Query Rewriting for Horn-$SHI\Omega$ plus Rules

Thomas Eiter$^1$, Magdalena Ortiz$^1$, Mantas Šimkus$^1$, Trung-Kien Tran$^2$ and Guohui Xiao$^1$

$^1$Institute of Information Systems, Vienna University of Technology
$\{\text{eiter—ortiz—xiao}\}@\text{krtu.wien.ac.at simkus}@\text{dbai.tuwien.ac.at}$

$^2$STARLab, Vrije Universiteit Brussel
$\text{trantr@vub.ac.be}$

Abstract

Query answering over Description Logic (DL) ontologies has become a vibrant field of research. Efficient realizations often exploit database technology and rewrite a given query to an equivalent SQL or Datalog query over a database associated with the ontology. This approach has been extensively studied for conjunctive query answering in the $DL\text{-Lite}$ and $EL$ families, but is much less explored for more expressive DLs and queries. We present a rewriting-based algorithm for conjunctive query answering over Horn-$SHI\Omega$ ontologies, possibly extended with recursive rules under limited recursion as in $DL+\log$. This setting not only subsumes both $DL\text{-Lite}$ and $EL$, but also yields an algorithm for answering (limited) recursive queries over Horn-$SHI\Omega$ ontologies (an undecidable problem for full recursive queries). A prototype implementation shows its potential for applications, as experiments exhibit efficient query answering over full Horn-$SHI\Omega$ ontologies and benign downsampling to $DL\text{-Lite}$, where it is competitive with comparable state of the art systems.

Introduction

Description Logics (DLs) are the primary tool for representing and reasoning about knowledge given by an ontology. They are mostly fragments of first-order logic with a clear-cut semantics, convenient syntax and decidable reasoning, performed by quite efficient algorithms. This has led to important applications of DLs in areas like Ontology Based Data Access (OBDA), Data Integration and the Semantic Web, where the OWL standard is heavily based on DLs.

An important reasoning task in DLs is query answering similar as in databases, where a database-style query is evaluated over an ontology, viewing it as an enriched database.

Example 1. Consider the following sociopolitical ontology. The Human Development Index (HDI) of certain territories $T$, whose value $V$ may be low, medium, high or very high (as in the UN Development Programme) is given by facts hasHDI$(T, V)$. Further facts classify territories as cities, countries, etc. and relate their locations. The facts are shown in the two left columns of Table 1. The axioms (a)–(e) on the right hand side provide a terminology (in DL syntax) stating that: (a) the isLocatedIn relation is transitive; (b) the capital of a territory is located in that territory; (c) every country has a capital; (d) only cities can be capitals; and (e) only one capital can be located in each country. The query $q_1$ can be used to retrieve disadvantaged territories that lie in countries with high HDI but have a low HDI themselves. Observe that if we evaluate $q_1$ over the database (i.e., the facts), it returns no answer: indeed, Mexico is the only country with high HDI, and there is no fact isLocatedIn$(X, \text{Mexico})$ such that territory $X$ has low HDI. However, if we evaluate $q_1$ over the full ontology, we can infer from axiom (a) that Carichi is located in Mexico, and return (Carichi, Mexico) as an answer. The query $q_2$, which retrieves countries whose capital city has a high HDI, would also have an empty answer over the database, but from the axioms (b)–(e) we can infer that Brasilia is the capital of Brazil and Islamabad the capital of Pakistan, and return both countries as an answer to the query.

To supply this reasoning service, a number of challenges must be faced. Conjunctive queries (CQs) have typically much higher complexity than standard reasoning in a DL, and recursive Datalog queries are undecidable even in very weak DLs, including the ones considered here (Levy and Rousset 1998). For reasoning with large instance data, translating queries into database query languages has proved to be efficient. Calvanese et al. (2007) introduced a query rewriting technique for the $DL\text{-Lite}$ family of DLs, where the terminological information is incorporated into the query in such a way that it can be straight evaluated over the database facts. For example, a rewriting of query $q_1$ in Table 1 should include, among other queries,

$$\text{disadvantagedTerritory}(x, y) \leftarrow \text{hasHDI}(x, \text{low}), \text{country}(y), \text{hasCapital}(y, x), \text{hasHDI}(y, \text{high})$$

which adds all tuples $(x, y)$ to the query answer that can be inferred using axiom (b). Such rewriting approaches have been developed for answering CQs in DLs of the $DL\text{-Lite}$ family, and to a lesser extent for $EL$, but they are practically unexplored for more expressive DLs and queries (see Related Work for details).

In this paper we present a rewriting-based method for query answering over ontologies in Horn-$SHI\Omega$ (the disjunction-free fragment of $SHI\Omega$). This DL extends $DL\text{-Lite}$ and $EL$, two prominent DLs closely related to the OWL 2 QL and the OWL 2 EL profiles, respectively, which offer different expressiveness while allowing for tractable reasoning. For example, axiom (b) is allowed in most DLs of the $DL\text{-Lite}$ family but not in $EL$, while (c) is allowed in
$\mathcal{L}$ but not in $DL$-Lite. Axioms (a), (d) and (e) are not expressible in either of them, but they are expressible in Horn-$SHIQ$. Despite the increase in expressivity, reasoning in Horn-$SHIQ$ is still tractable in data complexity.

In this paper, we make the following contributions:

- We provide a practical algorithm for rewriting queries over Horn-$SHIQ$ ontologies. It first applies a special resolution calculus, and then rewrites the query w.r.t. the saturated TBox into a Datalog program ready for evaluation over any ABox. It runs in polynomial time in data complexity, and thus is worst-case optimal.

- It can handle CQs and the more general weakly $DL$-safe Datalog queries in the style of $DL+log$ (Rosati 2006), where only existentially quantified variables may be bound to 'anonymous' domain elements implied by axioms.

- It is, to our knowledge, the first rewriting algorithm that supports transitive roles, which are considered relevant in practice (Sattler 2000), although challenging for query answering (Glimm et al. 2006, Eiter et al. 2009). It simultaneously allows for full existential quantification, inverse roles, and number restrictions, covering and extending the OWL2 profiles QL and RL, as well as a large fragment of EL.

- A prototype implementation for CQ answering (without transitive roles) shows that our approach behaves well in practice. In experiments it worked efficiently and it scaled down nicely to $DL$-Lite, where it is competitive with state of the art query rewriting systems.

Due to space constraints, we only provide proof sketches of the central results, and refer the reader to (Eiter et al. 2012b) for more details.

