# Query Answering in the Description Logic Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$ * 

Thomas Eiter ${ }^{1}$, Georg Gottlob ${ }^{2}$, Magdalena Ortiz $^{1}$, and Mantas Šimkus ${ }^{1}$<br>${ }^{1}$ Institute of Information Systems, TU Vienna, Austria<br>(eiter|ortiz|simkus)@kr.tuwien.ac.at<br>${ }^{2}$ Computing Laboratory, Oxford University, United Kingdom<br>georg.gottlob@comlab.ox.ac.uk


#### Abstract

We provide an ExpTime algorithm for answering conjunctive queries (CQs) in Horn-SHIQ, a Horn fragment of the well-known Description Logic $\mathcal{S H} \mathcal{I} \mathcal{Q}$ underlying the OWL-Lite standard. The algorithm employs a domino system for model representation, which is constructed via a worst-case optimal tableau algorithm for Horn- $\mathcal{S H I Q}$; the queries are answered by reasoning over the domino system. Our algorithm not only shows that CQ answering in Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$ is not harder than satisfiability testing, but also that it is polynomial in data complexity, making Horn- $\mathcal{S H} \mathcal{H} \mathcal{Q}$ an attractive expressive Description Logic.


## 1 Introduction

Driven by the development of semantically enhanced systems, as in the context of the Semantic Web and of Enterprise Application Integration, query answering in Description Logics (DLs) has emerged as an important topic. A variety of algorithms have been proposed for this problem and, aiming at different applications, aspects like combined and data complexity have been guiding their development. The former characterizes the cost of query answering in the general case, while the latter in the case when the query and the knowledge base except the factual part are fixed. Data complexity is especially important for applications in which DLs are used to formalize rich data models for data repositories, as in such context the model is static as compared to the data contents, and typical user queries are known. For querying DLs, conjunctive queries (CQs) have been most widely considered and three major settings have been addressed:

- very expressive DLs for which standard reasoning tasks, like satisfiability testing or instance checking, are intractable both in data and combined complexity. As query answering is at least as hard, the problem is trivially intractable in
 plexity [13] and coNP-complete data complexity [6].
- tailored DLs, like DL-Lite [5], which aim at lower complexity at the price of limited expressiveness. In DL-Lite, CQ answering is coNP-complete in combined complexity, but polynomial if the query is fixed, and has very low data complexity (reducible to FOL, thus inside logarithmic space).

[^0]- weak DLs, like $\mathcal{E L}$ [1], for which standard reasoning and CQ answering under data complexity are P-complete, while CQ answering under combined complexity is intractable [17]. Several extensions of $\mathcal{E} \mathcal{L}$ sharing this property (e.g., $\mathcal{E L H}$, $\left.\mathcal{E} \mathcal{L} \mathcal{I}^{f}, \mathcal{E} \mathcal{L}^{+}, \mathcal{E} \mathcal{L}^{++}\right)$can be found in $[17,12,10]$.

Since CQ answering is intractable under combined complexity already over very simple knowledge bases, fragments of expressive DLs with tractable data complexity are of particular interest. Ideally, such fragments should also allow for CQ answering with combined complexity not higher than of standard reasoning.
 troduced in [9] as a Horn fragment of $\mathcal{S H \mathcal { L }}$, in which the syntax is restricted in a way that disjunction is not expressible. While standard reasoning in Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$ is ExpTime-complete in general [11], it is polynomial if the taxonomy is fixed [9].

Our main contributions and results are briefly summarized as follows:

- We provide an ExpTime algorithm for answering CQs in Horn-SHIQ. The algorithm is based on answering tree-shaped queries over particular trees (that capture a universal model of the KB) finitely represented by domino systems. The latter are relatives of saturated mosaic sets known in other branches of logic, and the recent knot sets [16] and domino sets [18] in DLs.
- For constructing domino systems, we exploit a dedicated tableaux-based algorithm for consistency checking in Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$, which is of independent interest. It adapts the standard $\mathcal{S H} \mathcal{I} \mathcal{Q}$ tableaux [8] (using, e.g., anywhere blocking [15] and a kind of lazy unfolding [7]) to terminate in deterministic single exponential time, yielding a representation of a universal model of $\mathcal{K}$ such that each CQ over $\mathcal{K}$ can be answered on it. This may also be exploited to precompile $\mathcal{K}$ into a (query-independent) domino system for on-line query answering.
- Based on our algorithm, we show that CQs in Horn-SHIQ have ExpTimecomplete combined complexity and P-complete data complexity. We also present a fragment of Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$ for which CQs are easier. In Horn- $\mathcal{S H} \mathcal{Q}^{-}$, which forbids inverse roles and existential projection on the left hand side of containment axioms, the combined complexity of CQs is lowered to PSPACE-completeness.

As Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$ is an expressive fragment of OWL-Lite, our results are relevant for the Semantic Web context. They extend in fact from CQs to the class of positive (existential) queries. Our result on the combined complexity of CQs in Horn-SHIQ is of particular interest. Firstly, it reveals another expressive DL for which CQ answering is not harder than standard reasoning (cf. [16, 13, 14]). Secondly, it suggests that the exponential jump in combined complexity of CQs by adding inverse roles to $\mathcal{A L C}$, which was found by Lutz [13], relies on their interaction with disjunction. In other words, if disjunction is eliminated, then inverse roles do not make CQ answering harder.

## 2 Preliminaries

The Description Logics $\mathcal{S H \mathcal { H } \mathcal { Q }}$ and Horn- $\mathcal{S H \mathcal { H } \mathcal { Q }}$ We assume countably infinite sets $\mathbf{C}, \mathbf{R}$ and $\mathbf{I}$ of concept names, role names, and individuals respectively, where $\mathbf{C}$ contains special concepts names $\top$ and $\perp$. Roles are expressions
$R$ and $R^{-}$, where $R \in \mathbf{R}$, and their inverses are $\operatorname{Inv}(R)=R^{-}$and $\operatorname{Inv}\left(R^{-}\right)=R$, respectively. For any roles $R$ and $S, R \sqsubseteq S$ is a role inclusion axiom (RI), and $\operatorname{Trans}(R)$ is a transitivity axiom (TA). For any set $\mathcal{R}$ of RIs and TAs, $\sqsubseteq_{\mathcal{R}}^{*}$ denotes the reflexive transitive closure of $\{(R, S) \mid R \sqsubseteq S \in \mathcal{R}$ or $\operatorname{Inv}(R) \sqsubseteq \operatorname{Inv}(S) \in \mathcal{R}\}$; we write $\operatorname{Trans}_{\mathcal{R}}\left(R^{\prime}\right)$ if $R^{\prime} \sqsubseteq_{\mathcal{R}}^{*} R$ and $R \sqsubseteq_{\mathcal{R}}^{*} R^{\prime}$ for some $R$ s.t. $\operatorname{Trans}(R) \in \mathcal{R}$ or $\operatorname{Trans}\left(R^{-}\right) \in \mathcal{R}$. A role $R$ is simple w.r.t. $\mathcal{R}$, if there is no $S \sqsubseteq_{\mathcal{R}}^{*} R$ with $\operatorname{Trans}_{\mathcal{R}}(S)$.

Concepts are inductively defined as follows: (a) each $A \in \mathbf{C}$ is a concept, and (b) if $C, D$ are concepts, $R$ is a role, and $S$ is a simple role, then $C \sqcap D, C \sqcup D$, $\neg C, \forall R . C, \exists R . C, \geq n S . C$ and $\leq n S . C$, for $n \geq 1$, are concepts.

An expression $C \sqsubseteq D$, where $C, D$ are concepts, is a general concept inclusion axiom (GCI), and expressions $a: A$ and $\langle a, b\rangle: R$, where $A \in \mathbf{C}, a$ and $b$ are individuals, and $R$ is a role, are concept and role assertions, respectively. A $\mathcal{S H I Q}$ knowledge base (KB) is a tuple $\mathcal{K}=\langle\mathcal{T}, \mathcal{R}, \mathcal{A}\rangle$, where the TBox $\mathcal{T}$ is a finite set of GCIs, the $R B o x \mathcal{R}$ is a finite set of RIs and TAs, and the $A B o x \mathcal{A}$ is a finite nonempty set of assertions. We denote by $\mathbf{C}(\mathcal{K}), \mathbf{R}(\mathcal{K})$ and $\mathbf{I}(\mathcal{K})$ the sets of concept names, role names, and individuals occurring in $\mathcal{K}$.

