From Horn-SRIQ to Datalog: A Data-Independent Transformation that Preserves Assertion Entailment

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-$\mathcal{SRIQ}$, the extension of Horn-$\mathcal{SHIQ}$ that supports non-trxansitive, complex roles—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 the our approach is more efficient than state-of-the-art DL reasoners when reasoning with data-intensive ontologies.

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 (Callahan, Cruz-Toledo, and Dumontier 2013; Vrandečić and Krötzsch 2014). For these tasks, both the concepts of an object 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 the inverse role or 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 complex roles, 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 $T_{ex}$ be the TBox with the following axioms modelling conflicts of interests between researchers.

\[
\begin{align*}
\text{ResearchGroup} & \sqsubseteq \forall \text{hasMember}\text{.Researcher} \\
\text{Researcher} & \sqsubseteq \exists \text{hasMember}^- \text{.ResearchGroup} \\
\text{collaborated} & \circ \text{hasMember}^- \circ \text{hasMember} \sqsubseteq \text{hasConflict} \\
\text{hasMember} & \circ \text{supervises} \sqsubseteq \text{hasMember}
\end{align*}
\]

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 $T_{ex}$, we can infer from the ABox $A_{ex} = \{\text{collaborated}(\text{gottlob, alonzo}), \text{supervises}(\text{alonzo, alan}), \text{Researcher}(\text{alonzo})\}$ the two assertions $\text{Researcher}(\text{alan})$ and $\text{hasConflict}(\text{gottlob, alan})$. Both entailments depend on the existence of a research group which has both alan and alonzo as a member, whose existence 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-$\mathcal{SRIQ}$—a DL fragment that supports complex roles and role conjunctions (Krötzsch, Rudolph, and Hitzler 2013)—, based on data-independent rewritings into Datalog rule sets. Specifically, given a TBox $T$, we describe how to construct a Datalog rule set $R_T$ s.t., for every ABox $A$ and assertion $\alpha$ defined only using symbols occurring in $T$, we have $\langle T, A \rangle \models \alpha$ iff $\langle R_T, A \rangle \models \alpha$.

To show practical feasibility, we implemented and evaluated our transformation, showing that Datalog rewritings for many real-world Horn-$\mathcal{SRIQ}$ 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 (Steigmiller, Liebig, and Glimm 2014)—considered as one of the leading DL reasoners (Parria et al. 2017)—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-$\mathcal{SRIQ}$ TBoxes into Datalog rule sets that preserves satisfiability and assertion entailment.

Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
• We show that the resulting rule sets can be transformed into equivalent DLP ontologies (Grosos et al. 2003)—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-SRIQ KBs, while the former three do not perform that well on data-intensive ontologies.

The resulting Datalog programs can be used to solve AR more efficiently than state-of-the-art DL reasoners when reasoning over data-intensive ontologies.

The formal proofs and arguments for the results in this paper are submitted as additional.

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. (2004) reduce AR for DLs which do not support role chains, and which are atomic, to reasoning over Horn-SHIQ ontologies. A similar method tailored for the DL Horn-ACChTQ is presented by Carral et al. (Carral, Dragoste, and Krötzsch 2018). Recently, Shqpjona et al. (2016) proposed Datalog rewritings to perform instance queries over ACCH TO KBs extended with closed predicates.

State-of-the-art DL reasoners such as Fact++ (Tsarkov and Horrocks 2006), HermiT (Motik, Scharer, and Horrocks 2009), Pellet (Sirin et al. 2007) and Konclude (Steigmiller, Liebig, and Glumm 2014) support SROIQ KBs. However, while the former three do not perform that well on data-intensive ontologies (Parsia et al. 2017), Konclude does not support 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 EL++ that can deal with role chains (Krötzsch 2011), leading profile reasoners such as ELK (Kazakov, Krötzsch, and Simancik 2014) do not support this expressive feature yet.

Preliminaries

We consider logical theories based on finite signatures consisting of mutually disjoint sets N_c of concepts (unary predicates), N_r of roles (binary predicates), N_v of variables, and N_i of individuals (constants), as well as an unbounded set N_0 of nulls disjoint with all of the above. There is a bijective and irreflexive function : N_i → N_r with R = R for all R ∈ N_r and ⊥ T ∈ N_c. For a formula or set thereof , we use ( ) to denote the set of all concepts and roles in . The sets of terms and ground terms are N_t = 2_N_i U N_0 U N_v and N_gt = 2^N_i U N_0, respectively. The use of 2^N_i rather than

\[ \bigwedge_{i=1}^{n} A_i(x) \rightarrow B(x) \quad \bigcap_{i=1}^{n} A_i \subseteq B \quad (\land) \]

\[ A(x) \wedge R(x, y) \rightarrow B(y) \quad A \subseteq \forall R.B \quad (\forall) \]

\[ A(x) \rightarrow \exists y.R(x, y) \wedge B(y) \quad A \subseteq \exists R.B \quad (\exists) \]

\[ A(x) \wedge R(x, y) \wedge B(y) \wedge R(z, z) \wedge B(z) \rightarrow y \approx z \quad A \subseteq \leq R.B \quad (\leq) \]

\[ \bigwedge_{i=1}^{n} R_i(x_{i-1}, x_i) \rightarrow S(x_0, x_n) \quad R_1 \circ \ldots \circ R_n \subseteq S \quad (\circ) \]

\[ \bigwedge_{i=1}^{n} R_i(x, y) \rightarrow S(x, y) \quad \bigcap_{i=1}^{n} R_i \subseteq S \quad (\cap) \]

Figure 1: Horn-SRIQ KB Axioms, where A_i, B, ∈ N_c, R_i, S ∈ N_r, x_i, y, z ∈ N_v, and m > 1

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 ∈ N_c with the singleton set {a}.

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

An existential rule is a formula of one of the forms:

\[ \forall \vec{x}, \vec{z}. (B[\vec{x}, \vec{z}] \rightarrow \exists y.H[\vec{x}, \vec{y}]) \rightarrow \] (→)

\[ \forall \vec{x}. (B[\vec{x}] \rightarrow x \approx y) \rightarrow (\approx) \]

Where B and H are non-empty, null-free conjunctions of atoms, and x, y ∈ \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 indentical up to bijective renaming of nulls. A knowledge base (KB) is a tuple (R, A) with R a rule set and 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 \mathbb{T}, we assume that \{ A(x) \rightarrow \top(x) \mid A \in N_c \} ∪ \{ R(x, y) \rightarrow \top(x) \wedge \top(y) \mid R \in N_r \} ⊆ R for every rule set R.

The DL Horn-SRIQ KB. Without loss of generality (Krötzsch, Rudolph, and Hitzler 2013), we define Horn-SRIQ 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 right hand side of Figure 1, and alternate between these two syntaxes whenever this is convenient.

For an axiom set R, let \prec_R be the minimal transitive relation over roles s.t. R \prec_R S iff R^+ \prec^+ R S; for every axiom in R of Type (\cap), R_i \prec_R S for all i ∈ [1, m]; and, for every axiom in R of Type (\circ),

• if n = 1 and R_1 \neq S^+, then R_1 \prec_R S,

• if n > 1 and R_1 \circ \ldots \circ R_n \neq S \circ S, then
  - if R_n = S, then R_i \prec_R S for all i ∈ \{1, \ldots, n − 1\},
sequence of our Datalog rewritings
the chase, which will later be useful to show completeness

\[
\sigma
\]


\[
\tau
\]


\[
\rho
\]


\[
\alpha
\]


\[
\beta
\]


\[
\gamma
\]


\[
\delta
\]


\[
\epsilon
\]


\[
\theta
\]


\[
\varphi
\]


\[
\psi
\]


\[
\chi
\]


\[
\nu
\]


\[
\xi
\]


\[
\omicron
\]


\[
\pi
\]


\[
\sigma
\]


\[
\tau
\]


\[
\rho
\]


\[
\alpha
\]


\[
\beta
\]


\[
\gamma
\]


\[
\delta
\]


\[
\epsilon
\]


\[
\theta
\]


\[
\varphi
\]


\[
\psi
\]


\[
\chi
\]


\[
\nu
\]


\[
\xi
\]


\[
\omicron
\]


\[
\pi
\]


\[
\sigma
\]


\[
\tau
\]


\[
\rho
\]


\[
\alpha
\]


\[
\beta
\]


\[
\gamma
\]


\[
\delta
\]


\[
\epsilon
\]


\[
\theta
\]


\[
\varphi
\]


\[
\psi
\]


\[
\chi
\]


\[
\nu
\]


\[
\xi
\]


\[
\omicron
\]


\[
\pi
\]
• if \( n = 1 \) and \( R_1 = S^- \), then \( q \rightarrow S^- \hat{q} \in N_T(V) \).
• if \( n = 2, R_1 = S, \) and \( R_2 = S, \) then \( \hat{q} \rightarrow q \in N_T(V) \).
Otherwise,
- if \( R_1 \neq S = R_n, \) then \( q \rightarrow q_0 \rightarrow R_1 \rightarrow R_2 \rightarrow \ldots \rightarrow R_n \rightarrow \hat{q} \in N_T(V), \)
- if \( R_1 = S = R_n, \) then \( \hat{q} \rightarrow q_1 \rightarrow R_2 \rightarrow \ldots \rightarrow R_n \rightarrow q \hat{q} \in N_T(V), \)
- if \( R_1 \neq S \neq R_n, \) then \( \hat{q} \rightarrow q_0 \rightarrow R_1 \rightarrow R_2 \rightarrow \ldots \rightarrow R_n \rightarrow q \hat{q} \in N_T(V). \)

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

Our definition of NFA coincides with that from (Horrocks, Kutz, and Sattler 2005) in the sense that the resulting NFA \( N_T(R) \) for any \( R \in \mathbb{N} \) does recognize 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 \( O^i \) is closed under the application of axioms of Type (\( \gamma \)), there is a binary fact \( R(t, u) \in O^i \) iff there are some \( S_1(t, t_1), \ldots, S_n(t_{n-1}, u) \in D(O^i) \) with \( S_1 \ldots S_n \in N_T(R). \)

Given a path \( P = R_1 \ldots R_n \) with \( R_1, \ldots, R_n \in \mathbb{N} \), we write \( q \rightarrow P \hat{q} \in N_T(R) \) (resp. \( P \in N_T(R) \)) to indicate that there is a path \( P \) from \( q \) to \( \hat{q} \) (resp. \( i \) to \( j \)) in \( N_T(R) \).