### Description Logic Horn-$SHIQ$

As usual, we assume countably infinite sets $\mathbb{N}_C \supset \{\top, \bot\}$ and $\mathbb{N}_R$ of atomic concepts and role names, respectively. $\mathbb{N}_R \cup \{\neg r \mid r \in \mathbb{N}_R\}$ is the set of roles. If $r \in \mathbb{N}_R$, then $\text{inv}(r) = r^{-}$ and $\text{inv}(r^{-}) = r$. Concepts are inductively defined as follows: (a) each $A \in \mathbb{N}_C$ is a concept, and (b) if $A \subseteq C \in \mathbb{N}_C$, $\neg C \in \mathbb{N}_C$, $\forall r.C \in \mathbb{N}_C$, $\exists r.C \in \mathbb{N}_C$, $n \in \mathbb{N}$, then $A \subseteq C \in \mathbb{N}_C$, is a concept. An expression $C \subseteq D$, where $C, D$ are concepts, is a general concept inclusion axiom (GCI). An expression $r \subseteq_\mathbb{T} s$, where $r, s$ are roles, is a role inclusion (RI). A transitivity axiom is an expression $\text{trans}(r)$, where $r$ is a role. A TBox $\mathcal{T}$ is a finite set of GCI, RI and transitivity axioms. We let $\subseteq_\mathbb{T}$ denote the reflexive transitive closure of $\{(r, s) \mid r \subseteq_\mathbb{T} s \in \mathcal{T} \lor \text{inv}(r) \subseteq_\mathbb{T} \text{inv}(s) \in \mathcal{T}\}$. A role $r$ is transitive in $\mathcal{T}$ if $\text{trans}(s) \in \mathcal{T}$ or $\text{trans}(s^{-}) \in \mathcal{T}$. A role $s$ is simple in $\mathcal{T}$ if there is no transitive role $r$ in $\mathcal{T}$ s.t. $r \not\supseteq_\mathbb{T} s$. $\mathcal{T}$ is a $SHIQ$ terminology if roles in concepts of the form $\equiv r.C$ and $\equiv n \cdot r.C$ are simple. The semantics for TBoxes is given by interpretations $I = (\Delta^I, \mathcal{T})$. We write $I \models \mathcal{T}$ if $I$ is a model of $\mathcal{T}$. See (Baader et al. 2007) for more details.

A TBox $\mathcal{T}$ is a Horn-$SHIQ$ TBox (in normalized form), if every GCI in $\mathcal{T}$ takes one of the following forms:

$$
\begin{align*}
(F1) & \quad A_1 \sqcap \ldots \sqcap A_n \sqsubseteq B, \\
(F2) & \quad A_1 \sqsubseteq \forall r.B, \\
(F3) & \quad A_1 \sqsubseteq \exists r.B, \\
(F4) & \quad A_1 \sqsubseteq \equiv 1 \cdot r.B,
\end{align*}
$$

where $A_1, \ldots, A_n, B$ are concept names and $r$ is a role. Axioms (F2) are called existential. W.l.o.g. we treat here only Horn-$SHIQ$ TBoxes in normalized form; our results generalize to full Horn-$SHIQ$ by means of TBox normalization; see e.g. (Kazakov 2009; Krötzsch, Rudolph, and Hitzler 2007) for a definition and normalization procedures.

An Horn-$ALCHIQ$ TBox is a Horn-$SHIQ$ TBox with no transitivity axioms. Horn-$ALCHIQ$ TBoxes are obtained by allowing role conjunction $r_1 \sqcap r_2$, where $r_1, r_2$ are roles (we use it for a similar purpose as Glimm et al. (2008)). In any interpretation $I$, $(r_1 \sqcap r_2)^I = r_1^I \cap r_2^I$. We let $\text{inv}(r_1 \sqcap r_2) = \text{inv}(r_1) \sqcap \text{inv}(r_2)$ and assume w.l.o.g. that for each role inclusion $r \sqsubseteq s$ of an Horn-$ALCHIQ$ TBox $\mathcal{T}$, (i) $\text{inv}(r) \sqsubseteq \text{inv}(s)$ in $\mathcal{T}$, and (ii) $s \in \{p, p^{-}\}$ for a role name $p$. For a set $W$ and a concept or role conjunction $\Gamma = \gamma_1 \sqcap \ldots \sqcap \gamma_m$, we write $\Gamma \sqsubseteq W$ for $\{\gamma_1, \ldots, \gamma_m\} \subseteq W$.

### Ontologies and Knowledge Bases

Following (Levy and Roussel 1998) we now define knowledge bases (KBs). Let $N_C$, $N_R$, and $N_P$ be countable infinite sets of constants (or, individuals), variables and Datalog relations, respectively; we assume these sets as well as $N_C$ and $N_P$ are mutually disjoint. Each $\sigma \in N_P$ has an associated non-negative integer arity. An atom is an expression $p^I(\bar{t})$, where $\bar{t} \in (N_I)^n \cup (N_V)^n$, and (i) $p \in N_C$ and $n = 1$, (ii) $p \in N_R$ and $n = 2$, or (iii) $p \in N_P$ and $n$ is the arity of $p$. If $\bar{t} \in (N_I)^n$, then $p^I(\bar{t})$ is ground. Ground atoms $A(a)$ and $r(a, b)$, where $A \subseteq \mathbb{N}_C$ and $r$ is a role, are concept and role assertions, respectively. An ABox $\mathcal{A}$ is a finite set of ground atoms. A rule $\rho$ is an expression of the form

$$h(\bar{u}) \leftarrow p_1(\bar{v}_1), \ldots, p_m(\bar{v}_m) \quad (1)$$

where $h(\bar{u})$ is an atom (the head), $\{p_1(\bar{v}_1), \ldots, p_m(\bar{v}_m)\}$ are also atoms (the body atoms, denoted $\text{body}(\rho)$), and $\bar{u}, \bar{v}_1, \ldots, \bar{v}_m$ are tuples of variables. The variables in $\bar{u}$ are distinguished. A KB is a tuple $K = (\mathcal{T}, \mathcal{A}, \mathcal{P})$, where $\mathcal{T}$ is a TBox, $\mathcal{A}$ is an ABox, and $\mathcal{P}$ is a set of rules (a program).

