt{NamedEdges}\) holds in a number of situations, e.g., when \(x \in \Delta_0, y = x\) and \(\exists R.\texttt{Self} \in \texttt{Label}(x)\);
- the case when \(x \notin \Delta_0 \cup \Delta_2, (\leq 1.S.A) \in \texttt{Label}(x), x = \text{Succ}(y, E) = \text{Succ}(y', E'), R \in \texttt{Roles}(E), R' \in \texttt{Roles}(E'), R \subseteq R, R' \subseteq R, S, A \in \texttt{Label}(y)\) and \(A \in \texttt{Label}(y')\) does not require any treatment, because it occurs just as a side effect of global caching and \(x \notin \Delta_0 \cup \Delta_2\) means \(x\) is cloneable (which allows to use one copy of \(x\) for \(y\) and another one for \(y'\)).

Let us explain the special rule \((Q_7)\) (the other rules among \((Q_1)-(Q_8)\) deal with simpler cases in the usual way). Suppose \(x \notin \Delta_0, (\leq 1.S.A) \in \texttt{Label}(x), \text{Succ}(y, E) = x, \text{Succ}(x, E') = y', R \in \texttt{Roles}(E), R' \in \texttt{Roles}(E'), R \subseteq R, R' \subseteq R, S, A \in \texttt{Label}(y)\) and \(A \in \texttt{Label}(y')\). With global caching, we cannot merge nodes in the usual way. We cannot merge \(y\) and \(y'\) together nor simulate merging \(y\) to \(y'\) because \(y'\) may be used as a successor of other nodes than \(x\). What we can do is to keep \(y'\) (for the sake of the other nodes that have \(y'\) as a successor) and simulate merging \(y'\) to \(y\) (for the sake of \(y\)). The simulation is done as follows:
− If Label(y′) ⊈ Label(y) then we execute ExtendLabel(y, Label(y′)), which simulates changing y by using a replacement y′, and what further should be done is a matter of y′.

− Else, with \( E_2 = E \cup \{ R' | R' \in \text{Roles}(E') \} \) and \( X = \text{Trans}(\text{Label}(y), E_2) \cup \{ \forall S.B \mid B \in \text{Label}(y), B \text{ occurs in } T \} \), if \( E \subsetneq E_2 \) or \( X \subsetneq \text{Label}(x) \) then we undefine Succ(y, E) and set \( \text{Succ}(y, E) := \text{Find}(\text{Satr}(\text{Label}(x) \cup X)) \). This can be understood as follows: merging y′ to y would add the inverses of the roles of \( \text{Roles}(E') \) into E and add the concepts of X into Label(x) (concepts \( \forall S.B \), where \( B \in \text{Label}(y) \) and \( B \text{ occurs in } T \), can be added into Label(x) because y would be the only “S-neighbor” of x). But we do not want to modify Label(x), so such a change is only simulated.

− In the remaining case, we do nothing. The intuition is that when constructing a model for the considered knowledge base and following the paths going through y and x we will simply ignore the edge specified by Succ(x, E′) = y′. All the conditions of this case guarantee that ignoring the edge specified by Succ(x, E′) = y′ does not cause any problem (for checking premises of TBox axioms at x by using CheckPremise). In particular, \( \text{Next}(x, \exists S.T) \) may be y′, but the presence of all the concepts \( \forall S.B \) in Label(x) where \( B \in \text{Label}(y) \) and \( B \text{ occurs in } T \) makes sure that checking whether a condition \( \forall \exists S.B \) “holds” at x is done correctly.

The rule (Q\(_9\)) in Table 3 is used to update the set of “unnamed noncloneable nodes”: if \( x \in \Delta_0 \cup \Delta_2, (\leq 1.S.A) \in \text{Label}(x), x = \text{Succ}(y, E), R \in \text{Roles}(E), \overline{R} \sqsubseteq \forall S.R, y \notin \Delta_0 \cup \Delta_2 \) and \( A \in \text{Label}(y) \) then we add y into \( \Delta_2 \).

When realizing requirements that are nominals, we may have to replace a named individual by another one, which may cause some nodes in \( \Delta \setminus \Delta_0 \) to have the same label. In that case, we compact the graph by Procedure CompactGraph (on page 17).

Algorithm 1 (on page 15) is our formal algorithm for checking satisfiability of a clausal Horn-DL knowledge base. It initializes a Horn-DL graph and expands it by using the rules in Tables 2 and 3. It returns \( \text{false} \) (unsatisfiable) when the attribute Status receives value unsat. When the graph cannot be expanded anymore, the answer \( \text{true} \) (satisfiable) is returned.

