LTCS-Report

From Horn-$\mathcal{SRIQ}$ to Datalog:
A Data-Independent Transformation that Preserves Assertion Entailment (Extended Version)

David Carral, Larry González, Patrick Koopmann

Contents

From Horn-$\mathcal{SRIQ}$ to Datalog:
A Data-Independent Transformation that Preserves Assertion Entailment (Extended Version)

David Carral, Larry González, Patrick Koopmann

November 26, 2018

Abstract

Ontology-based access to large data-sets has recently gained a lot of attention. To access data efficiently, one approach is to rewrite the ontology into Datalog, and then use powerful Datalog engines to compute implicit entailments. Existing rewriting techniques support Description Logics (DLs) from $\mathcal{ELH}$ to Horn-$\mathcal{SHIQ}$. We go one step further and present one such data-independent rewriting technique for Horn-$\mathcal{SRIQ}_\sqcap$, the extension of Horn-$\mathcal{SHIQ}$ that supports role chain axioms, an expressive feature prominently used in many real-world ontologies. We evaluated our rewriting technique on a large known corpus of ontologies. Our experiments show that the resulting rewritings are of moderate size, and that our approach is more efficient than state-of-the-art DL reasoners when reasoning with data-intensive ontologies.

1 Introduction

Assertion retrieval (AR)—i.e., the task of inferring implicit assertions from a Description Logics (DL) knowledge base (KB)—is an important reasoning task with many applications in knowledge representation and data management. For instance, the computation of AR can be used to solve SPARQL query answering, and to compute statistics on the implicit inferences of data-intensive ontologies such as in [3, 23]. For these tasks, both the concepts an object satisfies and the relations between objects are relevant. Typical DL ontologies focus on providing axioms about concepts, but expressive ontologies also allow to make inferences about roles, e.g., through the use of logical constructors such as inverse roles and role chains.

Efficient AR on large datasets requires the use of "one-pass" algorithms that compute the full set of entailed assertions as part of a saturation procedure. Although many customised algorithms and implementations of this type have been developed in the past, to the best of our knowledge, either these procedures do not support role chains, or they are not complete for deriving role assertions. Indeed, the retrieval of roles in the presence of role chains is a rather challenging task, as it may require reasoning about paths involving objects not explicit in the data.

Example 1. *Let $\mathcal{T}_{ex}$ be the TBox with the following axioms modelling conflicts of interests*

between researchers.

$$\textsf{ResearchGroup} \sqsubseteq \forall\textsf{hasMember}.\textsf{Researcher}$$
$$\textsf{Researcher} \sqsubseteq \exists\textsf{hasMember}^-.\textsf{ResearchGroup}$$
$$\textsf{collaborated} \circ \textsf{hasMember}^- \circ \textsf{hasMember} \sqsubseteq \textsf{hasConflict}$$
$$\textsf{hasMember} \circ \textsf{supervises} \sqsubseteq \textsf{hasMember}$$

Here, the third axiom uses a role chain to express that, if a researcher collaborated with someone who is a member of a research group, then he has a conflict of interest with everyone from that group. Using $\mathcal{T}_{ex}$, we can infer from the ABox

$$\mathcal{A}_{ex} = \{\textsf{collaborated}(\textsf{gottlob}, \textsf{alonzo}), \textsf{supervises}(\textsf{alonzo}, \textsf{alan}), \textsf{Researcher}(\textsf{alonzo})\}$$

the two assertions $\textsf{Researcher}(\textsf{alan})$ *and* $\textsf{hasConflict}(\textsf{gottlob}, \textsf{alan})$. *Both entailments depend on the existence of a research group which has both* alan *and* alonzo *as members, the existence of which is implied but not explicit. Specifically,* gottlob *has a conflict of interest with* alan *because there is a path via* alonzo *and this research group connecting* gottlob *with* alan*, which corresponds to the role chain in the third axiom.*

We propose a technique for AR from KBs formulated in Horn-$\mathcal{SRIQ}_\sqcap$—a DL fragment that supports complex roles and role conjunctions [15]—based on data-independent rewritings into Datalog rule sets. Specifically, given a TBox $\mathcal{T}$, we describe how to construct a Datalog rule set $\mathcal{R}_\mathcal{T}$ s.t., for every ABox $\mathcal{A}$ and assertion α only using symbols occurring in $\mathcal{T}$, we have $\langle\mathcal{T}, \mathcal{A}\rangle \models \alpha$ iff $\langle\mathcal{R}_\mathcal{T}, \mathcal{A}\rangle \models \alpha$.

To show practical feasibility, we implemented and evaluated our transformation, showing that Datalog rewritings for many real-world Horn-$\mathcal{SRIQ}_\sqcap$ TBoxes are of moderate size. Moreover, we computed our Datalog rewritings for two real-world ontologies, and performed AR over the resulting Datalog KBs. Our results show that our approach can outperform *Konclude* [21]—considered as one of the leading DL reasoners [19]—when solving AR over data-intensive ontologies. This is rather noteworthy, since (unlike Konclude) our rewritings are complete for role retrieval.

In summary, our contributions are as follows.

- We present a worst-case optimal transformation of Horn-$\mathcal{SRIQ}_\sqcap$ TBoxes into Datalog rule sets that preserves satisfiability and assertion entailment.

- We show that the resulting rule sets can be transformed into equivalent DLP ontologies [7]—the DL fragment underlying the OWL RL standard.

- We empirically show that our rewriting technique produces Datalog rule sets of moderate size for many real-world Horn-$\mathcal{SRIQ}_\sqcap$ TBoxes.

- We empirically show that the resulting Datalog programs can be used to solve AR more efficiently than DL reasoners when dealing with data-intensive ontologies.

Formal proofs and arguments for the results in this paper, as well as evaluation details, are in the appendix.

1.1 Related Work

Even though there are many algorithms and implementations for AR on DL KBs, we find that none of them can satisfactorily handle role retrieval, i.e., the retrieval of role assertions, in the presence of role chains.

There are many approaches that can efficiently perform AR for DLs which do not support role chains, and which are similar in spirit to our approach. Hustadt et al. [10] reduce standard reasoning tasks in the DL $\mathcal{SHIQ}^-$ to reasoning over disjunctive query Datalog programs. Eiter et al. [6] propose a method that combines materialisation—a step that can be repurposed to solve role retrieval—and rewriting to solve conjunctive query answering over Horn-$\mathcal{SHIQ}$ ontologies. A similar method tailored for the DL Horn-$\mathcal{ALCHOIQ}$ is presented by Carral et al. [4]. Recently, Ahmetaj et al. [1] proposed Datalog rewritings to perform instance queries over $\mathcal{ALCHIO}$ KBs extended with closed predicates.

State-of-the-art DL reasoners such as Fact++ [22], HermiT [18], Pellet [20] and Konclude [21] support $\mathcal{SROIQ}$ KBs. However, while the former three do not perform that well on data-intensive ontologies [19], Konclude does not support role retrieval as part of its one-pass algorithm. As our results indicate, Datalog rewritings have the potential to outperform all these approaches.

Regarding less expressive DLs, despite the fact that there are theoretical algorithms for $\mathcal{EL}^{++}$ that can deal with role chains [13], leading profile reasoners such as ELK [12] do not support this expressive feature yet.

2 Preliminaries

We consider logical theories based on finite signatures consisting of mutually disjoint sets $\mathsf{N_c}$ of *concepts* (unary predicates), $\mathsf{N_r}$ of *roles* (binary predicates), $\mathsf{N_v}$ of *variables*, and $\mathsf{N_i}$ of *individuals* (constants), as well as an unbounded set $\mathsf{N_0}$ of *nulls* disjoint with all of the above. There is a bijective and irreflexive function $\cdot^- : \mathsf{N_r} \to \mathsf{N_r}$ with $R^{--} = R$ for all $R \in \mathsf{N_r}$, and $\bot, \top \in \mathsf{N_c}$. For a formula or set thereof φ, we use $\mathsf{sig}(\varphi)$ to denote the set of all concepts and roles in φ. The sets of *terms* and *ground terms* are $\mathsf{N_t} = 2^{\mathsf{N_i}} \cup \mathsf{N_0} \cup \mathsf{N_v}$ and $\mathsf{N_{gt}} = 2^{\mathsf{N_i}} \cup \mathsf{N_0}$, respectively. The use of $2^{\mathsf{N_i}}$ rather than $\mathsf{N_i}$ in the definition of terms is for convenience of the definition of the chase later in this section. Thus, we henceforth identify every $a \in \mathsf{N_i}$ with the singleton set $\{a\}$.

2.1 Existential Rules

We write tuples of terms $t_1, \ldots, t_n$ as $\vec{t}$, and treat such tuples as sets when the order is irrelevant. An *atom* is a formula of the form $C(t)$ or $R(t, u)$ with $C \in \mathsf{N_c}$, $R \in \mathsf{N_r}$, and $t, u \in \mathsf{N_t}$. We identify a binary atom $R(t, u)$ with $R^-(u, t)$. A formula or set thereof is *ground* if it only contains ground terms. For a formula φ, we write $\varphi[\vec{x}]$ to indicate that $\vec{x}$ is the set of all free variables occurring in φ.

An *(existential) rule* is a formula of one of the forms:

$$\forall \vec{x}, \vec{z}. \big(B[\vec{x}, \vec{z}] \to \exists \vec{y}. H[\vec{x}, \vec{y}] \big) \tag{$\to$}$$

$$\forall \vec{x}. \big(B[\vec{x}] \to x \approx y \big) \tag{$\approx$}$$

Where B and H are non-empty, null-free conjunctions of atoms, and $x, y \in \vec{x}$. A *Datalog rule* is a rule without existentially quantified variables. A *fact* is a ground atom. We identify facts and sets thereof if they are identical up to bijective renaming of nulls. A *knowledge base (KB)* is a tuple $\langle \mathcal{R}, \mathcal{A} \rangle$ with $\mathcal{R}$ a rule set and $\mathcal{A}$ an *ABox*—a set of facts without nulls, i.e., *assertions*. We treat KBs as first-order theories and define semantical notions such as entailment and satisfiability in the usual way. To axiomatise the semantics of $\top$, we assume that $\{A(x) \to \top(x) \mid A \in \mathsf{N_c}\} \cup \{R(x, y) \to \top(x) \wedge \top(y) \mid R \in \mathsf{N_r}\} \subseteq \mathcal{R}$ for every rule set $\mathcal{R}$.

$$\bigwedge_{i=1}^{n} A_i(x) \to B(x) \qquad\qquad \prod_{i=1}^{n} A_i \sqsubseteq B \qquad\qquad (\sqcap)$$

$$A(x) \wedge R(x,y) \to B(y) \qquad\qquad A \sqsubseteq \forall R.B \qquad\qquad (\forall)$$

$$A(x) \to \exists y.R(x,y) \wedge B(y) \qquad\qquad A \sqsubseteq \exists R.B \qquad\qquad (\exists)$$

$$A(x) \wedge R(x,y) \wedge B(y)$$
$$\wedge\, R(x,z) \wedge B(z) \to y \approx z \qquad\qquad A \sqsubseteq\, \leqslant 1\, R.B \qquad\qquad (\leq)$$

$$\bigwedge_{i=1}^{n} R_i(x_{i-1}, x_i) \to S(x_0, x_n) \qquad\qquad R_1 \circ \ldots \circ R_n \sqsubseteq S \qquad\qquad (\circ)$$

$$\bigwedge_{i=1}^{m} R_i(x,y) \to S(x,y) \qquad\qquad \prod_{i=1}^{m} R_i \sqsubseteq S \qquad\qquad (\sqcap_{\mathsf{r}})$$

Figure 1: Horn-$\mathcal{SRIQ}_{\sqcap}$ Axioms, where $A_{(i)}, B \in \mathsf{N_c}$, $R_{(i)}, S \in \mathsf{N_r}$, $x_{(i)}, y, z \in \mathsf{N_v}$, $n \geq 1$, and $m > 1$

2.2 The DL Horn-$\mathcal{SRIQ}_{\sqcap}$

Without loss of generality [15], we define Horn-$\mathcal{SRIQ}_{\sqcap}$ using a restricted set of normalised axioms, which we introduce in the right hand side of Figure 1. We identify each of these axioms with the corresponding rule in the left hand side of Figure 1, and alternate between these two syntaxes whenever this is convenient.

For an axiom set $\mathcal{R}$, let $\prec_{\mathcal{R}}^{+}$ be the minimal transitive relation over roles s.t. $R \prec_{\mathcal{R}}^{+} S$ iff $R^- \prec_{\mathcal{R}}^{+} S$; for every axiom in $\mathcal{R}$ of Type ($\sqcap_{\mathsf{r}}$), $R_i \prec_{\mathcal{R}}^{+} S$ for all $i \in [\![1, m]\!]$; and, for every axiom in $\mathcal{R}$ of Type ($\circ$),

- if $n = 1$ and $R_1 \neq S^-$, then $R_1 \prec_{\mathcal{R}}^{+} S$, and

- if $n > 1$ and $R_1 \circ \ldots \circ R_n \neq S \circ S$, then

  - if $R_n = S$, then $R_i \prec_{\mathcal{R}}^{+} S$ for all $i \in \{1, \ldots, n-1\}$,
  - if $R_1 = S$, then $R_i \prec_{\mathcal{R}}^{+} S$ for all $i \in \{2, \ldots, n\}$, and
  - if $R_1 \neq S \neq R_n$, then $R_i \prec_{\mathcal{R}}^{+} S$ for all $i \in \{1, \ldots, n\}$.

A role V is *complex* wrt. $\mathcal{R}$ if there is an axiom in $\mathcal{R}$ of Type ($\circ$) with $n > 1$ and $S \prec_{\mathcal{R}}^{*} V$ with $\prec_{\mathcal{R}}^{*}$ the reflexive closure of $\prec_{\mathcal{R}}^{+}$. Otherwise, V is *simple*.

Definition 1. *An axiom set $\mathcal{T}$ is a (Horn-$\mathcal{SRIQ}_{\sqcap}$) TBox if $\prec_{\mathcal{T}}^{+}$ is irreflexive, and all roles occurring in an axiom of Type ($\leq$), or in the left hand side of an axiom of Type ($\sqcap_{\mathsf{r}}$) in $\mathcal{T}$ are simple. A KB $\langle \mathcal{T}, \mathcal{A} \rangle$ is Horn-$\mathcal{SRIQ}_{\sqcap}$ if $\mathcal{T}$ is a Horn-$\mathcal{SRIQ}_{\sqcap}$ TBox.*

2.3 The Chase

A well-known way of characterising entailments from KBs is the chase, which we introduce next.

A *substitution* σ is a partial function over $\mathsf{N_t}$. We use $[t_1/u_1, \ldots, t_n/u_n]$ to denote the substitution σ s.t. $\sigma(t_i) = u_i$ for all $i \in [\![1, n]\!]$. For a formula φ, we write $\varphi\sigma$ to denote the formula obtained by replacing all occurrences of a term t in φ with $\sigma(t)$ if t is in the domain of σ. For a tuple $\vec{t}$ of terms, $\sigma_{\vec{t}} \subseteq \sigma$ is the restriction of σ to the domain $\vec{t}$.
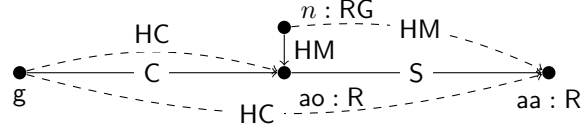
Figure 2: Chase of $\mathcal{O}_x = \langle \mathcal{T}_{ex}, \mathcal{A}_{ex} \rangle$ from Example 1

To handle rules of Type ($\approx$), we represent individuals as sets, which is why we used $2^{\mathsf{N_i}}$ in the definition of terms. For a given substitution σ and two variables x, y, we define $\sigma_{x,y}^{\mathsf{rn}}$ by $\sigma_{x,y}^{\mathsf{rn}}(x) = \sigma_{x,y}^{\mathsf{rn}}(y) = \sigma(x)$ if $\sigma(x), \sigma(y) \in \mathsf{N_0}$, and $\sigma_{x,y}^{\mathsf{rn}}(x) = \sigma_{x,y}^{\mathsf{rn}}(y) = (\sigma(x) \cup \sigma(y)) \cap \mathsf{N_i}$ otherwise. Intuitively, $\sigma_{x,y}^{\mathsf{rn}}$ is the substitution identifying $\sigma(x)$ and $\sigma(y)$.

A tuple $\langle \rho, \sigma \rangle$ with $\rho = B[\vec{x}, \vec{z}] \to \exists \vec{y}.H[\vec{x}, \vec{y}]$ a rule and σ a substitution is *applicable* to a set of facts $\mathcal{F}$ if $B\sigma \subseteq \mathcal{F}$, and $H\sigma' \not\subseteq \mathcal{F}$ for all $\sigma' \supseteq \sigma_{\vec{x}}$. The *application* of $\langle \rho, \sigma \rangle$ on $\mathcal{F}$, written $\mathcal{F}\langle \rho, \sigma \rangle$, is the set of facts $\mathcal{F} \cup H\sigma'$ with $\sigma' \supseteq \sigma_{\vec{x}}$ a substitution mapping every variable in $\vec{y}$ to a fresh null. If ρ is of the form $B[\vec{x}] \to x \approx y$, then $\langle \rho, \sigma \rangle$ is applicable to $\mathcal{F}$ if $B\sigma \subseteq \mathcal{F}$ and $\sigma(x) \neq \sigma(y)$. In this case, the application of $\langle \rho, \sigma \rangle$ on $\mathcal{F}$, also denoted by $\mathcal{F}\langle \rho, \sigma \rangle$, is the set $\mathcal{F}\sigma_{x,y}^{\mathsf{rn}}$.

We introduce this non-standard approach of rule applications with equality to ensure that the forest-model property of Horn-$\mathcal{SRIQ}_\sqcap$ ontologies is reflected in the structure of the chase, which will later be useful to show completeness of our Datalog rewritings

Definition 2. *A* chase sequence *for a KB $\mathcal{K} = \langle \mathcal{R}, \mathcal{A} \rangle$ is a sequence $\mathcal{F}^0 = \mathcal{A}, \mathcal{F}^1, \ldots$ of sets of facts s.t.*

- *for all $i \geq 1$, $\mathcal{F}^i = \mathcal{F}^{i-1}\langle \rho, \sigma \rangle$ for a rule $\rho \in \mathcal{R}$ and some substitution σ s.t. $\langle \rho, \sigma \rangle$ is applicable, and*

- *for all $\langle \rho, \sigma \rangle$ with $\rho \in \mathcal{R}$, there is some $k \geq 0$ s.t. $\langle \rho, \sigma \rangle$ is not applicable to $\mathcal{F}^i$ for all $i \geq k$ (**fairness**).*

The chase *of $\mathcal{K}$, denoted by $\mathcal{K}^\infty$, is the is the union of all sets in some (arbitrarily chosen) chase sequence of $\mathcal{K}$.*

For the rest of the paper, we fix a Horn-$\mathcal{SRIQ}_\sqcap$ KB $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ and some (possibly infinite) chase sequence $\mathcal{O}^0, \mathcal{O}^1, \ldots$ for $\mathcal{O}$. For all $i \geq 1$, let $\rho_i \in \mathcal{T}$ be an axiom and σ_i a substitution s.t. $\mathcal{O}^i = \mathcal{O}^{i-1}\langle \rho_i, \sigma_i \rangle$. By abuse of notation, we write $P(a_1, \ldots, a_n) \in \mathcal{F}$, with $\mathcal{F}$ a set of facts, $P \in \mathsf{N_c} \cup \mathsf{N_r}$, and $a_1, \ldots, a_n \in \mathsf{N_i}$, if $P(b_1, \ldots, b_n) \in \mathcal{F}$ for some $b_1, \ldots, b_n \in 2^{\mathsf{N_i}}$ with $a_i \in b_i$ for all $i \in [\![1, n]\!]$.

Theorem 1. *A KB $\mathcal{K}$ is satisfiable iff $\bot(t) \notin \mathcal{K}^\infty$ for all $t \in \mathsf{N_{gt}}$. If $\mathcal{K}$ is satisfiable, $\mathcal{K} \models \alpha$ iff $\alpha \in \mathcal{K}^\infty$ for every assertion α.*

We later show that the every chase step in a chase sequence of a Horn-$\mathcal{SRIQ}_\sqcap$ ontology reflects the "forest-shaped" when we restrict to facts containing at least one null, which corresponds to the well-known forest-model property of Horn-$\mathcal{SRIQ}_\sqcap$. In the presence of complex roles, the forest-model property is not entirely apparent in the chase steps of an ontology. To characterise this property, we distinguish binary facts in the chase that are not produced via the application of axioms of the Type ($\circ$) with $n \geq 2$, or the propagation of such facts.