<table>
<thead>
<tr>
<th>village(Carichi)</th>
<th>hasHDl(Carichi, low)</th>
<th>(a) trans(isLocatedIn)</th>
<th>(c) country $\sqsubseteq \exists$hasCapital.capital</th>
</tr>
</thead>
<tbody>
<tr>
<td>state(Chihuahua)</td>
<td>hasHDl(Mexico, high)</td>
<td>(b) hasCapital $\sqsubseteq$ isLocatedIn$^{-}$</td>
<td>(d) country $\sqsubseteq \exists$hasCapital.city</td>
</tr>
<tr>
<td>country(Mexico)</td>
<td>hasHDl(Islamabad, high)</td>
<td>(e) country $\sqsubseteq \exists$isLocatedIn$^{-}$capital</td>
<td></td>
</tr>
<tr>
<td>capital(China)</td>
<td>hasHDl(China, high)</td>
<td>(q1) disadvantagedTerritory(x, y) $\leftarrow$ hasHDl(x, low), isLocatedIn(x, y), country(y), hasHDl(y, high)</td>
<td></td>
</tr>
<tr>
<td>country(China)</td>
<td>isLocatedIn(Carichi, Chihuahua)</td>
<td>(q2) hasDevelopedCapital(x) $\leftarrow$ country(x), hasCapital(x, y), city(y), hasHDl(y, high)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: An example ontology and queries

727
The semantics for a KB $\mathcal{K} = (T, \mathcal{A}, \mathcal{P})$ is given by extending an interpretation $\mathcal{I}$ to symbols in $N_0 \cup N_0$. For any $c \in N_1$ and $p \in N_0$ of arity $n$, we have $c^2 \in \Delta^2$ and $p^2 \subseteq (\Delta^2)^n$. A match in $\mathcal{I}$ for a rule $\rho$ of the form (1) is a mapping from variables in $p$ to elements in $\Delta^2$ such that $\pi(\vec{t}) \in p^2$ for each body atom $p(\vec{t})$ of $\rho$. We define:

(a) $\mathcal{I} \models \rho$ if $\pi(\vec{u}) \in h^2$ for every match $\pi$ for $\rho$ in $\mathcal{I}$.
(b) $\mathcal{I} \models \emptyset$ if $\mathcal{I} \models \rho$ for each $\rho \in \mathcal{P}$.
(c) $\mathcal{I} \models \mathcal{A}$ if $(\vec{c}) \in p^2$ for all $p(\vec{c}) \in \mathcal{A}$.
(d) $\mathcal{I} \models \mathcal{K}$ if $\mathcal{I} \models T$, $\mathcal{I} \models \mathcal{A}$, and $\mathcal{I} \models \emptyset$.

Finally, given a ground atom $p(\vec{c})$, $\mathcal{K} \models \rho(\vec{c})$ if $(\vec{c}) \in p^2$ for all models $\mathcal{I}$ of $\mathcal{K}$. We recall weak DL-safety (Rosati 2006). A KB $\mathcal{K} = (T, \mathcal{A}, \mathcal{P})$ is weakly DL-safe if each rule $\rho \in \mathcal{P}$ satisfies the next condition: every distinguished variable $x$ of $\rho$ occurs in some body atom $p(\vec{t})$ of $\rho$ such that $p \in N_0$. We make the Unique Name Assumption (UNA).

A KB $\mathcal{K} = (T, \mathcal{A}, \emptyset)$ is an ontology (for brevity, we use $\mathcal{O} = (T, \mathcal{A})$). A conjunctive query (CQ) $q$ over $\mathcal{O}$ is a rule of the form (1) such that $h$ does not occur in $\mathcal{O}$. By ans($\mathcal{I}, q$) we denote the set of all $\vec{c} \in N_1[\vec{c}]$ such that there is a match $\pi$ for $q$ with $\pi(\vec{u}) = (\vec{c})^2$. By ans($\mathcal{O}, q$) we denote the answer to $q$ over $\mathcal{O}$ defined as the set of all $\vec{c} \in N_1[\vec{c}]$ such that $\pi(\vec{c}) \in \mathcal{I}$ for every model $\mathcal{I}$ of $\mathcal{O}$.

Note that, for a KB $\mathcal{K} = (\emptyset, \mathcal{A}, \mathcal{P})$, $\mathcal{A} \cup \mathcal{P}$ is an ordinary Datalog program with constraints (cf. (Dantsin et al. 2001)). By models of Datalog programs, we mean Herbrand models, and we recall that a consistent $\mathcal{A} \cup \mathcal{P}$ has a unique least (Herbrand) model $MM(\mathcal{A} \cup \mathcal{P})$.

We also consider programs $\mathcal{P}$ with atoms $r^-(x, y)$, $r \in N_R$. Its semantics is given by the semantics of the program $\mathcal{P}'$ obtained by replacing each $r^-(x, y)$ by $r(y, x)$.

**Elimination of Transitivity Axioms** It is handy to eliminate transitivity axioms from $\mathcal{SHIQ}$ TBoxes (see, e.g., Hustadt et al. 2007)). We use the transformation from (Kazakov 2009), which ensures the resulting TBox is in normal form.

**Definition 1.** Let $T^*$ be the Horn-$\mathcal{ALCHIQ}$ TBox obtained from a Horn-$\mathcal{SHIQ}$ TBox $T$ by (i) adding for every $A \subseteq \forall s. B \in T$ and every transitive role $r$ with $r^+ \subseteq s$, the axioms $A \sqsubseteq \forall r. B$, $B^r \sqsubseteq \forall r. B^r$ and $B^r \subseteq B$, where $B^r$ is a fresh concept name; and (ii) removing all transitivity axioms.

As the transformation does not preserve answers to CQs, we relax the notion of TBoxes.

**Definition 2.** Let $T$ be a Horn-$\mathcal{SHIQ}$ TBox. A T-match for a query $q$ in an interpretation $\mathcal{I}$ is a mapping $\pi$ from variables of $q$ to elements in $\Delta^2$ such that:

(a) if $\alpha = p(\vec{t})$ is a body atom in $q$, where $p \in N_0 \cup N_0$, $p$ is a simple role in $\mathcal{I}$, then $\pi(\vec{t}) \in p^2$, and
(b) if $\alpha = s(x, y)$ with $s$ non-simple, then there exist $d_1 \in \Delta^2$, $\ldots$, $d_k \in \Delta^2$ and a transitive $r^+ \subseteq s$ s.t. $d_1 = \pi(x)$, $d_k = \pi(y)$, and $(d_i, d_{i+1}) \in r^+$ for all $1 \leq i < k$.

The sets ans$^T(\mathcal{I}, q)$ and ans$^T(\mathcal{O}, q)$ are defined as ans$^T(\mathcal{I}, q)$ and ans$^T(\mathcal{O}, q)$ but using T-matches instead of matches. The next claim follows from known techniques, see e.g. (Eiter, Ortiz, and Simkus 2012) for a similar result.

**Proposition 1.** For any Horn-$\mathcal{SHIQ}$ ontology $\mathcal{O} = (T, \mathcal{A})$ and CQ $q$, we have ans$^T(\mathcal{O}, q) = \text{ans}^T((T^*, \mathcal{A}), q)$.