When expanding the graph, we need to pay attention only to the nodes reachable from \( \Delta_1 \). The rules (\( \perp_1 \))–(\( \perp_5 \)) are used to detect inconsistency caused by ABox assertions and role assertions. To make these rules simple, we expand \( \mathcal{R} \) to \( \mathcal{R}' \) as in Procedure InitializeGraph and define the saturation operator Satr appropriately (for dealing with Self). Apart from the mentioned procedures/functions, our algorithm also uses Function CheckPremise (on page 14) and Procedure Merge (on page 17). Notice that all the expansion rules in Table 3 are devoted to number restrictions, and Procedures Merge and CompactGraph are devoted to nominals. In the case without number restrictions and nominals, our algorithm would be much simpler.

**Theorem 4.2.** Algorithm 1 correctly checks satisfiability of clausal Horn-DL knowledge bases and has \( \text{PTIME} \) data complexity.

This theorem follows from Lemmas A.1, A.2 and Corollary A.4, which are presented and proved in the appendix. The following corollary immediately follows from Theorem 4.2 and Proposition 2.5.

**Corollary 4.3.** The problem of checking satisfiability of Horn-DL knowledge bases has \( \text{PTIME} \) data complexity.

## 5 Illustrative Examples

This section contains examples illustrating Algorithm 1. They show our techniques of dealing with automaton-modal operators, inverse roles and the concept constructor \( \forall \exists R.C \). For simplicity, the examples are designed without the other additional features like number restrictions, nominals, the universal role and the concept constructor \( \exists R.\text{Self} \). For these examples, we also ignore the node \( v_0 = \text{Next}_U(\exists R.T) \) and the nominal \( \{ a \} \in \text{Label}(a) \) for each named individual a. This does not affect the output of Algorithm 1.
Example 5.1. Let $\mathbf{R}_t = \{r, s\}$, $\mathbf{C} = \{A, B, C, D, E\}$, $\mathbf{I} = \{a, b, c\}$,

$$\mathcal{R} = \{r \circ r \subseteq r, \ r \circ \neg r \subseteq s, \ \neg r \circ \neg r \subseteq s, \ r \circ \neg r \subseteq s\},$$

and let $\mathcal{T}$ be the TBox consisting of the following axioms:

\begin{align*}
A \sqsubseteq \exists r.C & \quad \text{(16)} \\
C \sqsubseteq \forall r . D & \quad \text{(17)} \\
D \sqsubseteq C & \quad \text{(18)} \\
A \sqcap \forall \exists r . C \sqsubseteq E & \quad \text{(19)} \\
A \sqcap \exists s . B \sqsubseteq E & \quad \text{(20)}
\end{align*}

Let the RIA-automaton-specification $\mathbf{A}$ of $\mathcal{R}$ be specified by:

\begin{align*}
\mathbf{A}_r &= (\mathbf{R}, \{0, 1\}, 0, \{(0, r, 1), \{r, 1\}, \{0\})} \\
\mathbf{A}_\tau &= (\mathbf{R}, \{0, 1\}, 1, \{(1, \tau, 0), \{1, \tau, 1\}, \{0\})} \\
\mathbf{A}_s &= (\mathbf{R}, \{0, 1, 2, 3\}, 0, \{(0, s, 1), \{0, r, 2\}, \{2, r, 2\}, \{2, \tau, 3\}, \{3, \tau, 3\}, \{1, 3\})} \\
\mathbf{A}_\pi &= (\mathbf{R}, \{0, 1, 2, 3\}, 0, \{(0, \pi, 1), \{0, r, 2\}, \{2, r, 2\}, \{2, \tau, 3\}, \{3, \tau, 3\}, \{1, 3\})}.
\end{align*}

Note that $\mathbf{A}_\pi = (\mathbf{A}_p)_0$, $\mathbf{A}_\tau = (\mathbf{A}_r)_1$, $\mathbf{A}_s = (\mathbf{A}_r)_2$ and $\mathbf{A}_\pi = (\mathbf{A}_\pi)_0$.