All binary facts in $\mathcal{O}^0$ are *direct*. For all $i \geq 1$, a binary fact $\phi \in \mathcal{O}^i \setminus \mathcal{O}^{i-1}$ is *direct* iff ρ_i is of Type ($\exists$) or ($\sqcap_{\mathsf{r}}$); ρ_i is of Type ($\circ$) with $n = 1$ and $R_1(\sigma_i(x_0), \sigma_i(x_1)) \in \mathcal{O}^{i-1}$ is direct; or ρ_i is

of Type ($\leq$), and there is a direct fact $\phi' \in \mathcal{O}^{i-1}$ s.t. $\phi'(\sigma_i)^{\mathsf{rn}}_{x,y} = \phi$. For $i \geq 0$, we write $\mathsf{D}(\mathcal{O}^i)$ to denote the set of all direct facts in $\mathcal{O}^i$.

Example 2. *Consider the TBox $\mathcal{T}_{ex}$ and ABox $\mathcal{A}_{ex}$ from Example 1. The chase of $\mathcal{O}_x = \langle \mathcal{T}_{ex}, \mathcal{A}_{ex} \rangle$ is depicted in Figure 2, where direct and not direct facts are represented using full and dashed arrows, respectively. Note that n is a null introduced by the chase.*

If we consider only the direct facts that occur in the chase sequence of an ontology, we can establish the "forest model property" reflected in every chase step of this sequence. For all $i \geq 0$, let $\mathsf{F}(\mathcal{O}^i)$ be the graph s.t. every $a \in 2^{\mathsf{N}_i}$ in $\mathcal{O}^i$ is a node in $\mathsf{F}(\mathcal{O}^i)$, and $t_{n-1} \to t_n \in \mathsf{F}(\mathcal{O}^i)$ if there is a sequence of facts $R_1(t_0, t_1), \ldots, R_n(t_{n-1}, t_n) \in \mathsf{D}(\mathcal{O}^i)$ with $t_0 \in 2^{\mathsf{N}_i}$ and $t_i \neq t_j$ for all $0 \leq i < j \leq n$.

Lemma 1. *For all $i \geq 0$,*

- *all nulls in $\mathcal{O}^i$ occur as nodes in $\mathsf{F}(\mathcal{O}^i)$, and*

- *$\mathsf{F}(\mathcal{O}^i)$ is a rooted forest where every individual node is a root, and every null node is not.*

2.4 Non-Deterministic Automata

In our approach, we need to trace the paths of complex roles in the chase of a Horn-$\mathcal{SRIQ}_\sqcap$ KB that traverse only direct facts. To do so, we make use of well-known automata techniques from [9, 11]. Here, we use non-deterministic finite automata (NFAs) in a rather informal way, and use the notation $p \to_R q \in \mathcal{N}$ to denote that, in the NFA $\mathcal{N}$, there is a transition from a state p to a state q with the letter R, instead of introducing transition relations formally.

Definition 3. *For a TBox $\mathcal{T}$, let $\mathcal{T}_- \supseteq \mathcal{T}$ be the TBox with $R_n^- \circ \ldots \circ R_1^- \sqsubseteq S^- \in \mathcal{T}_-$ for every axiom of Type ($\circ$) in $\mathcal{T}$.*

For every $V \in \mathsf{N}_r$, the NFA $\mathcal{N}_\mathcal{T}(V)$ is the smallest NFA s.t. $i_V \to_V f_V \in \mathcal{N}_\mathcal{T}(V)$ with i_V and f_V the only initial and final states; and for every transition $q \to_S \hat{q} \in \mathcal{N}_\mathcal{T}(V)$ and every axiom in $\mathcal{T}_-$ of the form ($\circ$), we have

- *if $n = 1$ and $R_1 = S^-$, then $q \to_{S^-} \hat{q} \in \mathcal{N}_\mathcal{T}(V)$,*

- *if $n = 2$, $R_1 = S$, and $R_2 = S$, then $\hat{q} \to_\epsilon q \in \mathcal{N}_\mathcal{T}(V)$,*

- *Otherwise,*

  - *if $R_1 \neq S = R_n$, then $q \to_\epsilon q_0 \to_{R_1} q_1 \to_{R_2} q_2 \to_{R_3} \cdots \to_{R_{n-1}} q_{n-1} \to_\epsilon q \in \mathcal{N}_\mathcal{T}(V)$,*
  - *if $R_1 = S \neq R_n$, then $\hat{q} \to_\epsilon q_1 \to_{R_2} q_2 \to_{R_3} q_3 \to_{R_4} \cdots \to_{R_n} q_n \to_\epsilon \hat{q} \in \mathcal{N}_\mathcal{T}(V)$, and*
  - *if $R_1 \neq S \neq R_n$, then $q \to_\epsilon q_0 \to_{R_1} q_1 \to_{R_2} q_2 \to_{R_3} \cdots \to_{R_n} q_n \to_\epsilon \hat{q} \in \mathcal{N}_\mathcal{T}(V)$.*

In the above, states q_i are assumed to be fresh and distinct.

Our definition of NFA coincides with that from [9] in the sense that the resulting NFA $\mathcal{N}_\mathcal{T}(R)$ for any $R \in \mathsf{N}_r$ does recognise the same language. With analogous arguments to those presented by Horrocks et al., we can show the following claim.

Lemma 2. *For all $i \geq 0$, if $\mathcal{O}^i$ is closed under the application of axioms of Type ($\sqcap_r$), there is a binary fact $R(t, u) \in \mathcal{O}^i$ iff there are some $S_1(t, t_1), \ldots, S_n(t_{n-1}, u) \in \mathsf{D}(\mathcal{O}^i)$ with $S_1 \cdot \ldots \cdot S_n \in \mathcal{N}_\mathcal{T}(R)$.*
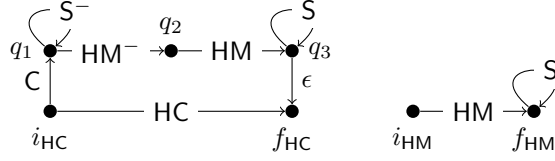
Figure 3: The NFA $\mathcal{N}_{\mathcal{T}_{ex}}(\mathsf{HC})$ and $\mathcal{N}_{\mathcal{T}_{ex}}(\mathsf{HM})$

Given a $\mathsf{P} = R_1 \cdot \ldots \cdot R_n$ with $R_1, \ldots, R_n \in \mathsf{N_r}$, we write $q \rightarrow^*_{\mathsf{P}} \hat{q} \in \mathcal{N}_{\mathcal{T}}(R)$ (resp. $\mathsf{P} \in \mathcal{N}_{\mathcal{T}}(R)$) to indicate that there is a path P from q to $\hat{q}$ (resp. i_R to f_R) in $\mathcal{N}_{\mathcal{T}}(R)$.

Example 3. *Consider $\mathcal{O}_x = \langle \mathcal{T}_{ex}, \mathcal{A}_{ex} \rangle$ with TBox $\mathcal{T}_{ex}$ and ABox $\mathcal{A}_{ex}$ from Example 1. The NFA $\mathcal{N}_{\mathcal{T}_{ex}}(\mathsf{HC})$ and $\mathcal{N}_{\mathcal{T}_{ex}}(\mathsf{HM})$ are depicted in Figure 3 (for the sake of clarity, we have removed some ϵ-transitions). As implied by Lemma 2 and since $\mathsf{HC}(\mathsf{g}, \mathsf{aa})$, we have $\mathsf{C}(\mathsf{g}, \mathsf{ao})$, $\mathsf{HM}^-(\mathsf{ao}, n)$, $\mathsf{HM}(n, \mathsf{ao}), \mathsf{S}(\mathsf{ao}, \mathsf{aa}) \in \mathsf{D}(\mathcal{O}_x^\infty)$ such that $\mathsf{C} \cdot \mathsf{HM}^- \cdot \mathsf{HM} \cdot \mathsf{S} \in \mathcal{N}_{\mathcal{T}_{ex}}(\mathsf{HC})$ (see Figure 2).*

3   Datalog Rewritings in Horn-$\mathcal{SRIQ}_\sqcap$

In this section, we define the Datalog AR-rewriting $\mathcal{R}_\mathcal{T}$ for the TBox $\mathcal{T}$ and discuss complexity results.

Definition 4. *A rule set $\mathcal{R}$ is an AR-rewriting for $\mathcal{T}$ iff, for every ABox $\mathcal{A}$ and assertion α over $\mathsf{sig}(\mathcal{T})$, $\langle \mathcal{T}, \mathcal{A} \rangle$ and $\langle \mathcal{R}, \mathcal{A} \rangle$ are equi-satisfiable and $\langle \mathcal{T}, \mathcal{A} \rangle \models \alpha$ iff $\langle \mathcal{R}, \mathcal{A} \rangle \models \alpha$.*

Let $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ and $\mathcal{K}_\mathcal{O} = \langle \mathcal{R}_\mathcal{T}, \mathcal{A} \rangle$. By Theorem 1, $\mathcal{R}_\mathcal{T}$ is an AR-rewriting only if the chase of $\mathcal{K}_\mathcal{O}$ coincides with the chase of $\mathcal{O}$ on all assertions over $\mathsf{sig}(\mathcal{T})$. The challenge in constructing Datalog AR-rewritings is that assertions in the $\mathcal{O}^\infty$ might be introduced by rule applications on facts with nulls, whilst no Datalog rule can introduce such terms.

Example 4. *Let $\mathcal{O}_x$ be the ontology from Example 1. Then, the assertion $\mathsf{HC}(\mathsf{g}, \mathsf{aa})$ is in $\mathcal{O}_x^\infty$ because $\mathsf{HC}(\mathsf{g}, \mathsf{ao}), \mathsf{HM}(n, \mathsf{ao}), \mathsf{HM}(n, \mathsf{aa}) \in \mathcal{O}_x$ (see Figure 2). Analogously, $\mathsf{R}(\mathsf{aa}) \in \mathcal{O}_x^\infty$ because $\mathsf{RG}(n), \mathsf{HM}(n, \mathsf{aa}) \in \mathcal{O}_x^\infty$. Note that the facts $\mathsf{HM}(n, \mathsf{ao}), \mathsf{HM}(n, \mathsf{aa})$, and $\mathsf{RG}(n)$ cannot occur in the case of a Datalog AR-rewriting, since $n \in \mathsf{N_0}$.*

To replicate assertion entailments in $\mathcal{K}_\mathcal{O}^\infty$ such as the ones highlighted in the previous example, we encode information in $\mathcal{K}_\mathcal{O}^\infty$ about the null successors of an individual in $\mathcal{O}^\infty$ using fresh concepts and roles. For all $R \in \mathsf{N_r}$ and states $q, \hat{q} \in \mathcal{N}_\mathcal{T}(R)$, we introduce the fresh concepts A_q and $R_{q,\hat{q}}$, and the fresh role R_q. Intuitively, these are used to encode the following information about $\mathcal{O}^\infty$ in $\mathcal{K}_\mathcal{O}^\infty$.

1. If $A_q(a) \in \mathcal{K}_\mathcal{O}^\infty$, then there are some $A(t_0) \in \mathcal{O}^\infty$, and some $R_1(t_0, t_1)$, $\ldots$, $R_n(t_{n-1}, a) \in \mathsf{D}(\mathcal{O}^\infty)$ with $q \rightarrow^*_{R_1 \cdot \ldots \cdot R_n} \hat{q} \in \mathcal{N}_\mathcal{T}(R)$.

2. If $R_{q,\hat{q}}(a) \in \mathcal{K}_\mathcal{O}^\infty$, then there are some $R_1(a, t_1), \ldots, R_n(t_{n-1}, a) \in \mathsf{D}(\mathcal{O}^\infty)$ with $t_1, \ldots, t_{n-1} \in \mathsf{N_0}$ and $q \rightarrow^*_{R_1 \cdot \ldots \cdot R_n} \hat{q} \in \mathcal{N}_\mathcal{T}(R)$.

3. If $R_q(a, b) \in \mathcal{K}_\mathcal{O}^\infty$, then $S_1(a, t_1), \ldots, S_n(t_{n-1}, b) \in \mathsf{D}(\mathcal{O}^\infty)$ with $i_R \rightarrow^*_{S_1 \cdot \ldots \cdot S_n} q \in \mathcal{N}_\mathcal{T}(R)$.

Note that all terms t_i may possibly be nulls that do not appear in the chase of $\mathcal{K}_\mathcal{O}$.

$$\frac{\mathbb{A} \sqsubseteq \exists \mathbb{R}.(\mathbb{B} \sqcap \mathbb{C}) \qquad \mathbb{C} \sqsubseteq A}{\mathbb{A} \sqsubseteq \exists \mathbb{R}.(\mathbb{B} \sqcap \mathbb{C} \sqcap A)} \tag{1}$$

$$\frac{\mathbb{A} \sqsubseteq \exists(\mathbb{R} \sqcap \mathbb{S}).\mathbb{B} \qquad \mathbb{S} \sqsubseteq R}{\mathbb{A} \sqsubseteq \exists(\mathbb{R} \sqcap \mathbb{S} \sqcap R).\mathbb{B}} \tag{2}$$

$$\frac{\mathbb{A} \sqsubseteq \exists \mathbb{R}.(\mathbb{B} \sqcap \bot)}{\mathbb{A} \sqsubseteq \bot} \tag{3}$$

$$\frac{\mathbb{A} \sqsubseteq \exists(\mathbb{R} \sqcap R).\mathbb{B} \qquad A \sqsubseteq \forall R.B}{\mathbb{A} \sqcap A \sqsubseteq \exists(\mathbb{R} \sqcap R).(\mathbb{B} \sqcap B)} \tag{4}$$

$$\frac{\mathbb{A} \sqsubseteq \exists(\mathbb{R} \sqcap R^-).(\mathbb{B} \sqcap A) \qquad A \sqsubseteq \forall R.B}{\mathbb{A} \sqsubseteq B} \tag{5}$$

$$\frac{\begin{array}{l} \mathbb{A} \sqsubseteq \exists(\mathbb{R} \sqcap R).(\mathbb{B} \sqcap B) \qquad A \sqsubseteq \, \leqslant 1\, R.B \\ \mathbb{C} \sqsubseteq \exists(\mathbb{S} \sqcap R).(\mathbb{D} \sqcap B) \end{array}}{\mathbb{A} \sqcap \mathbb{C} \sqcap A \sqsubseteq \exists(\mathbb{R} \sqcap \mathbb{S} \sqcap R).(\mathbb{B} \sqcap \mathbb{D} \sqcap B)} \tag{6}$$

$$\frac{\begin{array}{l} \mathbb{A} \sqsubseteq \exists(\mathbb{R} \sqcap R^-).(\mathbb{B} \sqcap \mathbb{C} \sqcap A) \qquad A \sqsubseteq \, \leqslant 1\, R.B \\ \mathbb{C} \sqsubseteq \exists(\mathbb{S} \sqcap R).(\mathbb{D} \sqcap B \sqcap C) \end{array}}{\begin{array}{l} \mathbb{A} \sqcap B \sqsubseteq C \\ \mathbb{A} \sqcap B \sqsubseteq \exists(\mathbb{R} \sqcap R^- \sqcap \mathbb{S}^-).(\mathbb{B} \sqcap \mathbb{C} \sqcap A) \end{array}} \tag{7}$$

Figure 4: Derivation Rules where $A, B \in \mathsf{N_c}$, $R \in \mathsf{N_r}$, and $\mathbb{A}, \mathbb{B}, \mathbb{C}, \mathbb{D}$ and $\mathbb{R}, \mathbb{S}$ are conjunctions of elements in $\mathsf{N_c}$ and $\mathsf{N_r}$, respectively

To ascertain when information about one of these predicates needs to be used in $\mathcal{K}_{\mathcal{O}}$, we make use of a sound saturation calculus from [6], shown in Figure 4, which we also use to infer further axioms relevant to our Datalog program. Since this calculus was originally designed for Horn-$\mathcal{SHIQ}$, we first need to extend our input TBox $\mathcal{T}$ to a TBox $\mathcal{T}_+$ in which the behaviour of axioms of Type ($\circ$) is sufficiently simulated. For instance, if the calculus derives from $\mathcal{T}_+$ an axiom of the form $\mathbb{A} \sqsubseteq A_q$, then we can conclude that, for every term t s.t. $B(t) \in \mathcal{O}^\infty$ for every $B \in \mathbb{A}$, there is a set of direct facts $A(t_0), R_1(t_0, t_1), \ldots, R_n(t_{n-1}, a) \in \mathcal{O}^\infty$ with a corresponding path in the automata, irrespectively of the ABox $\mathcal{A}$. We further augment $\mathcal{T}_+$ to a TBox $\mathcal{T}_\times$ that allows us to trace paths in possible chases for $\mathcal{T}$. Using the inferences from this calculus, we then describe the rewriting $\mathcal{R}_{\mathcal{T}}$.

Definition 5. *Let $B(\mathcal{T})$ be the set of axioms that, for every axiom $\rho \in \mathcal{T}$ of Type ($\forall$), contains $A \sqsubseteq A_{i_R}, A_{f_R} \sqsubseteq B$, and $A_q \sqsubseteq \forall S.A_{\hat{q}} \in B(\mathcal{T})$ for every $q \rightarrow_S^* \hat{q} \in \mathcal{N}_{\mathcal{T}}(R)$ with $S \in \mathsf{N_r}$. Let $\mathcal{T}_+ = \mathcal{T}_- \cup B(\mathcal{T})$, and $\mathcal{T}_\times = \mathcal{T}_- \cup B(\mathcal{T} \cup \bigcup_{R \in \mathsf{N_r}} \{X \sqsubseteq \forall R.Y\})$, with X and Y fresh concepts.*

Then, $\mathcal{R}_{\mathcal{T}}$ is the Datalog rule set that contains every axiom in $\mathcal{T}_+$ that is not of Type ($\exists$), and every axiom that can be inferred using the implications described in Table 1.

Theorem 2. *The rule set $\mathcal{R}_{\mathcal{T}}$ is an AR-rewriting of $\mathcal{T}$.*

This result is a corollary of Lemmas 3, 16, and 11. Lemmas 16, and 11 are proven in the appendix.