<table>
<thead>
<tr>
<th>$M \subseteq S(N \cap N')$</th>
<th>$N \subseteq A$</th>
<th>$R \subseteq$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M \subseteq S(N \cap N' \cap A)$</td>
<td>$M \subseteq \exists S(N \cap N' \cap A)$</td>
<td>$M \subseteq \exists S(N \cap N' \cap A)$</td>
</tr>
<tr>
<td>$M \subseteq \exists S(N \cap N' \cap A)$</td>
<td>$M \subseteq \exists S(N \cap N' \cap A)$</td>
<td>$M \subseteq \exists S(N \cap N' \cap A)$</td>
</tr>
<tr>
<td>$M \subseteq \exists S(N \cap N' \cap A)$</td>
<td>$M \subseteq \exists S(N \cap N' \cap A)$</td>
<td>$M \subseteq \exists S(N \cap N' \cap A)$</td>
</tr>
</tbody>
</table>

Table 2: Inference rules. $M[\emptyset], N[\emptyset]$ (resp., $S[\emptyset]$) are conjunctions of atomic concepts (roles); $A, B$ are atomic concepts.

**Canonical Models**

A stepping stone for tailoring query answering methods for Horn DLs and languages like Datalog is the canonical model property (Eiter et al. 2008b; Ortiz, Rudolph, and Simkus 2011; Cali, Gottlob, and Lukasiewicz 2009). In particular, for a consistent Horn-$\mathcal{ALCHIQ}$ ontology $\mathcal{O} = (T, \mathcal{A})$, there exists a model $\mathcal{I}$ of $\mathcal{O}$ that can be homomorphically mapped into any other model $\mathcal{I}'$ of $\mathcal{O}$. We show that such an $\mathcal{I}$ can be built in three steps:

2. Close $\mathcal{A}$ under all but existential axioms of $T$.
3. Extend $\mathcal{A}$ by “applying” the existential axioms of $T$.

For Step (1), we tailor from the inference rules in (Kazakov 2009; Ortiz, Rudolph, and Simkus 2010) a calculus to support model building for Horn-$\mathcal{ALCHIQ}$ ontologies.

**Definition 3.** Given a Horn-$\mathcal{ALCHIQ}$ TBox $T$, $\Xi(T)$ is the TBox obtained from $T$ by exhaustively applying the inference rules in Table 2.

For Step (2), we use Datalog rules that express the semantics of GCIs, ignoring existential axioms.

**Definition 4.** Given a Horn-$\mathcal{ALCHIQ}$ TBox $T$, $cr(T)$ is the Datalog program described in Table 3.

Given a consistent Horn-$\mathcal{ALCHIQ}$ ontology $\mathcal{O} = (T, \mathcal{A})$, the least model $\mathcal{J}$ of the Datalog program $cr(T) \cup \mathcal{A}$ is almost a canonical model of $\mathcal{O}$; however, existential axioms may be violated. We deal with this in Step (3), by extending $\mathcal{J}$ with new domain elements as required by axioms $A \subseteq \exists r. N \in \Xi(T)$. This step is akin to the database chase (Maier and Mendelzon 1979), where fresh values and tuples are introduced to satisfy a given set of dependencies.

**Definition 5.** Let $T$ be a Horn-$\mathcal{ALCHIQ}$ TBox and $\mathcal{I}$ an interpretation. A GCI $M \subseteq \exists S.N$ is applicable at $e \in \Delta^2$ if

(a) $e \in M^2$.}
I obtained from (b) chase (a) chase (tology. Then for each atomic $A \subseteq \Lambda B \in \Xi(T)$ $r(x, y) \rightarrow r_1(x, y), \ldots, r_n(x, y)$ for all $r_1 \cap \ldots \cap r_n \subseteq r \in T$ $\perp(x) \rightarrow \langle A(x), r(x, y_1), \ldots, r(x, y_n) \rangle, B(y_1), B(y_2), y_1 \neq y_2$ for each $A \subseteq \Lambda r_1 B \in T$ $\Gamma \leftarrow A(x), A_1(x), \ldots, A_n(x), r(y, B(y))$ for all $A_1 \cap \ldots \cap A_n \subseteq \exists(r_1 \cap \ldots \cap r_n).B_1 \cap \ldots \cap B_n$ and $A \subseteq \Lambda r_1 B \in \Xi(T)$ such that $r=r_1$ and $B=B_1$ for some $i, j$ with $A \subseteq \Lambda \{B_1(y), \ldots, B_n(y), r_1(y, x), \ldots, r_n(x, y)\}$

Table 3: Completion rules $cr(T)$ for Horn-ALCHIQ\(^n\)

(b) there is no $e' \in \Delta^T$ with $(e, e') \in S^T$ and $e' \in N^T$, (c) there is no axiom $M' \sqsubseteq \exists S'.N' \in T$ such that $e \in (M')^T, S \subseteq S', N \subseteq N'$, and $S \subseteq S$ or $N \subseteq N'$.

An interpretation $\mathcal{J}$ obtained from $T$ by an application of an applicable axiom $M \sqsubseteq \exists S.N \in e \in \Delta^T$ is defined as:

- $\Delta^T = \Delta^T \cup \{d\}$ with $d$ a new element not present in $\Delta^T$ (we call $d$ a successor of $e$),
- For each atomic $A \in \mathcal{N}$ and $o \in \Delta^T$, we have $o \in A^T$ if (a) $o \in \Delta^T$ and $o \in A^T$; or (b) $o = d$ and $A \in N$.
- For each role name $r$ and $o, o' \in \Delta^T$, we have $(o, o') \in r^T$ if (a) $o, o' \in \Delta^T$ and $(o, o') \in r^T$; or (b) $(o, o') = (e, d)$ and $r \in S$; or (c) $(o, o') = (d, e)$ and $\text{inv}(r) \in S$.

We denote by $\text{chase}(T, T)$ a possibly infinite interpretation obtained from $T$ by applying the existential axioms in $T$. We require the application to be fair: the application of an applicable axiom cannot be indefinitely postponed.

We note that $\text{chase}(T, T)$ is unique up to renaming of domain elements. As usual in DLs, it can be seen as a ‘forest’: application of existential axioms simply attaches ‘trees’ to a possibly arbitrarily shaped $T$. The following statement can be shown similarly as in (Ortiz, Rudolph, and Simkus 2011).

**Proposition 2.** Let $\mathcal{O} = (T, A)$ be a Horn-ALCHIQ\(^n\) ontology. Then $\mathcal{O}$ is consistent if $A \cup cr(T)$ is consistent. Moreover, if $\mathcal{O}$ is consistent, then (a) $\text{chase}((M \cup \mathcal{A} \cup cr(T^*)), \Xi(T))$ is a model of $O$, and (b) $\text{chase}((M \cup \mathcal{A} \cup cr(T^*)), \Xi(T))$ can be homomorphically mapped into any model of $O$.

In database terms, this means that checking consistency of $\mathcal{O} = (T, A)$ reduces to evaluating the (plain) Datalog query $cr(T)$ over the database $A$. Note that $\Xi(T)$ can be computed in exponential time in size of $T$: the calculus only infers axioms of the form $M \subseteq B$ and $M \subseteq \exists S.N$, where $M$, $N$ are conjunctions of atomic concepts, $B$ is atomic and $S$ is a conjunction of roles. The number of such axiom is single exponential in the size of $T$.