Consider the Horn-DL knowledge base $KB = (\mathcal{R}, \mathcal{T}, \mathbf{A})$ with

$$\mathbf{A} = \{A(b), \ B(b), \ A(c), \ \neg E(c), \ r(b, a), \ s(b, c)\}.$$

Figure 3 illustrates the Horn-DL graph constructed by Algorithm 1 for $KB$. The nodes of the graph are $a, b, c, v_1, v_1', v_2, v_2', v_3, v_4$, where $\Delta_0 = \{a, b, c\}$. In each node, we display the concepts of the label of the node. The main steps of the run of the algorithm are numbered from 0 to
17. In the table representing a node $x \in \{a, b, c\}$, the number in the left cell in a row denotes the step at which the concepts in the right cell were added to the label of the node. For a node not belonging to $\mathcal{A}_0 = \{a, b, c\}$, the number before the name of the node denotes the step at which the node was created. A label $n : \exists r. \varphi$ displayed for an edge from a node $x$ to a node $y$ means that $\text{Next}(x, \exists r. \varphi) = y$ and the edge was created at the step $n$. A label $n : \text{deleted}$ beside a dashed edge means that the edge was deleted at the step $n$.

The steps of the run of Algorithm 1 for $KB$ are as follows:

0: Initialization.
1: Applying $(\forall)$ to the nodes $x = b$ and $y = a$.
2: Applying $(\forall)$ to the nodes $x = b$ and $y = a$.
3: Applying $(\exists)$ to the node $x = b$ using the clause (16).
4: Applying $(\exists)$ to the node $x = b$ using the clause (20).
5: Applying $(\exists)$ to the node $x = c$ using the clause (16).
6: Applying $(\exists)$ to the node $x = b$ and the concept $\exists r. C$.
7: Applying $(\exists)$ to the node $x = v_1$ using the clause (17).
8: Applying $(\forall)$ to the nodes $x = b$ and $y = v_1'$.
9: Applying $(\exists)$ to the node $x = b$ using the clause (18).
10: Applying $(\exists)$ to the node $x = b$ using the clause (17).
11: Applying $(\exists)$ to the node $x = c$ and the concept $\exists r. C$.
12: Applying $(\exists)$ to the node $x = v_2$ using the clause (17).
13: Applying $(\forall)$ to the nodes $x = c$ and $y = v_2'$.
14: Applying $(\exists)$ to the node $x = c$ using the clause (18).
15: Applying $(\exists)$ to the node $x = c$ using the clause (17).
16: Applying $(\exists)$ to the node $x = b$ and the concept $\exists r. T$.
17: Applying $(\exists)$ to the node $x = c$ and the concept $\exists r. T$.

As no more expansion rules are applicable to the graph and the Status of the graph is different from unsat (none of the rules $(\bot_1)$-$(\bot_7)$ was applicable), by Theorem 4.2, the knowledge base $KB$ is satisfiable. \hfill \triangleleft

**Example 5.2.** This example illustrates global caching. Let $R_x$, $C$, $I$, $\mathcal{R}$, $\mathcal{T}$, $A$ be as in Example 5.1. Consider the Horn-DL knowledge base $KB = \langle \mathcal{R}, \mathcal{T}, \mathcal{A}_2 \rangle$ with

$$\mathcal{A}_2 = \{ A(a), B(a), A(b), r(a, b) \}.$$ 

Figure 4 illustrates the Horn-DL graph constructed by Algorithm 1 for $KB_2$ in a similar way as Figure 3 does for $KB$. The steps of the run of Algorithm 1 for $KB_2$ are as follows:

0: Initialization.
1: Applying $(\forall)$ to the nodes $x = a$ and $y = b$.
2: Applying $(\forall)$ to the nodes $x = a$ and $y = b$.
3: Applying $(\exists)$ to the node $x = a$ using the clause (16).
4: Applying $(\exists)$ to the node $x = a$ using the clause (20).
5: Applying $(\exists)$ to the node $x = b$ using the clause (16).
6: Applying $(\exists)$ to the node $x = a$ and the concept $\exists r. C$.
7: Applying $(\exists)$ to the node $x = b$ and the concept $\exists r. C$.
8: Applying $(\exists)$ to the node $x = u$ using the clause (17).
9: Applying $(\forall)$ to the nodes $x = a$ and $y = u'$.
10: Applying $(\exists)$ to the node $x = a$ using the clause (18).
11: Applying $(\exists)$ to the node $x = a$ using the clause (17).
12: Applying $(\forall)$ to the nodes $x = b$ and $y = u'$.
13: Applying $(\exists)$ to the node $x = b$ using the clause (18).
14: Applying $(\exists)$ to the node $x = b$ using the clause (17).
15: Applying (∪) to the node \( x = b \) using the clause (20).
16: Applying (∃) to the node \( x = a \) and the concept \( ∃r.T \).
17: Applying (∃) to the node \( x = b \) and the concept \( ∃r.T \).