**Example 3.** Consider \( O_x = \{ T_{e_2}, A_{e_2} \} \) with \( T_{e_2} \) and \( A_{e_2} \) from Example 1. The NFA \( N_{e_2}(HC) \) and \( N_{e_2}(HM(HC)) \) are depicted in Figure 3 (for the sake of clarity, we have removed some \( \epsilon \)-transitions). As implied by Lemma 2 and since \( HC(g, aa) \), we have \( C(g, aa), HM_\sim(a, a, n), HM(n, aa), S(a, aa) \in D(O_{e_2}) \) such that \( C \cdot HM_\sim \cdot HM \cdot S \in \mathcal{N}_{HC}(T_{e_2}) \) (see Figure 2).

**Datalog Rewritings in Horn-\( SRIQ_\cap \)**

In this section, we define the Datalog AR-rewriting \( R_T \) for the TBox \( T \) and discuss complexity results.

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

Let \( O = \{ T, A \} \) and \( K_{O} = \{ R_T, A \} \). By Theorem 1, \( R_T \) is an AR-rewriting only if the chase of \( K_O \) coincides with the chase of \( O \) on all assertions over \( \text{sig}(T) \). The challenge in constructing Datalog AR-rewritings is that assertions in the \( O_{\infty} \) might be introduced by rule applications on facts with nulls, whilst no Datalog rule can introduce such terms.

**Example 4.** Let \( O_x \) be the ontology from Example 1. Then, the assertion \( HC(g, aa) \) is in \( O_{\infty} \) because \( HC(g, aa), HM(n, ao), HM(n, aa) \in O_x \) (see Figure 2). Analogously, \( R(aa) \in O_{\infty} \) because \( R(g), HM(n, aa) \in O_{\infty} \). Note that the facts \( HM(n, ao), HM(n, aa), \) and \( R(g) \) cannot occur in the case of a Datalog AR-rewriting, since \( n \in \mathbb{N} \).

To replicate assertion entailments in \( K_{\infty} \) such as the ones highlighted in the previous example, we encode information in \( K_{\infty} \) about the null successors of an individual in \( O_{\infty} \) using fresh concepts and roles. For all \( R \in \mathbb{N} \), and states \( q, \hat{q} \in N_T(R) \), 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 \( O_{\infty} \) in \( K_{\infty} \).

1. If \( A_q(a) \in K_{\infty} \), then there are some \( A(t_0) \in O_{\infty} \), and some \( R_1(t_0, t_1), \ldots, R_n(t_{n-1}, a) \in D(O_{\infty}) \) with \( q \rightarrow R_1, \ldots, R_n \hat{q} \in N_T(T) \).
2. If \( R_q(q) \in K_{\infty} \), then there are some \( R_1(a, t_1), \ldots, R_n(t_{n-1}, a) \in D(O_{\infty}) \) with \( t_1, \ldots, t_n \in \mathbb{N} \) and \( q \rightarrow R_1, \ldots, R_n \hat{q} \in N_T(T) \).
3. If \( R_q(a, b) \in K_{\infty} \), then \( S_1(a, t_1), \ldots, S_n(t_{n-1}, b) \in D(O_{\infty}) \) with \( i \rightarrow S_1, \ldots, S_n \hat{q} \in N_T(T) \).

Note that all terms \( t_i \) may possibly be nulls that do not appear in the chase of \( K_{\infty} \).

To ascertain when information about one of these predicates needs to be used in \( K_{\infty} \), we make use of a sound saturation calculus from (Eiter et al. 2012), 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-\( SRIQ_\cap \), we first need to extend our input TBox \( T \) to a TBox \( \hat{T} \) in which the behaviour of axioms of Type (\( \gamma \)) is sufficiently simulated. For instance, if the calculus derives from \( T_+ \) an axiom of the
form \( \mathcal{A} \sqsubseteq A_q \), then we can conclude that, for every term \( t \) s.t. \( B(t) \in \mathcal{O}^\infty \) for every \( B \in \mathcal{A} \), there is a set of direct facts \( A(t_0), R_1(t_0, t_1), \ldots, R_n(t_{n-1}, 0) \in \mathcal{O}^\infty \) with a corresponding path in the automata, irrespectively of the ABox \( \mathcal{A} \). We further augment \( \mathcal{T}_s \) to a TBox \( \mathcal{T}_x \) that allows us to trace paths in possible chases for \( \mathcal{T} \). Using the inferences from this calculus, we then describe the rewriting \( \mathcal{R}_T \).

**Definition 5.** Let \( \mathcal{B}(\mathcal{T}) \) be the set of axioms that, for every axiom \( \rho \in \mathcal{T} \) of Type \( \forall \) (\( \exists \)), contains \( A \sqsubseteq A_{i1}, A_{i2} \sqsubseteq B \), and \( A_q \sqsubseteq \mathcal{S}_A \mathcal{Q}_A \in \mathcal{B}(\mathcal{T}) \) for every \( q \rightarrow q', q \in \mathcal{N}_R(\mathcal{T}) \) with \( S \in \mathcal{N}_n \). Let \( \mathcal{T}_s = \mathcal{T}_s \cup \mathcal{B}(\mathcal{T}) \), and \( \mathcal{T}_x = \mathcal{T}_x \cup \mathcal{B}(\mathcal{T} \cup \bigcup_{R \in \mathcal{R}_n} \{ X \sqsubseteq Y, \mathcal{R}_Y \}) \), with \( X \) and \( Y \) fresh concepts.

Then, \( \mathcal{R}_T \) is the Datalog rule set that contains every axiom in \( \mathcal{T}_s \) 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 \( \mathcal{R}_T \) is an AR-rewriting of \( \mathcal{T} \).

This result is a corollary of Lemmas 3, 16, and 11.

**Example 5.** Let \( \mathcal{O}_s \) be the ontology from Example 1. Then, the Datalog rule set \( \mathcal{R}_{T_x} \) contains (amongst others) all the rules in \( \mathcal{T}_{T_s} \) that are not of Type \( \exists \), as well as the following.

1. \( R(x) \rightarrow R_{\text{new}}(x) \)
2. \( R_{\text{new}}(x) \rightarrow R(x) \)
3. \( C(x, y) \rightarrow HC_{q_1}(x, y) \)
4. \( HC_{q_1}(x, y) \rightarrow HC_{q_1}(x, y) \)
5. \( HC_{q_2}(x, y) \rightarrow HC_{\text{fuc}}(x, y) \)
6. \( HC_{\text{fuc}}(x, y) \rightarrow HC_{\text{fuc}}(x, y) \)

*The chase of \( \mathcal{K}_{\mathcal{O}_s} \) is depicted in Figure 5. Note that \( \mathcal{K}_{\mathcal{O}_s} \) 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 \( \mathcal{R}_T \). 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 sig(\( \mathcal{T} \)), \( \langle \mathcal{T}, \mathcal{A} \rangle \) is satisfiable if and only if \( \langle \mathcal{T}, \mathcal{A} \rangle \models \mathcal{F} \) iff \( \langle \mathcal{T}_s, \mathcal{A} \rangle \models \mathcal{F} \).

Since \( \mathcal{T}_s \supseteq \mathcal{T} \), the “if” direction of this lemma follows trivially from monotonicity of 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 \( \mathcal{T}_s \) 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 \( \alpha \) over sig(\( \mathcal{T} \)),

- if \( \bot(t) \in \langle \mathcal{T}_+, \mathcal{A} \rangle^\infty \) with \( t \in \mathcal{N}_{gt} \), then \( \bot(u) \in \langle \mathcal{R}_T, \mathcal{A} \rangle^\infty \) for some \( u \in \mathcal{N}_n \), and
- if \( \alpha \in \langle \mathcal{T}_+, \mathcal{A} \rangle^\infty \), then \( \alpha \in \langle \mathcal{R}, \mathcal{A} \rangle^\infty \).

Let \( \mathcal{O}_i^0, \mathcal{O}_i^1, \ldots \) be a chase sequence for the ontology \( \mathcal{O}_i = \langle \mathcal{T}_i, \mathcal{A} \rangle \) where axioms of Type \( \forall \) are applied with the highest priority. For every \( i \in [1, n] \), we select an axiom \( \rho_i \in \mathcal{T}_i \), and a substitution \( \sigma_i \) s.t. \( \mathcal{O}_i^0 = \mathcal{O}_i^{-1}(\rho_i, \sigma_i) \). To prove Lemma 4, we show via induction that for every \( i \geq 1 \) and every assertion \( \alpha \in \mathcal{O}_i^0 \), we have \( \alpha \in \mathcal{K}_{\mathcal{O}_i} \).

The base case of this induction is trivial, since \( \mathcal{O}_1^0 = \mathcal{A} \) and \( \mathcal{A} \subseteq \mathcal{K}_{\mathcal{O}_1} \) by Definition 2. For the induction step, we provide a thorough case analysis based on the type of the axiom \( \rho_i \), and the type of the elements occurring in the range of \( \sigma_i \). Since \( \alpha \in \mathcal{K}_{\mathcal{O}_i} \) for every assertion \( \alpha \in \mathcal{O}_i^0 \) by the induction hypothesis, many cases follow trivially. The more challenging cases are the following.

1. \( \rho_i \) is of Type (\( \forall \)), \( \sigma_i(x_0), \sigma_i(x_n) \in 2^{\mathcal{N}} \) and \( \sigma_i(x_j) \in \mathcal{N}_0 \) for \( j \in \{1, \ldots, n-1\} \).
2. \( \rho_i \) is of Type (\( \exists \)), \( \sigma_i(x) \in \mathcal{N}_0 \) and \( \sigma_i(y) \in \mathcal{N}_1 \).
3. \( \rho_i \) is of Type (\( \exists \)), \( \sigma_i(y) \in \mathcal{N}_0 \) and \( \sigma_i(x), \sigma_i(z) \in 2^{\mathcal{N}} \) and \( \sigma_i(y) \in \mathcal{N}_0 \).

In all of the challenging cases, the occurrence of facts containing nulls in \( \mathcal{O}_i^0 \) 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 (\( \exists \)) is derived by the calculus if there is a corresponding cyclic path along nulls in \( \mathcal{O}^\infty \).