Example 5. *Let $\mathcal{O}_x$ be the ontology from Example 1. Then, the Datalog rule set $\mathcal{R}_{\mathcal{T}_{ex}}$ contains*

9

$$\bigwedge_{D \in \mathbb{D}} D(x) \to A(x) \quad \Longleftarrow \quad \mathbb{D} \sqsubseteq A \in \Gamma(\mathcal{T}_\times) \tag{$\sqcap$}$$

$$\bigwedge_{D \in \mathbb{D}} D(x) \to R_{q,\hat{q}}(x) \quad \Longleftarrow \quad R \in \mathsf{N_r}, q, \hat{q} \in \mathcal{N}_R(\mathcal{T}), \text{ and } \mathbb{D} \sqcap X_q \sqsubseteq X_{\hat{q}} \in \Gamma(\mathcal{T}_\times) \tag{$\circlearrowright$}$$

$$A(x) \wedge \bigwedge_{D \in \mathbb{D} \cup A} D(x) \wedge R(x,y) \wedge B(y) \to C(y) \quad \Longleftarrow \quad A \sqsubseteq \leqslant 1\, R.B, \mathbb{D} \sqsubseteq \exists (\mathbb{R} \sqcap R).(\mathbb{A} \sqcap B \sqcap C) \in \Gamma(\mathcal{T}_\times) \tag{$\lhd\,1$}$$

$$A(x) \wedge \bigwedge_{D \in \mathbb{D}} D(x) \wedge R(x,y) \wedge B(y) \to S(x,y) \quad \Longleftarrow \quad A \sqsubseteq \leqslant 1\, R.B, \mathbb{D} \sqsubseteq \exists (\mathbb{R} \sqcap R \sqcap S).(\mathbb{A} \sqcap B) \in \Gamma(\mathcal{T}_\times) \tag{$\lhd\,2$}$$

$$S(x,y) \to R_q(x,y) \quad \Longleftarrow \quad R, S \in \mathsf{N_r} \text{ and } i_R \to_S^* q \in \mathcal{N}_\mathcal{T}(R) \tag{R 1}$$

$$R_{i_R,q}(x) \to R_q(x,x) \quad \Longleftarrow \quad R \in \mathsf{N_r} \text{ and } R_{i_R,q} \in \mathcal{R}_\mathcal{T} \tag{R 2}$$

$$R_q(x,y) \wedge S(y,z) \to R_{\hat{q}}(x,z) \quad \Longleftarrow \quad R, S \in \mathsf{N_r} \text{ and } q \to_S^* \hat{q} \in \mathcal{N}_\mathcal{T}(R) \tag{R 3}$$

$$R_q(x,y) \wedge R_{q,\hat{q}}(y) \to R_{\hat{q}}(x,y) \quad \Longleftarrow \quad R \in \mathsf{N_r} \text{ and } R_{q,\hat{q}} \in \mathcal{R}_\mathcal{T} \tag{R 4}$$

$$R_{f_R}(x,y) \to R(x,y) \quad \Longleftarrow \quad R \in \mathsf{N_r} \tag{R 5}$$

Table 1: Rules to construct $\mathcal{R}_\mathcal{T}$, where $\Gamma(\mathcal{T}_\times)$ is the saturation of $\mathcal{T}_\times$ by the rules in Figure 4 and all concepts A and B and those in the conjunctions $\mathbb{D}$ and $\mathbb{A}$ occur $\mathcal{T}_+$.
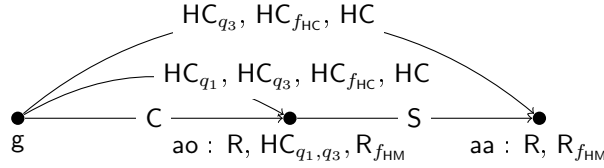

Figure 5: Representation of $\mathcal{K}_\mathcal{O}^\infty$ with $\mathcal{O}$ from Example 1

(amongst others) all the rules in $\mathcal{T}_{ex}$ that are not of Type ($\exists$), as well as the following.

$$\mathsf{R}(x) \to \mathsf{R}_{f_{\mathsf{HM}}}(x) \qquad \mathsf{R}_{f_{\mathsf{HM}}}(x) \to \mathsf{R}(x)$$

$$\mathsf{R}_{f_{\mathsf{HM}}}(x) \wedge \mathsf{S}(x,y) \to \mathsf{R}_{f_{\mathsf{HM}}}(y)$$

$$\mathsf{C}(x,y) \to \mathsf{HC}_{q_1}(x,y) \qquad \mathsf{R}(x) \to \mathsf{HC}_{q_1,q_3}(x)$$

$$\mathsf{HC}_{q_1}(x,y) \wedge \mathsf{HC}_{q_1,q_3}(x) \to \mathsf{HC}_{q_3}(x,y)$$

$$\mathsf{HC}_{q_3}(x,y) \to \mathsf{HC}_{f_{\mathsf{HC}}}(x,y) \qquad \mathsf{HC}_{f_{\mathsf{HC}}}(x,y) \to \mathsf{HC}(x,y)$$

$$\mathsf{HC}_{f_{\mathsf{HC}}}(x,y) \wedge \mathsf{S}(y,z) \to \mathsf{HC}_{f_{\mathsf{HC}}}(x,z)$$

The chase of $\mathcal{K}_{\mathcal{O}_x}$ is depicted in Figure 5. Note that $\mathcal{K}_{\mathcal{O}_x}^\infty$ contains every assertion in $\mathcal{O}^\infty$.

While we provide for full proofs of Theorem 2 in the appendix, we give an overview of some of the main technical ideas in this section. While showing soundness of our approach is not as challenging, we focus on the argument showing completeness of the AR-rewriting $\mathcal{R}_\mathcal{T}$. Before discussing this proof, we give an intermediate result.

Lemma 3. *For a TBox $\mathcal{T}$, an ABox $\mathcal{A}$ and a fact set $\mathcal{F}$ defined over $\mathsf{sig}(\mathcal{T})$, $\langle \mathcal{T}, \mathcal{A} \rangle$ is satisfiable iff $\langle \mathcal{T}_+, \mathcal{A} \rangle$ is, and $\langle \mathcal{T}, \mathcal{A} \rangle \models \mathcal{F}$ iff $\langle \mathcal{T}_+, \mathcal{A} \rangle \models \mathcal{F}$.*

Since $\mathcal{T}_+ \supseteq \mathcal{T}$, the "if" direction of this lemma follows trivially from monotonicity of logical entailment. The "only if" direction is proven in the appendix (cf. Lemma 13).

By Lemma 3, it suffices to show that our Datalog rewritings entail the same assertions as $\mathcal{T}_+$ in order to show completeness of our rewriting, which by Theorem 1 is consequence of the following lemma.

Lemma 4. *For a TBox $\mathcal{T}$, an ABox $\mathcal{A}$ and an assertion α over $\mathsf{sig}(\mathcal{T})$,*

- *if $\bot(t) \in \langle \mathcal{T}_+, \mathcal{A} \rangle^\infty$ with $t \in \mathsf{N_{gt}}$, then $\bot(u) \in \langle \mathcal{R}_\mathcal{T}, \mathcal{A} \rangle^\infty$ for some $u \in 2^{\mathsf{N_i}}$, and*

- *if $\alpha \in \langle \mathcal{T}_+, \mathcal{A} \rangle^\infty$, then $\alpha \in \langle \mathcal{R}, \mathcal{A} \rangle^\infty$.*

10

Let $\mathcal{O}_+^0, \mathcal{O}_+^1, \ldots$ be a chase sequence for the ontology $\mathcal{O}_+ = \langle \mathcal{T}_+, \mathcal{A} \rangle$ where axioms of Type $(\sqcap_r)$ are applied with the highest priority. For every $i \in [\![1, n]\!]$, we select an axiom $\rho_i \in \mathcal{T}_+$ and a substitution σ_i s.t. $\mathcal{O}_+^i = \mathcal{O}_+^{i-1} \langle \rho_i, \sigma_i \rangle$.

To prove Lemma 4, we show via induction that for every $i \geq 1$ and every assertion $\alpha \in \mathcal{O}_+^i$, we have $\alpha \in \mathcal{K}_{\mathcal{O}}^\infty$. The base case of this induction is trivial, since $\mathcal{O}_+^0 = \mathcal{A}$ and $\mathcal{A} \subseteq \mathcal{K}_{\mathcal{O}}^\infty$ by Definition 2. For the induction step, we provide a thorough case analysis based on the type of the axiom ρ_i, and the type of the elements occurring in the range of σ_i. Since $\alpha \in \mathcal{K}_{\mathcal{O}}^\infty$ for every assertion $\alpha \in \mathcal{O}^{i-1}$ by the induction hypothesis, many cases follow trivially. The more challenging cases are the following.

1. ρ_i is of Type $(\circ)$, $\sigma_i(x_0), \sigma_i(x_n) \in 2^{\mathsf{N}_i}$ and $\sigma_i(x_j) \in \mathsf{N}_0$ for $j \in \{1, \ldots, n-1\}$.

2. ρ_i is of Type $(\forall)$, $\sigma_i(x) \in \mathsf{N}_0$ and $\sigma_i(y) \in \mathsf{N}_i$.

3. ρ_i is of Type $(\leq)$, and a) $\sigma_i(y) \in 2^{\mathsf{N}_i}$ and $\sigma_i(x), \sigma_i(z) \in \mathsf{N}_0$, or b) $\sigma_i(x), \sigma_i(y) \in 2^{\mathsf{N}_i}$ and $\sigma_i(z) \in \mathsf{N}_0$.

Cases in which ρ_i is of Type $(\leq)$, and either $\sigma_i(z) \in 2^{\mathsf{N}_i}$ and $\sigma_i(x), \sigma_i(y) \in \mathsf{N}_0$, or $\sigma_i(x), \sigma_i(z) \in 2^{\mathsf{N}_i}$ and $\sigma_i(y) \in \mathsf{N}_0$, are also non-trivial, but analogous to Cases 3a) and 3b).

In all of the challenging cases, the occurrence of facts containing nulls in $\mathcal{O}_+^{i-1}$ results in the introduction of new assertions in $\mathcal{O}_+^i$—a situation previously illustrated in Example 4. To illustrate our completeness argument, we give a brief proof sketch that shows that induction step for Case (1). First, we introduce a preliminary lemma, which ensures that an axiom as used for Rule $(\circlearrowleft)$ is derived by the calculus if there is a corresponding cyclic path along nulls in $\mathcal{O}^\infty$.

Lemma 5. *Let $i \geq 1$, $R_1(t_0, t_1), \ldots, R_n(t_{n-1}, t_n) \in \mathsf{D}(\mathcal{O}_+^i)$, and $q, \hat{q} \in \mathcal{N}_{\mathcal{T}}(R)$ with $q \neq \hat{q}$. If*

- $q \rightarrow_{\mathsf{P}}^* \hat{q} \in \mathcal{N}_{\mathcal{T}}(R)$ *with* $\mathsf{P} = R_1 \cdot \ldots \cdot R_n$, *and*

- $t_0 \in 2^{\mathsf{N}_i}$, *and* $t_1, \ldots, t_{n-1} \in \mathsf{N}_0$;

then there exists $\mathbb{A} \subseteq \{A \mid A(t_0) \in \mathcal{O}_+^i\}$ *s.t.* $\mathbb{A} \sqcap X_q \sqsubseteq X_{\hat{q}} \in \Gamma(\mathcal{T}_\times)$.

This result can be shown via induction on the depth of the sequence $R_1(t_0, t_1), \ldots, R_n(t_{n-1}, t_n)$—the maximum minus the mimimum depth of a term in $t_0, \ldots, t_n$ in the rooted forest $\mathsf{F}(\mathcal{O}_+^i)$. We proceed with the proof for case (1).

Proof (Sketch). Let ρ_i be an axiom of the form $R_1 \circ \ldots \circ R_n \sqsubseteq S \in \mathcal{T}_+$. Then, $R_1(\sigma_i(x_0), \sigma_i(x_1)), \ldots, R_n(\sigma_i(x_{n-1}), \sigma_i(x_n)) \in \mathcal{O}_+^{i-1}$.

By Lemma 2 and the fact that $\mathcal{O}_+^i$ is closed under application of rules of Type $(\sqcap_r)$, there is a sequence $V_1(t_0, t_1), \ldots, V_m(t_{m-1}, t_m) \in \mathsf{D}(\mathcal{O}_+^{i-1})$ with $\sigma_i(x_{j-1}) = t_0$, $\sigma_i(x_j) = t_m$, and $V_1 \cdot \ldots \cdot V_m \in \mathcal{N}_{\mathcal{T}}(R_j)$ for every $j \in \{1, \ldots, n\}$ (note that possibly $m = 1$). By concatenating these sequences, we can construct a sequence $V_1(t_0, t_1), \ldots, V_m(t_{m-1}, t_m) \in \mathsf{D}(\mathcal{O}_+^{i-1})$ s.t. $\sigma_i(x_0) = t_0$, $\sigma_i(x_n) = t_m$, and $V_1 \cdot \ldots \cdot V_m \in \mathcal{N}_{\mathcal{T}}(S)$. Hence, there are states $q_0, \ldots, q_m$ s.t. $q_0 = i_V$, $q_m = f_V$, and $q_0 \rightarrow_{W_1} q_1 \rightarrow_{W_2} q_2 \ldots \rightarrow_{W_m} q_m \in \mathcal{N}_{\mathcal{T}}(V)$. Let $k_0, \ldots, k_o$ be the longest sorted sequence of natural numbers with $t_{k_j} \in 2^{\mathsf{N}_i}$ for all $j \in \{0, \ldots, o\}$. We show via induction that $S_{q_{k_j}}(t_0, t_{k_j}) \in \mathcal{K}_{\mathcal{O}}^\infty$ for all $j \in \{1, \ldots, o\}$. In turn, this implies $S(\sigma_i(x), \sigma_i(y)) \in \mathcal{K}_{\mathcal{O}}^\infty$ since $S_{q_{f_S}}(x, y) \rightarrow S(x, y) \in \mathcal{R}_{\mathcal{T}}$ as $k_0 = t_0 = \sigma_i(x)$, $t_{k_o} = t_m = \sigma_i(y)$, and $q_{k_m} = f_S$.

To show the base case, we check that $S_{q_{k_1}}(t_0, t_{k_1}) \in \mathcal{K}_{\mathcal{O}}^\infty$. We consider two possible cases a) and b) depending on whether $k_1 = 1$. a) Let $k_1 = 1$. Then, $W_1(t_0, t_1) \in \mathcal{K}_{\mathcal{O}}^\infty$ by the inductive

hypothesis. Since $W_1(x,y) \to S_{q_1}(x,y) \in \mathcal{R}_{\mathcal{T}}$, $S_{q_1}(t_0, t_1) \in \mathcal{K}_{\mathcal{O}}^\infty$. b) Let $k_1 > 1$. By Lemma 1, $t_{k_1} = t_0$. By Lemma 9, $\mathbb{A} \sqcap X_{i_S} \sqsubseteq X_{q_{k_1}} \in \Gamma(\mathcal{T}_\times)$ with $\mathbb{A} \subseteq \mathsf{N}_{\mathsf{c}}^{i-1}(t_0)$ and hence, $\mathbb{A}(x) \to S_{i_S, q_{k_1}}(x) \in \mathcal{R}_{\mathcal{T}}$. By the inductive hypothesis, $\mathbb{A}(t_0) \in \mathcal{K}_{\mathcal{O}}^\infty$ and hence, $S_{i_S, q_{k_1}}(t_0) \in \mathcal{K}_{\mathcal{O}}^\infty$. Since $S_{i_S, q_{k_1}}(x) \to S_{q_{k_1}}(x, x) \in \mathcal{R}_{\mathcal{T}}^\infty$, $S_{q_{k_1}}(t_0, t_{k_1}) \in \mathcal{K}_{\mathcal{O}}^\infty$.

To show the induction step, we verify that, for all $j \in \{2, \ldots, o\}$, $S_{q_{k_j}}(t_0, t_{k_j}) \in \mathcal{K}_{\mathcal{O}}^\infty$ provided that $S_{q_{k_{j-1}}}(t_0, t_{k_{j-1}}) \in \mathcal{K}_{\mathcal{O}}^\infty$. We consider two possible cases a) and b) depending on whether $k_1 = 1$. Let $k_j = k_{j-1} + 1$. Then, $W_{k_j}(t_{k_{j-1}}, t_{k_j}) \in \mathcal{K}_{\mathcal{O}}^\infty$ by the inductive hypothesis. Since $S_{q_{k_{j-1}}}(x, y) \wedge W_{k_j}(y, z) \to S_{q_{k_j}}(x, z) \in \mathcal{R}_{\mathcal{T}}$, $S_{q_{k_j}}(t_0, t_{k_j}) \in \mathcal{K}_{\mathcal{O}}^\infty$. Let $k_j > k_{j-1} + 1$. Then, $t_{k_j} = t_{k_{j-1}}$ by Lemma 1. This case is analogous to the second case considered in the proof of the base case. $\square$

In addition to showing correctness, we can show that our approach is worst-case optimal for Horn-$\mathcal{SRIQ}_\sqcap$ and even for less expressive DLs such as $\mathcal{ELH}$ and Horn-$\mathcal{SHIQ}$.

Definition 6. *An axiom set is a* Horn-$\mathcal{SHIQ}$ TBox *if, for every axiom $\rho \in \mathcal{T}$ of Type* $(\circ)$*, we have that a) $n = 1$ or b) $n = 2$, and $R_1 = R_2 = S$.*

A $\mathcal{ELH}$ TBox $\mathcal{T}$ is a set containing axioms of Type $(\sqcap)$, $(\exists)$, $(\circ)$, and of the form $\exists R.A \sqsubseteq B$ with $A, B \in \mathsf{N}_{\mathsf{c}}$ and $R \in \mathsf{N}_{\mathsf{r}}$ s.t. a) $n = 1$ for every axiom of the form $(\circ)$ and b) for every $R \in \mathsf{N}_{\mathsf{r}}$, $\mathcal{T}$ uses R or R^-, but not both.

Axioms of the form $\exists R.A \sqsubseteq B$ are equivalent to $A \sqsubseteq \forall R^-.B$, which is why $\mathcal{ELH}$ is included in Horn-$\mathcal{SRIQ}_\sqcap$.

Theorem 3. *Let $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ be an ontology. If $\mathcal{T}$ is Horn-$\mathcal{SRIQ}_\sqcap$/Horn-$\mathcal{SHIQ}$/$\mathcal{ELH}$, then we can compute $\mathcal{R}_{\mathcal{T}}$ and $\langle \mathcal{R}_{\mathcal{T}}, \mathcal{A} \rangle^\infty$ in* 2ExpTime/ExpTime/PTime*, respectively.*

Finally, we show that our rewritings can be transformed into DLP TBoxes. This feature may prove useful for users that want to produce KBs that are expressible using the OWL standard.

Definition 7. *A DLP TBox is an axiom set that a) does not contain axioms of Type $(\exists)$ and b) may contain axioms of the form $\prod_{i=1}^n A_i \sqsubseteq \exists R.Self$ with $A \in \mathsf{N}_{\mathsf{c}}$ and $R \in \mathsf{N}_{\mathsf{r}}$.*

Definition 8. *Given a TBox $\mathcal{T}$, the* DLP-rewriting $\mathcal{T}_{\mathrm{dlp}}$ *of $\mathcal{T}$ is the TBox containing every DLP axiom in $\mathcal{R}_{\mathcal{T}}$ which additionally satisfies all of the following.*

1. *If $\bigwedge_{A \in \mathbb{A}} A(x) \wedge R(x,y) \wedge B(y) \to C(y) \in \mathcal{R}_{\mathcal{T}}$, then $\mathbb{A} \sqsubseteq X_{\mathbb{A}}$, $X_{\mathbb{A}} \sqsubseteq \forall R.X_{R^-, \mathbb{A}}$, $X_{R^-, \mathbb{A}} \sqcap B \sqsubseteq C \in \mathcal{T}_{dlp}$.*

2. *If $\bigwedge_{A \in \mathbb{A}} A(x) \wedge R(x,y) \wedge B(y) \to S(x,y) \in \mathcal{R}_{\mathcal{T}}$, then $\mathbb{A} \sqsubseteq \exists W_{\mathbb{A}}.Self$, $B \sqsubseteq \exists W_B.Self$, $W_{\mathbb{A}} \circ R \circ W_B \sqsubseteq S \in \mathcal{T}_{dlp}$.*

3. *If $R_q(x,y) \wedge R_{q,\hat{q}}(y) \to R_{\hat{q}}(x,y) \in \mathcal{R}_{\mathcal{T}}$, then $R_{q,\hat{q}} \sqsubseteq \exists W_{q,\hat{q}}.Self$, $R_q \circ W_{q,\hat{q}} \sqsubseteq R_{\hat{q}} \in \mathcal{T}_{dlp}$.*

In the above, all $X_{\mathbb{A}}$ and $R.X_{R^-, \mathbb{A}}$ are fresh concepts unique for every $\mathbb{A} \subseteq \mathsf{N}_{\mathsf{c}}$ and $R \in \mathsf{N}_{\mathsf{r}}$, and all $W_{\mathbb{A}}$ and $W_{q,\hat{q}}$ are fresh roles unique for every $W \in \mathsf{N}_{\mathsf{r}}$ and the states q and $\hat{q}$.

The rules introduced in (1)–(3) in Definition 8 correspond to consequence-preserving transformations from rules to axioms described in [14]. From this, it follows that $\mathcal{T}_{\mathrm{dlp}}$ is an AR-rewriting of $\mathcal{T}$.
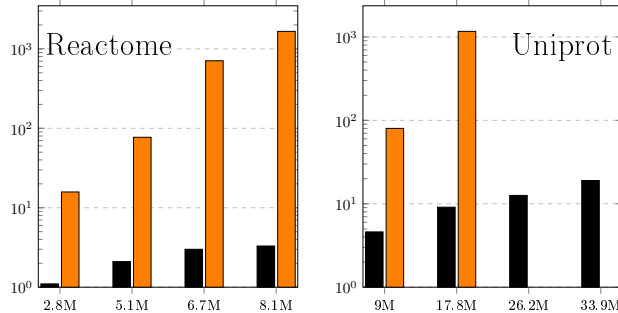
Figure 6: Times in seconds for RDFox (dark) and Konclude (bright), each over four ABoxes with increasing numbers of assertions.