An interpretation $\mathcal{I}=\left\langle\Delta^{\mathcal{I}},{ }^{\mathcal{I}}\right\rangle$ for a $\mathrm{KB} \mathcal{K}$ consists of a nonempty domain $\Delta^{\mathcal{I}}$ and a valuation function. ${ }^{\mathcal{I}}$ that maps each individual $c \in \mathbf{I}(\mathcal{K})$ to an element $c^{\mathcal{I}} \in \Delta^{\mathcal{I}}$, each concept name $C \in \mathbf{C}(\mathcal{K})$ to a subset $C^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$, and each role name $R \in \mathbf{R}(\mathcal{K})$ to a subset $R^{\mathcal{I}}$ of $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$, in such a way that $\top^{\mathcal{I}}=\Delta^{\mathcal{I}}$ and $\perp^{\mathcal{I}}=\emptyset$. The function ${ }^{\mathcal{I}}$ is extended to all concepts and roles in the standard way (see, e.g., [11]), and satisfaction of $\mathcal{K}$ by $\mathcal{I}(\mathcal{I} \models \mathcal{K})$, i.e. modelhood, is also standard.

The DL Horn-SHIQ was introduced [9] as a fragment of $\mathcal{S H \mathcal { H } \mathcal { Q } \text { . The main }}$ idea is to restrict the syntax in a way that $\sqcup$ is not expressible, establishing a correspondence to a Horn fragment of first-order logic with equality. Without loss of generality, we focus here on a normal form of Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$ in [11], to which each Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q} \mathrm{KB}$ can be efficiently rewritten while preserving the answers to arbitrary CQs (as follows from [11]).
Definition 1. (Normal) Horn-SHIQ KBs contain only GCIs of the forms

$$
\begin{array}{rll}
A \sqcap B \sqsubseteq C & A \sqsubseteq \forall R . B & A \sqsubseteq \geq m S . B \\
\exists R . A \sqsubseteq B & A \sqsubseteq \exists R . B & A \sqsubseteq \leq 1 S . B
\end{array}
$$

where $A, B, C$ are concept names, $R$ is a role, $S$ is a simple role, and $m \geq 1$.
Example 1. Assume two Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$ KBs $\mathcal{K}_{1}=\langle\mathcal{T}, \emptyset, \mathcal{A}\rangle$ and $\mathcal{K}_{2}=\langle\mathcal{T}, \mathcal{R}, \mathcal{A}\rangle$, where $\mathcal{T}=\{A \sqsubseteq \exists R . A, B \sqsubseteq \exists P . C\}, \mathcal{R}=\{P \sqsubseteq R\}$, and $\mathcal{A}=\{a: A, a: B\}$. Note that both are consistent, however adding $A \sqsubseteq \forall R . \perp$ to $\mathcal{T}$ makes them inconsistent.
Conjunctive Queries Let $\mathbf{V}$ be a countably infinite set of variables. A (Boolean) conjunctive query ( CQ , or query) over a $\mathrm{KB} \mathcal{K}$ is a finite set $q$ of atoms of the form $A(x)$ or $R(x, y)$, where $A$ is a concept name, $R$ is a role and $x, y \in \mathbf{V}{ }^{3}$ By $\mathbf{V}(q)$ we denote the variables occurring in the atoms of $q$. The query graph of $q$ is the directed graph $G^{q}$ over nodes $\mathbf{V}(q)$ with an edge between nodes $x$ and $y$ iff $R(x, y) \in q$ for some role $R$. The query $q$ is tree-shaped if $G^{q}$ is a tree.

[^1]A match for $q$ in an interpretation $\mathcal{I}$ is a mapping $\theta: \mathbf{V}(q) \rightarrow \Delta^{\mathcal{I}}$ s.t. (i) $\theta(x) \in A^{\mathcal{I}}$ for each $A(x) \in q$, and (ii) $\langle\theta(x), \theta(y)\rangle \in R^{\mathcal{I}}$ for each $R(x, y) \in q$. We say that $\mathcal{I}$ satisfies $q(\mathcal{I} \models q)$, if $q$ has a match in $\mathcal{I}$, and that $\mathcal{K}$ entails $q(\mathcal{K} \models q)$, if $q$ has a match in each model $\mathcal{I}$ of $\mathcal{K}$.
Example 2. Assume the queries $t q_{1}=\{A(x), R(x, y), A(y), R(x, z), C(z)\}$ and $t q_{2}=\{B(x), R(x, y), A(y), P(x, z), C(z)\}$. As easily seen, $\mathcal{K}_{1} \not \models t q_{1}$ and $\mathcal{K}_{1} \models t q_{2}$, while $\mathcal{K}_{2} \models t q_{1}$ and $\mathcal{K}_{2} \models t q_{2}$. Note that both queries are tree-shaped.

## 3 Conjunctive Queries Over Domino Trees

This section describes an algorithm for answering CQs over trees induced by domino systems, which is exploited in the next sections for deciding CQ entailment in Horn-SHIQ KBs. A domino system finitely represents a possibly infinite tree-shaped interpretation that be can built by connecting matching dominoes.

Definition 2. $A$ domino is a tuple $\left\langle c, r, c^{\prime}\right\rangle$ where $c, c^{\prime}$ are sets of concepts names and $r$ is a set of roles (w.r.t. an underlying alphabet). A domino system is a tuple $\langle D, \triangleright, \mathcal{R}\rangle$, where $D$ is a set of dominoes, $\triangleright \subseteq D \times D$ is a direct successor relation with $c_{1}^{\prime}=c_{2}$ whenever $\left\langle c_{1}, r_{1}, c_{1}^{\prime}\right\rangle \triangleright\left\langle c_{2}, r_{2}, c_{2}^{\prime}\right\rangle$, and $\mathcal{R}$ is an RBox. We also require that for each $\left\langle c, r, c^{\prime}\right\rangle \in D$, the set $r$ is closed under role inclusions in $\mathcal{R}$, i.e., $R \in r$ and $R \sqsubseteq R^{\prime} \in \mathcal{R}$ imply $R^{\prime} \in r$. Furthermore, $D$ contains one designated initial domino of the form $\left\langle\emptyset, \emptyset, c^{\prime}\right\rangle$.
Following the terminology in [14], we define next the tree-shaped interpretation induced by a domino system. Its domain is represented by a prefix-closed set of words; for a word $w=e_{1} \cdots e_{n}$, let $\left\langle w \mid e_{n+1}\right\rangle$ denote the word $e_{1} \cdots e_{n} \cdot e_{n+1}$.
Definition 3. The tree base of a domino system $\mathcal{D}=\langle D, \triangleright, \mathcal{R}\rangle$ is the interpretation $\mathcal{I}=\left\langle\Delta^{\mathcal{I}}, \mathcal{I}^{\mathcal{I}}\right\rangle$ (w.r.t. the alphabet underlying $\mathcal{D}$ ) defined as follows:

1. The domain $\Delta^{\mathcal{I}}$ is the smallest set of words over dominoes such that:

- if $t \in D$ is the initial domino, then $t \in \Delta^{\mathcal{I}}$;
- if $t_{1} \cdots t_{n} \in \Delta^{\mathcal{I}}$ and $t_{n} \triangleright t_{n+1}$, then $t_{1} \cdots t_{n} \cdot t_{n+1} \in \Delta^{\mathcal{I}}$.