**Lemma 5.** Let \( i \geq 1 \), \( R_1(t_0, t_1), \ldots, R_n(t_{n-1}, t_n) \in \mathcal{D}(\mathcal{O}_i^0) \), and \( q, \hat{q} \in \mathcal{N}_R(\mathcal{T}) \) with \( q \neq \hat{q} \). If

- \( q \rightarrow \hat{q} \), then \( \hat{q} \in \mathcal{N}_R(\mathcal{T}) \)
- \( \mathcal{T}_0 \in \mathcal{N}_n \)

then there exists \( \mathcal{K} \subseteq \{ A \mid A(t_0) \in \mathcal{O}_i^0 \} \) s.t. \( \mathcal{N}_q \subseteq X_q \in \mathcal{N}_R(\mathcal{T}) \).

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 minimum depth of a term in \( t_0, \ldots, t_n \) in the root forest \( \mathcal{F}(\mathcal{O}_i^0) \). We proceed with the proof for case (1).

**Proof (Sketch).** Let \( \rho_i \) be an axiom of the form \( \rho_i \circ \ldots \circ R_n \sqsubseteq S \in \mathcal{T}_s \). Then, \( R_1(\sigma_i(x_0), \sigma_i(x_1)), \ldots, R_n(\sigma_i(x_{n-1}), \sigma_i(x_n)) \in \mathcal{O}_i^{0-1} \).

By Lemma 2 and the fact that \( \mathcal{O}_i^0 \) is closed under application of rules of Type (\( \forall \)), there is a sequence \( V_i(t_0, t_1), \ldots, V_m(t_{m-1}, t_m) \in \mathcal{D}(\mathcal{O}_i^{0-1}) \) with \( \sigma_i(x_{j-1}) = t_0, \sigma_i(x_j) = t_m, \) and \( V_i \cdot \ldots \cdot V_m \in \mathcal{N}_R(\mathcal{T}) \) for every \( j \in \{1, \ldots, n\} \) (note that possibly \( m = 1 \)). By concatenating these sequences, we can construct a sequence
\[ \bigwedge_{D \in D} D(x) \rightarrow A(x) \iff \mathbb{D} \subseteq A \in \Gamma(T_x) \] (7)
\[ \bigwedge_{D \in D} D(x) \rightarrow R_{q,\check{q}}(x) \iff R \in \mathbb{N}, q, \check{q} \in N_R(T), \text{ and } \mathbb{D} \cap X_q \subseteq X_{\check{q}} \in \Gamma(T_x) \] (8)
\[ A(x) \land \bigwedge_{D \in D} D(x) \land R(x, y) \land B(y) \rightarrow C(y) \iff A \subseteq \{1 \rightarrow R.B, \mathbb{D} \subseteq \exists (R \land R). (A \land B \cap C) \in \Gamma(T_x) \} \quad (<1) \]
\[ A(x) \land \bigwedge_{D \in D} D(x) \land R(x, y) \land B(y) \rightarrow S(x, y) \iff A \subseteq \{1 \rightarrow R.B, \mathbb{D} \subseteq \exists (R \land R \cap S). (A \land B) \in \Gamma(T_x) \} \quad (<2) \]
\[ S(x, y) \rightarrow R_q(x, y) \iff R, S \in \mathbb{N}, \text{ and } i_R \rightarrow_{S}^{q} q \in N_T(R) \] (R 1)
\[ R_{q,\check{q}}(x) \rightarrow R_q(x, x) \iff R \in \mathbb{N}, \text{ and } R_{q,\check{q}} \in \mathbb{R}_T \] (R 2)
\[ R_q(x, y) \land S(y, z) \rightarrow R_q(x, z) \iff R, S \in \mathbb{N}, \text{ and } q \rightarrow_{S}^{z} q \in N_T(R) \] (R 3)
\[ R_q(x, y) \land R_{q,\check{q}}(y) \rightarrow R_q(x, y) \iff R \in \mathbb{N}, \text{ and } R_{q,\check{q}} \in \mathbb{R}_T \] (R 4)
\[ R_{q,R}(x) \rightarrow R_q(x, y) \iff R \in \mathbb{N}, \] (R 5)

Table 1: Rules to construct \( \mathcal{R}_T \), where \( \Gamma((T_x)) \) is the saturation of \( T_x \) by the rules in Figure 4 and all concepts \( A \) and \( B \) and those in the conjunctions \( \mathbb{D} \) and \( \mathbb{A} \) occur \( \{0, \ldots, n\} \).

To show the base case, we check that \( S_{q,\check{q}}(t_0, t_0) \in K^\infty \). We consider two possible cases a) and b) depending on whether \( k_1 = 1 \). a) Let \( k_1 = 1 \). Then, \( W_{q_0}(t_0, t_1) \in K^\infty \)
by the inductive hypothesis. Since \( W_{q_0}(t_0, t_1) \in \mathbb{R}_T \), \( S_{q_0}(t_0, t_1) \in K^\infty \). b) Let \( k_1 > 1 \). By Lemma 1, \( t_{k_1} = t_0 \). By Lemma 9, \( \mathbb{A} \subseteq X_{q_0} \subseteq X_{q_0} \in \Gamma(T_x) \)
with \( \mathbb{A} \subseteq N_T(t_0) \) and hence, \( A(x) \rightarrow S_{q_0}(q_0, x) \in \mathbb{R}_T \).

To show the base case, we consider that \( S_{q_0}(t_0, t_0) \in K^\infty \).

In addition to showing correctness, we can show that our approach is worst-case optimal for Horn-\( \mathcal{SHIQ} \)- 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 \( p \in T \) of Type (c), we have that a) \( n = 1 \) or b) \( n = 2 \), and \( R_1 = R_2 = S \).

A \( \mathcal{ELH} \) TBox \( T \) is a set containing axioms of Type (7), (8), (c), and of the form \( \exists R.A \subseteq B \) with \( A, B \in \mathbb{N}_R \), \( R \in \mathbb{N} \), s.t. a) \( n = 1 \) for every axiom of the form (c) and b) for every \( R \in \mathbb{N}_R \), \( T \) uses \( R \) or \( R^- \), but not both.

Axioms of the form \( \exists R.A \subseteq B \) are equivalent to \( A \subseteq \forall R^- B \), which is why \( \mathcal{ELH} \) is included in Horn-\( \mathcal{SHIQ} \).

Theorem 3. Let \( O = (T, A) \) be an ontology. If \( T \) is Horn-\( \mathcal{SHIQ} \)-Horn-\( \mathcal{SHIQ} \)/\( \mathcal{ELH} \), then we can compute \( \mathcal{R}_T \) as well as \( \langle \mathcal{R}_T, A \rangle^\infty \) in \( 2\text{ExpTime/ExpTime/PTIME} \), respectively.

Finally, we show that, unlike other approaches such as the one proposed by Shqiponi et al. (2016), our rewritings can be transformed into DLP TBoxes.

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 \( \exists^1 A_i \subseteq \exists^1 R.Self \) with \( A_i \in \mathbb{N}_C \) and \( R \in \mathbb{N}_R \).

Definition 8. Let \( T_{dpl} \) be the TBox containing every DLP axiom in \( \mathcal{R}_T \) which additionally satisfies all of the following.

1. If \( \bigwedge_{A \in \mathbb{A}} A(x) \land R(x, y) \land B(y) \rightarrow C(y) \in \mathbb{R}_T \), then \( A \subseteq X_A, X_A \subseteq \forall R.X_{R^-}, X_{R^-} \cap B \subseteq C \in T_{dpl} \).

2. If \( \bigwedge_{A \in \mathbb{A}} A(x) \land R(x, y) \land B(y) \rightarrow S(x, y) \in \mathbb{R}_T \), then \( A \subseteq \exists W_A.Self, B \subseteq \exists W_B.Self, W_A \circ R \circ W_B \subseteq S \in T_{dpl} \).

3. If \( R_q(x, y) \land R_{q,\check{q}}(y) \rightarrow R_q(x, y) \in \mathbb{R}_T \), then \( R_{q,\check{q}} \subseteq \exists W_{q,\check{q}}.Self, R_q \circ W_{q,\check{q}} \subseteq R_{q,\check{q}} \in T_{dpl} \).

In the above, all \( X_A \) and \( R.X_{R^-} \) are fresh concepts unique for every \( A \subseteq \mathbb{N}_C \) and \( R \in \mathbb{N}_R \), and all \( W_A \) and \( W_{q,\check{q}} \) are fresh roles unique for every \( W \in \mathbb{N}_R \), and the states \( q \) and \( \check{q} \).

The rules introduced in (1)–(3) in Definition 8 correspond to consequence-preserving transformations from rules to axioms described in (Krotzsch, Rudolph, and Hitzler 2008). From this, it follows that \( T_{dpl} \) is an AR-rewriting of \( T \).

Evaluation

We implement our rewriting technique in Java using the OWL-API (Horridge and Bechhofer 2011) to handle OWL ontology files, and Clipper (Eiter et al. 2012) to apply the calculus from Figure 4. We performed two different experiments to validate the practical usefulness of our approach.

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 PAGOQA (Zhou et al. 2015). We have normalised these ontologies and removed axioms not expressible in Horn-SRIQ\(\ominus\). Also, we enriched each of these ontologies with three axioms of Type \(\ominus\), as neither ontology contained axioms of this form. 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. (2015). All ontology files used in the evaluation are available in an online repository.\(^1\)

The rewritten Datalog programs for the Reactome and Uniprot TBoxes contained 539 and 367 rules, respectively. We used RDFox (SVN version 2776) as Datalog engine for computing the chase of our rewritings (Motik et al. 2014), and compared its performance with that of Konclude v0.6.2. We performed all experiments 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 evaluations.\(^1\)

The sizes of the successful rewritings are shown in Figure 7, where the red bars correspond to the number of axioms in the input ontology, and the blue bars to the number of rules in the resulting Datalog rewritings. We see that despite some outliers, the rewritings were usually not substantially larger than the input ontology. In 51.4% of the cases, the rewritings were less than 20% larger than the input, while in 80.1% of the cases, the size increased by less than 50% larger than the input. On average, the size increased by 51.4%.