4 Evaluation

We implement our rewriting technique in Java using the OWL-API [8] to handle OWL ontology files, and Clipper [6] to apply the calculus from Figure 4. We performed two different experiments to validate the practical usefulness of our approach. All files used in the evaluation (the implemented system, the ontologies, and the tools compared to) are available online.[1]

AR on Data-Intensive Ontologies We compared the performance of performing AR using our Datalog rewritings versus using the DL reasoner Konclude. We considered two real-world, data-intensive ontologies from the biological domain, *Reactome* and *Uniprot*, which were used in the evaluation of PAGOdA [24]. We have normalised these ontologies and removed axioms not expressible in Horn-$\mathcal{SRIQ}_\sqcap$. Also, we enriched Reactome and Uniprot with three and five axioms of Type ($\circ$), respectively, as neither ontology contained axioms of this form. These axioms are listed in the last section of the appendix. The resulting ontologies contained 485 (Reactome), and 304 (Uniprot) terminological axioms, respectively. For each ontology, we considered ABoxes of various sizes, generated by sampling the real-world ABoxes using the method by Zhou et al. [24].

The rewritten Datalog programs for the Reactome and Uniprot TBoxes contained 539 and 367 rules, and were computed in 221 and 182 seconds, respectively. We used RDFox (SVN version 2776) as Datalog engine for computing the chase of our rewritings [17], and compared its performance with that of Konclude v0.6.2. We performed all experiments and computed both rewritings on a MacBook Pro with a 2,4 GHz Intel Core i5 and 8GB of RAM. Figure 6 shows the wall-clock times measured in this experiment, ignoring the time used for parsing and loading, in logarithmic scale. While Konclude reports detailed times, for RDFox we have measured the time from within our prototype. For more information, see the logs with the resulting evaluation can be found online. Konclude timed-out (with a one hour time limit) for the two largest of the Uniprot samples. Hence no times are reported there. Note that our implementation is performing full AR, whilst Konclude only performed class retrieval.

Size of Rewritings Computed To get an idea on how our approach would perform on other data-intensive real-world ontologies, we computed rewritings for a selected set of TBoxes from MOWLCorp [16]. From each ontology in this corpus of DL ontologies, we removed axioms that, after normalisation, were not in Horn-$\mathcal{SRIQ}_\sqcap$, and selected from the resulting ontology set those which contained role chain axioms, and removed TBoxes with more than 1,000 axioms,

---

[1]https://lat.inf.tu-dresden.de/horn-sriq-rewriting/aaai-evaluation-files.tar.gz
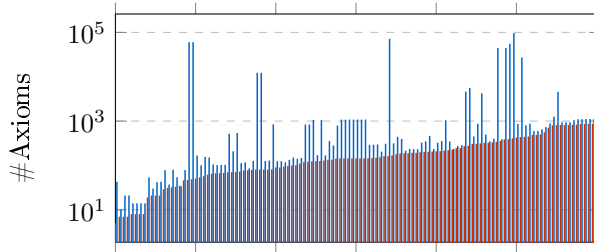
Figure 7: Sizes of TBoxes and their rewritings.

since TBoxes with smaller sizes are more likely to be used on large data sets. Furthermore, we removed all those ontologies which belong to any of the profiles OWL EL, OWL RL, and OWL QL, since they admit polynomial reasoning even without a Datalog rewriting. This resulted in a set of 187 ontologies on which we applied our implemented rewriting procedure.

For 121 ontologies, rewritings could be computed without memory errors. Often, memory errors were caused by complex role chains in the TBox which lead to an explosion of the resulting automata. For instance, we found one degenerate ontology in the corpus with only 10 axioms, 4 of which were role chain axioms with 8 roles each. For this TBox, $\mathcal{T}_\times$ contained 86,264 axioms, which Clipper could not handle. We believe that ontologies of this form are unlikely to be used in practice to reason about large ABoxes.

The sizes of the successful rewritings are shown in Figure 7, where the red bars correspond to the number of axioms in the input ontologues, and the blue bars to the number of rules in the resulting Datalog rewritings. For some ontologies, the rewritings got substantially larger. This was expected, and in theory unavoidable, due to the double exponential time complexity of assertion entailment in Horn-$\mathcal{SRIQ}_\sqcap$: for Datalog, this complexity is only polynomial, which is why our rewritings are in the worst case double exponential in the size of the input. Our evaluation confirms that these blow-ups are not only of theoretical nature, but do happen for the considered ontologies. On the other hand, in a lot of cases, the size of the computed rule sets was still of similar dimensions: in 59% of cases, the the increase was at most by 100%, and in 74% of cases, it was at most by 200%.

5 Conclusions and Future Work

To the best of our knowledge, we present the first data-independent Datalog transformation for Horn-$\mathcal{SRIQ}_\sqcap$, an expressive DL that allows for the use of the role chain constructor. Furthermore, we show that our transformation is worst-case optimal for $\mathcal{ELH}$, Horn-$\mathcal{SHIQ}$, and Horn-$\mathcal{SRIQ}_\sqcap$, and that the resulting Datalog programs can be translated into DLP ontologies. We empirically show that a) the use of Datalog rewritings can outperform state-of-the-art reasoners and that b) we can construct rewritings of moderate sizes for many real-world ontologies.

As for future work, we aim to develop a rewriting technique for expressive DLs language that allows for the use of non-deterministic role constructors and role chains based on the calculi from [5, 2]. Also, we intend to further optimise our prototype implementation, in order to produce even smaller rewritings and show that these can be efficiently computed.

6 Acknowledgements

References

[1] Shqiponja Ahmetaj, Magdalena Ortiz, and Mantas Šimkus. Polynomial datalog rewritings for expressive dls with closed predicates. In *Proc. IJCAI'16*, pages 878–885. AAAI Press, 2016.

[2] Andrew Bate, Boris Motik, Bernardo Cuenca Grau, František Simančík, and Ian Horrocks. Extending consequence-based reasoning to $\mathcal{SRIQ}$. In *Proc. KR'16*, pages 187–196. AAAI Press, 2016.

[3] Alison Callahan, Jose Cruz-Toledo, and Michel Dumontier. Ontology-based querying with Bio2RDF's linked open data. *J. Biomedical Semantics*, 4(S-1):S1, 2013.

[4] David Carral, Markus Krötzsch, Maximilian Marx, Ana Ozaki, and Sebastian Rudolph. Preserving constraints with the stable chase. In *Proc. ICDT'18*, pages 12:1–12:19. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018.

[5] David Tena Cucala, Bernardo Cuenca Grau, and Ian Horrocks. Consequence-based reasoning for description logics with disjunction, inverse roles, number restrictions, and nominals. In *Proc. IJCAI'18*, pages 1970–1976. AAAI Press, 2018.

[6] Thomas Eiter, Magdalena Ortiz, Mantas Šimkus, Trung-Kien Tran, and Guohui Xiao. Query rewriting for Horn-$\mathcal{SHIQ}$ plus rules. In *Proc. AAAI'12*. AAAI Press, 2012.

[7] Benjamin N. Grosof, Ian Horrocks, Raphael Volz, and Stefan Decker. Description logic programs: combining logic programs with description logic. In *Proc. WWW'03*, pages 48–57. ACM, 2003.

[8] Matthew Horridge and Sean Bechhofer. The OWL API: A java API for OWL ontologies. *Semantic Web*, 2(1):11–21, 2011.

[9] Ian Horrocks, Oliver Kutz, and Ulrike Sattler. The irresistible $\mathcal{SRIQ}$. In *Proc. OWLED'05*, volume 188 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2006.

[10] Ullrich Hustadt, Boris Motik, and Ulrike Sattler. Reducing $\mathcal{SHIQ}^-$ description logic to disjunctive datalog programs. In *Proc. KR'2004*, pages 152–162. AAAI Press, 2004.

[11] Yevgeny Kazakov. An extension of complex role inclusion axioms in the description logic $\mathcal{SROIQ}$. In *Proc. IJCAR'10*, volume 6173 of *Lecture Notes in Computer Science*, pages 472–486. Springer, 2010.

[12] Yevgeny Kazakov, Markus Krötzsch, and František Simančík. The incredible ELK: From polynomial procedures to efficient reasoning with $\mathcal{EL}$ ontologies. *J. Autom. Reasoning*, 53(1):1–61, 2014.

[13] Markus Krötzsch. Efficient rule-based inferencing for OWL EL. In *Proc. IJCAI'11*, pages 2668–2673. AAAI Press/IJCAI, 2011.

[14] Markus Krötzsch, Sebastian Rudolph, and Pascal Hitzler. Description logic rules. In *Proc. ECAI'08*, pages 80–84. IOS Press, 2008.

[15] Markus Krötzsch, Sebastian Rudolph, and Pascal Hitzler. Complexities of Horn dls. *ACM Trans. Comput. Logic*, 14(1):2:1–2:36, 2013.

[16] Nicolas Matentzoglu, Daniel Tang, Bijan Parsia, and Uli Sattler. The Manchester OWL repository: System description. In *Proc. ISWC'2014 (Posters & Demonstrations)*, volume 1272 of *CEUR Workshop Proceedings*, pages 285–288. CEUR-WS.org, 2014.

[17] Boris Motik, Yavor Nenov, Robert Piro, Ian Horrocks, and Dan Olteanu. Parallel material-isation of Datalog programs in centralised, main-memory RDF systems. In *Proc. AAAI'14*, pages 129–137. AAAI Press, 2014.

[18] Boris Motik, Rob Shearer, and Ian Horrocks. Hypertableau reasoning for description logics. *J. of Artificial Intelligence Research*, 36:165–228, 2009.

[19] Bijan Parsia, Nicolas Matentzoglu, Rafael S. Gonçalves, Birte Glimm, and Andreas Steigmiller. The OWL reasoner evaluation (ORE) 2015 competition report. *J. of Autom. Reasoning*, 59(4):455–482, 2017.

[20] Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur, and Yarden Katz. Pellet: A practical OWL-DL reasoner. *J. of Web Semantics*, 5(2):51–53, 2007.

[21] Andreas Steigmiller, Thorsten Liebig, and Birte Glimm. Konclude: System description. *J. of Web Semantics*, 27:78–85, 2014.

[22] Dmitry Tsarkov and Ian Horrocks. FaCT++ description logic reasoner: System description. In *Proc. IJCAR'06*, volume 4130 of *Lecture Notes in Computer Science*, pages 292–297. Springer, 2006.

[23] Denny Vrandečić and Markus Krötzsch. Wikidata: A free collaborative knowledgebase. *Commun. ACM*, 57(10), 2014.

[24] Yujiao Zhou, Bernardo Cuenca Grau, Yavor Nenov, Mark Kaminski, and Ian Horrocks. PAGOdA: Pay-as-you-go ontology query answering using a Datalog reasoner. *J. of Artif. Intell. Res.*, 54:309–367, 2015.

A   Forest-Model Property

In this section we show a preliminary result and Lemma 1.

Lemma 6. *Let $\mathcal{O}^0, \mathcal{O}^1, \ldots$ be a chase sequence for some ontology $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ and let S be some simple role with respect to $\mathcal{T}$. For all $i \geq 1$ and all binary facts of the form $S(t, u) \in \mathcal{O}^i$, we have that $S(t, u) \in \mathsf{D}(\mathcal{O}^i)$.*

Proof. We show the lemma via induction on the chase sequence $\mathcal{O}^0, \mathcal{O}^1, \ldots$ The base case trivially holds, since every binary fact in $\mathcal{O}^0$ is also contained in $\mathsf{D}(\mathcal{O}^0)$. To prove the inductive step (IS), we show that the lemma holds for $\mathcal{O}^i$ with $i \geq 1$ provided that (IH) it holds for $\mathcal{O}^{i-1}$. Let $\rho \in \mathcal{T}$ and σ be some axiom and substitution such that $\mathcal{O}^i = \mathcal{O}^{i-1}\langle \rho, \sigma \rangle$. If the axiom ρ is of Type ($\sqcap$), ($\forall$), ($\exists$), or ($\sqcap_\mathsf{r}$), then all binary facts in $\mathcal{O}^i$ are in $\mathsf{D}(\mathcal{O}^i)$ and the IS holds. We proceed to show that the IS also holds when the axiom ρ is of Type ($\leq$) or ($\circ$).

Case ($\leq$) Suppose for a contradiction that there is some $S(t,u) \in \mathcal{O}^i \setminus \mathcal{O}^{i-1}$ such that $S(t,u) \notin \mathsf{D}(\mathcal{O}^i)$ and S is a simple role. Then, there must be some fact $S(t',u') \in \mathcal{O}^{i-1}$ such that $S(t',u')\sigma_{x,y}^{\mathsf{rn}} = S(t,u)$. Since $S(t,u) \notin \mathsf{D}(\mathcal{O}^i)$, $S(t',u') \notin \mathsf{D}(\mathcal{O}^{i-1})$. This implies a contradiction, as we have that $S(t',u') \in \mathsf{D}(\mathcal{O}^{i-1})$ by IH.

Case ($\circ$) Then, $R_1(\sigma(x_0),\sigma(x_1)),\dots,R_n(\sigma(x_{n-1}),\sigma(x_n)) \in \mathcal{O}^{i-1}$. Let us suppose for a contradiction that $S(\sigma(x_0),\sigma(x_n)) \in \mathcal{O}^i$, S is a simple role, and $S(t,u) \notin \mathsf{D}(\mathcal{O}^i)$. Then, $n = 1$, R_1 is simple, and $R_1(\sigma(x_0),\sigma(x_1)) \notin \mathsf{D}(\mathcal{O}^{i-1})$. This implies a contradiction, since $R_1(\sigma(x_0),\sigma(x_1)) \in \mathsf{D}(\mathcal{O}^{i-1})$ by IH. $\qquad\square$

Lemma 1. *For all $i \geq 0$,*

- *all nulls in $\mathcal{O}^i$ occur as nodes in $\mathsf{F}(\mathcal{O}^i)$, and*

- *$\mathsf{F}(\mathcal{O}^i)$ is a rooted forest where every individual node is a root, and every null node is not.*

Proof. The lemma can be shown via induction on the sequence $\mathcal{O}^0, \mathcal{O}^1, \dots$ It is clear that the base case holds, as $\mathsf{F}(\mathcal{O}^0)$ does not contain any edges, and $\mathcal{O}^0$ does not contain any nulls. To show the inductive step, we show that the lemma holds for $\mathsf{F}(\mathcal{O}^i)$ with $i \geq 1$ provided that (IH) it holds for $\mathsf{F}(\mathcal{O}^{i-1})$. If the axiom ρ_i is of Type ($\sqcap$), ($\forall$), ($\circ$), or ($\sqcap_{\mathsf{r}}$), then the set of nulls in $\mathcal{O}^i$ coincides with the set of nulls in $\mathcal{O}^{i-1}$ and $\mathsf{F}(\mathcal{O}^i) = \mathsf{F}(\mathcal{O}^{i-1})$. Therefore, the lemma holds by IH for all of these cases. We proceed to show that the lemma also holds when ρ_i is of Type ($\exists$) or ($\leq$).

Case ($\exists$) Then, $A(\sigma_i(x)) \in \mathcal{O}^{i-1}$ and $\mathcal{O}^i = \mathcal{O}^{i-1} \cup \{R(\sigma_i(x),n), B(n)\}$ for some fresh null n. We consider two possible cases.

- Let $\sigma_i(x) \in 2^{\mathsf{N}_i}$. By IH, $\sigma_i(x)$ is a root in $\mathsf{F}(\mathcal{O}^{i-1})$.

- Let $\sigma_i(x) \in 2^{\mathsf{N}_0}$. By IH, there is some sequence of nodes $t_0,\dots,t_n \in \mathsf{F}(\mathcal{O}^{i-1})$ such that $t_0 \in 2^{\mathsf{N}_i}$ is a root, all $t_1,\dots,t_n \in \mathsf{N}_0$ are not, and $t_n = \sigma_i(x)$.

In either case, n only occurs in the edge $\sigma_i(x) \to n \in \mathsf{F}(\mathcal{O}^i)$ since n only occurs in facts $R(\sigma_i(x),n), B(n) \in \mathcal{O}^i$.

Case ($\leq$) In this case, no fresh nulls are introduced in $\mathcal{O}^i$ and hence, the first part of the lemma holds. Note that, R is simple by Definition 1. We consider four possible cases.

- $\sigma_i(y), \sigma_i(z) \in 2^{\mathsf{N}_i}$. Then, $\mathsf{F}(\mathcal{O}^i)$ results from replacing the roots $\sigma_i(y) \in 2^{\mathsf{N}_i}$ and $\sigma_i(z) \in 2^{\mathsf{N}_i}$ in $\mathsf{F}(\mathcal{O}^{i-1})$ with the fresh root $\sigma_i(y) \cup \sigma_i(z) \in 2^{\mathsf{N}_i}$.

- $\sigma_i(y) \in \mathsf{N}_0$ and $\sigma_i(z) \in 2^{\mathsf{N}_i}$. By Lemma 6, $R(\sigma_i(x),\sigma_i(y)) \in \mathsf{D}(\mathcal{O}^{i-1})$ and hence, $\sigma_i(x)$ is the predecessor of $\sigma_i(y)$ in $\mathsf{F}(\mathcal{O}^{i-1})$. We study two possible cases.

  - $\sigma_i(x) \in 2^{\mathsf{N}_i}$ is a root. Then, $\mathsf{F}(\mathcal{O}^i)$ results from replacing all occurrences of $\sigma_i(y)$ in $\mathsf{F}(\mathcal{O}^{i-1})$ with the root $\sigma_i(z)$, and then erasing all edges from the root $\sigma_i(x)$ to the root $\sigma_i(z)$.

  - $\sigma_i(x) \in \mathsf{N}_0$ is not a root. Then, $\mathsf{F}(\mathcal{O}^i)$ results from replacing all occurrences of the non-root $\sigma_i(y)$ in $\mathsf{F}(\mathcal{O}^{i-1})$ with the predecessor of its predecessor, i.e., $\sigma_i(z)$.

- $\sigma_i(y) \in 2^{\mathsf{N}_i}$ and $\sigma_i(z) \in \mathsf{N}_0$. Analogous to the previous case.

- $\sigma_i(y), \sigma_i(z) \in \mathsf{N}_0$. By Lemma 6, $R(\sigma_i(x), \sigma_i(y)), R(\sigma_i(x), \sigma_i(z)) \in \mathsf{D}(\mathcal{O}^{i-1})$. Since $\sigma_i(y) \neq \sigma_i(z)$, three possible cases arise.

  - $\sigma_i(y)$ is the predecessor of $\sigma_i(x)$ and $\sigma_i(x)$ is the predecessor of $\sigma_i(z)$. Then, $\mathsf{F}(\mathcal{O}^i)$ results from replacing all occurrences of the non-root $\sigma_i(z)$ in $\mathsf{F}(\mathcal{O}^{i-1})$ with the predecessor of its predecessor, i.e., $\sigma_i(y)$; and then erasing all edges from $\sigma_i(x)$ to $\sigma_i(z)$.

  - $\sigma_i(z)$ is the predecessor of $\sigma_i(x)$ which is the predecessor of $\sigma_i(y)$. Analogous to the previous case.

  - $\sigma_i(x)$ is the predecessor of $\sigma_i(y)$ and $\sigma_i(z)$. Then, $\mathsf{F}(\mathcal{O}^i)$ results from replacing all occurrences of the non-root $\sigma_i(z)$ by its sibling $\sigma_i(y)$.

In either case, we can verify that $\mathsf{F}(\mathcal{O}^i)$ is a rooted forest where every individual node is a root, and every null node is not. $\qquad\square$

B   Completeness

In this section, we show Lemma 11, from which Lemma 4 directly follows. Prior to stating and proving this lemma, we introduce some preliminary notions and intermediate results.

Consider some ontology $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$. Furthermore, consider a chase sequence $\mathcal{O}^0_+, \mathcal{O}^1_+, \ldots$ for $\mathcal{O}_+ = \langle \mathcal{T}_+, \mathcal{A} \rangle$, a sequence of axioms $\rho_1, \rho_2, \ldots \in \mathcal{T}_+$, and a sequence of substitutions such that all of the following conditions hold for all $i \geq 1$.