As no more expansion rules are applicable to the graph and the Status of the graph is different from unsat (none of the rules \( ⊑ \)-(⊥) was applicable), by Theorem 4.2, the knowledge base \( KB_2 \) is satisfiable.

\[Example 5.3.\] Let us continue Example 5.2. It can be seen from Figure 4 that \( KB_2 \models E(b) \). To formally show this, let \( A'_2 = A_2 \cup \{¬E(b)\} \) and consider the Horn-DL knowledge base \( KB'_2 = \langle R, T, A'_2 \rangle \). There exists a short run of Algorithm 1 for \( KB'_2 \) as demonstrated in Figure 5. The steps are as follows:

0: Initialization.
1: Applying (∨) to the nodes \( x = a \) and \( y = b \).
2: Applying (∪) to the node \( x = b \) using the clause (16).
3: Applying (∃) to the node \( x = b \) and the concept \( ∃r.C \).
4: Applying (∨) to the nodes \( x = b \) and \( y = u \).
5: Applying (∃) to the node \( x = b \) using the clause (20).
6: Applying (⊥) to the node \( x = b \) to establish Status = unsat.

Thus, by Theorem 4.2, the knowledge base \( KB'_2 \) is unsatisfiable, and hence \( KB_2 \models E(b) \).

\[6 Conclusions and Discussion\]

The rule language Horn-DL developed by us for the Semantic Web has PTime data complexity and, as discussed in Sections 1.3 and 3, is strictly and essentially richer than Horn-\( R\)eg\(^I\) [29] and...
the well-known rule languages Horn-\texttt{SHIQ} [15] and Horn-\texttt{SROIQ} [31]. We have developed the first algorithm with \textsc{PTime} data complexity for checking satisfiability of Horn-DL knowledge bases.

Our algorithm has been implemented in C++ by a student [16]. The implemented program receives as input an Horn-DL knowledge base $KB$ that need not to be in the clausal form. It automatically translates $KB$ to the clausal form. The program can check whether $KB$ is satisfiable and answer queries of the form $C(a)$ (i.e., whether $a$ is an instance of $C$ w.r.t. $KB$), where $C$ is an LHS concept. Possible applications of Horn-DL are similar to applications of the profiles OWL 2 EL, OWL 2 QL and OWL 2 RL of OWL 2. An ontology in Horn-DL, possibly using an OWL-like syntax, can be used for the Semantic Web in the usual way as a knowledge base. Having \textsc{PTime} data complexity, Horn-DL is a query language whose processing is scalable w.r.t. the set of individual assertions (i.e. data). One can implement our algorithm and its naive extension for answering conjunctive queries as Java API and incorporate them into the Jena framework. The naive extension just systematically instantiates variables in a conjunctive query by individual names. A direct implementation like [16] may be a bit time-consuming, but should not be a hard task. A real problem is to develop an efficient algorithm for processing conjunctive queries in Horn-DL. This remains as an open problem.\footnote{A conjunctive query $\varphi$ to a knowledge base $KB$ is a conjunction of formulas of the form $C(x)$ or $R(x,y)$, where $x$ and $y$ are variables for individual names. An answer for the query is a substitution $\theta$ that maps each variable to an individual name such that $KB \models \varphi\theta$.}

Acknowledgments

This work was supported by the Polish National Science Centre (NCN) under Grants No. 2011/01/B/ST6/02759 and 2011/01/B/ST6/02769.

References


\footnote{With respect to efficiency, our algorithm for Horn-DL is like a naive evaluation method of Datalog. It is expected to develop an advanced algorithm for Horn-DL like the one for Datalog that combines the magic-sets transformation with the improved semi-naive evaluation method.}
A Proofs

Define $\text{closure}_A(T')$ to be the smallest set of concepts and TBox axioms such that:

- TBox axioms of $T'$ belong to $\text{closure}_A(T')$,
- concepts and subconcepts occurring in $T'$ belong to $\text{closure}_A(T')$,
- if $R \in \mathcal{R}$ then $\{\top, \exists R. \text{Self}, [A_R] \exists R. T\} \subseteq \text{closure}_A(T')$,
- if $[R]C \in \text{closure}_A(T')$ then $[A_R]C \in \text{closure}_A(T')$,
- if $A \in \text{closure}_A(T')$ and $R \in \mathcal{R}$ then $[A_R]A \in \text{closure}_A(T')$,
- if $[A]C \in \text{closure}_A(T')$ and $q \in Q_A$ then $[A_q]C \in \text{closure}_A(T')$,
- if $\geq n.S.A \in \text{closure}_A(T')$ then $\exists S.A \in \text{closure}_A(T')$,
- if $S$ is a simple role and $A \in \text{closure}_A(T')$ then $\forall S.A \in \text{closure}_A(T')$.