**Conclusions and Future Work**

To the best of our knowledge, we present the first data-independent Datalog transformation for Horn-SRIQ\(\ominus\), 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-SRIQ\(\ominus\), and Horn-SRIQ\(\ominus\), 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 when dealing with data-intensive ontologies 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 (Cucala, Cuenca Grau, and Horrocks 2018; Bate et al. 2016). Also, we intend to further optimise our prototype implementation, in order to produce even smaller rewritings and show that these can be efficiently computed.

---

\(^1\)https://tinyurl.com/yawvg8bw
References


Forest-Model Property

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

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

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

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

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

\[ \square \]

Lemma 1. For all \( i \geq 0 \),
- all nulls in \( O^i \) occur as nodes in \( F(O^i) \), and
- \( F(O^i) \) 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 \( O^0, O^1, \ldots \). It is clear that the base case holds, as \( F(O^0) \) does not contain any edges, and \( O^0 \) does not contain any nulls. To show the inductive step, we show that the lemma holds for \( F(O^i) \) with \( i \geq 1 \) provided that (IH) it holds for \( F(O^{i-1}) \). If the axiom \( \rho_i \) is of Type \( (\cap), (\lor), (\exists) \), or \( (\forall) \), then the set of nulls in \( O^i \) coincides with the set of nulls in \( O^{i-1} \) and \( F(O^i) = F(O^{i-1}) \). Therefore, the lemma holds by IH for all of these cases. We proceed to show that the lemma also holds when \( \rho_i \) is of Type \( (\exists) \) or \( (\leq) \).

Case \( (\exists) \) Then, \( A(\sigma_i(x)) \in O^{i-1} \) and \( O^i = O^{i-1} \cup \{ R(\sigma_i(x), n), B(n) \} \) for some fresh null \( n \). We consider two possible cases.
- Let \( \sigma_i(x) \in 2^{N_i} \). By IH, \( \sigma_i(x) \) is a root in \( F(O^{i-1}) \).
- Let \( \sigma_i(x) \in 2^{N_0} \). By IH, there is some sequence of nodes \( t_0, \ldots, t_n \in F(O^{i-1}) \) such that \( t_0 \in 2^{N_i} \) is a root, all \( t_1, \ldots, t_n \in N_0 \) are not, and \( t_n = \sigma_i(x) \).

In either case, \( n \) only occurs in the edge \( \sigma_i(x) \rightarrow n \in F(O^i) \) since \( n \) only occurs in facts \( R(\sigma_i(x), n), B(n) \in O^i \).

Case \( (\leq) \) In this case, no fresh nulls are introduced in \( O^i \) and hence, the first part of the lemma holds. Note that, \( R \) is simple by Definition 1. We consider four possible cases.
- \( \sigma_i(y), \sigma_i(z) \in 2^{N_i} \). Then, \( F(O^i) \) results from replacing the roots \( \sigma_i(y) \in 2^{N_i} \) and \( \sigma_i(z) \in 2^{N_i} \) in \( F(O^{i-1}) \) with the fresh root \( \sigma_i(y) \cup \sigma_i(z) \in 2^{N_i} \).
- \( \sigma_i(y) \in N_0 \) and \( \sigma_i(z) \in 2^{N_i} \). By Lemma 6, \( R(\sigma_i(x), \sigma_i(y)) \in D(O^{i-1}) \) and hence, \( \sigma_i(x) \) is the predecessor of \( \sigma_i(y) \) in \( F(O^{i-1}) \). We study two possible cases.
2. If there is some axiom of Type $(\exists)$, then $F(O)$ results from replacing all occurrences of $\sigma_i(y)$ in $F(O^{i-1})$ with the root $\sigma_i(x)$, and then erasing all edges from the root $\sigma_i(x)$ to the root $\sigma_i(z)$.

- $\sigma_i(x) \in N_0$ is not a root. Then, $F(O)$ results from replacing all occurrences of the non-root $\sigma_i(y)$ in $F(O^{i-1})$ with the predecessor of its predecessor, i.e., $\sigma_i(z)$.

- $\sigma_i(y) \in N_0$. Analagous to the previous case.

- $\sigma_i(x)$ is the predecessor of $\sigma_i(y)$ and $\sigma_i(z)$. Then, $F(O)$ results from replacing all occurrences of the non-root $\sigma_i(z)$ in $F(O^{i-1})$ 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 $\sigma_i(y)$. Analagous to the previous case.

- $\sigma_i(x)$ is the predecessor of $\sigma_i(y)$ and $\sigma_i(z)$. Then, $F(O)$ 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 $F(O)$ is a rooted forest where every individual node is a root, and every null node is not.

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 $O = \langle T, A \rangle$. Furthermore, consider a chase sequence $O^0_+, O^1_+, \ldots$ for $O_0 = \langle T_+, A \rangle$, a sequence of axioms $\rho_1, \rho_2, \ldots \in T_+$, and a sequence of substitutions such that all of the following conditions hold for all $i \geq 1$.

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

2. If there is some axiom of Type $(\forall)$ or $(\forall \forall)$ in $T_+$ that is applicable to $O^{i-1}_+$, then $\rho_i$ is of Type $(\forall)$ or $(\forall \forall)$.

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

Because of conditions (1), (2), and (3), we can show the following.

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

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

We introduce some notation used in across this section.

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

- Let $\land = A_1 \land \ldots \land A_n$ be a conjunction of concepts, $x \in N_0$, and $t \in N_{gt}$. Then, we write $\land(x)$ as a shortcut for $A_1(x) \land \ldots \land A_n(x)$ and $\land(t) \in F$ as a shortcut for $A_1(t), \ldots, A_n(t) \in F.$
Lemma 8. Let \( i \geq 0 \) and let \( t, u \in \mathbb{N}_{gt} \) be some terms in \( \mathcal{O}_+^i \). If \( t \) is the predecessor of \( u \) in \( F(\mathcal{O}_+^i) \), then \( A \subseteq \exists N_r^i(t, u).N_r^i(u) \in \Gamma(T_x) \) for some \( A \subseteq N_r^i(t) \).

Proof. We verify this result via induction on the chase sequence \( \mathcal{O}_+^0, \mathcal{O}_+^1, \ldots \). Since \( F(\mathcal{O}_+^0) \) 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 \( \rho_i \). In the following enumeration, we consider all cases that do not automatically follow IH.

(\( \forall \)) 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, \( A' \subseteq \exists R. B \in \Gamma(T_x) \) with \( A' \subseteq N_r^{i-1}(t), R = N_r^{i-1}(t, \sigma_i(x)), \) and \( B = N_r^{i-1}(\sigma_i(x)) \). Since \( \prod_{j=1}^n A_j \in B \in \Gamma(T_x) \) and \( A_1, \ldots, A_n \in B, A' \subseteq \exists R.(B \cap B) \in \Gamma(T_x) \).

(\( \forall \)) Then, \( A(\sigma_i(x)), R(\sigma_i(x), \sigma_i(y)) \in \mathcal{O}_+^{i-1} \). By Lemma 7, \( R(\sigma_i(x), \sigma_i(y)) \in D(\mathcal{O}_+^{i-1}) \). We assume that \( \sigma_i(x) \) is the predecessor of \( \sigma_i(y) \), as otherwise this case holds by IH. By IH, \( A' \subseteq \exists R. B \in \Gamma(T_x) \) with \( A' \subseteq N_r^{i-1}(\sigma_i(x)), R = N_r^{i-1}(\sigma_i(x), \sigma_i(y)), \) and \( B = N_r^{i-1}(\sigma_i(y)) \). Since \( A \subseteq \forall R.B \in T_x \) and \( R \in \mathbb{R}, A' \subseteq \exists R.(B \cap B) \in \Gamma(T_x) \).

(\( \exists \)) 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, \( R(\sigma_i(x), \sigma_i(y)), R(\sigma_i(x), \sigma_i(z)) \in D(\mathcal{O}_+^{i-1}) \) by Lemma 6. We consider three possible cases.

\begin{itemize}
  \item \( \sigma_i(x) \) is the predecessor of both \( \sigma_i(y) \) and \( \sigma_i(z) \). By Lemma 1, \( \sigma_i(y), \sigma_i(z) \in \mathbb{N}_0 \) and hence, by IH, \( A' \subseteq \exists R. B, A'' \subseteq \exists R. C \in \Gamma(T_x) \) with \( A', A'' \subseteq N_r^{i-1}(\sigma_i(x)), R = N_r^{i-1}(\sigma_i(x), \sigma_i(y)), B = N_r^{i-1}(\sigma_i(y)), R' = N_r^{i-1}(\sigma_i(x), \sigma_i(z)), B' = N_r^{i-1}(\sigma_i(z)) \). Since \( A \subseteq \exists R.B \in T_x, R \in \mathbb{R} \cup \mathbb{S}, \) and \( B \in B \cap B \subseteq \exists R.(B \cap B) \subseteq \Gamma(T_x) \).
  \item \( \sigma_i(x) \) is the predecessor of \( \sigma_i(y) \) and \( \sigma_i(z) \) is the predecessor of \( \sigma_i(y) \). By Lemma 1, \( \sigma_i(y), \sigma_i(z) \in \mathbb{N}_0 \) and hence, by IH, \( A' \subseteq \exists R. B, B' \subseteq \exists R. C \in \Gamma(T_x) \) with \( A', A'' \subseteq N_r^{i-1}(\sigma_i(x)), R = N_r^{i-1}(\sigma_i(x), \sigma_i(y)), B' = N_r^{i-1}(\sigma_i(x)), B = N_r^{i-1}(\sigma_i(y)), C = N_r^{i-1}(\sigma_i(z)) \). Since \( A \subseteq \exists R.B \in T_x, R \in \mathbb{R} \cup \mathbb{S}, \) and \( R \in R \cap S \subseteq \exists R.(R \cap S) \subseteq \Gamma(T_x) \).
  \item \( \sigma_i(y) \) is the predecessor of \( \sigma_i(x) \) and \( \sigma_i(x) \) is the predecessor of \( \sigma_i(z) \) in \( \mathbb{N}_0 \). Analogous to the previous case.
\end{itemize}

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

\begin{itemize}
  \item \( \sigma_i(x_0) \) is the predecessor of \( \sigma_i(x_1) \). By IH, \( A' \subseteq \exists R. B \in \Gamma(T_x) \) with \( A' \subseteq N_r^{i-1}(\sigma_i(x_0)), R = N_r^{i-1}(\sigma_i(x_0), \sigma_i(x_1)), B = N_r^{i-1}(\sigma_i(x_1)) \). Since \( R_1 \subseteq S \in T_x \) and \( R_1 \subseteq \exists R.(R \cap S) \subseteq \Gamma(T_x) \).
  \item \( \sigma_i(x_1) \) is the predecessor of \( \sigma_i(x_0) \). By IH, \( A' \subseteq \exists R. B \in \Gamma(T_x) \) with \( A' \subseteq N_r^{i-1}(\sigma_i(x_1)), R = N_r^{i-1}(\sigma_i(x_1), \sigma_i(x_0)), B = N_r^{i-1}(\sigma_i(x_0)) \). Since \( R_1 \subseteq S \in T_x \) and \( R_1 \subseteq \exists R.(R \cap S) \subseteq \Gamma(T_x) \).
\end{itemize}

(\( \exists \)) Analogous to the previous case.