1. The set $\mathcal{O}^i_+$ is the application of $\langle \rho_i, \sigma_i \rangle$ on $\mathcal{O}^{i-1}_+$.

2. If there is some axiom of Type ($\sqcap$) or ($\sqcap_{\mathsf{r}}$) in $\mathcal{T}_+$ that is applicable to $\mathcal{O}^{i-1}_+$, then ρ_i is of Type ($\sqcap$) or ($\sqcap_{\mathsf{r}}$).

3. If there are not any axioms of Type ($\sqcap$) or ($\sqcap_{\mathsf{r}}$) in $\mathcal{T}_+$ applicable to $\mathcal{O}^{i-1}_+$, and there is an axiom $\rho \in \mathcal{T}_+$ of the form $A \sqsubseteq \forall R.B$ and a substitution σ such that $A(\sigma(x)) \in \mathcal{O}^{i-1}_+$ and $R(\sigma(x), \sigma(y)) \in \mathsf{D}(\mathcal{O}^{i-1}_+)$; then the axiom ρ_i is of the form $C \sqsubseteq \forall S.D$ and $S(\sigma_i(x), \sigma_i(y)) \in \mathsf{D}(\mathcal{O}^{i-1}_+)$.

Because of conditions (1), (2), and (2), we can show the following.

Lemma 7. *For all $i \geq 1$, if ρ_i is of Type ($\forall$), then $R(\sigma_i(x), \sigma_i(y)) \in \mathsf{D}(\mathcal{O}^{i-1}_+)$.*

Proof. Suppose for a contradiction that ρ_i is of the form $A \sqsubseteq \forall R.B$ and $R(\sigma_i(x), \sigma_i(y)) \notin \mathsf{D}(\mathcal{O}^{i-1}_+)$. By (1) above, $A(\sigma_i(x)), R(\sigma_i(x), \sigma_i(y)) \in \mathcal{O}^{i-1}_+$ and hence, by Lemma 2, there are some $R_1(t_0, t_1), \ldots, R_n(t_{n-1}, t_n) \in \mathsf{D}(\mathcal{O}^{i-1}_+)$ such that $t_0 = \sigma_i(x)$, $t_n = \sigma_i(y)$, and $R_1 \cdot \ldots \cdot R_n \in \mathcal{N}_{\mathcal{T}_+}(R)$. Note that, Lemma 2 is applicable because there are no axioms of Type $\sqcap_{\mathsf{r}}$ applicable to $\mathcal{O}^{i-1}_+$ by 2. Therefore, there must be some states $q_0, \ldots, q_n$ such that $q_0 = i_R$, $q_n = f_R$, and $q_{j-1} \to^*_{R_j} q_j \in \mathcal{N}_{\mathcal{T}_+}(R)$ for all $j \in [\![1, n]\!]$. Hence, $A \sqsubseteq A_{q_0}, A_{q_0} \sqsubseteq \forall R_1.A_{q_1}, \ldots, A_{q_{n-1}} \sqsubseteq \forall R_n.A_{q_n}, A_{q_n} \sqsubseteq B \in \mathcal{T}_+$. By conditions (2) and (3) above, $A_{q_0}(t_0), \ldots, A_{q_n}(t_n), B(\sigma_i(y)) \in \mathcal{O}^{i-1}_+$ (note that $t_n = \sigma_i(y)$). Therefore, $\mathcal{O}^i_+ = \mathcal{O}^{i-1}_+$ and $\langle \rho_i, \sigma_i \rangle$ is not applicable to $\mathcal{O}^{i-1}_+$. This contradicts Definition 2. $\qquad\square$

We introduce some notation used in across this section.

- For all $i \geq 0$ and $t, u \in \mathsf{N}_{\mathsf{gt}}$, let $\mathsf{N}^i_{\mathsf{c}}(t) = \{A \mid A(t) \in \mathcal{O}^i_+\}$ and $\mathsf{N}^i_{\mathsf{r}}(t, u) = \{R \mid R(t, u) \in \mathsf{D}(\mathcal{O}^i_+)\}$.

- Let $\mathbb{A} = A_1 \sqcap \ldots \sqcap A_n$ be a conjunction of concepts, $x \in \mathsf{N_v}$, and $t \in \mathsf{N_{gt}}$. Then, we write $\mathbb{A}(x)$ as a shortcut for $A_1(x) \wedge \ldots \wedge A_n(x)$ and $\mathbb{A}(t) \in \mathcal{F}$ as a shortcut for $A_1(t), \ldots, A_n(t) \in \mathcal{F}$.

Lemma 8. *Let $i \geq 0$ and let $t, u \in \mathsf{N_{gt}}$ be some terms in $\mathcal{O}_+^i$. If t is the predecessor of u in $\mathsf{F}(\mathcal{O}_+^i)$, then $\mathbb{A} \sqsubseteq \exists \mathsf{N_r^i}(t,u).\mathsf{N_c^i}(u) \in \Gamma(\mathcal{T}_\times)$ for some $\mathbb{A} \subseteq \mathsf{N_c^i}(t)$.*

Proof. We verify this result via induction on the chase sequence $\mathcal{O}_+^0, \mathcal{O}_+^1, \ldots$ Since $\mathsf{F}(\mathcal{O}_+^0)$ is empty, the base case trivially holds. To show the induction step, we check that the lemma holds for any $i \geq 1$ irrespectively of the type of axiom ρ_i. In the following enumeration, we consider all cases that do not automatically follow IH.

($\sqcap$) Then, $A_1(\sigma_i(x)), \ldots, A_n(\sigma_i(x)) \in \mathcal{O}_+^{i-1}$. Let us assume that $\sigma_i(x)$ is the successor of some term t occurring in $\mathcal{O}_+^{i-1}$, as otherwise the case holds by IH. By IH, $\mathbb{A}' \sqsubseteq \exists \mathbb{R}.\mathbb{B} \in \Gamma(\mathcal{T}_\times)$ with $\mathbb{A}' \subseteq \mathsf{N_c^{i-1}}(t)$, $\mathbb{R} = \mathsf{N_r^{i-1}}(t, \sigma_i(x))$, and $\mathbb{B} = \mathsf{N_c^{i-1}}(\sigma_i(x))$. Since $\bigsqcap_{j=1}^n A_j \sqsubseteq B \in \Gamma(\mathcal{T}_\times)$ and $A_1, \ldots, A_n \in \mathbb{B}$, $\mathbb{A}' \sqsubseteq \exists \mathbb{R}.(\mathbb{B} \sqcap B) \in \Gamma(\mathcal{T}_\times)$.

($\forall$) Then, $A(\sigma_i(x))), R(\sigma_i(x), \sigma_i(y)) \in \mathcal{O}_+^{i-1}$. By Lemma 7, $R(\sigma_i(x), \sigma_i(y)) \in \mathsf{D}(\mathcal{O}_+^{i-1})$. We assume that $\sigma_i(x)$ is the predecessor of $\sigma_i(y)$, as otherwise this case holds by IH. By IH, $\mathbb{A}' \sqsubseteq \exists \mathbb{R}.\mathbb{B} \in \Gamma(\mathcal{T}_\times)$ with $\mathbb{A}' \subseteq \mathsf{N_c^{i-1}}(\sigma_i(x))$, $\mathbb{R} = \mathsf{N_r^{i-1}}(\sigma_i(x), \sigma_i(y))$, and $\mathbb{B} = \mathsf{N_c^{i-1}}(\sigma_i(y))$. Since $A \sqsubseteq \forall R.B \in \mathcal{T}_\times$ and $R \in \mathbb{R}$, $\mathbb{A}' \sqcap A \sqsubseteq \exists \mathbb{R}.(\mathbb{B} \sqcap B) \in \Gamma(\mathcal{T}_\times)$.

($\exists$) Then, $A(\sigma_i(x)) \in \mathcal{O}_+^{i-1}$. Since $A \sqsubseteq \exists R.B \in \mathcal{T}_\times$, $A \sqsubseteq \exists R.B \in \Gamma(\mathcal{T}_\times)$.

($\leq$) Then, $A(\sigma_i(x)), R(\sigma_i(x), \sigma_i(y)), B(\sigma_i(y)), R(\sigma_i(x), \sigma_i(z)), B(\sigma_i(z)) \in \mathcal{O}_+^{i-1}$. By Definition 1, the role R is simple and hence, $R(\sigma_i(x), \sigma_i(y)), R(\sigma_i(x), \sigma_i(z)) \in \mathsf{D}(\mathcal{O}_+^{i-1})$ by Lemma 6. We consider three possible cases.

  - $\sigma_i(x)$ is the predecessor of both $\sigma_i(y)$ and $\sigma_i(z)$. By Lemma 1, $\sigma_i(y), \sigma_i(z) \in \mathsf{N_0}$ and hence, by IH, $\mathbb{A}' \sqsubseteq \exists \mathbb{R}.\mathbb{B}, \mathbb{A}'' \sqsubseteq \exists \mathbb{S}.\mathbb{C} \in \Gamma(\mathcal{T}_\times)$ with $\mathbb{A}', \mathbb{A}'' \subseteq \mathsf{N_c^{i-1}}(\sigma_i(x))$, $\mathbb{R} = \mathsf{N_r^{i-1}}(\sigma_i(x), \sigma_i(y))$, $\mathbb{B} = \mathsf{N_c^{i-1}}(\sigma_i(y))$, $\mathbb{S} = \mathsf{N_r^{i-1}}(\sigma_i(x), \sigma_i(z))$, and $\mathbb{C} = \mathsf{N_c^{i-1}}(\sigma_i(z))$. Since $A \sqsubseteq \leqslant 1\, R.B \in \mathcal{T}_\times$, $R \in \mathbb{R} \cap \mathbb{S}$, and $B \in \mathbb{B} \cap \mathbb{C}$, $\mathbb{A}' \sqcap \mathbb{A}'' \sqcap A \sqsubseteq \exists (\mathbb{R} \sqcap \mathbb{S}).(\mathbb{B} \sqcap \mathbb{C}) \in \Gamma(\mathcal{T}_\times)$.

  - $\sigma_i(z)$ is the predecessor of $\sigma_i(x)$ and $\sigma_i(x)$ is the predecessor of $\sigma_i(y)$. By Lemma 1, $\sigma_i(x), \sigma_i(y) \in \mathsf{N_0}$ and hence, by IH, $\mathbb{A}' \sqsubseteq \exists \mathbb{R}.\mathbb{B}, \mathbb{B}' \sqsubseteq \exists \mathbb{S}.\mathbb{C} \in \Gamma(\mathcal{T}_\times)$ with $\mathbb{A}' \subseteq \mathsf{N_c^{i-1}}(\sigma_i(z))$, $\mathbb{R} = \mathsf{N_r^{i-1}}(\sigma_i(z), \sigma_i(x))$, $\mathbb{B}' \subseteq \mathsf{N_c^{i-1}}(\sigma_i(x)) = \mathbb{B}$, $\mathbb{S} = \mathsf{N_r^{i-1}}(\sigma_i(x), \sigma_i(y))$, and $\mathbb{C} = \mathsf{N_c^{i-1}}(\sigma_i(y))$. Since $A \sqsubseteq \leqslant 1\, R.B \in \mathcal{T}_\times$, $R^- \in \mathbb{R}$, $A \in \mathbb{B} \cup \mathbb{C}$, and $R \in \mathbb{S}$, $\mathbb{A}' \sqcap B \sqsubseteq \exists (\mathbb{R} \sqcap \mathbb{S}^-).\mathbb{B} \in \Gamma(\mathcal{T}_\times)$.

  - $\sigma_i(y)$ is the predecessor of $\sigma_i(x) \in \mathsf{N_0}$ and $\sigma_i(x)$ is the predecessor of $\sigma_i(z) \in \mathsf{N_0}$. Analogous to the previous case.

($\circ$) Then, $R_1(\sigma_i(x_0), \sigma_i(x_1)), \ldots, R_{n-1}(\sigma_i(x_{n-1}), \sigma_i(x_n)) \in \mathcal{O}_+^{i-1}$. We assume that $n = 1$ and $R_1(\sigma_i(x_0), \sigma_i(x_1)) \in \mathsf{D}(\mathcal{O}_+^{i-1})$, as otherwise $S(\sigma_i(x_0), \sigma_i(x_n)) \notin \mathsf{D}(\mathcal{O}_+^{i-1})$. We consider two possible cases.

  - $\sigma_i(x_0)$ is the predecessor of $\sigma_i(x_1)$. By IH, $\mathbb{A}' \sqsubseteq \exists \mathbb{R}.\mathbb{B} \in \Gamma(\mathcal{T}_\times)$ with $\mathbb{A}' \subseteq \mathsf{N_c^{i-1}}(\sigma_i(x_0))$, $\mathbb{R} = \mathsf{N_r^{i-1}}(\sigma_i(x_0), \sigma_i(x_1))$, and $\mathbb{B} = \mathsf{N_c^{i-1}}(\sigma_i(x_1))$. Since $R_1 \sqsubseteq S \in \mathcal{T}_\times$ and $R_1 \in \mathbb{R}$, $\mathbb{A}' \sqsubseteq \exists (\mathbb{R} \sqcap S).\mathbb{B} \in \Gamma(\mathcal{T}_\times)$.

  - $\sigma_i(x_1)$ is the predecessor of $\sigma_i(x_0)$. By IH, $\mathbb{A}' \sqsubseteq \exists \mathbb{R}.\mathbb{B} \in \Gamma(\mathcal{T}_\times)$ with $\mathbb{A}' \subseteq \mathsf{N_c^{i-1}}(\sigma_i(x_1))$, $\mathbb{R} = \mathsf{N_r^{i-1}}(\sigma_i(x_1), \sigma_i(x_0))$, and $\mathbb{B} = \mathsf{N_c^{i-1}}(\sigma_i(x_0))$. Since $R_1^- \sqsubseteq S^- \in \mathcal{T}_\times$ and $R_1^- \in \mathbb{R}$, $\mathbb{A}' \sqsubseteq \exists (\mathbb{R} \sqcap S^-).\mathbb{B} \in \Gamma(\mathcal{T}_\times)$.

($\sqcap_\mathsf{r}$) Analogous to the previous case. $\qquad\square$

To structure some of the induction arguments below, we introduce the notion of depth of a term and a sequence of direct fact. Note that, we consider the roots in a rooted graph to have depth 0.

Definition 9. *For $i \geq 0$ and $t \in \mathsf{N}_{\mathsf{gt}}$ a term occurring in $\mathcal{O}_+^i$, let $\mathsf{dep}_i(t)$ be the depth of t in the rooted forest $\mathsf{F}(\mathcal{O}_+^i)$. For a sequence $\mathcal{F} = R_1(t_0, t_1), \ldots, R_n(t_{n-1}, t_n)$, $\mathsf{dep}_i(\mathcal{F}) = max(\mathsf{dep}_i(t_1), \ldots, \mathsf{dep}_i(t_n)) - min(\mathsf{dep}_i(t_1), \ldots, \mathsf{dep}_i(t_n))$.*

Lemma 9. *Consider $i \geq 1$, $R_1(t_0, t_1), \ldots, R_n(t_{n-1}, t_n) \in \mathsf{D}(\mathcal{O}_+^i)$, and states q and $\hat{q}$ in the NFA $\mathcal{N}_\mathcal{T}(R)$ with $q \neq \hat{q}$. If $q \rightarrow_\mathsf{P}^* \hat{q} \in \mathcal{N}_\mathcal{T}(R)$ with $\mathsf{P} = R_1 \cdot \ldots \cdot R_n$ and all $t_1, \ldots, t_{n-1}$ are descendants of t_0 in $\mathsf{F}(\mathcal{O}_+^i)$, then $\mathbb{A}' \sqcap X_q \sqsubseteq X_{\hat{q}} \in \Gamma(\mathcal{T}_\times)$ for some $\mathbb{A}' \subseteq \mathsf{N}_\mathsf{c}^i(t_0)$.*

Proof. We prove the lemma via induction on the depth of the sequence $\mathcal{F} = R_1(t_0, t_1), \ldots, R_n(t_{n-1}, t_n)$. Before proceeding with this inductive argument, we derive some conclusions from the premise of the lemma. Since $q \rightarrow_\mathsf{P}^* \hat{q} \in \mathcal{N}_\mathcal{T}(R)$, there are some states $q_0, \ldots, q_n$ such that $q_0 = q$, $q_n = \hat{q}$, and $q_{j-1} \rightarrow_{R_j}^* q_j \in \mathcal{N}_\mathcal{T}(R)$ for all $j \in [\![1, n]\!]$. Hence, $X_{q_{j-1}} \sqsubseteq \forall R_j.X_{q_j} \in \mathcal{T}_\times$ for all $j \in [\![1, n]\!]$.

To show the base case, we check that the lemma holds if $\mathsf{dep}_i(\mathcal{F}) = 1$. In this case, $n = 2$ by Lemma 1 and the fact that $t_j \neq t_0$ for all $j \in \{1, \ldots, n-1\}$. By Lemma 8, $\mathbb{A}' \sqsubseteq \exists \mathbb{R}.\mathbb{B} \in \Gamma(\mathcal{T}_\times)$ with $\mathbb{A}' \subseteq \mathsf{N}_\mathsf{c}^i(t_0)$, $\mathbb{R} = \mathsf{N}_\mathsf{r}^i(t_0, t_1)$, and $\mathbb{B} = \mathsf{N}_\mathsf{c}^i(t_1)$. Since $R_1, R_2^- \in \mathbb{R}$, $\mathbb{A}' \sqcap X_q \sqsubseteq \exists \mathbb{R}.(\mathbb{B} \sqcap X_{q_1})$, $\mathbb{A}' \sqcap X_q \sqsubseteq X_{\hat{q}} \in \Gamma(\mathcal{T}_\times)$ (note that $q = q_0$ and $\hat{q} = q_2$).

To show the inductive step, we verify that the lemma holds if $\mathsf{dep}_i(\mathcal{F}) \geq 2$ assuming that (IH) it holds for every sequence of facts $R_k(t_{k-1}, t_k), \ldots, R_\ell(t_{\ell-1}, \ell)$ with $k > 1$ and $\ell < n$. Note that, since every $t_1, \ldots, t_{n-1}$ is a descendant of t_0, every such sequence has lesser depth than $\mathcal{F}$. Let $k_0, \ldots, k_m \in \mathbb{N}$ be the longest sorted sequence of numbers such that $t_{k_j} = t_1$ for all $j \in [\![0, n]\!]$. By repeated application of the IH, we conclude that, for all $j \in [\![1, n]\!]$ there is some $\mathbb{B}_j' \subseteq \mathsf{N}_\mathsf{c}^i(t_1)$ with $\mathbb{B}_j' \sqcap A_{q_{k_{j-1}}} \sqsubseteq A_{q_{k_j}} \in \Gamma(\mathcal{T}_\times)$. By Lemma 8, $\mathbb{A}' \sqsubseteq \exists \mathbb{R}.\mathbb{B}$ with $\mathbb{A}' \subseteq \mathsf{N}_\mathsf{c}^i(t_0)$, $\mathbb{R} = \mathsf{N}_\mathsf{r}^i(t_0, t_1)$, and $\mathbb{B} = \mathsf{N}_\mathsf{c}^i(t_1)$. Hence, $\mathbb{A}' \sqcap A_{q_0} \sqsubseteq \exists \mathbb{R}.(\mathbb{B} \sqcap A_{q_1}) \in \Gamma(\mathcal{T}_\times)$ since $R_1 \in \mathbb{R}$ and $X_{q_0} \sqsubseteq \forall R_1.X_{q_1} \in \mathcal{T}_\times$. Therefore, $\mathbb{A}' \sqcap A_{i_R} \sqsubseteq \exists \mathbb{R}.(\mathbb{B} \sqcap A_{q_{k_1}} \sqcap A_{q_2} \sqcap \ldots \sqcap A_{q_{k_{m-1}}} \sqcap A_{q_{n-1}}), \in \Gamma(\mathcal{T}_\times)$ (note that $i_R = q_0$, $q_{k_1} = q_1$, and $q_{k_m} = q_{n-1}$). Since $R_n^- \in \mathbb{R}$ and $A_{q_{n-1}} \sqsubseteq \forall R.A_{f_R}$, $\mathbb{A}' \sqcap A_{q_{i_R}} \sqsubseteq A_{f_R} \in \Gamma(\mathcal{T}_\times)$ (note that $f_R = q_n$). □

Lemma 10. *Consider $i \geq 1$, $R_1(t_0, t_1), \ldots, R_n(t_{n-1}, t_n) \in \mathsf{D}(\mathcal{O}^i)$ with $n \geq 2$, $A \sqsubseteq \leqslant 1 \, R.B \in \mathcal{T}_\times$, $S \in \mathsf{N}_\mathsf{r}$, and states q and $\hat{q}$ in the NFA $\mathcal{N}_\mathcal{T}(S)$. If*

- $R^-(t_0, t), A(t), R(t, t_n), B(t_n) \in \mathcal{O}^i$ *with t the predecessor of t_n and successor of t_0,*

- $q \rightarrow_\mathsf{P}^* \hat{q} \in \mathcal{N}_\mathcal{T}(S)$ *with $\mathsf{P} = R_1 \cdot \ldots \cdot R_n$, and*

- $t_0 \in 2^{\mathsf{N}_\mathsf{i}}$ *and $t_1, \ldots, t_n \in \mathsf{N}_0$,*

then, $\mathbb{A}' \sqcap B \sqcap X_q \sqsubseteq X_{\hat{q}} \in \Gamma(\mathcal{T}_+)$ for some $\mathbb{A}' \subseteq \mathsf{N}_\mathsf{c}^i(t_0)$

Proof. By the premise of the lemma, there are some states $q_0, \ldots, q_n$ such that $q_0 = q$, $q_n = \hat{q}$, and $q_{j-1} \rightarrow_{R_j}^* q_j \in \mathcal{N}_\mathcal{T}(S)$ for all $j \in [\![1, n]\!]$. By the definition of $\mathcal{T}_+$ and $\mathcal{T}_\times$, $X_{q_{j-1}} \sqsubseteq \forall R_j.X_{q_j} \in \mathcal{T}_\times$ for all $j \in [\![1, n]\!]$.