**Lemma A.1.** Algorithm 1 runs in polynomial time in the size of $A$ (when assuming that $\mathcal{R}$ and $T$ are fixed).

*Proof.* Let $n$ be the size of $A$. Since $\mathcal{R}$ and $T$ are fixed, the size of $\text{closure}_A(T')$ is bounded by a constant. Observe that, at any step during an execution of Algorithm 1, for $x \in \Delta \setminus \Delta_0$, $\text{Label}(x) \subseteq \text{closure}_A(T')$, and for $a \in \Delta_0$, $\text{Label}(a) \setminus \{C \mid C(a) \in A\} \subseteq \text{closure}_A(T')$. Hence the sizes of these two sets are also bounded by a constant. Since each $x \in \Delta \setminus \Delta_0$ has a unique $\text{Label}(x) \subseteq \text{closure}_A(T')$, the set $\Delta \setminus \Delta_0$ contains only $O(1)$ elements. Hence, the size of $\Delta$ is of rank $O(n)$. If $\text{Succ}(x, E) = y$ then $E$ contains at least one concept of the form $\exists R.A$ or $\exists R.T$, which depends only on $T$, but not $A$. By (15), for each $x \in \Delta$, the numbers of $E \in \text{EdgeLabels}$ such that $\text{Succ}(x, E)$ is defined is bounded by a constant. Consequently, it can be seen that each iteration of the main loop of Algorithm 1 runs in polynomial time in $n$.

Observe that each iteration of the main loop of Algorithm 1:

- either extends the label of a node $x \in \Delta_0$,
- or adds a new node to $\Delta$,
- or creates a new “edge”,
- or combines edges to a “bigger” one,
- or merges named individuals and hence reduces the number of named individuals occurring
  in the constructed graph.

Due to the mentioned “bounds” on labels of nodes, the number of nodes in $\Delta \setminus \Delta_0$, and “edges”,
it follows that the number of iterations the main loop of Algorithm 1 executes is bounded by a polynomial in $n$. Therefore, Algorithm 1 runs in polynomial time in $n$. $\blacksquare$

**Lemma A.2.** If Algorithm 1 returns true then the knowledge base $\langle \mathcal{R}, T, A \rangle$ is satisfiable.

*Proof.* Suppose Algorithm 1 returns true for $\langle \mathcal{R}, T, A \rangle$. We will refer to the data structures used by that run of Algorithm 1. A model for $\langle \mathcal{R}, T, A \rangle$ will be constructed by starting from $\Delta_1$, then “unfolding” the remaining part of the graph constructed by Algorithm 1, and then completing the interpretation of basic roles and named individuals. For that we define a counterpart $\Delta'$ of $\Delta$, a counterpart $\text{Next}'$ of $\text{Next}$, a set $\text{Edges} \subseteq \Delta' \times \mathcal{R} \times \Delta'$, a mapping $f : \Delta' \rightarrow \Delta$ and a queue $\text{unresolved}$ of elements of $\Delta'$ in Algorithm 2.

Algorithm 2 does not have the termination property. What we want is that the “results” of Algorithm 2 are well defined although they may be infinite; and we have it.

Let $I$ be the interpretation with $\Delta^I = \Delta'$, specified by:

- for each $a \in I$, if $a \in \Delta_0$ then $a^I = a$ else $a^I = \text{Replacement}(a)$;
- for each $A \in C$, $A^I = \{u \in \Delta' \mid A \in \text{Label}(f(u))\}$;
- for all $R \in \mathcal{R}$, $R^I$ are the least relations satisfying the following conditions:
  - if $\langle u, R, v \rangle \in \text{Edges}$ then $\langle u, v \rangle \in R^I$,
  - for every word $S_1 \ldots S_k$ accepted by $A_R$, $S_1^I \circ \cdots \circ S_k^I \subseteq R^I$. 