2. The valuation function ${ }^{\mathcal{I}}$ is defined as follows:

- For each atomic concept $A, A^{\mathcal{I}}=\left\{\langle s \mid t\rangle \in \Delta^{\mathcal{I}} \mid t=\left\langle c, r, c^{\prime}\right\rangle \wedge A \in c^{\prime}\right\}$.
- For each role name $R$,

$$
\begin{aligned}
R^{\mathcal{I}}= & \left\{(s,\langle s \mid t\rangle) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \mid t=\left\langle c, r, c^{\prime}\right\rangle \wedge R \in r\right\} \cup \\
& \left\{(\langle s \mid t\rangle, s) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \mid t=\left\langle c, r, c^{\prime}\right\rangle \wedge \operatorname{Inv}(R) \in r\right\} .
\end{aligned}
$$

The domino tree $\mathcal{T}_{\mathcal{D}}=\left\langle\Delta^{\mathcal{T}}, \cdot^{\mathcal{T}}\right\rangle$ of $\mathcal{D}$ is the interpretation identical to $\mathcal{I}$ except that, for each role $R$, we have $R^{\mathcal{T}}=R^{\mathcal{I}} \cup \bigcup_{S \sqsubseteq_{\mathcal{R}}^{*} R \wedge T r a n s_{\mathcal{R}}(S)}\left(S^{\mathcal{I}}\right)^{+}$.
Query entailment in a domino system is naturally defined via the existence of matches in the represented domino tree. We first provide a procedure to verify the existence of special ordered matches for tree-shaped queries, and we then extend the result to all CQs via the standard method of query treeification.
Definition 4. A domino system $\mathcal{D}$ entails $a C Q q(\mathcal{D} \models q)$, if there is a match for $q$ in $\mathcal{I}_{\mathcal{D}}$. A match $\pi$ for a tree-shaped $C Q$ tq in $\mathcal{T}_{\mathcal{D}}$ is ordered if, for each $x, y \in \mathbf{V}(t q), \pi(x)$ is a proper prefix of $\pi(y)$ whenever $R(x, y) \in t q$ for some $R$. We write $\mathcal{D} \models_{o} t q$, if there is some ordered match for $t q$ in $\mathcal{T}_{\mathcal{D}}$.

```
function checkRoleSucc
input: \(\mathcal{D}=\langle D, \triangleright, \mathcal{R}\rangle\); dominoes \(t_{1}=\left\langle c_{1}, r_{1}, c_{1}^{\prime}\right\rangle, t_{3}=\left\langle c_{3}, r_{3}, c_{3}^{\prime}\right\rangle\) from \(D\); role set \(r \neq \emptyset\);
output: true iff \(t_{3}\) is an \(r\)-successor of \(t_{1}\)
if \(t_{1}\) has no direct successor in \(\mathcal{D}\), then return false;
Initialize \(t\) with some direct successor \(\left\langle c_{2}, r_{2}, c_{2}^{\prime}\right\rangle\) of \(t_{1}\);
\(s:=r_{2} ; i:=0\)
repeat
    if \(t=t_{3}\) and \(r \subseteq s\) then return true;
    if \(r \nsubseteq s\) or \(t\) has no direct successor in \(\mathcal{D}\), then return false;
    Reassign \(t\) to some direct successor \(\left\langle c_{2}, r_{2}, c_{2}^{\prime}\right\rangle\) of \(t\);
    Set \(s^{\prime}\) to the smallest set closed under the following rule:
    if \(S \sqsubseteq_{\mathcal{R}}^{*} R, S \in s, S \in r_{2}\), and \(\operatorname{Trans}_{\mathcal{R}}(S)\), then \(R \in s^{\prime}\);
    \(s:=s^{\prime}\)
until \(i>|D|\); return false
```

Fig. 1. Verifying Successor Dominoes.
Let $t q$ be a fixed tree-shaped query and $\mathcal{D}=\langle D, \triangleright, \mathcal{R}\rangle$ a domino system with tree $\mathcal{T}_{\mathcal{D}}$. To obtain a procedure for deciding $\mathcal{D} \models_{o} t q$, we provide some necessary and sufficient conditions for the existence of ordered matches that can be verified without building $\mathcal{I}_{\mathcal{D}}$ explicitly. Roughly, we search for an association of dominoes from $\mathcal{D}$ with the variables of $t q$. As the association must witness an ordered match, the domino $t_{x}$ associated with variable $x$ must encode the concept names needed to satisfy each unary atom $A(x) \in t q$, while for each role atom $R(x, y) \in$ $t q$, the domino $t_{x}$ must 'reach' the domino $t_{y}$ via an $R$-path.

Definition 5. For two dominoes $t_{1}=\left\langle c_{1}, r_{1}, c_{1}^{\prime}\right\rangle$ and $t_{3}=\left\langle c_{3}, r_{3}, c_{3}^{\prime}\right\rangle$ in $D$ and a set of roles $r \neq \emptyset$, we say $t_{3}$ is a $r$-successor of $t_{1}$ if one of the following holds:
(a) $t_{1} \triangleright t_{3}$ and $r \subseteq r_{3}$, or
(b) for some role set $r^{\prime}, D$ contains an $r^{\prime}$-successor $t_{2}$ of $t_{1}$ such that $t_{2} \triangleright t_{3}$ and for each $R \in r$ there exists $S \in r^{\prime}$ with $\operatorname{Trans}_{\mathcal{R}}(S), S \sqsubseteq_{\mathcal{R}}^{*} R$ and $S \in r_{3}$.
We are ready to define domino associations, which characterize the $\models_{o}$ relation.
Definition 6. $A$ domino association for $t q$ is a mapping $\mu$ that assigns to each $z \in \mathbf{V}(t q)$ a domino $\mu(z) \in D$ in a way such that, for each pair $x, y \in \mathbf{V}(t q)$ :
(a) if $A(x) \in t q$, then $A \in c^{\prime}$, where $\mu(x)=\left\langle c, r, c^{\prime}\right\rangle$; and
(b) if $r=\{R \mid R(x, y) \in t q\}$ is not empty, then $\mu(y)$ is an $r$-successor of $\mu(x)$.

Example 3. Consider the domino system $\mathcal{D}=\langle D, \triangleright, \mathcal{R}\rangle$ in Figure 2, where the


Fig. 2: Domino $\mathcal{D}$. successors of $t_{2}$ as $R$ is transitive, and that $t q_{1}$ has no domino association for $\mathcal{R}=\emptyset$. The query $t q_{2}$ in Example 2 has a domino association even in this case, witnessed by $\mu_{2}$ with $\mu_{2}(x)=t_{3}, \mu_{2}(y)=t_{2}$, and $\mu_{2}(z)=t_{4}$. The following is immediate from the definition of $\models_{o}$ and Definition 6 .

Theorem 1. $\mathcal{D} \models_{o}$ tq iff there exists a domino association for $t q$.
By Theorem 1, we can decide $\mathcal{D} \models_{o} t q$ by deciding existence of a domino association. We exploit for the latter the procedure checkRoleSucc in Figure 1, which nondeterministically checks whether a domino $t_{2}$ is an $r$-successor of a domino $t_{1}$.
Proposition 1. Let $t_{1}, t_{2}$ be dominoes of $\mathcal{D}$, and $r$ a role set. Then $t_{2}$ is an $r$-successor of $t_{1}$ iff some run of checkRoleSucc $\left(\mathcal{D}, t_{1}, t_{2}, r\right)$ returns true.
Now the following simple procedure assocDominoes $(\mathcal{D}, t q)$ non-deterministically decides the existence of a domino association for $t q$ w.r.t. $\mathcal{D}$ : (1) guess a mapping $\mu$ from $\mathbf{V}(t q)$ to dominoes of $\mathcal{D}$, and (2) check satisfaction of the conditions (a) and (b) in Definition 6; to check (b), call checkRoleSucc for each arc in $t q$.
Theorem 2. $\mathcal{D} \models_{o}$ tq iff some run of assocDominoes $(\mathcal{D}, t q)$ returns true.
Having a procedure to decide $\mathcal{D} \models_{o} t q$ for tree-shaped queries $t q$, we now settle deciding $\mathcal{D} \models q$ for arbitrary CQs $q$. Following $[4,6,14]$, we consider query treeifications, i.e., tree-shaped queries whose matches induce matches for $q$. ${ }^{4}$
Definition 7. For every $C Q$ q, let $q^{\mathcal{R}}$ be the smallest query such that: (a) $q \subseteq q^{\mathcal{R}}$, (b) $R(x, y) \in q^{\mathcal{R}}$ and $R \sqsubseteq P \in \mathcal{R}$ implies $P(x, y) \in q^{\mathcal{R}}$, (c) $R(x, y) \in q^{\mathcal{R}}$, $R(y, z) \in q^{\mathcal{R}}$ and $\operatorname{Trans}_{\mathcal{R}}(R)$ imply $R(x, z) \in q^{\mathcal{R}}$, and (d) $R(x, y) \in q^{\mathcal{R}}$ implies $\operatorname{Inv}(R)(y, x) \in q^{\mathcal{R}}$. A treeification of $q$ is a tree-shaped query $q^{\prime}$ such that $\left|q^{\prime}\right| \leq 2|q|$ and there exists a mapping $\theta$ from $\mathbf{V}(q)$ to $\mathbf{V}\left(q^{\prime}\right)$ fulfilling
a) $A(x) \in q$ implies $A(\theta(x)) \in q^{\prime}$, and
b) $R(x, y) \in q$ implies $R(\theta(x), \theta(y)) \in\left(q^{\prime}\right)^{\mathcal{R}}$.