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 \( t \in \mathbb{N}_{gt} \) a term occurring in \( \mathcal{O}_+^i \), let \( \text{dep}_i(t) \) be the depth of \( t \) in the rooted forest \( F(\mathcal{O}_+^i) \). For a sequence \( \mathcal{F} = R_1(t_0, t_1), \ldots, R_n(t_{n-1}, t_n), \) \( \text{dep}_i(\mathcal{F}) = \max(\text{dep}_i(t_1), \ldots, \text{dep}_i(t_n)) - \min(\text{dep}_i(t_1), \ldots, \text{dep}_i(t_n)) \).
Lemma 9. Consider \( i \geq 1 \), \( R_1(t_0, t_1), \ldots, R_n(t_{n-1}, t_n) \in D(O^\ell) \), and states \( q \) and \( \hat{q} \) in the NFA \( N_T(R) \) with \( q \neq \hat{q} \). If \( q \xrightarrow{\ell} \hat{q} \in N_T(R) \) with \( P = R_1 \cdot \ldots \cdot R_n \) and all \( t_1, \ldots, t_{n-1} \) are descendants of \( t_0 \) in \( F(O^\ell) \), then \( \mathcal{A}' \cap X_q \subseteq X_{\hat{q}} \in \Gamma(T_x) \) for some \( \mathcal{A}' \subseteq N^\varepsilon_i(t_0) \).

Proof. We prove the lemma via induction on the depth of the sequence \( F = R_1(t_0, t_1), \ldots, R_n(t_{n-1}, t_n) \). Before proceeding with this inductive argument, we derive some conclusions from the premise of the lemma. Since \( q \xrightarrow{\ell} \hat{q} \in N_T(R) \), there are some states \( q_0, \ldots, q_n \) such that \( q_0 = q \), \( q_n = \hat{q} \), and \( q_{j-1} \xrightarrow{\ell} q_j \in N_T(R) \) for all \( j \in [1, n] \). Hence, \( X_{q_{j-1}} \subseteq \forall R_j X_{q_j} \subseteq \Gamma(T_x) \) for all \( j \in [1, n] \).

To show the base case, we check that the lemma holds if \( \text{dep}_i(F) = 1 \). 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, \( \mathcal{A}' \subseteq \exists \mathcal{R}.B \in \Gamma(T_x) \) with \( \mathcal{A}' \subseteq N^\varepsilon_1(t_0), \mathcal{R} = N^\varepsilon_1(t_0, t_1), \) and \( \mathcal{B} = N^\varepsilon_2(t_1) \). Since \( R_1, R_2 \in \mathcal{R}, \mathcal{A}' \cap X_q \subseteq \exists \mathcal{R}.(B \cap X_{q_1}), \mathcal{A}' \cap X_q \subseteq X_{\hat{q}} \subseteq \Gamma(T_x) \) (note that \( q = q_0 \) and \( \hat{q} = q_2 \)).

To show the inductive step, we verify that the lemma holds if \( \text{dep}_i(F) \geq 2 \) assuming that (IH) it holds for every sequence of facts \( R_k(t_{k-1}, t_k), \ldots, R_{\ell}(t_{\ell-1}, t_\ell) \) 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 \( F \). 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 \( B'_j \subseteq N^\varepsilon_j(t_1) \) with \( B'_j \cap A_{k_{j-1}} \subseteq A_{k_j} \in \Gamma(T_x) \).

By Lemma 8, \( \mathcal{A}' \subseteq \exists \mathcal{R}.B \) with \( \mathcal{A}' \subseteq N^\varepsilon_1(t_0), \mathcal{R} = N^\varepsilon_1(t_0, t_1), \) and \( \mathcal{B} = N^\varepsilon_2(t_1) \). Hence, \( \mathcal{A}' \cap A_{q_0} \subseteq \exists \mathcal{R}.(B \cap A_{q_1}), \mathcal{A}' \cap X_q \subseteq \Gamma(T_x) \) since \( R_1 \in \mathcal{R} \) and \( X_{q_0} \subseteq \forall R_1 X_{q_1} \subseteq \Gamma(T_x) \). Therefore, \( \mathcal{A}' \cap \forall R_1 X_{q_1} \subseteq \exists \mathcal{R}.(B \cap A_{q_1}) \subseteq A_{q_1} \subseteq \Gamma(T_x) \) (note that \( i_R = q_0, q_{k_1} = q_1, \) and \( q_{k_m} = q_{n-1} \)). Since \( R^{-} = \mathcal{R} \) and \( A_{q_{n-1}} \subseteq \forall R.A_{k_j} \mathcal{A}' \cap A_{k_j} \subseteq A_{k_j} \subseteq \Gamma(T_x) \) (note that \( f_R = q_{n-1} \)).

Lemma 10. Consider \( i \geq 1 \), \( R_1(t_0, t_1), \ldots, R_n(t_{n-1}, t_n) \in D(O^\ell) \) with \( n \geq 2, A \subseteq 1 \) \( R.B \in \text{Tr}_x, S \in N_q, \) and states \( q \) and \( \hat{q} \) in the NFA \( N_S(S) \). If

- \( R^{-}(t_0, t), A(t), R(t, t_n), B(t_n) \in O^\ell \) with \( t \) the predecessor of \( t_n \) and successor of \( t_0 \),
- \( q \xrightarrow{\ell} \hat{q} \in N_T(S) \) with \( P = R_1 \cdot \ldots \cdot R_n \), and
- \( t_0 \in O^\ell \) and \( t_1, \ldots, t_n \in N_q \),

then \( \mathcal{A}' \cap \forall B \cap X_q \subseteq X_{\hat{q}} \subseteq \Gamma(T_x) \) for some \( \mathcal{A}' \subseteq N^\varepsilon_i(t_0) \).

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 \( q_{j-1} \xrightarrow{\ell} q_j \in N_T(S) \) for all \( j \in [1, n] \). By the definition of \( T_x \), \( X_{q_{j-1}} \subseteq \forall R_j X_{q_j} \subseteq \Gamma(T_x) \) for all \( j \in [1, n] \).

By Lemma 8, \( \mathcal{A}' \subseteq \exists \mathcal{R}.B', B' \subseteq \exists \mathcal{C} \subseteq \Gamma(T_x) \) with \( \mathcal{A}' \subseteq N^\varepsilon_1(t_0), \mathcal{R} = N^\varepsilon_1(t_0, t_1), B' \subseteq N^\varepsilon_2(t_1) = B, S = N^\varepsilon_1(t_1, t_n), \) and \( \mathcal{C} = N^\varepsilon_2(t_n) \). Hence, \( \mathcal{A}' \cap X_q \subseteq \exists \mathcal{R}.(B \cap X_{q_1}) \subseteq \Gamma(T_x) \) since \( R_1 \in \mathcal{R} \) 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 \( B'_j \subseteq B \) such that \( B'_j \cap X_{q_{k_{j-1}}} \subseteq X_{q_{k_j}} \subseteq \Gamma(T_x) \) for all \( j \in [1, n] \). Hence, \( \mathcal{A}' \cap X_q \subseteq \exists \mathcal{R}.(B \cap X_{q_1} \cap \cdots \cap X_{q_{k_m-1}} \cap X_{q_{k_{m+1}}} \cap X_{q_{k_m}}) \subseteq \Gamma(T_x) \) (note that \( k_0 = 1 \)). By Lemma 1, \( t_{k_m+1} = t_n \) (note that possibly \( k_m+1 \neq 1 \)). Since \( R_{k_{m+1}} \in S, B' \cap X_{q_{k_m}} \subseteq \exists \mathcal{C}(C \cap X_{q_{k_{m+1}}}) \subseteq \Gamma(T_x) \).

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 \( C'_{j} \subseteq C \) such that \( C'_{j} \cap X_{q_{\ell_{j-1}}} \subseteq X_{\hat{q}_{j}} \subseteq \Gamma(T_x) \) for all \( j \in [1, o] \). Hence, \( B' \cap X_{q_{k_m}} \subseteq \exists \mathcal{C}(C \cap X_{q_{\ell_0}} \cap \cdots \cap X_{q_{\ell_{o-1}}} \cap X_{q_{\ell_o}}) \subseteq \Gamma(T_x) \) (note that \( q_{k_m+1} = q_{\ell_1} \) and \( q_{\ell_o} = \hat{q} \)). Since \( \mathcal{R}^{-} = S, A, B \subseteq B, R \subseteq R, \mathcal{A}' \cap X_q \subseteq \exists \mathcal{R}.(B \cap A \cap X_{q_0} \cap \cdots X_{q_{k_m}}), B' \cap X_{q_{k_m}} \subseteq \exists \mathcal{C}(C \cap B \cap X_{q_0} \cap \cdots X_{q_{k_m-1}} X_{q_{k_m}}) \subseteq \Gamma(T_x) \), we have that \( \mathcal{A}' \cap B \cap X_q \subseteq X_{\hat{q}} \subseteq \Gamma(T_x) \).
Lemma 11. If $O_+ = \langle T_+, A \rangle$ entails some fact over $\bot$, then so does $K_\alpha = \langle \mathcal{R}_T, A \rangle$. For every assertion $\alpha$ defined over $\text{sig}(T)$, $O_+ \models \alpha$ implies $K_\alpha \models \alpha$.