**Rewriting Rules and Programs**

The following is immediate from Propositions 1 and 2:

**Theorem 3.** Let $\mathcal{O} = (T, A)$ be a Horn-SH\IQ\(^n\) ontology. Then $A \cup cr(T^*)$ is consistent iff $\mathcal{O}$ is consistent. Moreover, if $\mathcal{O}$ is consistent, then $\text{ans}(\mathcal{O}, q) = \text{ans}^T(\mathcal{I}_\mathcal{O}, q)$, where $\mathcal{I}_\mathcal{O} = \text{chase}(MM(A \cup cr(T^*)), \Xi(T^*))$.

Computing $\text{ans}^T(\mathcal{I}_\mathcal{O}, q)$ is still tricky because $\mathcal{I}_\mathcal{O}$ can be infinite. So we rewrite $q$ into a set $Q$ of CQs such that $\text{ans}^T(\mathcal{I}_\mathcal{O}, q) = \bigcup_{q' \in Q} \text{ans}^T(MM(A \cup cr(T^*)), q')$. This yields an algorithm for answering $q$ over $\mathcal{O}$, using only the finite $MM(A \cup cr(T^*))$.

The rewriting of $q$ is done in steps with the following intuition. Suppose $q$ has a $T$-match in $\mathcal{I}_\mathcal{O}$. A rewrite step clips off some variable $x$ such that $\pi(x)$ has no descendant in the image of $\pi$, merges the variables that are mapped to the predecessor of $\pi(x)$, and adds concept atoms to the resulting $q'$ that ensure that $q$ has a $T$-match whenever $q'$ does.

**Definition 6.** For a rule $\rho$ and a Horn-SH\IQ\(^n\) TBox $T$, we write $\rho \rightarrow_T \rho'$ if $\rho'$ is obtained from $\rho$ by the following steps:

(S1) Select an arbitrary non-distinctly ordered variable $x$ in $\rho$.
(S2) Replace each role atom $r(x, y)$ in $\rho$, where $y$ is arbitrary and $s$ is non-simple, either leave $\alpha$ untouched or replace it by two atoms $r(y, u_s), r(u_s, x)$, where $u_s$ is a fresh variable and $r$ is a transitive role with $r \subseteq_T s$.
(S3) For each atom $\rho = s(y, x) \in \rho$, where $y$ is arbitrary and $s$ is non-simple, either leave $\alpha$ untouched or replace it by two atoms $r(y, u_s), r(u_s, x)$, where $u_s$ is a fresh variable and $r$ is a transitive role with $r \subseteq_T s$.
(S4) Form some partitioning of the set $\{y \mid \exists \alpha \text{ path of length } \geq 2 \text{ from } y \}$ into sets $V_x$ and $V_y$, in such a way that $x \in V_x$ and $V_y$ has no distinguished variables.
(S5) Select some $M \sqsubseteq \exists S.N \in \Xi(T^*)$ such that (a) $\{y \mid \exists \alpha \text{ path of length } \geq 2 \text{ from } y \} \subseteq S$, (b) $\{A \mid A(z) \in \text{body}(\rho) \land z \in V_y \} \subseteq N$, and (c) for each variable $z \in V_y$ and $r(z, x)$ in $\text{body}(\rho)$ there is a transitive $s \subseteq_T r$ such that (i) $\{s, s^{-1}\} \subseteq S$, or (ii) there is an axiom $M' \sqsubseteq \exists S'.N' \in \Xi(T^*)$ such that $M' \subseteq N$ and $\{s, s^{-1}\} \subseteq S'$.
(S6) Drop each atom from $\rho$ containing a variable from $V_x$.
(S7) Rename each $y \in V_x$ of $\rho$ by $x$.
(S8) Add the atoms $\{A(x) \mid A \in M\}$ to $\rho$.

We write $\rho \rightarrow_T \rho'$ if $\rho'$ can be obtained from $\rho$ by finitely many rewrite iterations. We let $\text{rew}_T(\rho) = \{\rho' \mid \rho \rightarrow_T \rho' \}$.

For a set $\mathcal{P}$ of rules, $\text{rew}_T(\mathcal{P}) = \bigcup_{\rho \in \mathcal{P}} \text{rew}_T(\rho)$.

In (S1) we guess the variable $x$ and, for technical reasons, in (S2) we invert all atoms of the form $r(x, y)$. For all atoms $s(y, x)$ where $s$ is not simple, there must exist a transitive $r \subseteq_T s$ such that there is an $r$-path from $\pi(y)$ to $\pi(x)$. For the atoms where this path has length at least 2, we use in (S3) the role $r$ and introduce an ‘intermediate’ variable $u$ that can be mapped to the parent $p$ of $\pi(x)$. Now we know that for each atom $r(y, x)$ the ‘neighbor’ variable $y$ must mapped to (a) the parent $p$ of $\pi(x)$, or (b) in case $r$ is non-simple, possibly to $\pi(x)$. In (S4) we guess a partition of the neighbor variables into $V_x$ and $V_y$, which correspond to cases (a) and (b), respectively. In (S5) we select some axiom that would cause the existence of $\pi(x)$ when applied at $p$. Then we can clip off $x$ in (S6), merge all variables of $V_y$ (and, for technical reasons, we rename them to $x$, which does not occur in $\rho$ anymore) in (S7), and add to $\rho$ atoms for the concepts on the left hand side of the selected axiom.
Example 2. In our first example, illustrated in Figure 1a, all roles are simple. Let \( \rho : q(x_1) \leftarrow C(x_1), B(x_2), r_1(x_1, x_2), r_2(x_1, x_2), r_3(x_2, x_4), \) and assume \( A \sqsubseteq \exists (r_1 \cap r_2 \cap r_3). B \in \Xi (T^*). \) In (S1) we select the non-distinguished variable \( x_2. \) Next, in (S2), we replace \( r_2(x_2, x_4) \) by \( r_2(x_4, x_2). \) Since all roles are simple, we do nothing in (S3). In (S4) we choose \( V_{x_2} = \{ x_2 \} \) and rename them to \( x_2 \) in (S5), and add \( A(x_2) \) in (S6). Then we clip off \( x_2 \) in (S6), merge all variables in \( V_{x_2} \) and rename them to \( x_2 \) in (S7), and add \( A(x_2) \) in (S8), to obtain \( \rho' : q(x_1) \leftarrow C(x_1), r_1(x_1, x_2), A(x_2). \)

This rewriting also provides a reduction of query answering to reasoning in Datalog, provided that the completion rules take transitivity into account.