As easily shown, each match for a treeification $q^{\prime}$ of $q$ in $\mathcal{T}_{\mathcal{D}}$ is also a match for $q$. On the other hand, from each match for $q$ in $\mathcal{T}_{\mathcal{D}}$, we can obtain a treeification $q^{\prime}$ that is mappable into $\mathcal{I}_{\mathcal{D}}$ via an ordered match. ${ }^{5}$ Hence, we obtain:
Theorem 3. For any $C Q q, \mathcal{D} \models q$ iff $\mathcal{D} \models_{o} q^{\prime}$, for some treeification $q^{\prime}$ of $q$.
It follows that we can decide $\mathcal{D} \models q$ by listing all treeifications of $q$ and posing them separately over $\mathcal{D}$. Note that there are finitely many treeifications of $q$.

## 4 A Tableaux Algorithm for Horn-SHIQ

In order to exploit the results of Section 3 for query answering in Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$, we first provide a tableau algorithm for KB satisfiability. Like the standard $\mathcal{S H I Q}$ tableau [8], it uses a completion forest to represent a model; in the next section we extract from it a domino system that can be used for query answering.
 concepts names; $D, E$ to denote (arbitrary) concepts; $R, R^{\prime}$ to denote a role; $S$ a simple role; and $a, b$ to denote individuals.

[^2]Most of the following definitions are based on [8], while 9 and 14 follow [15] and are related to anywhere blocking. Definition 8 is simplified since only normalized KBs are considered, and the $\approx$ relation from [8] is omitted in Definition 10. ${ }^{6}$
Definition 8. (concept closure) We define $\mathbf{C l}(\mathcal{K})$ as the smallest set of concepts closed under subconcepts such that (i) $D, E \in \mathbf{C l}(\mathcal{K})$ for every $D \sqsubseteq E \in \mathcal{T}$; and (ii) if $\forall R . A \in \mathbf{C l}(\mathcal{K})$, $\operatorname{Trans}_{\mathcal{R}}\left(R^{\prime}\right)$ and $R^{\prime} \sqsubseteq_{\mathcal{R}}^{*} R$ for some $R^{\prime}$, then $\forall R^{\prime} . A \in \mathbf{C l}(\mathcal{K})$.

Definition 9. ((named/unnamed) nodes) We assume a countably infinite set $\mathbf{N}$ of nodes and a strict total order $\lessdot$ on $\mathbf{N}$. Each $a \in \mathbf{I}(\mathcal{K})$ is associated with one fixed node $n_{a} \in \mathbf{N}$; the nodes $n_{a}$ are named, all other nodes are unnamed.
Definition 10. (completion forest) $A$ completion forest for a $\mathrm{KB} \mathcal{K}$ is a tuple $\mathcal{F}=\langle\mathcal{N}, \mathcal{E}, \mathcal{L}, \not \approx\rangle$ where $\mathcal{N} \subseteq \mathbf{N}$ and $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ define a directed graph; $\mathcal{L}$ is a labeling function assigning each $n \in \mathcal{N}$ a subset of $\mathbf{C l}(\mathcal{K})$ and each pair $u, u^{\prime} \in \mathcal{N} \times \mathcal{N}$ to a set of roles (over $\mathcal{K}$ ), in such a way that $\mathcal{L}\left(u, u^{\prime}\right)=\emptyset$ for all $\left(u, u^{\prime}\right) \notin \mathcal{E} ;$ and $\not \approx \subseteq \mathcal{N} \times \mathcal{N}$ is a binary relation, tacitly assumed to be symmetric.
Definition 11. (successor, neighbor) For a completion forest $\mathcal{F}=\langle\mathcal{N}, \mathcal{E}, \mathcal{L}, \not \not\rangle\rangle$ and a pair $u, u^{\prime} \in \mathcal{N}, u^{\prime}$ is a successor of $u$ if $\left(u, u^{\prime}\right) \in \mathcal{E}$. The inverse of successor is called predecessor; the transitive closures of successor and predecessor are ancestor and descendant respectively. For all $R, u^{\prime}$ is an $R$-successor of $u$ if $R^{\prime} \in \mathcal{L}\left(u, u^{\prime}\right)$ for some $R^{\prime}$ with $R^{\prime} \sqsubseteq_{\mathcal{R}}^{*} R$. We call $u^{\prime}$ an $R$-neighbor of $u$, if $u^{\prime}$ is an $R$-successor of $u$, or if $u$ is an $\operatorname{Inv}(R)$-successor of $u^{\prime}$.
Definition 12. (clash-free completion forest) A completion forest $\mathcal{F}$ contains $a$ clash, if (i) for some $u \in \mathcal{N}, \perp \in \mathcal{L}(u)$; or (ii) for some $u \in \mathcal{N}$ with $\leq$ $n S . C \in \mathcal{L}(u)$, $u$ has $n+1 S$-neighbors $w_{0}, \ldots, w_{n}$ such that, for all $0 \leq i<$ $j \leq n, C \in \mathcal{L}\left(w_{i}\right)$ and $w_{i} \not \approx w_{j} \in \mathcal{F}$. If $\mathcal{F}$ contains no clash, then $\mathcal{F}$ is clash-free.
Definition 13. (initial completion forest) The initial completion forest $\mathcal{F}_{\mathcal{A}}$ for $\mathcal{K}$ has the named node $n_{a}$ labeled with $\mathcal{L}\left(n_{a}\right)=\{A \in \mathbf{C}(\mathcal{K}) \mid a: A \in \mathcal{A}\}$ for each individual $a \in \mathbf{I}(\mathcal{K})$, and an edge $\left(n_{a}, n_{b}\right) \in \mathcal{E}$ labeled $\mathcal{L}\left(n_{a}, n_{b}\right)=\{R \mid\langle a, b\rangle$ : $R \in \mathcal{A}\}$ for each pair $a, b \in \mathbf{I}(\mathcal{K})$; the relation $\not \approx$ is empty.
Definition 14. (blocking) For a completion forest $\mathcal{F}=\langle\mathcal{N}, \mathcal{E}, \mathcal{L}, \not \approx\rangle$, a node $u \in \mathcal{N}$ is blocked if $u$ is unnamed and $u$ is either directly or indirectly blocked; $u$ is indirectly blocked if one of its ancestors is blocked; $u$ is directly blocked if none of its ancestors is blocked and there is some $u^{\prime} \lessdot u$ such that $u, u^{\prime}$ are unnamed nodes, $\mathcal{L}(u)=\mathcal{L}\left(u^{\prime}\right), \mathcal{L}(v)=\mathcal{L}\left(v^{\prime}\right)$, and $\mathcal{L}(v, u)=\mathcal{L}\left(v^{\prime}, u^{\prime}\right)$, where $v$ and $v^{\prime}$ are the predecessors of $u$ and $u^{\prime}$ respectively.
The expansion rules are given in Figure 3, where a node $u \in \mathbf{N}$ is new in $\mathcal{F}$ if $u^{\prime} \lessdot u$ for every $u^{\prime} \in \mathcal{N}$. The $\leq$-rule calls the operation $\operatorname{merge}(u, N)$ described in Figure 4. The rules are similar to those in [8], except for the first three, which (lazily) ensure the satisfaction of the TBox axioms. Also, the restricted form of at-most number restrictions allows us to have just one $\leq$-rule and a deterministic $\operatorname{merge}(u, N)$ that simultaneously merges all nodes in $N$ into one. ${ }^{7}$