By Lemma 8, $\mathbb{A}' \sqsubseteq \exists \mathbb{R}.\mathbb{B}, \mathbb{B}' \sqsubseteq \exists \mathbb{S}.\mathbb{C} \in \Gamma(\mathcal{T}_\times)$ with $\mathbb{A}' \subseteq \mathsf{N}_\mathsf{c}^i(t_0)$, $\mathbb{R} = \mathsf{N}_\mathsf{r}^i(t_0, t_1)$, $\mathbb{B}' \subseteq \mathsf{N}_\mathsf{c}^i(t_1) = \mathbb{B}$, $\mathbb{S} = \mathsf{N}_\mathsf{r}^i(t_1, t_n)$, and $\mathbb{C} = \mathsf{N}_\mathsf{c}^i(t_n)$. Hence, $\mathbb{A}' \sqcap X_q \sqsubseteq \exists \mathbb{R}.(\mathbb{B} \sqcap X_{q_1}) \in \Gamma(\mathcal{T}_\times)$ since $R_1 \in \mathbb{R}$ and $t_1 = t$ by Lemma 1 (note that $q_0 = q$).

Let $k_0, \ldots, k_m \in \mathbb{N}$ be the longest sorted sequence such that $t_{k_j} = t_1$ for all $j \in [\![0, n]\!]$. By Lemma 9, there is some $\mathbb{B}_j' \subseteq \mathbb{B}$ such that $\mathbb{B}_j' \sqcap X_{q_{k_{j-1}}} \sqsubseteq X_{q_{k_j}} \in \Gamma(\times)$ for all $j \in [\![1, n]\!]$. Hence,

20

$\mathbb{A}' \sqcap X_q \sqsubseteq \exists \mathbb{R}.(\mathbb{B} \sqcap X_{q_{k_0}} \sqcap \ldots \sqcap X_{q_{k_m}}) \in \Gamma(\times)$ (note that $k_0 = 1$). By Lemma 1, $t_{k_m+1} = t_n$ (note that possibly $k_m + 1 \neq n$). Since $R_{k_m+1} \in \mathbb{S}$, $\mathbb{B}' \sqcap X_{q_{k_m}} \sqsubseteq \exists \mathbb{S}.(\mathbb{C} \sqcap X_{q_{k_m+1}}) \in \Gamma(\times)$.

Let $\ell_0, \ldots, \ell_o \in \mathbb{N}$ be the longest sorted sequence where $\ell_0 = k_m+1$ and $t_{\ell_j} = t_n$ for all $j \in [\![1, o]\!]$. By Lemma 9, there is some $\mathbb{C}'_j \subseteq \mathbb{C}$ such that $\mathbb{C}'_j \sqcap X_{q_{\ell_{j-1}}} \sqsubseteq X_{q_{\ell_j}} \in \Gamma(\times)$ for all $j \in [\![1, o]\!]$. Hence, $\mathbb{B}' \sqcap X_{q_{k_m}} \sqsubseteq \exists \mathbb{S}.(\mathbb{C} \sqcap X_{q_{\ell_0}} \sqcap \ldots \sqcap X_{q_{\ell_{o-1}}} \sqcap X_{\hat{q}}) \in \Gamma(\times)$ (note that $q_{k_m+1} = q_{\ell_1}$ and $q_{\ell_o} = \hat{q}$). Since $R^- \in \mathbb{S}$, $A, B \in \mathbb{B}$, $R \in \mathbb{R}$, $\mathbb{A}' \sqcap X_q \sqsubseteq \exists \mathbb{R}.(\mathbb{B} \sqcap A \sqcap X_{q_{k_0}} \sqcap \ldots \sqcap X_{q_{k_m}})$, $\mathbb{B}' \sqcap X_{q_{k_m}} \sqsubseteq \exists \mathbb{S}.(\mathbb{C} \sqcap B \sqcap X_{q_{\ell_0}} \sqcap \ldots \sqcap X_{q_{\ell_{o-1}}} \sqcap X_{\hat{q}})$, we have that $\mathbb{A}' \sqcap B \sqcap X_q \sqsubseteq X_{\hat{q}} \in \Gamma(\times)$. $\square$

Lemma 11. *If $\mathcal{O}_+ = \langle \mathcal{T}_+, \mathcal{A} \rangle$ entails some fact over $\bot$, then so does $\mathcal{K}_{\mathcal{O}} = \langle \mathcal{R}_{\mathcal{T}}, \mathcal{A} \rangle$. For every assertion α defined over $\mathsf{sig}(\mathcal{T})$, $\mathcal{O}_+ \models \alpha$ implies $\mathcal{K}_{\mathcal{O}} \models \alpha$.*

Proof. We show the lemma via induction on the chase sequence $\mathcal{O}_+^0, \mathcal{O}_+^1, \ldots$

- Base Case: We show that $\alpha \in \mathcal{K}_{\mathcal{O}}^\infty$ for every assertion $\alpha \in \mathcal{O}^0$.

- Induction step (IS): For every $i \geq 1$, we show that the following claims hold provided that the induction hypothesis also holds.

  - If $\alpha \in \mathcal{K}_{\mathcal{O}}^\infty$ for every assertion $\alpha \in \mathcal{O}^i$.
  - If $\bot(t) \in \mathcal{O}^i$ with $t \in \mathsf{N_{gt}}$, then $\bot(u) \in \mathcal{K}_{\mathcal{O}}^\infty$ for some $u \in 2^{\mathsf{N_i}}$.

- Induction hypothesis (IH): $\alpha \in \mathcal{K}_{\mathcal{O}}^\infty$ for every assertion $\alpha \in \mathcal{O}^{i-1}$.

The base case holds since $\mathcal{O}_+^0 = \mathcal{A}$ and $\mathcal{A} \subseteq \mathcal{K}_{\mathcal{O}}^\infty$ by Definition 2. We show that the IS does hold for any $i \geq 1$ irrespectively of the type of the axiom ρ_i and the type of the terms that occur in the range of σ_i. Some cases will not be explicitly included in this analysis, because of the following reasons.

1. We altogether ignore cases in which the set $\mathcal{O}^i \setminus \mathcal{O}^{i-1}$ does not contain any assertions nor facts over $\bot$, as these trivially hold.

2. All cases where $\rho_i = B \rightarrow H$ is a Datalog rule with H an equality-free atom, and the range of σ_i is a subset of $2^{\mathsf{N_i}}$ can be shown with the following argument.

   - By IH, $B\sigma_i \subseteq \mathcal{K}_{\mathcal{O}}^\infty$.
   - Since $\rho_i \in \mathcal{R}_{\mathcal{T}}$, $H\sigma_i \subseteq \mathcal{K}_{\mathcal{O}}^\infty$.

   Therefore, we do not include these cases in the case by case analysis below.

3. To further reduce the number of cases that need to be considered, we assume without loss of generality that $\bot$ may only occur in the right-hand side of axioms of Type ($\sqcap$).

Case ($\sqcap$)   Let ρ_i be an axiom of the form $\bigsqcap_{j=1}^n A_j \sqsubseteq B$. Then, $A_1(\sigma_i(x)), \ldots, A_n(\sigma_i(x)) \in \mathcal{O}_+^{i-1}$. By (1)-(3), we only need to consider the case where $B = \bot$ and $\sigma_i(x) \in \mathsf{N_0}$. By Lemma 1, there is a sequence $t_0, \ldots, t_m$ of terms in $\mathcal{O}_+^{i-1}$ such that $t_0 = a \in 2^{\mathsf{N_i}}$, $t_1, \ldots, t_m \in \mathsf{N_0}$, $t_m = \sigma_i(x)$, and t_{j-1} is the predecessor of t_j for all $j \in [\![1, m]\!]$. For all $j \in [\![1, m]\!]$, $\mathbb{A}'_{j-1} \sqsubseteq \exists \mathbb{R}_j.\mathbb{A}_j \in \Gamma(\mathcal{T}_\times)$ with $\mathbb{A}'_{j-1} \subseteq \mathsf{N}_{\mathsf{c}}^{i-1}(t_{j-1})$, $\mathbb{R}_j = \mathsf{N}_{\mathsf{r}}^{i-1}(t_{j-1}, t_j)$, and $\mathbb{A}_j = \mathsf{N}_{\mathsf{c}}^{i-1}(t_j)$ by Lemma 8. Since $\bigsqcap_{j=1}^n A_j \sqsubseteq \bot \in \mathcal{T}$, $\mathbb{A}'_{m-1} \sqsubseteq \exists \mathbb{R}_m.(\mathbb{A}_m \sqcap \bot) \in \Gamma(\mathcal{T}_\times)$ and therefore, $\mathbb{A}'_j \sqsubseteq \bot \in \Gamma(\mathcal{T}_\times)$ for all $j \in [\![0, m-1]\!]$. Since $\mathbb{A}'_0(x) \rightarrow \bot(x) \in \mathcal{R}_{\mathcal{T}}$ and $\mathbb{A}'_0(\sigma_i(x)) \in \mathcal{K}_{\mathcal{O}}^\infty$ by IH, $\bot(\sigma_i(x)) \in \mathcal{K}_{\mathcal{O}}^\infty$.

Case ($\forall$)　Let ρ_i be an axiom of the form $A \sqsubseteq \forall R.B \in \mathcal{T}$. Then, $A(\sigma_i(x)), R(\sigma_i(x), \sigma_i(y)) \in \mathcal{O}_+^{i-1}$. By (1)-(3), we only need to consider cases where $\sigma_i(x) \in \mathsf{N_0}$ and $\sigma_i(y) \in 2^{\mathsf{N_i}}$. By Lemma (7), $R^-(\sigma_i(y), \sigma_i(x)) \in \mathsf{D}(\mathcal{O}_+^{i-1})$ and hence, $\sigma_i(y)$ is the predecessor of $\sigma_i(x)$ by Lemma 1. By Lemma 8, $\mathbb{A}' \sqsubseteq \exists \mathbb{R}.\mathbb{B} \in \Gamma(\mathcal{T}_\times)$ with $\mathbb{A}' \subseteq \mathsf{N_c^{i-1}}(\sigma_i(y))$, $\mathbb{R} = \mathsf{N_r^{i-1}}(\sigma_i(y), \sigma_i(x))$, and $\mathbb{B} = \mathsf{N_c^{i-1}}(\sigma_i(x))$. Since $R^- \in \mathbb{R}$ and $A \sqsubseteq \forall R.B \in \mathcal{T}_\times$, $\mathbb{A}'(x) \to B(x) \in \mathcal{R}_\mathcal{T}$. Since $\mathbb{A}'(\sigma_i(y)) \in \mathcal{K}_\mathcal{O}^\infty$ by IH, $B(\sigma_i(y)) \in \mathcal{K}_\mathcal{O}^\infty$.

Case ($\circ$)　Let ρ_i be an axiom of the form $R_1 \circ \ldots \circ R_n \sqsubseteq S \in \mathcal{T}_+$. Then, $R_1(\sigma_i(x_0), \sigma_i(x_1)), \ldots, R_n(\sigma_i(x_{n-1}), \sigma_i(x_n)) \in \mathcal{O}_+^{i-1}$. Because of (1)-(3), we only need to consider the case where $\sigma_i(x_0), \sigma_i(x_n) \in 2^{\mathsf{N_i}}$.

By Lemma 2, there is a sequence $V_1(t_0, t_1), \ldots, V_m(t_{m-1}, t_m) \in \mathsf{D}(\mathcal{O}_+^{i-1})$ with $\sigma_i(x_{j-1}) = t_0$, $\sigma_i(x_j) = t_m$, and $V_1 \cdot \ldots \cdot V_m \in \mathcal{N}_\mathcal{T}(R_j)$ for every $j \in [\![1, n]\!]$ (note that possibly $m = 1$). Note that, this lemma is applicable because $\mathcal{O}_+^{i-1}$ is closed under the application of axioms of the Type $\sqcap_r$ by the definition of $\mathcal{O}_+^0, \mathcal{O}_+^1, \ldots$. By concatenating the sequences above, we construct a sequence $V_1(t_0, t_1), \ldots, V_m(t_{m-1}, t_m) \in \mathsf{D}(\mathcal{O}_+^{i-1})$ such that $\sigma_i(x_0) = t_0$, $\sigma_i(x_n) = t_m$, and $V_1 \cdot \ldots \cdot V_m \in \mathcal{N}_\mathcal{T}(S)$. Hence, there are some states $q_0, \ldots, q_m$ such that $q_0 = i_S$, $q_m = f_S$, and $q_0 \to_{V_1} q_1 \to_{V_2} q_2 \ldots \to_{V_m} q_m \in \mathcal{N}_\mathcal{T}(S)$. Let $k_0, \ldots, k_o$ be the longest sorted sequence of natural numbers with $t_{k_j} \in 2^{\mathsf{N_i}}$ for all $j \in \{0, \ldots, o\}$. We show via induction that $S_{q_{k_j}}(t_0, t_{k_j}) \in \mathcal{K}_\mathcal{O}^\infty$ for all $j \in [\![1, o]\!]$. In turn, this implies $S(\sigma_i(x), \sigma_i(y)) \in \mathcal{K}_\mathcal{O}^\infty$ since $S_{q_{f_S}}(x, y) \to S(x, y) \in \mathcal{R}_\mathcal{T}$ (note that $t_0 = \sigma_i(x)$, $t_{k_o} = t_m = \sigma_i(y)$, and $q_{k_m} = f_S$).

To show the base case, we prove that $S_{q_{k_1}}(t_0, t_{k_1}) \in \mathcal{K}_\mathcal{O}^\infty$. We consider two possible cases.

- Let $k_1 = 1$. Then, $V_1(t_0, t_1) \in \mathcal{K}_\mathcal{O}^\infty$ by IH. Since $V_1(x, y) \to S_{q_1}(x, y) \in \mathcal{R}_\mathcal{T}$, $S_{q_1}(t_0, t_1) \in \mathcal{K}_\mathcal{O}^\infty$.

- Let $k_1 > 1$. Then, $t_{k_1} = t_0$ by Lemma 1. By Lemma 9, $\mathbb{A} \sqcap X_{i_S} \sqsubseteq X_{q_{k_1}} \in \Gamma(\mathcal{T}_\times)$ for some $\mathbb{A} \subseteq \mathsf{N_c^{i-1}}(t_0)$ and hence, $\mathbb{A}(x) \to S_{i_S, q_{k_1}}(x) \in \mathcal{R}_\mathcal{T}$. By IH, $\mathbb{A}(t_0) \in \mathcal{K}_\mathcal{O}^\infty$ and hence, $S_{i_S, q_{k_1}}(t_0) \in \mathcal{K}_\mathcal{O}^\infty$. Since $S_{i_S, q_{k_1}}(x) \to S_{q_{k_1}}(x, x) \in \mathcal{R}_\mathcal{T}^\infty$, $S_{q_{k_1}}(t_0, t_1) \in \mathcal{K}_\mathcal{O}^\infty$.

To show the induction step, we verify that, for all $j \in \{2, \ldots, o\}$, $S_{q_{k_j}}(t_0, t_{k_j}) \in \mathcal{K}_\mathcal{O}^\infty$ provided that $S_{q_{k_{j-1}}}(t_0, t_{k_{j-1}}) \in \mathcal{K}_\mathcal{O}^\infty$. We consider two possible cases.

- Let $k_j = k_{j-1} + 1$. Then, $V_{k_j}(t_{k_{j-1}}, t_{k_j}) \in \mathcal{K}_\mathcal{O}^\infty$ by IH. Since $S_{q_{k_{j-1}}}(x, y) \wedge V_{k_j}(y, z) \to S_{q_{k_j}}(x, z) \in \mathcal{R}_\mathcal{T}$, $S_{q_{k_j}}(t_0, t_{k_j}) \in \mathcal{K}_\mathcal{O}^\infty$.

- Let $k_j > k_{j-1} + 1$. Then, $t_{k_j} = t_{k_{j-1}}$ by Lemma 1. By Lemma 9, $\mathbb{A} \sqcap X_{q_{k_{j-1}}} \sqsubseteq X_{q_{k_j}} \in \Gamma(\mathcal{T}_\times)$ for some $\mathbb{A} \subseteq \mathsf{N_c^{i-1}}(t_{k_j})$ and hence, $\mathbb{A}(x) \to S_{q_{k_{j-1}}, q_{k_j}}(x) \in \mathcal{R}_\mathcal{T}$. By IH, $\mathbb{A}(t_{k_j}) \in \mathcal{K}_\mathcal{O}^\infty$ and hence, $S_{q_{k_{j-1}}, q_{k_j}}(t_{k_j}) \in \mathcal{K}_\mathcal{O}^\infty$. Since $S_{q_{k_{j-1}}}(x, y) \wedge S_{q_{k_{j-1}}, q_{k_j}}(y) \to S_{q_{k_j}}(x, y) \in \mathcal{R}_\mathcal{T}$, $S_{q_{k_j}}(t_0, t_{k_j}) \in \mathcal{K}_\mathcal{O}^\infty$.

Case ($\leq$)　Let ρ_i be an axiom of the form $A \sqsubseteq {\leqslant} 1\, R.B$. Then, $A(\sigma_i(x)), R(\sigma_i(x), \sigma_i(y)), B(\sigma_i(y))$, $R(\sigma_i(x), \sigma_i(z)), B(\sigma_i(z)) \in \mathcal{O}_+^{i-1}$. By Definition 1, the role R is simple and hence, $R(\sigma_i(x), \sigma_i(y))$, $R(\sigma_i(x), \sigma_i(z)) \in \mathsf{D}(\mathcal{O}^{i-1})$ by Lemma 6. Depending upon the type of the terms in the range of σ_i, we consider six possible cases cases.

1. Let $\sigma_i(x), \sigma_i(y), \sigma_i(z) \in 2^{\mathsf{N_i}}$. By IH, every assertion in $\mathcal{O}_+^{i-1}$ containing $\sigma_i(y)$ or $\sigma_i(z)$ is also in $\mathcal{K}_\mathcal{O}^\infty$. Hence, every assertion in $\mathcal{O}_+^i$ is also in $\mathcal{K}_\mathcal{O}^\infty$.

2. Let $\sigma_i(x) \in \mathsf{N_0}$ and $\sigma_i(y), \sigma_i(z) \in 2^{\mathsf{N_i}}$. By Lemma 1, both $\sigma_i(y)$ and $\sigma_i(z)$ are the predecessors of $\sigma_i(x)$ and hence, $\sigma_i(y) = \sigma_i(z)$. This is a contradiction by Definition 2 and hence, this case may not occur.

3. Let $\sigma_i(x), \sigma_i(y) \in \mathsf{N_0}$ and $\sigma_i(z) \in 2^{\mathsf{N_i}}$. By Lemma 1 and the fact that $\sigma_i(z) \neq \sigma_i(y)$, $\sigma_i(z)$ is the predecessor of $\sigma_i(x)$ which, in turn, is the predecessor of $\sigma_i(y)$.