**Proof.** We show the lemma via induction on the chase sequence $O_+^0, O_+^1, \ldots$

- **Base Case:** We show that $\alpha \in K_\alpha^\infty$ for every assertion $\alpha \in O^0$.
- **Induction step (IS):** For every $i \geq 1$, we show that the following claims hold provided that the induction hypothesis also holds.
  - If $\alpha \in K_\alpha^\infty$ for every assertion $\alpha \in O^i$.
  - If $\bot(t) \in O^i$ with $t \in N_{gt}$, then $\bot(u) \in K_\alpha^\infty$ for some $u \in 2^N$.
- **Induction hypothesis (IH):** $\alpha \in K_\alpha^\infty$ for every assertion $\alpha \in O^{i-1}$.

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

1. We altogether ignore cases in which the set $O^i \setminus O^{i-1}$ 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 $\sigma_i$ is a subset of $2^N$ can be shown with the following argument.
   - By IH, $B\sigma_i \subseteq K_\alpha^\infty$.
   - Since $\rho_i \in \mathcal{R}_T$, $H\sigma_i \subseteq K_\alpha^\infty$.

Therefore, we do not include these cases in the base 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 $\land$.

**Case (\land)**

Let $\rho_i$ be an axiom of the form $\prod_{j=1}^m A_j \sqsubseteq B$. Then, $A_1(\sigma_i(x)), \ldots, A_n(\sigma_i(x)) \in O_+^{i-1}$.

By (1)-(3), we only need to consider the case where $B = \bot$ and $\sigma_i(x) \in N_0$. By Lemma 1, there is a sequence $t_0, \ldots, t_m$ of terms in $O_+^{i-1}$ such that $t_0 = a \in 2^N$, $t_1, \ldots, t_m \in N_0$, $t_m = \sigma_1(x)$, and $t_{j-1}$ is the predecessor of $t_j$ for all $j \in [1, m]$. For all $j \in [1, m]$, $A_{j-1} \subseteq \mathcal{T}_j, A_j \in \mathcal{E}(\mathcal{T}_x)$ with $A_{j-1} \subseteq N_0^{i-1}(t_{j-1})$, $\mathcal{R}_j = N_0^{i-1}(t_{j-1}, t_j)$, and $A_j \subseteq N_0^{i-1}(t_j)$ by Lemma 8. Since $\prod_{j=1}^m A_j \subseteq \bot \in \mathcal{E}(\mathcal{T}_x)$ and $A_{m+1} \subseteq \bot \in \mathcal{E}(\mathcal{T}_x)$ for all $j \in [0, m - 1]$. Since $A_0'(x) \rightarrow \bot(x) \in \mathcal{R}_T$ and $A_0'(\sigma_i(x)) \subseteq K_\alpha^\infty$ by IH, $A_0(x) \subseteq K_\alpha^\infty$.

**Case (\forall)**

Let $\rho_i$ be an axiom of the form $A \sqsubseteq \forall R.B \in \mathcal{T}$. Then, $A(\sigma_i(x)), R(\sigma_i(x), \sigma_i(y)) \in O_+^{i-1}$.

By (1)-(3), we only need to consider cases where $\sigma_i(x) \in N_0$ and $\sigma_i(y) \in 2^N$.

By Lemma (7), $R^-(\sigma_i(y), \sigma_i(x)) \in D(O_+^{i-1})$ and hence, $\sigma_i(y)$ is the predecessor of $\sigma_i(x)$ by Lemma 1. By Lemma 8, $\mathcal{A}' \subseteq \exists \mathcal{R}.\mathcal{B} \in \mathcal{E}(\mathcal{T}_x)$ with $\mathcal{A}' \subseteq N_0^{i-1}(\sigma_i(y))$, $\mathcal{R} = N_0^{i-1}(\sigma_i(y), \sigma_i(x))$, and $\mathcal{B} = N_0^{i-1}(\sigma_i(x))$. Since $R^\in \mathcal{R}$ and $A \subseteq \forall R.B \in \mathcal{T}_x$, $A'(x) \rightarrow B(x) \in \mathcal{R}_T$. Since $\mathcal{A}'(\sigma_i(y)) \subseteq K_\alpha^\infty$ by IH, $B(\sigma_i(y)) \subseteq K_\alpha^\infty$.

**Case (\circ)**

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

By Lemma 2, there is a sequence $V_1(t_0, t_1), \ldots, V_m(t_m-1, t_m) \in D(O_+^{i-1})$ with $\sigma_i(x_{j-1}) = t_0, \sigma_i(x_j) = t_m$, and $V_1 \cdots V_m \in \mathcal{N}_T(R_j)$ for every $j \in [1, n]$ (note that possibly $m = 1$). Note that, this lemma is applicable because $O_+^{i-1}$ is closed under the application of axioms of the Type $\land$ by the definition of $O_+^0, O_+^1, \ldots$. By concatenating the
sequences above, we construct a sequence $V_1(t_0, t_1), \ldots, V_m(t_{m-1}, t_m) \in D(O^{i-1})$ such that $\sigma_i(x_0) = t_0, \sigma_i(x_n) = t_m$, and $V_1, \ldots, V_m \in \mathcal{N}_T(S)$. Hence, there are some states $q_0, \ldots, q_m$ such that $q_0 = t_S, q_m = f_S$, and $q_0 \rightarrow v_1, q_1 \rightarrow v_2, \ldots, \rightarrow v_m, q_m \in \mathcal{N}_S(T)$. Let $k_0, \ldots, k_o$ be the longest sorted sequence of natural numbers with $t_{k_j} \in 2^N$ for all $j \in \{0, \ldots, o\}$. We show via induction that $S_{q_k}((t_0, t_{k_j})) \in \mathbb{K}_\infty$ for all $j \in [1, o]$. In turn, this implies $S(\sigma_i(x), \sigma_i(y)) \in \mathcal{K}_\infty$ since $S_{f_S}((x, y)) \rightarrow S(x, y) \in \mathcal{R}_T$ (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 $S_{q_k_1}((t_0, t_{k_1})) \in \mathbb{K}_\infty$. We consider two possible cases.
• Let $k_1 = 1$. Then, $V_1(t_0, t_1) \in \mathbb{K}_\infty$ by IH. Since $V_i(x, y) \rightarrow S_q_i(x, y) \in \mathcal{R}_T$, $S_{q_1}(t_0, t_1) \in \mathbb{K}_\infty$.

• Let $k_1 > 1$. Then, $t_{k_1} = t_0$ by Lemma 1. By Lemma 9, $A \cap X_{q_0} \subseteq X_{q_k_1} \in \Gamma(T_x)$ for some $A \subseteq N^{i-1}_c(t_0)$ and hence, $A(x) \rightarrow S_{q_0}(q_{k_1})(x) \in \mathcal{R}_T$. By IH, $A(t_0) \in \mathbb{K}_\infty$ and hence, $S_{q_0}(q_{k_1})(t_0) \in \mathbb{K}_\infty$. Since $S_{q_0}(q_{k_1})(x) \rightarrow S_{q_0}(t_0, t_{k_1}) \in \mathbb{K}_\infty$.

To show the induction step, we verify that for all $j \in \{2, \ldots, o\}$, $S_{q_k_1}(t_0, t_{k_1}) \in \mathbb{K}_\infty$ provided that $S_{q_k_j}((t_0, t_{k_{j-1}})) \in \mathbb{K}_\infty$. We consider two possible cases.

• Let $k_j = k_{j-1} + 1$. Then, $V_{k_j}(t_{k_1-1}, t_{k_j}) \in \mathbb{K}_\infty$ by IH. Since $S_{q_k_j}(x, y) \wedge V_{k_j}(y, z) \rightarrow S_{q_k_j}(x, z) \in \mathcal{R}_T$, $S_{q_k_j}(t_0, t_{k_j}) \in \mathbb{K}_\infty$.

• Let $k_j > k_{j-1} + 1$. Then, $t_{k_j} = t_{k_{j-1}}$ by Lemma 1. By Lemma 9, $A \cap X_{q_k_{j-1}} \subseteq X_{q_{k_j}} \in \Gamma(T_x)$ for some $A \subseteq N^{i-1}_c(t_{k_{j-1}})$ and hence, $A(x) \rightarrow S_{q_k_{j-1}, q_k_j}(x) \in \mathcal{R}_T$. By IH, $A(t_{k_{j-1}}) \in \mathbb{K}_\infty$ and hence, $S_{q_k_{j-1}, q_k_j}(t_{k_{j-1}}) \in \mathbb{K}_\infty$. Since $S_{q_k_{j-1}, q_k_j}(x, y) \wedge S_{q_k_{j-1}, q_k_j}(y) \rightarrow S_{q_k_j}(x, y) \in \mathcal{R}_T$, $S_{q_k_j}(t_0, t_{k_j}) \in \mathbb{K}_\infty$.

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

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

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

Let $\sigma_i(x), \sigma_i(y) \in N_0$ and $\sigma_i(z) \in 2^N$. 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 $C(\sigma_i(z)) \in \mathbb{K}_\infty$ if $C(\sigma_i(y)) \in O^{i-1}_\infty$ for some $C \in N_c$. By Lemma 8, $A' \subseteq \exists R B, B' \subseteq S C \in \Gamma(T_x)$ with $A' \subseteq N^{i-1}_c(\sigma_i(z)), \forall R = N^{i-1}_c(\sigma_i(z), \sigma_i(x)), B' \subseteq N^{i-1}_c(\sigma_i(x)) \in B, S = N^{i-1}_c(\sigma_i(x), \sigma_i(y)), \text{ and } C = N^{i-1}_c(\sigma_i(y))$. Since $R' \subseteq R$, $A \in B, R \in S, B \in C, A' \cap B \subseteq C' \in \Gamma(T_x)$ and hence, $A' \wedge B \rightarrow C \in \mathcal{R}_T$. Since $A' \sigma_i(z)) \in \mathbb{K}_\infty$ by IH, $C(\sigma_i(z)) \in \mathbb{K}_\infty$.

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

• $t_1, \ldots, t_n \in N_0$. By Lemma 1, $t_0 = a$. By Lemma 10, $A' \cap X_{q_k_s} \subseteq X_{f_s} \in \Gamma(T_x)$ for some $A' \subseteq N^{i-1}_c(t_0)$ and therefore, $A' \rightarrow S_{f_s}(x) \in \mathcal{R}_T$. By IH, $A'(t_0) \in \mathbb{K}_\infty$.
We consider two possible cases. Let $\sigma(x, y) \in N_0$ and $\sigma_i(y) \in 2^N$. Analogous to the previous case.