[^3]```
\(\mathcal{T}\)-rule: if \(\quad A \sqsubseteq D \in \mathcal{T}, A \in \mathcal{L}(u)\), and \(D \notin \mathcal{L}(u)\),
    then \(\mathcal{L}(u):=\mathcal{L}(u) \cup\{D\}\).
\(\overline{\mathcal{T}_{\Pi}-\text { rule: if }} \quad A \sqcap B \sqsubseteq C \in \mathcal{T},\{A, B\} \subseteq \mathcal{L}(u)\),
        \(u\) is not indirectly blocked, and \(C \notin \mathcal{L}(u)\),
    then \(\mathcal{L}(u):=\mathcal{L}(u) \cup\{C\}\).
\(\overline{\mathcal{T}_{\exists} \text {-rule: if } \quad \exists R . A \sqsubseteq B \in \mathcal{T}, B \notin \mathcal{L}(u), u \text { is not indirectly blocked, and }}\)
        \(u\) has an \(R\)-neighbor \(u^{\prime}\) with \(A \in \mathcal{L}\left(u^{\prime}\right)\),
then \(\mathcal{L}(u):=\mathcal{L}(u) \cup\{B\}\).
\(\exists\)-rule: if \(\quad \exists R . A \in \mathcal{L}(u), u\) is not blocked,
        and \(u\) has no \(R\)-neighbor \(u^{\prime}\) with \(A \in \mathcal{L}\left(u^{\prime}\right)\),
    then set \(\mathcal{N}=\mathcal{N} \cup\left\{u^{\prime}\right\}, \mathcal{E}=\mathcal{E} \cup\left\{\left(u, u^{\prime}\right)\right\}, \mathcal{L}\left(u, u^{\prime}\right):=\{R\}\)
        and \(\mathcal{L}\left(u^{\prime}\right):=\{A\}\) for some \(u^{\prime}\) new in \(\mathcal{F}\).
\(\forall\)-rule: if \(\quad \forall R . A \in \mathcal{L}(u), u\) is not indirectly blocked, and
        \(u\) has an \(R\)-neighbour \(u^{\prime}\) with \(A \notin \mathcal{L}\left(u^{\prime}\right)\),
    then \(\mathcal{L}\left(u^{\prime}\right):=\mathcal{L}\left(u^{\prime}\right) \cup\{A\}\).
\(\nabla_{+}\)-rule: if \(\quad \forall R . A \in \mathcal{L}(u), u\) is not indirectly blocked,
        there is some \(R^{\prime}\) with \(\operatorname{Trans}_{\mathcal{R}}\left(R^{\prime}\right)\) and \(R^{\prime} \sqsubseteq_{\mathcal{R}}^{*} R\),
        and there is an \(R^{\prime}\)-neighbour \(u^{\prime}\) of \(u\) with \(\forall R^{\prime} . A \notin \mathcal{L}\left(u^{\prime}\right)\),
    then \(\mathcal{L}\left(u^{\prime}\right):=\mathcal{L}\left(u^{\prime}\right) \cup\left\{\forall R^{\prime} . A\right\}\).
\(\geq\)-rule: if \(\quad \geq m S . A \in \mathcal{L}(u), u\) is not blocked,
        and there are no \(m S\)-neighbours \(u_{1}, \ldots, u_{m}\) of \(u\)
        such that \(A \in \mathcal{L}\left(u_{i}\right)\) and \(u_{i} \not \approx u_{j}\) for \(1 \leq i<j \leq m\),
    then set \(\mathcal{N}=\mathcal{N} \cup\left\{u_{1}, \ldots, u_{m}\right\}, \mathcal{E}=\mathcal{E} \cup\left\{\left(u, u_{1}\right), \ldots,\left(u, u_{m}\right)\right\}\),
        \(\mathcal{L}\left(u, u_{i}\right):=\{S\}, \mathcal{L}\left(u_{i}\right):=\{A\}\) and \(u_{i} \not \approx u_{j}\)
        for \(1 \leq i<j \leq m\) and \(u_{1}, \ldots, u_{m}\) new in \(\mathcal{F}\).
s-rule: if \(\quad \leq 1 S . A \in \mathcal{L}(u), u\) is not indirectly blocked,
        \(N\) is the set of all \(S\)-neighbours \(u^{\prime}\) of \(u\) with \(A \in \mathcal{L}\left(u^{\prime}\right)\),
        \(|N|>1\) and there is no pair \(u^{\prime}, u^{\prime \prime}\) in \(N\) with \(u^{\prime} \not \approx u^{\prime \prime}\),
        then merge \((u, N)\).
```

Fig. 3. Tableaux expansion rules

The initial $\mathcal{F}_{\mathcal{A}}$ is expanded by exhaustively applying the rules in Figure 3. The expansion stops, if a clash is reached; otherwise, it continues until the forest is complete, i.e., no rule is applicable. It can be shown similarly as in $[8,15]$ that this algorithm is a decision procedure for KB satisfiability in Horn- $\mathcal{S H} \mathcal{I}$ Q.

Theorem 4. Let $\mathcal{K}$ be a Horn-SHIQ KB. Then $\mathcal{K}$ is satisfiable iff a complete and clash-free completion forest for $\mathcal{K}$ can be obtained.

Note that after applying any rule from Figure 3, the resulting forest is uniquely determined up to renaming of nodes. The only source of differences in the resulting forests lies in possibly different orderings of rule applications (this could be eliminated, e.g., using $\lessdot$ and any fixed ordering on $\mathbf{C l}(K)$ and on the rules). However, these differences are not relevant: each $\mathcal{F}$ represents a universal model $\mathcal{I}_{\mathcal{F}}$ (defined as its standard unravelling [8]) that is embeddable into every model of $\mathcal{K}$, and can be used for query answering. The following is shown by a straightforward induction on the construction of $\mathcal{I}_{\mathcal{F}}$ :
(1) let $u_{0}$ be the $\lessdot-$ minimal element of $N$;
(2) let $N^{\prime}=N \backslash\left\{u_{0}\right\}$; let $N^{\prime \prime}$ be the minimal set containing $N^{\prime}$, each unnamed successor $u^{\prime}$ of a node in $N^{\prime}$, and all descendants of $u^{\prime}$;
(3) if $\left(u^{\prime}, n\right) \in \mathcal{E}$ for some $u^{\prime} \in N^{\prime}$ and some named $n$, then $\mathcal{E}:=\mathcal{E} \cup\left(u_{0}, n\right)$ and $\mathcal{L}\left(u_{0}, n\right):=\mathcal{L}\left(u_{0}, n\right) \cup \mathcal{L}\left(u^{\prime}, n\right)$; if $\left(n, u^{\prime}\right) \in \mathcal{E}$ for some $u^{\prime} \in N^{\prime}$ and some named $n$, then $\mathcal{E}:=\mathcal{E} \cup\left(n, u_{0}\right)$ and $\mathcal{L}\left(n, u_{0}\right):=\mathcal{L}\left(n, u_{0}\right) \cup \mathcal{L}\left(n, u^{\prime}\right) ;$
(4) set $\mathcal{N}:=\mathcal{N} \backslash N^{\prime \prime}, \mathcal{E}:=\mathcal{E} \backslash\left\{\left(v, u^{\prime}\right) \mid u^{\prime} \in N^{\prime \prime}\right\}$, restrict $\mathcal{L}$ and $\not \approx$ to the new $\mathcal{N}, \mathcal{E}$;
(5) add $u_{0} \not \approx v$ for every $v \in \mathcal{N}$ such that $u^{\prime} \not \approx v$ for some $u^{\prime} \in N^{\prime}$;
(6) set $\mathcal{L}\left(u_{0}\right):=\mathcal{L}\left(u_{0}\right) \cup \mathcal{L}\left(N^{\prime}\right)$, where $\mathcal{L}\left(N^{\prime}\right)=\bigcup_{u^{\prime} \in N^{\prime}} \mathcal{L}\left(u^{\prime}\right)$;
(7) if $\left(u_{0}, u\right) \in \mathcal{E}$ then $\mathcal{L}\left(u_{0}, u\right):=\mathcal{L}\left(u_{0}, u\right) \cup \mathcal{L}\left(N^{\prime}, u\right)$, else $\mathcal{L}\left(u, u_{0}\right):=\mathcal{L}\left(u, u_{0}\right) \cup \mathcal{L}\left(u, N^{\prime}\right)$, where $\mathcal{L}\left(u, N^{\prime}\right)=\bigcup_{u^{\prime} \in N} \mathcal{L}\left(u, u^{\prime}\right)$ and $\mathcal{L}\left(N^{\prime}, u\right)=\left\{\operatorname{Inv}(R) \mid R \in \mathcal{L}\left(u, N^{\prime}\right)\right\}$.

Fig. 4. The $\operatorname{merge}(u, N)$ operation on $\mathcal{F}=\langle\mathcal{N}, \mathcal{E}, \mathcal{L}, \not \approx\rangle$
Theorem 5. Let $\mathcal{I}$ be a model of $\mathcal{K}$, let $\mathcal{F}$ be a complete and clash-free completion forest for $\mathcal{K}$, and let $\mathcal{I}_{\mathcal{F}}$ be the model of $\mathcal{K}$ represented by $\mathcal{F}$. Then there is a homomorphic embedding of $\mathcal{I}_{\mathcal{F}}$ into $\mathcal{I}$. Hence, for any $C Q q, \mathcal{K} \models q$ iff $\mathcal{I}_{\mathcal{F}} \models q$.

## 5 Conjunctive Queries over Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$

To answer CQs over Horn-SHIQ KBs, we exploit the method for answering tree-shaped queries over domino systems. For this section, we assume that $\mathcal{K}=$