   We first show that $C(\sigma_i(z)) \in \mathcal{K}_{\mathcal{O}}^{\infty}$ if $C(\sigma_i(y)) \in \mathcal{O}_+^{i-1}$ for some $C \in \mathsf{N_c}$. By Lemma 8, $\mathbb{A}' \sqsubseteq \exists \mathbb{R}.\mathbb{B}, \mathbb{B}' \sqsubseteq \mathbb{S}.\mathbb{C} \in \Gamma(\mathcal{T}_\times)$ with $\mathbb{A}' \subseteq \mathsf{N_c}^{i-1}(\sigma_i(z))$, $\mathbb{R} = \mathsf{N_r}^{i-1}(\sigma_i(z), \sigma_i(x))$, $\mathbb{B}' \subseteq \mathsf{N_c}^{i-1}(\sigma_i(x)) = \mathbb{B}$, $\mathbb{S} = \mathsf{N_r}^{i-1}(\sigma_i(x), \sigma_i(y))$, and $\mathbb{C} = \mathsf{N_c}^{i-1}(\sigma_i(y))$. Since $R^- \in \mathbb{R}$, $A \in \mathbb{B}$, $R \in \mathbb{S}$, $B \in \mathbb{C}$, $\mathbb{A}' \sqcap B \sqsubseteq C \in \Gamma(\mathcal{T}_\times)$ and hence, $\mathbb{A}'(x) \wedge B(x) \to C(x) \in \mathcal{R}_\mathcal{T}$. Since $\mathbb{A}'(\sigma_i(z)) \in \mathcal{K}_{\mathcal{O}}^{\infty}$ by IH, $C(\sigma_i(z)) \in \mathcal{K}_{\mathcal{O}}^{\infty}$.

   Furthermore, we show that $S(a, \sigma_i(z)) \in \mathcal{K}_{\mathcal{O}}^{\infty}$ if $S(a, \sigma_i(y)) \in \mathcal{O}_+^{i-1}$ for some $S \in \mathsf{N_r}$ and $a \in 2^{\mathsf{N_i}}$. By Lemma 2, there are some $R_1(t_0, t_1), \ldots, R_n(t_{n-1}, t_n) \in \mathsf{D}(\mathcal{O}_+^{i-1})$ and states $q_0, \ldots, q_n$ such that $a = t_0$, $\sigma_i(y) = t_n$, and $q_0 \to_{R_1} q_1 \ldots \to_{R_n} q_n \in \mathcal{N}_{\mathcal{T}_+}(R)$. We consider two possible cases.

   - $t_1, \ldots, t_n \in \mathsf{N_0}$. By Lemma 1, $t_0 = a$. By Lemma 10, $\mathbb{A}' \sqcap X_{q_{i_S}} \sqsubseteq X_{f_S} \in \Gamma(\mathcal{T}_\times)$ for some $\mathbb{A}' \subseteq \mathsf{N_c}^{i-1}(t_0)$ and therefore, $\mathbb{A}'(x) \to S_{i_S, f_S}(x) \in \mathcal{R}_\mathcal{T}$. By IH, $\mathbb{A}'(t_0) \in \mathcal{K}_{\mathcal{O}}^{\infty}$ and hence, $S_{i_S, f_S}(t_0) \in \mathcal{K}_{\mathcal{O}}^{\infty}$. Since $S_{i_S, f_S}(\sigma_i(z)) \to S_{f_S}(x, x), S_{f_S}(x, y) \to S(x, y) \in \mathcal{R}_\mathcal{T}$, $S(t_0, t_0) \in \mathcal{K}_{\mathcal{O}}^{\infty}$.
   - $t_j \in 2^{\mathsf{N_i}}$ for some $j \in \{1, \ldots, n-1\}$. With an analogous inductive argument to that from case ($\circ$), we can show that $S_{q_k}(a, \sigma_i(z)) \in \mathcal{K}_{\mathcal{O}}^{\infty}$. By Lemma 10, $\mathbb{A}' \sqcap X_{q_k} \sqsubseteq X_{f_n} \in \Gamma(\mathcal{T}_\times)$ for some $\mathbb{A}' \subseteq \mathbb{A}$ and hence, $\mathbb{A}'(x) \to S_{q_k, f_S}(x) \in \mathcal{R}_\mathcal{T}$. Since $\mathbb{A}'(\sigma_i(z)) \in \mathcal{K}_{\mathcal{O}}^{\infty}$ by IH, $S_{q_k, f_S}(\sigma_i(z)) \in \mathcal{K}_{\mathcal{O}}^{\infty}$. Since $S_{q_k}(x, y) \wedge S_{q_k, f_S}(y) \to S_{f_S}(x, y), S_{f_S}(x, y) \to S(x, y) \in \mathcal{R}_\mathcal{T}$, $S(a, \sigma_i(z)) \in \mathcal{K}_{\mathcal{O}}^{\infty}$.

4. Let $\sigma_i(x), \sigma_i(z) \in \mathsf{N_0}$ and $\sigma_i(y) \in 2^{\mathsf{N_i}}$. Analogous to the previous case.

5. Let $\sigma_i(x), \sigma_i(z) \in 2^{\mathsf{N_i}}$ and $\sigma_i(y) \in \mathsf{N_0}$. By Lemma 1, $\sigma_i(x)$ is the predecessor of $\sigma_i(y)$.

   We first show that $C(\sigma_i(z)) \in \mathcal{K}_{\mathcal{O}}^{\infty}$ if $C(\sigma_i(y)) \in \mathcal{O}_+^{i-1}$ for some $C \in \mathsf{N_c}$. By Lemma 8, $\mathbb{A}' \sqsubseteq \exists \mathbb{R}.\mathbb{B} \in \Gamma(\mathcal{T}_+)$ with $\mathbb{A}' \subseteq \mathsf{N_c}^{i-1}(\sigma_i(x))$, $\mathbb{R} = \mathsf{N_r}^{i-1}(\sigma_i(x), \sigma_i(y))$, $\mathbb{B} = \mathsf{N_c}^{i-1}(\sigma_i(y))$. Since $R \in \mathbb{R}$, $A(x) \wedge \mathbb{A}'(x) \wedge R(x, y) \wedge B(y) \to C(y) \in \mathcal{R}_\mathcal{T}$. Since $A(\sigma_i(x)), \mathbb{A}(\sigma_i(z)), R(\sigma_i(x), \sigma_i(z))$, $B(\sigma_i(z)) \in \mathcal{K}_{\mathcal{O}}^{\infty}$ by IH, $C(\sigma_i(z)) \in \mathcal{K}_{\mathcal{O}}^{\infty}$.

   We show that $S(a, \sigma_i(z)) \in \mathcal{K}_{\mathcal{O}}^{\infty}$ if $S(a, \sigma_i(y)) \in \mathcal{O}_+^{i-1}$ for some $S \in \mathsf{N_r}$ and $a \in 2^{\mathsf{N_i}}$. We consider two possible cases.

   - Let $S(a, \sigma_i(z)) \in \mathsf{D}(\mathcal{O}_+^{i-1})$. By Lemma 1, $a = \sigma_i(x)$. Since $S \in \mathbb{R}$, $A(x) \wedge \mathbb{A}'(x) \wedge R(x, y) \wedge B(y) \to S(x, y) \in \mathcal{R}_\mathcal{T}$ and $S(\sigma_i(x), \sigma_i(z)) \in \mathcal{K}_{\mathcal{O}}^{\infty}$.
   - Let $S(a, \sigma_i(z)) \notin \mathsf{D}(\mathcal{O}_+^{i-1})$. By Lemma 2, there are some $R_1(t_0, t_1), \ldots, R_n(t_{n-1}, t_n) \in \mathsf{D}(\mathcal{O}_+^{i-1})$ with $a = t_0$, $\sigma_i(y) = t_n$, and $R_1 \cdot \ldots \cdot R_n \in \mathcal{N}_\mathcal{R}(S)$. Let k be the largest number with $t_k = \sigma_i(x)$ and $t_k, \ldots, t_n \in \mathsf{N_0}$ for some $k \in [\![0, n-1]\!]$ (note that such a number must exist by Lemma 1). Also by Lemma 1, $t_{k+1} = \sigma_i(y)$ and hence, $R_{k+1} \in \mathbb{R}$, $A(x) \wedge \mathbb{A}'(x) \wedge R(x, y) \wedge B(y) \to R_{k+1}(x, y) \in \mathcal{R}_\mathcal{T}$, and $R_{k+1}(\sigma_i(x), \sigma_i(z)) \in \mathcal{K}_{\mathcal{O}}^{\infty}$. We consider two possible cases.
     - Let $k > 0$. With an analogous argument to that from case ($\circ$), we can show that $S_{q_k}(a, \sigma_i(x)) \in \mathcal{K}_{\mathcal{O}}^{\infty}$. Also, $S_{q_k}(x, y) \wedge R_{k+1}(y, z) \to S_{q_{k+1}}(x, z) \in \mathcal{R}_\mathcal{T}$.
     - Let $k = 0$. Then, $R_{k+1}(x, y) \to S_{q_{k+1}}(x, y) \in \mathcal{R}_\mathcal{T}$ (note that $q_{k+1} = i_S$ in this case).

     In either of these cases, $S_{q_{k+1}}(a, \sigma_i(z)) \in \mathcal{K}_{\mathcal{O}}^{\infty}$. Let $\ell_0, \ldots, \ell_m$ be the longest sorted sequence with $\ell_0 = k+1$ and $\sigma_i(y) = t_{\ell_j}$ for all $j \in [\![0, n]\!]$. By Lemma 9, for all $j \in [\![1, n]\!]$ there is some $\mathbb{B}'_j \subseteq \mathsf{N_c}^{i-1}(\sigma_i(y))$ such that $\mathbb{B}'_j \sqcap X_{q_{\ell_{j-1}}} \sqsubseteq X_{\ell_j} \in \Gamma(\mathcal{T}_\times)$. Hence, $\mathbb{B}'_j(x) \to S_{q_{\ell_{j-1}}, q_{\ell_j}}(x) \in \mathcal{R}_\mathcal{T}$ for all $j \in [\![1, n]\!]$. Since $A(x) \wedge \mathbb{A}'(x) \wedge R(x, y) \wedge$

$B(y) \rightarrow \mathbb{B}(y) \in \mathcal{R}_\mathcal{T}$, $\mathbb{B}(\sigma_i(z)), S_{q_{\ell_{j-1}}, q_{\ell_j}}(\sigma_i(z)) \in \mathcal{K}_\mathcal{O}^\infty$ for all $j \in [\![1, n]\!]$. For all $j \in [\![1, n]\!]$, $S_{q_{\ell_{j-1}}}(x, y) \wedge S_{q_{\ell_{j-1}}, q_{\ell_j}}(y) \rightarrow S_{q_{\ell_j}}(x, y) \in \mathcal{R}_\mathcal{T}$. Hence, $S_{f_S}(a, \sigma_i(z)) \in \mathcal{K}_\mathcal{O}^\infty$ (note that $f_S = q_{\ell_m}$). Since $S_{f_S}(x, y) \rightarrow S(x, y) \in \mathcal{R}_\mathcal{T}$, $S(a, \sigma_i(z)) \in \mathcal{K}_\mathcal{O}^\infty$.

6. Let $\sigma_i(x), \sigma_i(y) \in 2^{\mathsf{N}_i}$, and $\sigma_i(z) \in \mathsf{N}_0$. Analogous to the above case. $\qquad \square$

C   Soundness

To show that our Datalog AR-rewriting is sound, we have to show that for every ABox $\mathcal{A}$ s.t. $\mathsf{sig}(\mathcal{A}) \subseteq \mathsf{sig}(\mathcal{A})$ and assertion α s.t. $\mathsf{sig}(\alpha) \subseteq \mathsf{sig}(\mathcal{T})$, we have $\langle \mathcal{R}_\mathcal{T}, \mathcal{A} \rangle \models \alpha$ only if $\langle \mathcal{T}, \mathcal{A} \rangle \models \alpha$. Note that the case where $\langle \mathcal{T}, \mathcal{A} \rangle$ is unsatisfiable is trivial, since in that case $\langle \mathcal{T}, \mathcal{A} \rangle \models \alpha$ holds for every assertion α. We therefore silently assume that $\langle \mathcal{T}, \mathcal{A} \rangle$ is satisfiable in every of the following lemmata.

We note that we can generalise soundness and completeness results of the chase to arbitrary sets of ground facts: For a given set $\mathcal{A}$ of ground facts (which may include facts over nulls) and a TBox $\mathcal{T}$, we call $\langle \mathcal{T}, \mathcal{A} \rangle$ a generalised KB, and define chase in the same way as for classical KBs. It is easy to see that Theorem 1 also applies to generalised KBs, as we can simply replace every null by a fresh individual name to reduce to entailment of normal KBs. We first show that our introduced names indeed fulfil their specified role.

Lemma 12. *Let $\mathcal{T}$ be any Horn-$\mathcal{SRIQ}_\sqcap$-TBox, $\rho = A \sqsubseteq \forall R.B \in \mathcal{T}$, and $\mathcal{T}' = \mathsf{B}(\mathcal{T}, \rho)$. Further, let $\mathcal{A}$ be any set of ground facts s.t. $\mathsf{sig}(\mathcal{A}) \subseteq \mathsf{sig}(\mathcal{T})$, q, $\hat{q}$ two states in $\mathcal{N}_\mathcal{T}(R)$ and t, u two ground terms. Let $\mathcal{A}' = \mathcal{A} \cup \{A_q(u)\}$. Then, $\langle \mathcal{T}', \mathcal{A}' \rangle \models A_{\hat{q}}(t)$ implies that there is a path P in the chase of $\langle \mathcal{T}, \mathcal{A} \rangle$ connecting u to t s.t. $q \rightarrow_\mathsf{P}^* \hat{q} \in \mathcal{N}_\mathcal{T}(R)$.*

Proof. Let $\mathcal{T}$, ρ, $\mathcal{T}'$, q, $\mathcal{A}$, $\mathcal{A}'$ and u be as in the lemma. Let $\mathcal{A} = \mathcal{F}^0, \mathcal{F}^1, \dots$ be the chase of $\langle \mathcal{T}', \mathcal{A}' \rangle$. By Theorem 1, if $\langle \mathcal{T}', \mathcal{A} \rangle \models A_{\hat{q}}(t)$, there exists some $i \geq 0$ s.t. $A_{\hat{q}}(t) \in \mathcal{F}^i$. It therefore suffices to show that for every $i \geq 0$, every state $\hat{q}$ in $\mathcal{N}_\mathcal{T}(R)$ and every ground term t , if $A_{\hat{q}}(t) \in \mathcal{F}^i$, then there is a path P in the chase of $\langle \mathcal{T}, \mathcal{A} \rangle$ connecting u to t s.t. $q \rightarrow_\mathsf{P}^* \hat{q} \in \mathcal{N}_\mathcal{T}(R)$. We do the proof by induction on i. Consider $i = 0$. Since $\mathsf{sig}(\mathcal{A}) \subseteq \mathcal{T}$ and A_q is fresh, we cannot have $A_q(t) \in \mathcal{A}$. Consequently, $A_{\hat{q}}(t) \in \mathcal{A}' = \mathcal{F}^0$ only if $\hat{q} = q$ and $t = u$, in which case the inductive hypothesis trivally holds.

Let $i > 0$, and assume the inductive hypothesis holds for $i-1$. The only interesting case is where $A_{\hat{q}}(t)$ is introduced in $\mathcal{F}^i$. Let ρ_i be the axiom applied for computing $\mathcal{F}^i$. The only axioms in $\mathcal{T}'$ in which A_q occurs positively are of the form $A_{q'} \sqsubseteq \forall S.A_{\hat{q}}$, where $q' \rightarrow_S^* \hat{q} \in \mathcal{N}_\mathcal{T}(R)$, which means that ρ_i must be of this form. This means that there exists some ground term v s.t. $A_{q'}(v), R(v, t) \in \mathcal{F}^{i-1}$. By the inductive hypothesis, there is a path P connecting u to v in the chase of $\langle \mathcal{T}, \mathcal{A} \rangle$ s.t. $q \rightarrow_\mathsf{P}^* q' \in \mathcal{N}_\mathcal{T}(R)$. This implies that there is the path $\mathsf{P} \cdot S$ connecting u to t, and that $i_R \rightarrow_{\mathsf{P} \cdot S}^* q \in \mathcal{N}_\mathcal{T}(R)$. $\qquad \square$

Lemma 13. *Let $\mathcal{T}$ be a Horn-$\mathcal{SRIQ}_\sqcap$-TBox, $\mathcal{A}$ an ABox s.t. $\mathsf{sig}(\mathcal{A}) \subseteq \mathsf{sig}(\mathcal{T})$ and α an assertion s.t. $\mathsf{sig}(\rho) \subseteq \mathsf{sig}(\mathcal{T})$. Then, $\langle \mathcal{T}_+, \mathcal{A} \rangle \models \alpha$ only if $\langle \mathcal{T}, \mathcal{A} \rangle \models \alpha$.*

Proof. Let $\mathcal{F}^0 = \mathcal{A}, \mathcal{F}^1, \dots$ be the chase of $\langle \mathcal{T}_+, \mathcal{A} \rangle$. We have to show that for all $i \geq 0$ and every $\alpha \in F^i$ s.t. $\mathsf{sig}(\rho) \subseteq \mathsf{sig}(\mathcal{T})$, $\langle \mathcal{T}, \mathcal{A} \rangle \models \alpha$. We do so by induction over i. The case where $i = 0$ is trivial. For $i > 0$, the only interesting case is where α is introduced in i by application of an axiom $\rho \in \mathcal{T}_+ \setminus \mathcal{T}$. If this is the case, by the definition of $\mathcal{T}_+$ there is an axiom $A \sqsubseteq \forall R.B \in \mathcal{T}$ s.t. $\rho \in \mathsf{B}(A \sqsubseteq \forall R.B, \mathcal{T})$. Note that for every distinct axioms $\rho_1, \rho_2 \in \mathcal{T}$, the sets of fresh predicates in $\mathsf{B}(\rho_1, \mathcal{T})$ and $\mathsf{B}(\rho_2, \mathcal{T})$ are disjoint. The only axioms in $\mathsf{B}(A \sqsubseteq \forall R.B, \mathcal{T})$ that share a predicate name with $\mathcal{T}$ are $A \sqsubseteq A_{i_R}$ and $A_{f_R} \sqsubseteq B$. We obtain that $\rho = A_{f_R} \sqsubseteq B$,

24

$\alpha = B(b)$ for some $b \in \mathsf{N_i}$, and that $A_{f_R}(b) \in \mathcal{F}^{-1}$. We thus have to show that $\langle \mathcal{T}, \mathcal{A} \rangle \models B(b)$. By induction over the axioms in $\mathsf{B}(A \sqsubseteq \forall R.B)$ applied to derive $B(b)$, we further obtain that $A(a), A_{i_R}(a) \in \mathcal{F}^{i-1}$ for some individual a. Because i) $A_{i_R}(a), A_{f_R}(b) \in \mathcal{F}^{-1}$, ii) the fresh predicates in $\mathsf{B}(A \sqsubseteq \forall R.B)$ do not occur in $\mathcal{T}_+ \setminus \mathsf{B}(A \sqsubseteq \forall R.B)$, iii) the inductive hypothesis, and iv) by Lemma 12, we obtain that there is a path P in the chase of $\langle \mathcal{T}, \mathcal{A} \rangle$ connecting a and b s.t. $i_R \to_{\mathsf{P}}^* f_R \in \mathcal{N}_{\mathcal{T}}(R)$. By Lemma 2, this implies that $\langle \mathcal{T}, \mathcal{A} \rangle \models R(a, b)$, and since $A \sqsubseteq \forall R.B \in \mathcal{T}$, $\langle \mathcal{T}, \mathcal{A} \rangle \models B(b)$. □

Lemma 3 is a direct consequence of Lemma 13.