Definition 7. For a Horn-\( \mathcal{SHIQ} \) TBox \( T, \) \( cr(T) = cr(T^*) \cup \{ r(x, z) \leftarrow -r(x, y), r(y, z) | r \text{ is transitive in } T \}. \)

Now we can show the following:

Theorem 4. Assume a consistent Horn-\( \mathcal{SHIQ} \) ontology \( O = (T, A) \) and a conjunctive query \( q. \) Then \( ans(O, q) = \bigcup_{q' \in rew_T(q)} ans(T, A \cup cr(T), q'). \)

Proof. (Sketch) Using standard techniques, it is easy to show \( ans(T, A \cup cr(T), q) = ans(T, A \cup cr(T^*), q) \) for every ABox \( A \) and CQ \( q. \) By this and Theorem 3, we only need to show \( ans(T, q) = ans(T, T, rew_T(q)), \) where \( J = MM(A \cup cr(T^*)) \) and \( T_O = chase(T, \Xi(T^*)). \)

To show \( ans(T, T_O, q) \supseteq ans(T, T, rew_T(q)) \), then there is a query \( q' \in rew_T(q) \) and a \( \mathcal{SHIQ} \)-match \( \pi_{q'} \) for \( q' \) in \( T \) such that \( \hat{u} \in \pi_{q'}(\hat{x}). \) By the construction of \( T_O, \) we also have \( \hat{u} \in ans(T, T_O, q). \) If \( q' \neq q \) (otherwise we are done), there is \( n > 0 \) such that \( q_0 \rightarrow_T q_1, \ldots, q_{n-1} \rightarrow_T q_n \) with \( q_0 = q \) and \( q_n = q'. \) Thus to prove the claim it suffices to show that \( \hat{u} \in ans(T, T_O, q_i) \) implies \( \hat{u} \in ans(T, T_O, q_{i+1}), \) where \( 0 < i < n. \)

Assume that \( q_i \) was obtained from \( q_{i-1} \) by a rewriting step, where some variable \( x \) was chosen in (S1), some \( V_x \) and \( V_p \) in (S4), and some axiom \( M \sqsubseteq \exists S.N \in (S5). \) Suppose \( \pi_{q_i} \) is a \( \mathcal{SHIQ} \)-match for \( q_i \) in \( T_O \) with \( \bar{u} = \pi_{q_i}(\bar{x}), \) i.e. \( \bar{u} \in ans(T, T_O, q_i). \) Let \( d = \pi_{q_i}(x). \) Due to step (S8) in the rewriting and the fact that \( T_O \) is a model of \( O, \) we have \( d \in (\exists S.N)^T. \) Then there is \( d' \in \Delta^T \) such that \( (d, d') \in S \) and \( d' \neq d \). One can show that the mapping \( \pi_{q_{i-1}} \) defined as follows is a \( \mathcal{SHIQ} \)-match for \( q_{i-1} \) in \( T_O: \)

\[ \pi_{q_{i-1}}(z) = d' \] for all \( z \in V_z, \pi_{q_{i-1}}(u) = d \) for all \( u \in V_p, \) and \( \pi_{q_{i-1}}(z) = \pi_{q_i}(z) \) for the remaining variables \( z. \)

Thus the \( ans(T, T_O, q) \subseteq ans(T, T, rew_T(q)) \) we prescribe the naming of fresh domain elements introduced when chasing \( J \) w.r.t. \( \Xi(T^*), \) and use an element \( e \cdot n \) for some integer \( n \) for each successor of an element \( e. \) For \( d \in \Delta^T \) we let \(|d| = 0, \) and for \( w \cdot n \in \Delta^T \) we let \(|w \cdot n| = |w| + 1. \) Assume a tuple \( \bar{u} \in ans(T, q). \) By definition, there is \( \mathcal{T} \)-match \( \pi_q \) for \( q \) in \( T_O \) such that \( \bar{u} = \pi_q(\bar{x}). \) We have to show that there exists \( q' \in rew_T(q) \) and a \( \mathcal{T} \)-match \( \pi_{q'} \) for \( q' \) in \( J \) such that \( \bar{u} = \pi_{q'}(\bar{x}). \)

For a \( \mathcal{T} \)-match \( \pi_q \) in \( T_O, \) let \( deg(q') = \sum_{y \in var(q')} |\pi'_y(y)|. \) Then, given that \( q \in rew_T(q), \) to prove the claim it suffices to prove the following statement: if \( q_1 \in rew_T(q) \) has a \( \mathcal{T} \)-match \( \pi_{q_1}, \) for \( q_1 \in T_O \) such that \( \bar{u} = \pi_{q_1}(\bar{x}) \) and \( deg(\pi_{q_1}) > 0, \) then there exists \( q_2 \in rew_T(q) \) that has a \( \mathcal{T} \)-match \( \pi_{q_2} \) for \( q_2 \) in \( T_O \) such that \( \bar{u} = \pi_{q_2}(\bar{x}) \) and \( deg(\pi_{q_2}) < deg(\pi_{q_1}). \)

The query \( q_2 \) is obtained by selecting for \( (S1) \) an \( x \) such that \( \pi_{q_1}(x) \notin N_1 \) (which exists because \( deg(\pi_{q_1}) > 0) \) and there is no variable \( x' \) of \( q_1 \) with \( \pi_{q_1}(x) \) a prefix of \( \pi_q(x') \) (that is, \( x \) is a leaf in the subforest of \( T_O \) induced by the image of \( \pi_{q_1}. \) Let \( d_x = \pi_{q_1}(x), \) and let \( d_x \) be the parent of \( d_x, \) i.e. \( d_x = d_y \cdot n \) for some integer \( n. \) For Step (S3), we choose to rewrite the atoms \( \Gamma = \{ s(y), x \in q_1 \mid \pi_{q_1}(y) \neq d_x \wedge \pi_{q_1}(y) \neq d_x \}. \) By definition of \( \mathcal{T} \)-match, for each such atom there is a transitive role \( r \in \Delta^T \) such that there is an \( r \)-path from \( \pi_{q_1}(y) \) to \( \pi_{q_1}(x); \) this \( r \) can be used in the fresh atoms. Let \( V_x \) be the set of fresh variables introduced in this step. For Step (S4), let \( V_x = \{ z \in var(q_1) | \pi_{q_1}(z) = d_x \} \) and \( V_y = \{ z \in var(q_1) | \pi_{q_1}(z) = d_y \} \cup V_x. \) We know from the construction of \( T_O \) that \( d_x \) was introduced by an application of an axiom \( ax = M \subseteq \exists S.N \in \Xi(T^*) \) such that \( d_p \in M \subseteq \Xi(T^*); \) we choose this axiom for Step (S5). It is not hard to show that \((S5.a-c) \) hold. Finally, a \( \mathcal{T} \)-match \( \pi_{q_2} \) for \( q_2 \in T_O \) such that \( \bar{u} = \pi_{q_2}(\bar{x}) \) and \( deg(\pi_{q_2}) < deg(\pi_{q_1}) \) is obtained from \( \pi_{q_1} \) by setting \( \pi_{q_2}(z) = \pi_{q_1}(z) \) for all \( z \neq x, \) and \( \pi_{q_2}(x) = d_p. \)