 complete and clash-free completion forest $\mathcal{F}_{\mathcal{K}}$ for $\mathcal{K}$, we extract a domino system $\mathcal{D}_{\mathcal{F}_{\mathcal{K}}}$ that encodes a forest-shaped universal model of $\mathcal{K}$ for query answering. We then rewrite $q$ into a set of tree-shaped queries which can be posed separately over $\mathcal{D}_{\mathcal{F}_{\mathcal{K}}}$, such that $\mathcal{K} \models q$ iff one of the generated queries is entailed by $\mathcal{D}_{\mathcal{F}_{\mathcal{K}}}$.

The transformation of the completion forest into $\mathcal{D}_{\mathcal{F}_{\mathcal{K}}}$, which we now present, eliminates the 'graph part' of the forest by encoding it into the initial domino.
Definition 15. Let $\mathcal{F}=\langle\mathcal{N}, \mathcal{E}, \mathcal{L}, \not \approx\rangle$ be a complete and clash-free completion forest for $\mathcal{K}$. For every $u \in \mathcal{N}$, let $\mathcal{L}^{\prime}(u)=\mathcal{L}(u) \cap \mathbf{C}(\mathcal{K})$.

Let $t_{0}=\langle\emptyset, \emptyset, c\rangle$ be the domino where $c$ is the smallest set of fresh concept names such that Root $\in c$ and, for each pair $n_{a}, n_{b}$ of named nodes in $\mathcal{N}$, (i) $A \in \mathcal{L}^{\prime}\left(n_{a}\right)$ implies $A_{a} \in c$, (ii) if $n_{b}$ is an $R$-neighbour of $n_{a}$, then $R_{a, b} \in c$, (iii) $R_{a, b} \in c$ and $R \sqsubseteq R^{\prime} \in \mathcal{R}$ implies $R_{a, b}^{\prime} \in c$, (iv) $R_{a, b} \in c, R_{b, d} \in c$ and $\operatorname{Trans}_{\mathcal{R}}(R)$ implies $R_{a, d} \in c$, and (v) $R_{a, b} \in c$ implies $\operatorname{Inv}(R)_{b, a} \in c$.

For each named node $n_{a} \in \mathcal{N}$, let $t_{a}$ denote the domino $t_{a}=\left\langle c,\left\{Q_{a}\right\}, c^{\prime}\right\rangle$, where $Q_{a}$ is fresh and $c^{\prime}=\mathcal{L}^{\prime}\left(n_{a}\right)$. For a pair $\left(u, u^{\prime}\right) \in \mathcal{E}$, let $t\left(u, u^{\prime}\right)$ denote the domino $\left\langle\mathcal{L}^{\prime}(u),\left\{R \mid u^{\prime}\right.\right.$ is an $R$-neighbour of $\left.\left.u\right\}, \mathcal{L}^{\prime}\left(u^{\prime}\right)\right\rangle$. Then $\mathcal{D}_{\mathcal{F}}=\langle D, \triangleright, \mathcal{R}\rangle$ is the domino system with initial domino $t_{0}$, where

- $D$ is the smallest domino set containing (i) $t_{0}$, (ii) $t_{a}$ for each named $n_{a} \in \mathcal{N}$, and (iii) each $t\left(u, u^{\prime}\right)$ such that $\left(u, u^{\prime}\right) \in \mathcal{E}$ and $u^{\prime}$ is unnamed and not blocked.
- $\triangleright$ is the smallest relation s.t. (i) for all named $n_{a} \in \mathcal{N}, t_{0} \triangleright t_{a}$ and $t_{a} \triangleright t\left(n_{a}, u\right)$ for every $t\left(n_{a}, u\right) \in D$; and (ii) if $t\left(u, u^{\prime}\right), t\left(u^{\prime}, v\right) \in D$ for some $\left(u, u^{\prime}\right),\left(u^{\prime}, u^{\prime \prime}\right) \in \mathcal{E}$ such that either $u^{\prime \prime}=v$ is not blocked or $u^{\prime \prime}$ is blocked by $v$, then $t\left(u, u^{\prime}\right) \triangleright t\left(u^{\prime}, v\right)$.

[^4]Since the specific complete and clash-free $\mathcal{F}$ does not matter, we assume in what follows a fixed arbitrary $\mathcal{F}_{\mathcal{K}}$ and denote its domino system $\mathcal{D}_{\mathcal{F}_{\mathcal{K}}}$ simply by $\mathcal{D}_{\mathcal{K}}$. As easily seen, we can reconstruct a universal model of $\mathcal{K}$ from the domino tree of $\mathcal{D}_{\mathcal{K}}$. However, for querying $\mathcal{D}_{\mathcal{K}}$, we need to rewrite $q$ in order to handle the links between individuals encoded as concept names in the initial domino.

Definition 16. A link rewriting of $q$ w.r.t. $\mathcal{K}$ is a $C Q$ obtained from $q$ as follows:

1. Exhaustively replace, one by one, $R(y, x)$ by $\operatorname{Inv}(R)(x, y)$ whenever there are atoms of the form $R(y, x)$ and $S(x, y)$ in $q$.
2. Let $\mu: \mathbf{V}(q) \rightarrow \mathbf{I}(\mathcal{K})$ be a partial function, and let $\nu(x) \in\{r$ (root), $i$ (inside) $\}$ be a choice for each $x \in \operatorname{dom}(\mu)$. Let $\{z\} \cup\left\{x^{\prime} \mid x \in \mathbf{V}(q)\right\}$ be fresh variables. Then, for each $R(x, y) \in q$ with $\{x, y\} \subseteq \operatorname{dom}(\mu)$, let $S \sqsubseteq_{\mathcal{R}}^{*} R$ be arbitrary such that $\operatorname{Trans}_{\mathcal{R}}(S)$ holds if either $\nu(x)=i$ or $\nu(y)=i$, and (i) replace $R(x, y)$ in $q$ by $\operatorname{Root}(z), S_{a, b}(z)$, where $\mu(x)=a$ and $\mu(y)=b$, and (ii) add in $q$, depending on the choice $[\nu(x), \nu(y)]$, the following atoms:

$$
\begin{array}{lll}
{[r, r]:} & Q_{a}(z, x), Q_{b}(z, y) ; & {[i, i]:} \\
{[i, r]:} & Q_{a}\left(z, x^{\prime}\right), Q_{b}\left(z, y^{\prime}\right), S\left(x, x^{\prime}\right), S\left(y^{\prime}, y\right) ; \\
{\left[\begin{array}{l}
a \\
\left(z, x^{\prime}\right), \\
Q_{b}(z, y), S\left(x, x^{\prime}\right) ;
\end{array}\right.} & {[r, i]:} & Q_{a}(z, x), Q_{b}\left(z, y^{\prime}\right), S\left(y^{\prime}, y\right) .
\end{array}
$$

Roughly speaking, possible $R$-connections between matches for $x$ and $y$ in $\mathcal{I}_{\mathcal{F}_{\mathcal{K}}}$ via two individuals $a, b$ are reflected in the query by the atoms $\operatorname{Root}(z), S_{a, b}(z)$ and $[\nu(x), \nu(y)]$. For example, if $\mathcal{R}=\{\operatorname{Trans}(S), S \sqsubseteq R\}$, the $\mathrm{CQ} q^{\prime}=\{\operatorname{Root}(z)$, $\left.S_{a, b}(z), Q_{a}\left(z, x^{\prime}\right), S\left(x, x^{\prime}\right), A(x), Q_{b}(z, y), B(y)\right\}$ is a link rewriting of $q=\{A(x)$, $R(x, y), B(y)\}$, obtained by choosing $\mu(x)=a, \mu(y)=b, \nu(x)=i, \nu(y)=r$ and $S$. A match for $q^{\prime}$ in $\mathcal{I}_{\mathcal{D}_{\mathcal{K}}}$ corresponds to a match for $q$ in $\mathcal{I}_{\mathcal{F}_{\mathcal{K}}}$ mapping $x$ to a descendant of $a$ (i.e., inside the tree rooted at $a$ ) and $y$ to $b$ (i.e., the root of $b$ 's tree), which are connected via an $S$-path and thus in the extension of $R$. Note that, as $\nu(x)=i$ was chosen, the non-transitive $R$ can not link the matches of $x$ and $y$. Choosing $\mu(x)=a, \mu(y)=b, \nu(x)=r, \nu(y)=r$ and $R$ we obtain a rewriting $q^{\prime \prime}=\left\{\operatorname{Root}(z), R_{a, b}(z), Q_{a}(z, x), A(x), Q_{b}(z, y), B(y)\right\}$ that captures the matches for $q$ which map $x$ and $y$ to $a$ and $b$ if they are $R$-neighbors in $\mathcal{I}_{\mathcal{F}_{\mathcal{K}}}$.

Theorem 6. $\mathcal{K} \models q$ iff $\mathcal{D}_{\mathcal{K}} \models q^{\prime}$ for some link rewriting $q^{\prime}$ of $q$. Hence, due to Theorem 3, $\mathcal{K} \models q$ iff $\mathcal{D}_{\mathcal{K}} \models_{o}$ tq, for some treeification tq of a link rewriting of $q$.

Theorem 6 suggests a procedure for deciding $\mathcal{K} \models q$ : it suffices to verify the existence of a treeification of a link rewriting of $q$ that has an ordered match in the domino tree of $\mathcal{D}_{\mathcal{K}}$. The latter can be verified using the method from Section 3.

## 6 Computational Complexity

We now show that CQ entailment in Horn- $\mathcal{S H \mathcal { L } \mathcal { Q }}$ is decidable in exponential time. The presented method relies on the extraction of a domino system from a complete and clash-free completion forest. Hence, the following is important.