Lemma 3. *For a TBox $\mathcal{T}$, an ABox $\mathcal{A}$ and a fact set $\mathcal{F}$ defined over $\mathsf{sig}(\mathcal{T})$, $\langle \mathcal{T}, \mathcal{A} \rangle$ is satisfiable iff $\langle \mathcal{T}_+, \mathcal{A} \rangle$ is, and $\langle \mathcal{T}, \mathcal{A} \rangle \models \mathcal{F}$ iff $\langle \mathcal{T}_+, \mathcal{A} \rangle \models \mathcal{F}$.*

Lemma 14. *Let $\mathcal{T}$ be a Horn-$\mathcal{SRIQ}_{\sqcap}$-TBox, $\mathcal{A}$ be a set of ground facts s.t. $\mathsf{sig}(\mathcal{A}) \subseteq \mathsf{sig}(\mathcal{T})$, R a complex role, q, q' two states in $\mathcal{N}_{\mathcal{T}}(R)$ and u, t two ground terms. Let $\mathcal{T}'$ be the set of axioms in $\mathcal{T}$ plus all axioms generated by the Rule ($\circlearrowleft$).*

Assume further that there is a path P connecting u and t in the chase of $\langle \mathcal{T}, \mathcal{A} \rangle$ s.t. $i_R \to_{\mathsf{P}}^* q \in \mathcal{N}_{\mathcal{T}}(R)$. Then, $\langle \mathcal{T}', \mathcal{A} \rangle \models R_{q,\hat{q}}(t)$ implies the following.

- *There is a path P' from t to t in the chase of $\langle \mathcal{T}, \mathcal{A} \rangle$, and*

- *$q \to_{\mathsf{P}'}^* \hat{q} \in \mathcal{N}_{\mathcal{T}}(R)$.*

Proof. The only axioms in $\mathcal{T}'$ in which $R_{q,\hat{q}}$ occurs positively are the ones added due to Rule ($\circlearrowleft$) in Table 1. Therefore, there must exist a set $\mathbb{D}$ of concept names s.t. $\langle \mathcal{T}', \mathcal{A} \rangle \models \mathbb{D}(t)$, and that $\mathcal{T}_{\times} \models \mathbb{D} \sqcap X_q \sqsubseteq X_{\hat{q}}$.

Define a new set of ground facts $\mathcal{A}'$ by adding the ground fact $X(u)$ to $\mathcal{A}$. By induction on the axioms in $\mathsf{B}(\mathcal{T}, X \sqsubseteq \forall R.X) \subseteq \mathcal{T}_{\times}$ and the path connecting u to t, it is easy to show that $\langle \mathcal{T}_{\times}, \mathcal{A}' \rangle \models X_q(t)$. Since furthermore $\langle \mathcal{T}, \mathcal{A}' \rangle \models \mathbb{D}(t)$ and $\mathcal{T}_{\times} \models \mathbb{D} \sqcap X_q \sqsubseteq X_{\hat{q}}$, we obtain that $\langle \mathcal{T}, \mathcal{A}' \rangle \models X_{\hat{q}}(t)$. Inspection of the rules in $\mathcal{T}_{\times}$ further shows that the chase of $\langle \mathcal{T}, \mathcal{A}' \rangle$ does not contain additional edges compared to $\langle \mathcal{T}, \mathcal{A} \rangle$, since we only added the ground fact $X(u)$. Consequently, by Lemma 12, there is a path in the chase of $\langle \mathcal{T}, \mathcal{A} \rangle$ connecting t to itself s.t. $q \to_{\mathsf{P}'}^* \hat{q} \in \mathcal{N}_{\mathcal{T}}(R)$. □

Lemma 15. *Let $\mathcal{T}$ be a Horn-$\mathcal{SRIQ}_{\sqcap}$-TBox, $\mathcal{A}$ be an ABox s.t. $\mathsf{sig}(\mathcal{A}) \subseteq \mathsf{sig}(\mathcal{T})$, R a complex role, q a state in $\mathcal{N}_{\mathcal{T}}(R)$ and $a, b \in \mathsf{N_i}$. Let $\mathcal{T}' \subseteq \mathcal{R}_{\mathcal{T}}$ be the set of Datalog rules that either occur in $\mathcal{T}$, or that are generated by Rules ($\sqcap$), ($\circlearrowleft$) and (R1)–(R4).*

Then, $\langle \mathcal{T}', \mathcal{A} \rangle \models R_q(a, b)$ only if in the chase of $\langle \mathcal{T}, \mathcal{A} \rangle$, there exists a path P from a to b s.t. $i_R \to_{\mathsf{P}}^* q \in \mathcal{N}_{\mathcal{T}}(R)$.

Proof. Let $\mathcal{T}$, $\mathcal{A}$, R and a be as in the lemma. Let $\mathcal{A} = \mathcal{F}_0, \mathcal{F}_1, \ldots$ be the chase of $\langle \mathcal{T}', \mathcal{A} \rangle$. We show that for every $i \geq 0$, state q in $\mathcal{N}_{\mathcal{T}}(R)$ and $b \in \mathsf{N_i}$, $R_q(a, b) \in \mathcal{F}^i$ only if in the chase of $\langle \mathcal{T}, \mathcal{A} \rangle$, there exists a path P from a to b s.t. $i_R \to_{\mathsf{P}}^* q \in \mathcal{N}_{\mathcal{T}}(R)$. Since $\mathsf{sig}(\mathcal{A}) \subseteq \mathsf{sig}(\mathcal{T})$ and R_q is fresh, the base case holds trivially. For $i > 0$, the only interesting case is where $R_q(a, b)$ is introduced in $\mathcal{F}^i$. Let ρ_i be the axiom applied on $\mathcal{F}^{i-1}$ to generate $\mathcal{F}^i$. The only axioms in which R_q occurs positively are those introduced by (R1)–(R4). We distinguish the cases.

1. ρ_i was introduced by (R1). Then a and b are connected by S already in the chase of $\langle \mathcal{T}'', \mathcal{A} \rangle$, where $\mathcal{T}''$ contains all the Datalog rules in $\mathcal{T}_+$, since no other rule in $\mathcal{T}'$ contains S positively. By soundness of $\mathcal{T}_+$ (Lemma 13), this implies that a and b are connected by S in the chase of $\langle \mathcal{T}, \mathcal{A} \rangle$, and we have $i_R \to_S^* q$ by the side condition of (R1).

2. ρ_i was introduced by (R2). Then, $b = a$, and the inductive hypothesis follows from Lemma 14.

3. ρ_i was introduced by (R3). Then, there is a state $\hat{q}$ in $\mathcal{N}_{\mathcal{T}}(R)$ and an individual c s.t. $R_{\hat{q}}(a,c), S(c,b) \in \mathcal{F}^{-1}$ and $\hat{q} \rightarrow_S^* q \in \mathcal{N}_{\mathcal{T}}(R)$. By inductive hypothesis, there is then a path P from a to c in the chase of $\langle \mathcal{T}, \mathcal{A} \rangle$ s.t. $i_R \rightarrow_{\mathsf{P}}^* \hat{q} \in \mathcal{N}_{\mathcal{T}}(R)$. We obtain that there is the path $\mathsf{P} \cdot S$ connecting a to b s.t. $i_R \rightarrow_{\mathsf{P} \cdot S}^* \hat{q} \in \mathcal{N}_{\mathcal{T}}(R)$.

4. ρ_i was introduced by (R4). Then, there is a state $\hat{q}$ in $\mathcal{N}_{\mathcal{T}}(R)$ s.t. $R_{\hat{q}}(a,b), (_b,_) \in \mathcal{F}^{-1}$. By inductive hypothesis, there is then a path P from a to b in the chase of $\langle \mathcal{T}, \mathcal{A} \rangle$ s.t. $i_R \rightarrow_{\mathsf{P}}^* \hat{q} \in \mathcal{N}_{\mathcal{T}}(R)$. By Lemma 14, there is a path P' in the chase of $\langle \mathcal{T}, \mathcal{A} \rangle$ from b to b s.t. $\hat{q} \rightarrow_{\mathsf{P}'}^* q \in \mathcal{N}_{\mathcal{T}}(R)$. We obtain that the path $\mathsf{P} \cdot \mathsf{P}'$ connects a and b in the chase of $\langle \mathcal{T}, \mathcal{A} \rangle$, and that $i_R \rightarrow_{\mathsf{P} \cdot \mathsf{P}'}^* q \in \mathcal{N}_{\mathcal{T}}(R)$. □

Lemma 16 (Soundness). *Let $\mathcal{T}$ be a Horn-$\mathcal{SRIQ}_{\sqcap}$-TBox and $\mathcal{A}$ be an ABox s.t. $\mathsf{sig}(\mathcal{A}) \subseteq \mathsf{sig}(\mathcal{T})$. Then, for every assertion α s.t. $\mathsf{sig}(\alpha) \subseteq \mathsf{sig}(\mathcal{T})$, $\langle \mathcal{R}_{\mathcal{T}}, \mathcal{A} \rangle \models \alpha$ only if $\langle \mathcal{T}, \mathcal{A} \rangle \models \alpha$.*

Proof. Let $\mathcal{F}^0 = \mathcal{A}, \mathcal{F}^1, \ldots$ be the chase of $\langle \mathcal{R}_{\mathcal{T}}, \mathcal{A} \rangle$. We show that for every $i \geq 0$ and every assertion $\alpha \in F^i$ s.t. $\mathsf{sig}(\alpha) \subseteq \mathsf{sig}(\mathcal{T})$, $\langle \mathcal{T}, \mathcal{A} \rangle \models \alpha$. Note that, since $\mathcal{R}_{\mathcal{T}}$ is a Datalog rule set, no axiom in $\mathcal{R}_{\mathcal{T}}$ introduces nulls, so that for all $i \geq 0$, $\mathcal{F}^i$ is an ABox.

We do the proof by induction on i. The base case is trivial. Assume the inductive hypothesis holds for $i - 1$. The only interesting case is where α is introduced in $\mathcal{F}^i$ by an axiom ρ_i. We distinguish the cases based on the origin of ρ_i.

1. If $\rho_i \in \mathcal{T}_+$, then $\langle \mathcal{T}, \mathcal{A} \rangle \models \alpha$ directly follows from the inductive hypothesis and from Lemma 13.

2. If ρ_i is introduced by ($\sqcap$), by soundness of the Datalog calculus, $\mathcal{T}_+ \models \rho_i$. Therefore, $\langle \mathcal{T}, \mathcal{A} \rangle \models \alpha$ follows from the inductive hypothesis and from Lemma 13.

3. ρ_i cannot have been introduced by any of the Rules ($\circlearrowleft$) or (R1)–(R4) since $\mathsf{sig}(\alpha) \subseteq \mathsf{sig}(\mathcal{T})$.

4. If ρ_i is introduced by Rule (R5), then $R_{f_R}(a,b) \in \mathcal{F}^{i-1}$ for some $a, b \in \mathsf{N_i}$, and $\alpha = R(a,b)$. It then follows from Lemma 15 that there is a path P in the chase of $\langle \mathcal{T}, \mathcal{A} \rangle$ connecting a to b s.t. $i_R \rightarrow_{\mathsf{P}}^* f_R \in \mathcal{N}_{\mathcal{T}}(R)$. Consequently, by Lemma 2, we have $\langle \mathcal{T}, \mathcal{A} \rangle \models R(a,b)$.

5. If ρ_i is introduced by Rule ($\lhd 1$), we have $\alpha = C(b)$ for some $b \in \mathsf{N_i}$, $A(a), \mathbb{D}(a), R(a,b), B(b) \in \mathcal{F}^{i-1}$ for some $a \in \mathsf{N_i}$, and

$$\mathcal{T}_\times \models A \sqsubseteq \,\leqslant 1\, R.B, \mathbb{D} \sqsubseteq \exists(\mathbb{R} \sqcap R).(\mathbb{A} \sqcap B \sqcap C).$$

Note that, since $\mathsf{sig}(\mathcal{A}) \subseteq \mathsf{sig}(\mathcal{T})$, $\mathcal{A}$ cannot have any occurrences of the fresh concept name X. Inspection of the axioms in $\mathcal{T}_\times \setminus \mathcal{T}_+$ reveals that therefore that the above entailment in fact holds already for $\mathcal{T}_+$. Furthermore, the latter entailment can be weakened to $\mathcal{T}_+ \models \mathbb{D} \sqsubseteq \exists R.(B \sqcap C)$. Since $\mathbb{D}(a) \in \mathcal{F}^{i-1}$, $\langle \mathcal{T}_+, \mathcal{F}^{i-1} \rangle \models \exists y.(R(a,y) \wedge B(y) \wedge C(y))$, that is, there is some R-successor t of a in the chase of $\langle \mathcal{T}_+, \mathcal{F}^{i-1} \rangle$ that satisfies B and C. Since $\mathcal{T}_+ \models A \sqsubseteq \leq 1 R.B$, a can only have one R-successor satisfying B, so that in fact $t = a$, and $\langle \mathcal{T}_+, \mathcal{F}^{i-1} \rangle \models C(b)$. By the inductive hypothesis and Lemma 13, we obtain that $\langle \mathcal{T}, \mathcal{A} \rangle \models C(b)$.

6. If ρ_i is introduced by Rule ($\lhd 2$), we have $\alpha = S(a,b)$ for some $a, b \in \mathsf{N_i}$, $A(a), \mathbb{D}(a), R(a,b), B(b) \in \mathcal{F}^{i-1}$ for some $a \in \mathsf{N_i}$, and

$$\mathcal{T}_\times \models A \sqsubseteq \,\leqslant 1\, R.B, \mathbb{D} \sqsubseteq \exists(\mathbb{R} \sqcap R \sqcap S).(\mathbb{A} \sqcap B).$$

Similar as in the last case, we obtain that $\mathcal{T}_+ \models \mathbb{D} \sqsubseteq \exists(R \sqcap S).B$. Consequently, a has an R-successor in the chase of $\langle \mathcal{T}_+, \mathcal{F}^{i-1} \rangle$ that satisfies B and is also an S-successor of a. Since $A(a), R(a,b), B(b) \in \mathcal{F}^{i-1}$ and $\mathcal{T}_+ \models A \sqsubseteq {\leq} 1 R.B$, this successor has to be b. We obtain that $\langle \mathcal{T}_+, \mathcal{F}^{-1} \rangle \models S(a,b)$. By the inductive hypothesis and Lemma 13, we obtain that $\langle \mathcal{T}, \mathcal{A} \rangle \models S(a,b)$. $\qquad\square$

D   Complexity Results

Theorem 3. *Let* $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ *be an ontology. If* $\mathcal{T}$ *is Horn-$\mathcal{SRIQ}_\sqcap$/Horn-$\mathcal{SHIQ}$/$\mathcal{ELH}$, then we can compute* $\mathcal{R}_\mathcal{T}$ *and* $\langle \mathcal{R}_\mathcal{T}, \mathcal{A} \rangle^\infty$ *in* 2ExpTime/ExpTime/PTime*, respectively.*

Proof. We note that the number of states in each automaton $\mathcal{N}_\mathcal{T}(R)$ is exponentially bounded in the size of $\mathcal{T}$, so that the sizes of $\mathcal{T}_+$ and $\mathcal{T}_\times$ are also at most exponential in the input if $\mathcal{T}$ is in Horn-$\mathcal{SRIQ}_\sqcap$, and at polynomial if $\mathcal{T}$ is in Horn-$\mathcal{SHIQ}$ or $\mathcal{ELH}$. The calculus in Figure 4 adds one axiom to $\Gamma(\mathcal{T}_\times)$ in each step, the number of which is exponentially bounded in the size of $\mathcal{T}_\times$. This is so because every derived axiom contains at most one role conjunction and at most two concept conjunctions (one on the left-hand side, one on the right-hand side), and the number of concept and role names used for this is bounded by the size of $\mathcal{T}_\times$. It is also easy to see that each role application can be performed in polynomial time.

For $\mathcal{ELH}$, we note that from the calculus in Figure 4, only the Rules (1), (2), (4) and (5) apply. We recall that for every $R \in \mathsf{N_r}$, $\mathcal{T}$ contains one of R and R^- but not both. Since axioms of the form $\exists R.A \sqsubseteq B$ correspond to axioms $A \sqsubseteq \forall R^-.B$, we obtain that in fact Rule (4) also is never applied, so that we are left with (1), (2) and (5). We obtain that no rule derives an axiom whose left-hand-side does not occur already as left-hand-side in the input axiom set. Rule (5) does not derive axioms that are larger than its premises, and is the only rule that infers axioms without role restrictions, and the number of axioms it derives is polynomially bounded in the input. Rules (1) and (2) only infer axioms that are logically stronger than its premise, so that the premise can be removed from the axiom set after applying this rule. We therefore obtain that we can compute an axiom set equivalent to $\Gamma(\mathcal{T}_\times)$, contains all relevant axioms, and whose size is polynomially bounded in the size of $\mathcal{T}_\times$, and therefore in the size of $\mathcal{T}$, if and can be computed in polynomial time

Finally, $\mathcal{R}_\mathcal{T}$ can be generated by traversing $\Gamma(\mathcal{T}_\times)$ and each automata at most once, and its size is bounded by the size of $\mathcal{R}_{\mathcal{T}_\times}$ and the number of states occurring in all automata. We obtain that

- for Horn-$\mathcal{SRIQ}_\sqcap$, $\mathcal{R}_\mathcal{T}$ can be computed in 2ExpTime and is of at most double exponential size,

- for Horn-$\mathcal{SHIQ}$, $\mathcal{R}_\mathcal{T}$ can be computed in ExpTime and is of at most exponential size, and

- for $\mathcal{ELH}$, $\mathcal{R}_\mathcal{T}$ can be computed in PTime and is of at most polynomial size.

Now let m be the size of $\mathcal{R}_\mathcal{T}$. The complexity results now follow from the fact that the chase of $\langle \mathcal{R}_\mathcal{T}, \mathcal{A} \rangle$ can always be computed in time polynomial in m.

We first note that the number of elements in $\mathcal{R}_\mathcal{T}^\infty$ is polynomially bounded: since Datalog rules do not introduce nulls and the arity of each predicate is at most 2, the number of assertions in $\mathcal{R}_\mathcal{T}^\infty$ is bounded by the number of individual names and predicate names in $\mathcal{T}$ and $\mathcal{A}$. Since every step in the chase either introduces a new assertion or merges a pair of individuals, we obtain that the chase sequence is polynomially bounded as well. We therefore only have to

show that each rule application can be performed in polynomial time. Inspection of the axioms allowed in Horn-$\mathcal{SRIQ}_\sqcap$ and the Datalog rules included based on Table (1), we obtain that except for axioms of Type ($\circ$), every rule has at most three variables on the left-hand-side. Applicability of these rules can therefore be decided in polynomial time by simply iterating over all pairs of individuals. For rules of Type ($\circ$), we iterate over all pairs of individuals a, b and determine whether $S(a, b)$ can be inferred using graph-reachability. For this, we construct a directional graph G incrementally as follows: the initial graph contains a as a node, and for every i in $[\![1, n]\!]$ starting from $i = 1$, we add an edge labelled R_i connecting two nodes c and d if c is a node in the current graph and $R(c, d)$ occurs in the current fact base. Whether there is a path along $R_1, \ldots, R_n$ connecting a and b can then be decided by determining whether b is reachable from a in G, in which case the rule is applicable. Since this can be done in polynomial time, we obtain that each rule application can be performed in polynomial time, and as there are at most polynomially many necessary to compute $\langle \mathcal{R}_\mathcal{T}, \mathcal{A} \rangle^\infty$, we obtain that $\langle \mathcal{R}_\mathcal{T}, \mathcal{A} \rangle^\infty$ can be computed in time polynomial in m. $\qquad\square$

E   Added Role Chains in the Evaluation

In this section, we list the axioms of Type ($\circ$) that were added to the ontologies Reactome (8-12) and to Uniprot (13-15) for the experiments presented on the first part in the evaluation section.

$$\text{controlled} \circ \text{controlled}^- \sqsubseteq \text{coControlled} \tag{8}$$
$$\text{interactionScore} \circ \text{scoreSource} \sqsubseteq \text{interactionScoreProvenance} \tag{9}$$
$$\text{organism} \circ \text{CellVocabulary}^- \sqsubseteq \text{organismCellVocabulary} \tag{10}$$
$$\text{participant} \circ \text{dataSource} \sqsubseteq \text{participantDataSource} \tag{11}$$
$$\text{controlled} \circ \text{controller} \sqsubseteq \text{controlledBy} \tag{12}$$
$$\text{cellularComponent}^- \circ \text{orientation} \sqsubseteq \text{cellularOrientation} \tag{13}$$
$$\text{database}^- \circ \text{transcribedFrom} \sqsubseteq \text{transcriptionStoredIn} \tag{14}$$
$$\text{database}^- \circ \text{translatedTo} \sqsubseteq \text{translationStoredIn} \tag{15}$$