Algorithm 2: unfolding a Horn-DL graph for the proof of Lemma A.2

```plaintext
Δ' := Δ₁, Edges := NamedEdges, unresolved := Δ;
foreach x ∈ Δ₁ do f(x) := x;
while unresolved is not empty do
  extract an element u from unresolved;
  foreach E and y such that Succ(f(u), E) = y do
    if f(u) ∈ Δ₀ and there exist (≤ 1 S.A) ∈ Label(f(u)), ⟨u', R', u⟩ ∈ Edges, R ∈ Roles(E) such that
      R' ⊆ S and R ⊌ S then
        foreach ∃R', C ∈ Roles(E) do Next'(u, ∃R', C) := u';
    else if y ∈ Δ₁ then
      if y /∈ Δ' then add y into Δ' and unresolved and set f(y) := y;
      foreach R ∈ Roles(E) do add ⟨u, R, y⟩ into Edges;
      foreach ∃R.C ∈ Roles(E) do Next'(u, ∃R.C) := y;
    else
      if there exists (≥ n S.A) ∈ Label(f(u)) such that n ≥ 2 and ∃S.A ∈ E then
        let m be the largest among such numbers n;
        else let m = 1;
        add new elements v₁,...,vₘ into Δ' and unresolved;
      foreach 1 ≤ i ≤ m do
        f(vᵢ) := y;
      foreach R ∈ Roles(E) do add ⟨u, R, vᵢ⟩ into Edges;
      foreach ∃R.C ∈ Roles(E) do Next'(u, ∃R.C) := v₁;
  foreach ∃R.Self ∈ Label(f(u)) do add ⟨u, R, u⟩ into Edges;
```

Note that the steps 6 and 7 of Algorithm 2 correspond to a case discussed earlier of the rule (Q₁) when it does nothing. The steps 8–11 are related to the rule (Q₀). The partial mapping Next' has the property that: if ∃R.C ∈ Label(f(u)) then Next'(u, ∃R.C) is defined to be a v such that ⟨u, v⟩ ∈ R² and C ∈ Label(f(v)).

We show that I is a model of (R, T, A). As Procedure Merge is defined appropriately, it suffices to prove that, for every u ∈ Δ' and every ϕ ∈ Label(f(u)), u ∈ ϕᵀ. We prove this by induction on the structure of ϕ. Let u ∈ Δ' and suppose ϕ ∈ Label(f(u)).

- Case ϕ = A is trivial.
- Case ϕ = {a}: By the rule (O), we have that u ∈ Δ₀ and u = a. Hence u ∈ ϕᵀ.
- Case ϕ = ∃R.Self: By the step 21 of Algorithm 2 and the definition of I, ⟨u, u⟩ ∈ R². Hence u ∈ ϕᵀ.
- Case ϕ = (≤ 1 S.A): By the rules (Q₁)–(Q₉) and the steps 6–11 of Algorithm 2, we have that u ∈ ϕᵀ.
- Case ϕ = (≥ n S.A): Since ϕ ∈ Label(f(u)) and Label(f(u)) is saturated, we also have that ∃S.A ∈ Label(f(u)). Hence, there exists E ∈ EdgeLabels such that ∃S.A ∈ E and Succ(f(u), E) is defined. By the steps 13–20 of Algorithm 2, it follows that u ∈ ϕᵀ.
- Case ϕ = ∃U.A: For x = Nextᵦ(∃U.A), we have that x ∈ Δ₁ ⊆ Δ' and A ∈ Label(x). Hence u ∈ ϕᵀ.
- Case ϕ = ∀U.A: By the rule (∀U), we have that u ∈ ϕᵀ.
- Case ϕ = ∃R.C, where C ∈ C ∪ { ⊤ }: Since ϕ ∈ Label(f(u)), with v = Next'(u, ∃R.C), we have that ⟨u, v⟩ ∈ R² and C ∈ Label(f(v)). Hence u ∈ ϕᵀ.
- Case ϕ = ∀R.A: Let v be any element of Δᵀ such that ⟨u, v⟩ ∈ R². We show that v ∈ Aᵀ. Since ⟨u, v⟩ ∈ R², there exist a word S₁...Sₖ accepted by Aᵦ and elements u₀ = u, u₁,...,uₖ₋₁, uₖ = v such that, for every 1 ≤ i ≤ k, ⟨uᵢ₋₁, uᵢ⟩ ∈ Sᵦ, and ⟨uᵢ₋₁, Sᵦ, uᵢ⟩ ∈ Edges or ⟨uᵢ−1, Sᵦ, uᵢ−1⟩ ∈ Edges. Let A = Aᵦ. Since S₁...Sₖ is accepted by A, there exist states q₀ = q₀, q₁,...,qₖ such that qₖ ∈ FA and ⟨qᵢ₋₁, Sᵦ, qᵢ⟩ ∈ δᵤ for every 1 ≤ i ≤ k. Since ϕ ∈ Label(f(u)) and ϕ = ∀R.A, by saturation, we have that [Aᵦ]A ∈ Label(f(u)), which means [A]A ∈ Label(f(u)) and [Aᵦ,qᵦ]C ∈ Label(f(u₀)). For each i from 1 to k, since
\[ \langle u_{i-1}, S_i, u_i \rangle \in Edges \text{ or } \langle u_i, S_i, u_{i-1} \rangle \in Edges, \text{ it follows that } [A_q]A \in \text{Label}(f(u_i)) \text{.} \] Since \( q_k \in F_A \) and \( u_k = v \), it follows that \( A \in \text{Label}(f(v)) \). Hence, by the inductive assumption, \( v \in A^T \).