Theorem 7. The tableau algorithm for Horn-SHIQ in Section 4 decides consistency of Horn-SHIQ KBs in single exponential time. For a consistent KB, it constructs a complete and clash-free completion forest of at most exponential size.

Proof (Sketch). Definition 14 ensures that if a completion forest $\mathcal{F}$ contains two pairs of nodes with the same node-arc-node label combination, one of them is blocked. The number of such combinations, and thus of nodes in a forest, is single exponential in the input $\mathrm{KB} \mathcal{K}$ (in fact, it is bounded by $2^{2|\mathbf{C l}(\mathcal{K})| \times 2|\mathbf{R}(\mathcal{K})|}$ ). Using the usual arguments [8], it can be shown that the number of rule applications needed to generate $\mathcal{F}$ is polynomially bounded by the maximal number of nodes it can have, as the shrinking rules do not cause repeated rule applications.
We are ready to formulate the main complexity results of this paper.
Theorem 8. Conjunctive query entailment $\mathcal{K} \models q$ in Horn-SHIQ is ExpTimecomplete in combined complexity, i.e., in the size of the $K B \mathcal{K}$ and the query $q$.

Proof (Sketch). By Theorem 7, we can check the consistency of $\mathcal{K}$ using the tableau-based algorithm in exponential time. If $\mathcal{K}$ is found inconsistent, then $\mathcal{K} \models q$ trivially holds. Otherwise, we can extract the domino system $\mathcal{D}_{\mathcal{K}}$ from the completion forest $\mathcal{F}_{\mathcal{K}}$ that was constructed, in time polynomial in $\left|\mathcal{F}_{\mathcal{K}}\right|$.

Each link rewriting $q^{\prime}$ of $q$, as well as each treeification $t q$ of $q^{\prime}$, has size polynomial w.r.t. $|\mathcal{K}|+|q|$. The are at most exponentially many $q^{\prime}$ and, for each $q^{\prime}$, at most exponentially many $t q$; hence, there are at most exponentially many $t q$ in total, and they can be traversed in polynomial space. To show ExpTime membership of $\mathcal{K} \models q$, it is thus sufficient to show that $\mathcal{D}_{\mathcal{K}} \models_{o} t q$ is decidable in exponential time w.r.t. $|\mathcal{K}|+|q|$. Indeed, checkRoleSucc runs in NPSpace w.r.t. $|\mathcal{K}|+|q|$ if $\mathcal{D}_{\mathcal{K}}$ is precomputed (note that the counter $i$ needs only polynomial space). The procedure assocDominoes runs in NP (w.r.t. $|\mathcal{K}|+|q|)$ using checkRoleSucc as an oracle. Hence, $\mathcal{D}_{\mathcal{K}} \models_{o} t q$ is in NP ${ }^{\text {NPSPAcE }}=$ PSPACE w.r.t. $|\mathcal{K}|+|q|$, if $\mathcal{D}_{\mathcal{K}}$ is precomputed. As computing $\mathcal{D}_{\mathcal{K}}$ is feasible in exponential time, it follows that deciding $\mathcal{K} \models q$ is in ExpTime. The matching lower bound follows from the ExpTime-hardness of consistency checking in Horn-SHIQ [11].
The next result shows that CQs in Horn- $\mathcal{S H} \mathcal{H} \mathcal{Q}$ are tractable in data complexity.
Theorem 9. Conjunctive query entailment $\mathcal{K} \models q$ in Horn-SHIQ is P -complete in data complexity, i.e., in the size of the $A B o x \mathcal{A}$ of the $K B \mathcal{K}=\langle\mathcal{T}, \mathcal{R}, \mathcal{A}\rangle$.

Proof (Sketch). As in Theorem 8, we can check the consistency of $\mathcal{K}$ when we construct the completion forest $\mathcal{F}_{\mathcal{K}}$. As $\mathcal{T}$ and $\mathcal{R}$ are fixed, $\left|\mathcal{F}_{\mathcal{K}}\right|$ is polynomial w.r.t. $\mathcal{A}$, so $\mathcal{F}_{\mathcal{K}}$ and $\mathcal{D}_{\mathcal{K}}$ can be constructed in time polynomial w.r.t. $\mathcal{A}$. Next, for fixed $q, \mathcal{T}$ and $\mathcal{R}$, there are polynomially many treeifications $t q$ of link rewritings $q^{\prime}$ of $q$ w.r.t. $\mathcal{A}$, and they can be traversed in polynomial time. By Theorem 1, it remains to show that the existence of a domino association $\mu: \mathbf{V}(t q) \rightarrow D$, where $D$ is the domino set of $\mathcal{D}_{\mathcal{K}}$, is decidable in polynomial time w.r.t. $\mathcal{A}$. Since $|\mathbf{V}(t q)|$ is bounded by a constant w.r.t. $\mathcal{A}$ and $|D|$ is polynomial w.r.t. $\mathcal{A}$, there are polynomially many candidate $\mu$ w.r.t. $\mathcal{A}$. We can check $r$-successorship between dominoes $t_{1}, t_{2}$ of $\mathcal{D}_{\mathcal{K}}$ in time polynomial in $\left|\mathcal{D}_{\mathcal{K}}\right|$, i.e., polynomial w.r.t. $\mathcal{A}$. Hence, we can check whether $\mu$ satisfies Definition 6 in polynomial time w.r.t. $\mathcal{A}$. The resulting P membership bound is tight, as consistency checking in any DL allowing for conjunction on the left hand side and quantified universal restrictions on the right hand side of GCIs is P-hard in data complexity [3].

The source of ExpTime-hardness of consistency testing in Horn-SHIQ, and hence of query entailment, are inverse roles and concepts of the form $\exists R . A$ on the left hand side of the GCIs. Intuitively, both constructs allow to propagate information from a node to its ancestors in a completion forest; any one of them allows for an encoding of a generic Alternating PSpace Turing machine. If we disallow both, obtaining the DL Horn- $\mathcal{S H} \mathcal{Q}^{-}$, consistency testing and CQ entailment drop to PSpace-completeness. Roughly, the direct successors of a node in a completion forest can be inferred in polynomial time from its label. Hence, the existence of a complete and clash-free completion forest $\mathcal{F}$ is refutable in PSpace without building it, by non-deterministically following a path in $\mathcal{F}$ (of at most exponential length) that leads to a clash. CQ entailment is decidable in PSpace by supplying checkRoleSucc with a PSpace oracle for navigating the domino system $\mathcal{D}_{\mathcal{K}}$ (note that the explicit construction of $\mathcal{D}_{\mathcal{K}}$ may require exponential space). This procedure is worst-case optimal, as consistency checking in Horn- $\mathcal{S H} \mathcal{Q}^{-}$is PSpace-hard (provable, e.g., by a generic Turing machine encoding). Finally, also our P-completeness result for data complexity of CQs carries over from Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$ to Horn- $\mathcal{S H} \mathcal{Q}^{-}$. Details will be given in the full paper.

## 7 Related Work and Conclusion

As found recently, answering CQs in some expressive DLs, including $\mathcal{A L C H}$ [16] and $\mathcal{A L C H} \mathcal{Q}$ [14], is not harder than standard reasoning, and in fact ExpTimecomplete in combined complexity. Horn-SHIQ is another such DL but orthogonal to those mentioned, as it offers transitive and inverse roles but excludes disjunction (we note that one can infer from [2] the ExpTime-completeness result for the DL Horn- $\mathcal{A L C H}$ I, i.e., Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$ without transitive roles and number restrictions). Moreover, the data complexity of CQs is polynomial in Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$ but intractable in $\mathcal{A L C H}$ and $\mathcal{A L C H \mathcal { L }}$ (in fact, it is coNP-hard already for $\mathcal{A L}$ [3]).