Let $\sigma_i(x), \sigma_i(z) \in 2^N$ and $\sigma_i(y) \in N_0$. By Lemma 1, $\sigma_i(x)$ is the predecessor of $\sigma_i(y)$.

We first show that $C(\sigma_i(z)) \in K^\infty_\mathcal{O}$ if $C(\sigma_i(y)) \in O_{+}^{+1}$ for some $C \in N_0$. By Lemma 8, $\mathcal{A}' \subseteq \exists \mathcal{R} \mathcal{B} \in \Gamma(T_+)$ with $\mathcal{A}' \subseteq N_{-}^{i-1}(\sigma_i(x))$, $\mathcal{B} = N_{-}^{i-1}(\sigma_i(y))$. Since $R \in \mathcal{R}, A(x) \land A'(x) \land R(x, y) \land B(y) \rightarrow C(x) \in \mathcal{R}_T$.

We will consider two possible cases.

Let $S(a, \sigma_i(z)) \in D(O_{+}^{i-1})$. By Lemma 2, there are some $R_1(t_0, t_1), \ldots, R_n(t_{n-1}, t_n) \in D(O_{+}^{i-1})$ with $a = t_0, \sigma_i(y) = t_n$, and $R_1 \cdot \ldots \cdot R_n \in \mathcal{N}_R(S)$. Let $k$ be the largest number with $t_k = \sigma_i(x)$ and $t_0, \ldots, t_n \in N_0$ for some $k \subseteq \{0, n - 1\}$ (note that such a number must exist by Lemma 1). Also by Lemma 1, $t_{k+1} = \sigma_i(y)$ and hence, $R_{k+1} \in \mathcal{R}, A(x) \land A'(x) \land R(x, y) \land B(y) \rightarrow R_{k+1}(x, y) \in \mathcal{R}_T$. Let $R_{k+1}(x, y) \in \mathcal{R}_T$ (note that $\sigma_{k+1}$ is in this case).

In either of these cases, $S_{q_{k+1}}(a, \sigma_i(z)) \in K^\infty_\mathcal{O}$. Let $\ell_0, \ldots, \ell_m$ be the longest sorted sequence with $\ell_0 = k + 1$ and $\sigma_i(y) = \ell_j$ for all $j \subseteq \{0, n\}$. By Lemma 9, for all $j \subseteq \{0, n\}$ there is some $E_j \subseteq N_{-}^{i-1}(\sigma_i(y))$ such that $E_j \cap X_{q_{j+1}} \subseteq \Gamma(T_+)$. Hence, $E_j(x) \rightarrow S_{q_{j+1}}(x, y) \in \mathcal{R}_T$ for all $j \subseteq \{0, n\}$. Since $A(x) \land A'(x) \land R(x, y) \land B(y) \rightarrow E_j(y) \in \mathcal{R}_T, E_j(\sigma_i(z)), S_{q_{j+1}}(x, y) \in K_{+}^\infty_\mathcal{O}$ for all $j \subseteq \{0, n\}$. For all $j \subseteq \{0, n\}, S_{q_{j+1}}(x, y) \land \sigma_i(z) \rightarrow S_{q_{j+1}}(x, y) \in \mathcal{R}_T$. Hence, $S_{q_{m}}(a, \sigma_i(z)) \in K_{+}^\infty_\mathcal{O}$ (note that $f_{q_{m}} = q_{m}$). Since $S_{q_{m}}(x, y) \rightarrow S(x, y) \in \mathcal{R}_T, S(a, \sigma_i(z)) \in K^\infty_\mathcal{O}$.

Let $\sigma_i(x), \sigma_i(y) \in 2^N, and \sigma_i(z) \in N_0$. Analogous to the above case.

\[\square\]

**Soundness**

To show that our Datalog AR-rewriting is sound, we have to show that for every ABox $A$ s.t. $\text{sig}(A) \subseteq \text{sig}(T)$ and assertion $\alpha$ s.t. $\text{sig}(\alpha) \subseteq \text{sig}(T)$, we have $\langle \mathcal{R}_T, A \rangle \models \alpha$ only if $\langle T, A \rangle \models \alpha$. Note that the case where $\langle T, A \rangle$ is unsatisfiable is trivial, since in that case $\langle T, A \rangle \models \alpha$ holds for every assertion $\alpha$. We therefore silently assume that $\langle T, A \rangle$ 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 $A$ of ground facts (which may include facts over nulls) and a TBox $T$, we call $\langle T, A \rangle$ 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-$\mathcal{SRIQ}_\rho$-TBox, $\rho = A \sqsubseteq \forall R.B \in \mathcal{T}$, and $\mathcal{T}' = \mathcal{B}(\mathcal{T}, \rho)$. Further, let $A$ be any set of ground facts s.t. $\text{sig}(A) \subseteq \text{sig}(\mathcal{T})$, $q, \hat{q}$ two states in $N_R(\mathcal{T})$ and $t, u$ two ground terms. Let $A' = A \cup \{A_q(u)\}$. Then, $\langle \mathcal{T}', A' \rangle \models A_q(t)$ implies that there is a path $P$ in the chase of $\langle \mathcal{T}, A \rangle$ connecting $u$ to $t$ s.t. $q \rightarrow_{p}^s \hat{q} \in N_R(\mathcal{T})$. 

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

Let $i > 0$, and assume the inductive hypothesis holds for $i - 1$. The only interesting case is where $A_q(t)$ is introduced in $\mathcal{F}^i$. Let $\rho_i$ 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 $A_q \sqsubseteq \forall S.A_q$, where $q' \rightarrow_{e}^s \hat{q} \in N_R(\mathcal{T})$, which means that $\rho_i$ must be of this form. This means that there exists some ground term $v$ s.t. $A_q(v), R(v, t) \in \mathcal{F}^i$. By the inductive hypothesis, there is a path $P$ connecting $u$ to $v$ in the chase of $\langle \mathcal{T}, A \rangle$ s.t. $q \rightarrow_{P}^s \hat{q} \in N_R(\mathcal{T})$. This implies that there is the path $P \cdot S$ connecting $u$ to $t$, and that $i_R \rightarrow_{P}^s q \in N_R(\mathcal{T})$. □

Lemma 13. Let $\mathcal{T}$ be an Horn-$\mathcal{SRIQ}_\rho$-TBox, $\mathcal{A}$ an ABox s.t. $\text{sig}(\mathcal{A}) \subseteq \text{sig}(\mathcal{T})$ and $\alpha$ an assertion s.t. $\text{sig}(\alpha) \subseteq \text{sig}(\mathcal{T})$. Then, $\langle \mathcal{T}, \alpha \rangle \models \alpha$ 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 $\langle \mathcal{T}, \mathcal{A} \rangle$. We have to show that for all $i \geq 0$ and every $\alpha \in \mathcal{F}^i$ s.t. $\text{sig}(\alpha) \subseteq \text{sig}(\mathcal{T}), \langle \mathcal{T}, \mathcal{A} \rangle \models \alpha$. We do so by induction over $i$. The case where $i = 0$ is trivial. For $i > 0$, the only interesting case is where $\alpha$ is introduced in $i$ by application of an axiom $\rho \in \mathcal{T}_\rho \setminus \mathcal{T}$. If this is the case, by the definition of $\mathcal{T}_\rho$ there is an axiom $A \sqsubseteq \forall R.B \in \mathcal{T}$ s.t. $\rho \in \mathcal{B}(A \sqsubseteq \forall R.B, \mathcal{T})$. Note that for every distinct axioms $\rho_1, \rho_2 \in \mathcal{T}_\rho$, the sets of fresh predicates in $\mathcal{B}(\rho_1, \mathcal{T})$ and $\mathcal{B}(\rho_2, \mathcal{T})$ are disjoint.

The only axioms in $\mathcal{B}(A \sqsubseteq \forall R.B, \mathcal{T})$ that share a predicate name with $\mathcal{T}$ are $A \subseteq A_{i_{R}}$, and $A_{j_{R}} \subseteq B$. We obtain that $\rho = A_{j_{R}} \subseteq B, \alpha = B(b)$ for some $b \in N_0$, and that $A_{i_{R}}(b) \in \mathcal{F}^{-1}$. We thus have to show that $\langle \mathcal{T}, A \rangle \models B(b)$. By induction over the axioms in $\mathcal{B}(A \sqsubseteq \forall R.B)$ applied to derive $B(b)$, we further obtain that $A(a), A_{i_{R}}(a) \in \mathcal{F}^{-1}$ for some individual $a$. Because i) $A_{i_{R}}(a), A_{j_{R}}(b) \in \mathcal{F}^{-1}$, ii) the fresh predicates in $\mathcal{B}(A \sqsubseteq \forall R.B)$ do not occur in $\mathcal{T}_\rho \setminus \mathcal{B}(A \sqsubseteq \forall R.B)$, iii) the inductive hypothesis, and iv) by Lemma 12, we obtain that there is a path $P$ in the chase of $\langle \mathcal{T}, A \rangle$ connecting $a$ and $b$ s.t. $i_R \rightarrow_{P}^s f_R \in N_R(\mathcal{T})$. By Lemma 2, this implies that $\langle \mathcal{T}, A \rangle \models R(a, b)$, and since $A \subseteq \forall R.B \in \mathcal{T}, \langle \mathcal{T}, A \rangle \models B(b)$. □

Lemma 3. For a TBox $\mathcal{T}$, an ABox $\mathcal{A}$ and a fact set $\mathcal{F}$ defined over $\text{sig}(\mathcal{T})$, $\langle \mathcal{T}, \alpha \rangle$ is satisfiable if and only if $\langle \mathcal{T}, \alpha \rangle \models \mathcal{F}$ if $\langle \mathcal{T}, \mathcal{A} \rangle \models \mathcal{F}$.