- Case \( \varphi = (C \subseteq D) \) and \( C = C_1 \cap \ldots \cap C_k \): Suppose \( u \in C^T \). We prove that \( u \in D^T \). The last call \texttt{CheckPremise}(\( f(u), C \)) returned \texttt{true} because the following observations hold for every \( 1 \leq i \leq k \):
  \begin{itemize}
    \item Case \( C_i = A \): Since \( u \in C_i^T \), we have that \( A \in \text{Label}(f(u)) \).
    \item Case \( C_i = \{a\} \): Since \( u \in C_i^T \), we have that \( u = a \). Hence, \( u \in \Delta_0 \) and \( f(u) = u = a \).
    \item Case \( C_i = \exists S.f \ell \): Since \( u \in C_i^T \), \( \langle u, u \rangle \in s^T \). Observe that, due to the saturation operator and the definition of \texttt{NamedEdges}, \( \langle u, u \rangle \in s^T \) happens only when:
      \begin{itemize}
        \item \( u \notin \Delta_0 \) and \( \exists S.f \ell \in \text{Label}(f(u)) \), or
        \item \( u \in \Delta_0 \) and \( (u, s, u) \in \texttt{NamedEdges} \), which imply that \( f(u) = u \) and \( \langle f(u), s, f(u) \rangle \in \texttt{NamedEdges} \).
      \end{itemize}
    \item Case \( C_i = \exists U.A \): Since \( u \in C_i^T \), there exists \( v \in \Delta' \) such that \( A \in \text{Label}(f(v)) \). Clearly, \( y = f(v) \) is reachable from \( \Delta_1 \).
    \item Case \( C_i = \forall U.A \): For \( x = \text{Next}_y(\exists U.\top) \), we have that \( x \in \Delta \subseteq \Delta' \) and \( f(x) = x \). Since \( u \in C_i^T \), we have that \( x \in A^T \), hence \( A \in \text{Label}(f(x)) \), which means \( A \in \text{Label}(\text{Next}(\exists U.\top)) \).
  \end{itemize}

- Case \( C_i = \exists R.A \): Since \( u \in C_i^T \), there exist a word \( S_1 \ldots S_k \) accepted by \( A_R \) and elements \( u_0 = u, u_1, \ldots, u_{k-1}, u_k \) such that \( u_0 \in A^T \) and, for every \( 1 \leq i \leq k \), \( \langle u_{i-1}, S_i, u_i \rangle \in Edges \) or \( \langle u_i, S_i, u_{i-1} \rangle \in Edges \). Let \( A = A_R \). Since \( S_{k-1} \ldots S_1 \) is accepted by \( A \), there exist states \( q_k = q_{k-1}, \ldots, q_0 \) such that \( q_0 \in F_A \) and \( \langle q_i, S_i, u_{i-1} \rangle \in \Delta \) for every \( k \geq i \geq 1 \). Since \( u_k \in A^T \), we have that \( A \in \text{Label}(f(u_k)) \), and, by saturation, \( [A_R](A_R)A \in \text{Label}(f(u_k)) \), which means \( [A_{q_k}](A_R)A \in \text{Label}(f(u_k)) \). For each \( i \) from \( k \) to \( 1 \), since \( \langle u_{i-1}, S_i, u_i \rangle \in Edges \) or \( \langle u_i, S_i, u_{i-1} \rangle \in Edges \), it follows that \( [A_{q_{i-1}}](\exists R.\top) \in \text{Label}(f(u_{i-1})) \). Since \( q_0 \in F_A \) and \( u_0 = u \), it follows that \( (\exists R.\top)A \in \text{Label}(f(u)) \).
- Case \( Case(i = 1) \): Since \( u \in C_i^T \), we have that \( u \in (\forall R.A)^T \) and \( u \in (\exists R.A)^T \). Thus, there exist a word \( S_1 \ldots S_k \) accepted by \( A_R \) and elements \( u_0 = u, u_1, \ldots, u_{k-1}, u_k \) such that, for every \( 1 \leq i \leq k \), \( \langle u_{i-1}, u_i \rangle \in S_i^T \), and \( \langle u_i, S_i, u_{i-1} \rangle \in Edges \). Let \( A = A_R \). Since \( S_{k-1} \ldots S_1 \) is accepted by \( A \), there exist states \( q_k = q_{k-1}, \ldots, q_0 \) such that \( q_0 \in F_A \) and \( \langle q_i, S_i, u_{i-1} \rangle \in \Delta \) for every \( k \geq i \geq 1 \). By saturation, \( [A_R](\exists R.\top) \in \text{Label}(f(u_k)) \), which means \( [A_{q_k}](\exists R.\top) \in \text{Label}(f(u_k)) \). For each \( i \) from \( k \) to \( 1 \), since \( \langle u_{i-1}, S_i, u_i \rangle \in Edges \) or \( \langle u_i, S_i, u_{i-1} \rangle \in Edges \), it follows that \( [A_{q_{i-1}}](\exists R.\top) \in \text{Label}(f(u_{i-1})) \). Since \( q_0 \in F_A \) and \( u_0 = u \), it follows that \( (\exists R.A) \in \text{Label}(f(u)) \).
- Case \( \varphi = (\forall R.A)^T \): Clearly, we have that \( A \in \text{Label}(\text{Next}(f(u), (\exists R.\top))) \).
- Case \( \varphi = y \neq y' \): This involves with the rule \( (Q \tau) \) and the steps 6 and 7 of Algorithm 2. There must exist \( (\leq 1 S.A') \in \text{Label}(f(u)) \) such that \( S \subseteq R \) and \( \forall S.B \in \text{Label}(f(u)) \) for all \( B \in \text{Label}(f(u)) \). In particular, we have \( \forall S.A \in \text{Label}(f(u)) \). Consequently, by the saturation operator and the rules \( (\forall) \) and \( (\exists) \), we have \( A \in \text{Label}(\text{Next}(f(u), (\exists R.\top))) \).