Different approaches have been recently used to show that CQs have tractable data complexity in some DLs. A large class of such DLs are extensions of $\mathcal{E} \mathcal{L}$ [1], considered e.g. in $[17,12,10]$, of which $\mathcal{E} \mathcal{L H}, \mathcal{E} \mathcal{L} \mathcal{I}^{f}$, and $\mathcal{E} \mathcal{L}^{++}$are particularly noticeable. For $\mathcal{E L}$ and $\mathcal{E L H}$, which are subsumed by Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$ but not by Horn- $\mathcal{S H} \mathcal{Q}^{-}$(due to the absence of existential restrictions on the LHS of the GCIs), a reduction to Datalog has been given in [17]. In both $\mathcal{E L}$ and $\mathcal{E} \mathcal{L H}$, CQs have coNP-complete combined complexity and P -complete data complexity.
 qualified universal quantification on the RHS of axioms and more general number restrictions) an explicit (partial) construction of a universal model was used in [10]. Like in Horn- $\mathcal{S H \mathcal { L } \mathcal { Q }}$ and Horn- $\mathcal{S H} \mathcal{Q}^{-}$, CQs have P-complete data complexity in $\mathcal{E} \mathcal{L I} \mathcal{I}^{f}$. Finally, for $\mathcal{E} \mathcal{L}^{++}$, which is orthogonal to both Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$ and Horn- $\mathcal{S H} \mathcal{Q}^{-}\left(\mathcal{E} \mathcal{L}^{++}\right.$has nominals and regular role hierarchies, but lacks universal quantification), special proof-graphs with automata were used in [12]. Noticeably, CQs in $\mathcal{E} \mathcal{L}^{++}$have PSPACE-complete combined and P-complete data complexity respectively, and thus the same complexity as in Horn- $\mathcal{S H} \mathcal{Q}^{-}$.

Another prominent family for which data complexity has been deeply investigated is DL-Lite [5]. For the core DL-Lite and its extension with functionality and conjunction, which is subsumed by Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$ but not by Horn- $\mathcal{S H} \mathcal{Q}^{-}$, query
rewriting into first-order logic over relational databases has been been employed. CQ answering has very low data complexity (inside logarithmic space), and its coNP-complete combined complexity is also much lower than for Horn- $\mathcal{S H} \mathcal{Q}^{-}$.

Our ongoing and future work is devoted to the following issues. The first concerns richer query syntax. As the normal form and the universal model property of Horn-SHIQ carry over to unions of CQs and the more general positive existential queries (PQs), our results can be immediately extended to them. In fact, answering PQs in Horn- $\mathcal{S H I} \mathcal{I}$ is reducible to answering at most exponentially many (in the size of the PQ) CQs. Further, since the universal model property allows us to precompile a knowledge base $\mathcal{K}$ into a (query-independent) domino system $\mathcal{D}_{\mathcal{K}}$ for on-line query answering, the identification of cases in which $\mathcal{D}_{\mathcal{K}}$ is small would be beneficial. Finally, an obvious issue is a detailed study of other fragments of Horn- $\mathcal{S H} \mathcal{I} \mathcal{Q}$ besides Horn- $\mathcal{S H} \mathcal{Q}^{-}$. The effect of syntactic restrictions similarly as in [3] on data complexity is here of particular interest.

## References

1. F. Baader. Terminological cycles in a description logic with existential restrictions. In: IJCAI'03, pages 325-330, 2003.
2. A. Cali, G. Gottlob, M. Kifer. Taming the Infinite Chase: Query Answering under Expressive Relational Constraints. In: KR 2008, to appear, 2008.
3. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Data complexity of query answering in description logics. In: KR 2006, 2006.
4. D. Calvanese, G. De Giacomo, and M. Lenzerini. On the decidability of query containment under constraints. In: PODS'98, pages 149-158, 1998.
5. D. Calvanese, G. D. Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Tractable reasoning and efficient query answering in description logics: The $D L$-lite family. J. Autom. Reasoning, 39(3):385-429, 2007.
6. B. Glimm, I. Horrocks, C. Lutz, and U. Sattler. Conjunctive Query Answering for the Description Logic $\mathcal{S H} \mathcal{I} \mathcal{Q}$. In: IJCAI'07, pages 399-404, 2007.
7. I. Horrocks. Optimising Tableaux Decision Procedures for Description Logics. PhD thesis, University of Manchester, 1997.
 CAI 2005, pages 448-453, 2005.
8. U. Hustadt, B. Motik, and U. Sattler. Data complexity of reasoning in very expressive description logics. In: IJCAI 2005, pages 466-471, 2005.
9. A. Krisnadhi and C. Lutz. Data complexity in the $\mathcal{E} \mathcal{L}$ family of description logics. In: DL'07, Bressanone, Italy, CEUR-WS vol. 250, pages 88-99, 2007.
10. M. Krötzsch, S. Rudolph, and P. Hitzler. Complexity boundaries for Horn description logics. In: $A A A I^{\prime} 07$, pages 452-457, 2007.
11. M. Krötzsch, S. Rudolph, and P. Hitzler. Conjunctive queries for a tractable fragment of OWL 1.1. In: $I S W C / A S W C^{\prime} 07$, pages $310-323,2007$.
12. C. Lutz. Inverse roles make conjunctive queries hard. In: $D L^{\prime} 07,2007$.
13. C. Lutz. Two upper bounds for conjunctive query answering in $\mathcal{S H} \mathcal{I} \mathcal{Q}$. In: $D L^{\prime} 08$.
14. B. Motik, R. Shearer, and I. Horrocks. Optimized reasoning in description logics using hypertableaux. In: $C A D E^{\prime} 07$, LNCS 4603, pages 67-83. Springer, 2007.
15. M. Ortiz, M. Šimkus, and T. Eiter. Worst-case optimal conjunctive query answering in description logics without inverses. In: AAAI' 08 , to appear, 2008.
16. R. Rosati. On conjunctive query answering in $\mathcal{E} \mathcal{L}$. In: $D L^{\prime} 0^{\prime} y_{,} 2007$.
17. S. Rudolph, M. Krötzsch, and P. Hitzler. Terminological reasoning in $\mathcal{S H \mathcal { I } \mathcal { Q } \text { with }}$ ordered binary decision diagrams. In: $A A A I^{\prime} 08$, to appear, 2008.

[^0]:    * This work was partially supported by the Austrian Science Fund (FWF) grant P20840, Wolfgang Pauli Institute (WPI), and the Mexican National Council for Science and Technology (CONACYT) grant 187697.

[^1]:    ${ }^{3}$ W.l.o.g, no individuals occur in $q$; we can replace each $a$ in $q$ by a new variable $y$, add $C_{a}(y)$ to $q$ and $a: C_{a}$ to $\mathcal{A}$, where $C_{a}$ is a new concept name. Non-Boolean queries (i.e., with answer variables) can be reduced to Boolean queries as usual.

[^2]:    ${ }^{4}$ Unlike $[4,6,14]$, we do not use treeifications to reduce CQ entailment to concept satisfiability, as this would require the use of role conjunction and the decidability of this extension of Horn- $\mathcal{S H \mathcal { I } \mathcal { Q }}$ in ExpTime is not apparent.
    ${ }^{5}$ As implicit in [6], such a $q^{\prime}$ with $\left|q^{\prime}\right| \leq 2|q|$ exists: to obtain a treeification from a match, one replaces each atom $R(x, y) \in q$ with a pair of atoms in case $x, y$ are mapped (i) to the same node, or (ii) to nodes in different branches of the domino tree.

[^3]:    ${ }^{6}$ It is irrelevant for query answering, but could be emulated e.g. using node labels $C_{a}$.
    ${ }^{7}$ Note that the TBox internalization of [8] is not adequate for Horn- $\mathcal{S H I \mathcal { L }}$, and that the other rules of $[8]$ are not necessary due to the normal form of the KB.

[^4]:    ${ }^{8}$ Note that in case of inconsistent KBs, query entailment is trivial.