By the above, we can answer \( q \) over \( O = (T, A) \) by posing \( rew_T(q) \) over the Datalog program \( A \cup cr(T). \) The method also applies to KBs \( K = (T, A, P), \) where \( T \) is in Horn-\( \mathcal{SHIQ} \) and \( P \) is weakly DL-safe. The ground atomic consequences of \( K \) can be collected by fixed-point compu-
Theorem 5. For a ground atom \( \alpha \) over a KB \( K = (T, A, P) \) where \( T \) is a Horn-SHIQ TBox and \( P \) is weakly DL-safe, we have \((T, A, P) \models \alpha \) if \( cr(T) \cup rew_T(P) \cup A \models \alpha \).

Proof. (Sketch) Let \( P' = cr(T) \cup rew_T(P) \). For the “if” direction, the only interesting case is when \( \rho^* = \alpha \) is consistent. In this case it suffices to show that the rules of \( P' \) applied on \( A \) derive only atoms that are consequences of \((T, A, P) \). This is straightforward for all rules in \( cr(T) \), since the rules are already logical consequences of \( T \). For the rules in \( \rho' \in rew_T(P) \) it is a consequence of Theorem 4.

To show the “only if” direction, again the only interesting case is where \( cr(T) \cup rew_T(P) \cup A \) is consistent. In this case one can consider the set \( A' \) of all ground \( \alpha \) such that \( cr(T) \cup rew_T(P) \cup A \models \alpha \), and show that \( I = \text{chase}(A', \Xi(T')) \) is a model of \((T, A, P) \) such that \( I \nmodels \alpha \) for all \( \alpha \) such that \( cr(T) \cup rew_T(P) \cup A \nmodels \alpha \).

This reduction yields a worst-case optimal algorithm.

Theorem 6. For a ground atom \( \alpha \) over a KB \( K = (T, A, P) \) where \( T \) is a Horn-SH IQ TBox and \( P \) is weakly DL-safe, checking \((T, A, P) \models \alpha \) is ExpTime-complete in general, and PTime-complete in data complexity.

Proof. (Sketch) By Theorem 5, checking \((T, A, P) \models \alpha \) is equivalent to deciding \( cr(T) \cup rew_T(P) \cup A \models \alpha \). We analyze the computational cost of the latter check.

We recall that \( \Xi(T') \) can be computed in exponential time in size of \( T \) and is independent from \( A \). The program rewriting \( rew_T(P) \) is finite and computable in time exponential in the size of \( T \) and \( P \); rules in \( rew_T(\rho) \), where \( \rho \in P \), use only relation names and variables that occur in \( T \) and \( P \) (fresh variables introduced in \( S3 \) are eliminated in \( S6 \) and \( S7 \)). Hence, the size of each rule resulting from a rewrite step is of size polynomial in the size of \( T \) and \( P \), and thus the number of rules in \( rew_T(P) \) is at most exponential in the size of \( T \) and \( P \). The size of \( rew_T(P) \) is constant when data complexity is considered. Furthermore, the grounding of \( cr(T) \cup rew_T(P) \cup A \) is exponential in the size of \( K \), but polynomial for fixed \( T \) and \( P \). By the complexity of Datalog, it follows that the algorithm resulting from Theorem 5 is exponential in combined but polynomial in data complexity. These results are worst-case optimal, and apply already to plain conjunctive queries (Eiter et al. 2008b).

Both \( \Xi(T') \) and \( rew_T(P) \) can be of exponential size, but this worst-case complexity is only exhibited by some ‘hard’ instances. Our experimental results in the next Section show that both sets are of manageable size for many ontologies.

Implementation and Experiments

The results of the previous Section directly give an algorithm for answering a CQ \( q \) over a given Horn-SHIQ ontology \( O = (T, A) \), which works as follows: (1) we eliminate transitivity axioms in \( T \) to get \( T^* \); (2) we saturate \( T^* \) into

\[
\Xi(T') \text{ using the calculi in Table 2; (3) we obtain } rew_T(\rho) \text{ by exhaustively applying to } \rho \text{ the rewriting step in Definition 6 using the axioms } \Xi(T'); (4) we put together } A, \text{ the completion rules } cr(T), \text{ and } rew_T(\rho) \text{ into a Datalog program } P; \text{ and (5) we evaluate the program } P.
\]

To evaluate the feasibility of this algorithm, we have implemented a prototype system CliPer\(^1\) for answering CQs containing only simple roles. To the best of our knowledge, it is the first such system for full Horn-SHIQ (under the standard semantics of first-order logic), and in expressiveness subsumes similar DL-Lite and EL reasoning engines.

CliPer is implemented in Java and uses OWLAPI 3.2.2 (Horridge and Bechhofer 2011) to manage ontologies. It accepts an ontology \( O = (T, A) \) and a query \( q \) in SPARQL syntax as input. Initial preprocessing involves normalization of \( O \) and checking that \( T \) is Horn. Then it applies steps (1) – (5) above (with some minor optimizations). For the Datalog evaluation in step (5), it uses DLV-20101014 (Leone et al. 2006) or Clingo 3.0.3 (Gebser et al. 2011).

Experiments

We tested our CliPer system on a Pentium Core2 Duo 2.00GHZ with 2GB RAM under Ubuntu 10.04 and 512MB heap size for the Java VM. We conducted the following experiments.