We have shown that \texttt{CheckPremise}(\( f(u), C \)) returned \texttt{true}. It follows that \( D \in \text{Label}(f(u)) \), and by the inductive assumption, \( u \in D^T \).

If \( \mathcal{I} \) is an interpretation, \( C \subseteq D \) is a TBox axiom, and \( X \) is a set of concepts and TBox axioms then define \((C \subseteq D)^\mathcal{I} = (\neg C \cup D)^\mathcal{I} \) and \( X^\mathcal{I} = \bigcap \{ \varphi^\mathcal{I} \mid \varphi \in X \} \).

**Lemma A.3.** Let \( KB = (\mathcal{R}, \mathcal{T}, \cdot, A) \) be a clausal Horn-DL knowledge base. Suppose \( KB \) is satisfiable and \( \mathcal{I} \) is a model of \( KB \). Consider an execution of Algorithm 1 for \( KB \) and a moment either before the main loop (i.e., after the initialization) or after an iteration of the main loop of Algorithm 1. Let \( r = \{ \langle x, u \rangle \in \Delta \times \Delta^T \mid u \in (\text{Label}(x))^T \} \). Then:
1. $\mathcal{I}$ is a model of $\langle R', T', A' \rangle$,
2. for every $a \in \Delta_0$, $\langle a, a^I \rangle \in r$,
3. for every $a, b \in \mathcal{I}$ such that $\text{Replacement}(a) = b$, we have $a^I = b^I$,
4. for every $x \in \Delta_0 \cup \Delta_2$, there exists exactly one $u \in \Delta^I$ such that $\langle x, u \rangle \in r$,
5. for every $u \in \Delta^I$, $\langle \text{Next}_U(\exists U. \top), u \rangle \in r$,
6. for every $A \in C$ and $u \in A^I$, $\langle \text{Next}_U(\exists U. A), u \rangle \in r$,
7. for every $x, y \in \Delta$, $u, v \in \Delta^I$, $E \in \text{EdgeLabels}$ and $\exists R. C \in E$ such that $\text{Succ}(x, E) = y$, if $\langle x, u \rangle \in r$, $\langle u, v \rangle \in R^I$ and $v \in C^I$, then $\langle y, v \rangle \in r$ and, for every $S \in \text{Roles}(E)$, $\langle u, v \rangle \in S^I$,
8. for every $x \in \Delta$, there exists $u \in \Delta^I$ such that $\langle x, u \rangle \in r$.

This lemma can be proved by induction on the number of executed steps in a straightforward way. The following corollary is a consequence of the last assertion of this lemma.

**Corollary A.4.** If $KB = \langle R, T, A \rangle$ is a satisfiable clausal Horn-DL knowledge base then Algorithm 1 returns true for it.