Lemma 14. Let $\mathcal{T}$ be a Horn-$\mathcal{SRIQ}_\rho$-TBox, $A$ be a set of ground facts s.t. $\text{sig}(A) \subseteq \text{sig}(\mathcal{T})$, $R$ a complex role, $q, q'$ two states in $N_R(\mathcal{T})$ and $u, t$ two ground terms. Let $\mathcal{T}'$ be the set of axioms in $\mathcal{T}$ plus all axioms generated by the Rule (C). Assume further that there is a path $P$ connecting $u$ and $t$ in the chase of $\langle \mathcal{T}, A \rangle$ s.t. $i_R \rightarrow_{P}^s q \in N_R(\mathcal{T})$. Then, $\langle \mathcal{T}', A \rangle \models R_{q, q'}(t)$ implies the following.

- There is a path $P'$ from $t$ to $t$ in the chase of $\langle \mathcal{T}, A \rangle$, and
- $q \rightarrow_{P} \hat{q} \in N_R(\mathcal{T})$.

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


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

Lemma 15. Let $T$ be a Horn-$\mathbf{SRILQ}_{\forall\exists}$-TBox, $A$ be an ABox s.t. $\text{sig}(A) \subseteq \text{sig}(T)$, $R$ a complex role, $q$ a state in $N_R(T)$ and $a, b \in N_i$. Let $T' \subseteq R_T$ be the set of Datalog rules that either occur in $T$, or that are generated by Rules (\textcircled{1}), (\textcircled{3}) and (R1)–(R4).

Then, $\langle T', A \rangle \models R_q(a, b)$ only if in the chase of $\langle T, A \rangle$, there exists a path $P$ from $a$ to $b$ s.t. $i_R \rightarrow_P^* q \in N_R(T)$.

Proof. Let $T$, $A$, $R$ and $a$ be as in the lemma. Let $A = F_0, F_1, \ldots$ be the chase of $\langle T', A \rangle$. We show that for every $i \geq 0$, state $q$ in $N_R(T)$ and $b \in N_i$, $R_q(a, b) \in F_i$ only if in the chase of $\langle T, A \rangle$, there exists a path $P$ from $a$ to $b$ s.t. $i_R \rightarrow_P^* q \in N_R(T)$.

Since $\text{sig}(A) \subseteq \text{sig}(T)$ 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 $F_i$. Let $\rho_i$ be the axiom applied on $F_{i-1}$ to generate $F_i$. The only axioms in which $R_q$ occurs positively are those introduced by (R1)–(R4). We distinguish the cases.

1. $\rho_i$ was introduced by (R1). Then $a$ and $b$ are connected by $S$ already in the chase of $\langle T'', A \rangle$, where $T''$ contains all the Datalog rules in $T_x$, since no other rule in $T'$ contains $S$ positively. By soundness of $T_x$ (Lemma 13), this implies that $a$ and $b$ are connected by $S$ in the chase of $\langle T, A \rangle$, and we have $i_R \rightarrow_P^* q \in N_R(T)$.

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

3. $\rho_i$ was introduced by (R3). Then, there is a state $\hat{q}$ in $N_R(T)$ and an individual $c$ s.t. $R_q(a, c), S(c, b) \in F_{i-1}$ and $i_R \rightarrow_P^* \hat{q} \in N_R(T)$. By inductive hypothesis, there is then a path $P$ from $a$ to $c$ in the chase of $\langle T, A \rangle$ s.t. $i_R \rightarrow_P^* \hat{q} \in N_R(T)$. We obtain that there is the path $P \cdot S$ connecting $a$ to $b$ s.t. $i_R \rightarrow_P^* \hat{q} \in N_R(T)$.

4. $\rho_i$ was introduced by (R4). Then, there is a state $\hat{q}$ in $N_R(T)$ s.t. $R_q(a, b), (b) \in F_{i-1}$.

By inductive hypothesis, there is then a path $P$ from $a$ to $b$ in the chase of $\langle T, A \rangle$ s.t. $i_R \rightarrow_P^* \hat{q} \in N_R(T)$. By Lemma 14, there is a path $P'$ in the chase of $\langle T, A \rangle$ from $b$ to $b$ s.t. $i_R \rightarrow_P^* q \in N_R(T)$. We obtain that the path $P \cdot P'$ connects $a$ and $b$ in the chase of $\langle T, A \rangle$, and that $i_R \rightarrow_P^* \hat{q} \in N_R(T)$.

Lemma 16 (Soundness). Let $T$ be a Horn-$\mathbf{SRILQ}_{\forall\exists}$-TBox and $A$ be an ABox s.t. $\text{sig}(A) \subseteq \text{sig}(T)$. Then, for every assertion $\alpha$ s.t. $\text{sig}(\alpha) \subseteq \text{sig}(T)$, $\langle R_T, A \rangle \models \alpha$ only if $\langle T, A \rangle \models \alpha$.

Proof. Let $F_0 = A, F_1, \ldots$ be the chase of $\langle R_T, A \rangle$. We show that for every $i \geq 0$ and every assertion $\alpha \in F_i$ s.t. $\text{sig}(\alpha) \subseteq \text{sig}(T)$, $\langle T, A \rangle \models \alpha$. Note that, since $R_T$ is a Datalog rule set, no axiom in $R_T$ introduces nulls, so that for all $i \geq 0$, $F_i$ 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 $\alpha$ is introduced in $F_i$ by an axiom $\rho_i$. We distinguish the cases based on the origin of $\rho_i$.

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

2. If $\rho_i$ is introduced by (\textcircled{1}), by soundness of the Datalog calculus, $T_x \models \rho_i$. Therefore, $\langle T, A \rangle \models \alpha$ follows from the inductive hypothesis and from Lemma 13.

3. $\rho_i$ cannot have been introduced by any of the Rules (\textcircled{3}) or (R1)–(R4) since $\text{sig}(\alpha) \subseteq \text{sig}(T)$.
4. If $\rho_i$ is introduced by Rule (R5), then $R_{fR}(a,b) \in \mathcal{F}^{i-1}$ for some $a,b \in \mathbb{N}_i$, and $\alpha = R(fR)$. It then follows from Lemma 15 that there is a path $P$ in the chase of $\langle T, A \rangle$ connecting $a$ to $b$ s.t. $i_R \rightarrow_P fR \in \mathcal{N}_R(T)$. Consequently, by Lemma 2, we have $\langle T, A \rangle \models R(a,b)$. 

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

$$\mathcal{T}_x \models A \subseteq 1 R.B, D \subseteq \exists (R \cap R). (A \cap B).$$

Note that, since $\text{sig}(A) \subseteq \text{sig}(T)$, $A$ cannot have any occurrences of the fresh concept name $X$. Inspection of the axioms in $\mathcal{T}_x \setminus \mathcal{T}_+$ reveals that therefore that the above entailment in fact holds already for $\mathcal{T}_+$. Furthermore, the latter entailment can be weakened to $\mathcal{T}_+ \models D \subseteq \exists R.(B \cap C)$. Since $D(a) \in \mathcal{F}^{i-1}$, $\langle \mathcal{T}_+, \mathcal{F}^{i-1} \rangle \models \exists y.(R(a,y) \land B(y) \land C(y))$, 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 $\mathcal{T}_+ \models A \subseteq 1 R B$, $a$ 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 $\langle T, A \rangle \models C(b)$. 

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

$$\mathcal{T}_x \models A \subseteq 1 R.B, D \subseteq \exists (R \cap R \cap S). (A \cap B).$$

Similar as in the last case, we obtain that $\mathcal{T}_+ \models D \subseteq \exists (R \cap S). B$. 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 $\mathcal{T}_+ \models A \subseteq 1 R.B$, this successor has to be $b$. We obtain that $\langle \mathcal{T}_+, \mathcal{F}^{i-1} \rangle \models S(a,b)$. By the inductive hypothesis and Lemma 13, we obtain that $\langle T, A \rangle \models S(a,b)$. 

\[ \square \]

**Complexity Results**

**Theorem 3.** Let $\mathcal{O} = \langle T, A \rangle$ be an ontology. If $T$ is Horn-$\mathcal{SRIQ}_\text{r}/$Horn-$\mathcal{SHIQ}/\mathcal{ELH}$, then we can compute $\mathcal{R}_T$ as well as $\langle \mathcal{R}_T, \mathcal{A} \rangle^\infty$ in $2\text{ExpTime}/\text{ExpTime}/\text{PTime}$, 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 $\mathcal{T}_x$ are also at most exponential in the input if $\mathcal{T}$ is in Horn-$\mathcal{SRIQ}_r$, and at polynomial if $\mathcal{T}$ is in Horn-$\mathcal{SHIQ}$ or $\mathcal{ELH}$. The calculus in Figure 4 adds one axiom to $\Gamma(\mathcal{T}_x)$ in each step, the number of which is exponentially bounded in the size of $\mathcal{T}_x$. 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 $\mathcal{T}_x$. 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 $R \in \mathbb{N}_x$, $\mathcal{T}$ contains one of $R$ and $R^-$ but not both. Since axioms of the form $\exists R.A \subseteq B$ correspond to axioms $A \subseteq \forall R^- . B$, 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 $\Gamma(\mathcal{T}_x)$, contains all relevant axioms, and whose size is polynomially bounded in the size of $\mathcal{T}_x$, and therefore in the size of $\mathcal{T}$, if and can be computed in polynomial time.
Finally, $\mathcal{R}_T$ can be generated by traversing $\Gamma(T_x)$ and each automata at most once, and its size is bounded by the size of $\mathcal{R}_{T_x}$ and the number of states occurring in all automata. We obtain that

- for Horn-$\mathcal{SRIQ}$, $\mathcal{R}_T$ can be computed in 2ExpTime and is of at most double exponential size,
- for Horn-$\mathcal{SHIQ}$, $\mathcal{R}_T$ can be computed in ExpTime and is of at most exponential size, and
- for $\mathcal{ELH}$, $\mathcal{R}_T$ can be computed in PTIME and is of at most polynomial size.

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

We first note that the number of elements in $\mathcal{R}_T^\infty$ 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 $\mathcal{R}_T^\infty$ is bounded by the number of individual names and predicate names in $T$ and $A$. 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-$\mathcal{SRIQ}$ and the Datalog rules included based on Table (1), we obtain that except for axioms of Type (◦), 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 (◦), 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 $R_i(c, d)$ 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 $\langle \mathcal{R}_T, A \rangle^\infty$, we obtain that $\langle \mathcal{R}_T, A \rangle^\infty$ can be computed in time polynomial in $m$. □