\(^1\)http://www.kr.tuwien.ac.at/research/systems/cliPer/
Table 6: Queries over Horn-SHIQ ontology

<table>
<thead>
<tr>
<th>Query</th>
<th># Rules</th>
<th>Rewriting (ms)</th>
<th>Datalog (DLV) Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>2</td>
<td>68</td>
<td>80 / 320 / 560 / 830</td>
</tr>
<tr>
<td>Q2</td>
<td>3</td>
<td>63</td>
<td>90 / 330 / 560 / 830</td>
</tr>
<tr>
<td>Q3</td>
<td>9</td>
<td>96</td>
<td>90 / 320 / 570 / 810</td>
</tr>
<tr>
<td>Q4</td>
<td>172</td>
<td>143</td>
<td>230 / 830 / 1430 / 1580</td>
</tr>
<tr>
<td>Q5</td>
<td>16</td>
<td>91</td>
<td>90 / 330 / 570 / 820</td>
</tr>
<tr>
<td>Q6</td>
<td>255</td>
<td>177</td>
<td>250 / 890 / 1530 / 1800</td>
</tr>
<tr>
<td>Q7</td>
<td>8</td>
<td>89</td>
<td>80 / 320 / 570 / 820</td>
</tr>
<tr>
<td>Q8</td>
<td>175</td>
<td>146</td>
<td>230 / 830 / 1430 / 1580</td>
</tr>
<tr>
<td>Q9</td>
<td>175</td>
<td>145</td>
<td>230 / 820 / 1400 / 1600</td>
</tr>
<tr>
<td>Q10</td>
<td>2</td>
<td>64</td>
<td>80 / 330 / 570 / 830</td>
</tr>
</tbody>
</table>

Table 7: Experiment with UOBM Horn-SHIQ ontology

1. Downscaling test. We compared CLIPPER with other query rewriting systems for DL-Lite, viz. REQUIEM (Perez-Urbina et al. 2009) and PRESTO (Rosati and Almatelli 2010), and found that it is competitive and scales down well on DL-Lite ontologies. We used the ontologies ADOLENA (A), STOCK-EXCHANGE (S), VICODI (V), and UNIVERSITY (U), and queries (Q1–Q5) from the REQUIEM test suite, which have been widely used for system tests. In addition, we considered the queries in Table 4.

The number of rewritten queries and rewriting time are shown in Table 5 (loading and preprocessing times are excluded). CLIPPER and PRESTO generated in most cases rule sets of comparable size, and in short time; in a few cases PRESTO generated significantly less rules than CLIPPER. REQUIEM, in its G-version (which generates optimized rule sets), generated in several cases significantly more rules. This is largely explained by rule unfolding to produce CQs; still in many cases, in particular for S and U except Q7, the final result is small (but needed considerably more time).

For UNIVERSITY, we evaluated the rewritten queries over 4 different ABoxes with 67k to 320k assertions using DLV (the other ontologies in the suite don’t have ABoxes). Interestingly, in all cases the execution times for the three rewritings were very similar; the average runtime of each query on the 4 ABoxes is shown in brackets.

2. Full Horn-SHIQ. To test CLIPPER on a full Horn-SHIQ ontology, we modified the UOBM ontology (Ma et al. 2006), which is in SHOIN(D), by dropping or strengthening (in case of disjunctions) non-Horn-SHIQ TBox axioms; the final ontology has 171 TBox axioms. We used ABoxes $A_i$, $1 \leq i \leq 4$, with 20k, 80k, 140k and 200k assertions. The test queries in Table 6 were tailored to require reasoning with Horn-SHIQ constructs unavailable in DL-Lite and $\mathcal{E}L$. Table 7 presents the number of rewritten queries, rewriting time, and DLV running time of Data-

talog programs. The results show that CLIPPER answered all queries in reasonable time and scaled well (time printed $A_1 / A_2 / A_3 / A_4$). The rewriting times for all the queries are small and close. The high number of rules generated for Q4, Q6, Q8, and Q9 is due to many different possibilities to derive atoms in the query; e.g., for Professor(z) and worksFor(z,y) in Q4. However, the evaluation still performs well (it stays within a small factor).

Related Work and Conclusion

Since Calvanese et al. (2007) introduced query rewriting in their seminal work on DL-Lite, many query rewriting techniques have been developed and implemented, e.g. (Perez-Urbina et al. 2009, Rosati and Almatelli 2010, Chortaras et al. 2011, Gottlob et al. 2011), usually aiming at an optimized rewriting size. Some of them also go beyond DL-Lite; e.g., Perez-Urbina et al. cover $\mathcal{EL}I$, while Gottlob et al. consider Datalog$\pm$. Most approaches rewrite a query into a (union of) CQs: Rosati and Almatelli generate a non-recursive Datalog program, while Perez-Urbina et al. produce a CQ for $\mathcal{DL}$-Lite and a (recursive) Datalog program for DLs of the $\mathcal{EL}$ family. Our approach rewrites a CQ into a union of CQs, but generates (possible recursive) Datalog rules to capture the TBox.

Our technique resembles Rosati’s (2007) for CQs in $\mathcal{EL}$, which incorporates the input CQ into the TBox before saturation, after which the TBox is translated into Datalog. This is best-case exponential, which we avoid (a rewrite step occurs only if the TBox has an applicable existential axiom).

Rewriting approaches for more expressive DLs are less common. A notable exception is Hustadt et al.’s (2007) translation of SHIQ terminologies into disjunctive Data-

talog, which is implemented in the KAON2 reasoner. It can be used to answer queries over arbitrary ABoxes, but supports only instance queries. An extension to CQs (without transitive roles) (Hustadt et al., 2004) is not implemented. Ortiz et al. (2010) use a Datalog rewriting to establish complexity bounds of standard reasoning in the Horn fragments of SHIQ and SROIQ, but it does not cover CQs.

To our knowledge, CQ answering for Horn-SHIQ and beyond has not been implemented before. Algorithms for full SHIQ were first given in (Glimm et al. 2008, Calvanese et al. 2007), for Horn-SHIQ in (Eiter et al. 2008b), and for Horn-SHIQ in (Ortiz et al. 2011). They are of theoretical interest (to prove complexity results) but not suited for implementation due to prohibitive sources of complexity.

Outlook We presented a rewriting-based algorithm for CQ answering over Horn-SHIQ ontologies, possibly extended with weakly DL-safe rules. Our results can be generalized to the case without UNA using standard techniques: the current algorithm works correctly for TBoxes without axioms of the form $A \subseteq 1 \cdot B$, while for a full Horn-SHIQ TBox $\mathcal{T}$ an axiomatization of the equality predicate can be used to replace the constraints in $cr(\mathcal{T})$.

Our prototype implementation shows potential for practical applications, and further optimizations will improve it. Future versions of CLIPPER will support transitive roles in queries, and KBs with sets of weakly DL-safe rules.
An interesting application of our method is reasoning with DL-programs, which loosely couple rules and ontologies (Eiter et al. 2008a). The inline evaluation framework translates ontologies into rules to avoid the overhead caused by using two interacting reasoners (Heymans, Eiter, and Xiao 2010; Eiter et al. 2012a). The techniques in this paper can be faithfully integrated into it to efficiently evaluate DL-programs involving Horn-SHIQ ontologies.

Acknowledgements. This work was supported by the Austrian Science Fund (FWF) grants P20840 and T515.

References


Eiter, T.; Gottlob, G.; Ortiz, M.; and Simkus, M. 2008b. Query answering in the description logic Horn-SHIQ. In JELIA’08, 166–179. Springer.


Ortiz, M.; Rudolph, S.; and Simkus, M. 2011. Query answering in the Horn fragments of the description logics SHOIQ and SROIQ. In IJCAI’11, 1039–1044.


