Ontology Materialization by Abstraction Refinement in Horn $\mathcal{SHOIF}$

Supplemental Material

Abstract

Abstraction refinement is a recently introduced reasoning technique for materializing concept assertions in ontologies. Although the approach is sound for the very expressive Description Logic $\mathcal{SROIQ}$, it is complete only for Horn $\mathcal{ALCHOI}$. In this paper, we propose an extension of this method that is now complete for Horn $\mathcal{SHOIF}$ and also handles role- and equality-materialization. This is particularly challenging due to the sophisticated interaction between functionality, nominals and inverse roles. To show completeness, we use a tailored set of materialization rules that loosely decouple the ABox from the TBox, which is intrinsically interesting. An empirical evaluation demonstrates that the abstractions are significantly smaller than the original ontologies and the materializations can be computed efficiently.

Introduction

Description Logics (DLs) are popular languages for knowledge representation and reasoning. They are the underlying formalism for the standardized Web Ontology Language (OWL), which are widely used in many application areas. Recent years have also seen an increasing interest in ontology-based data access (OBDA), where a TBox with background knowledge, often expressed in a DL language, is used to enrich datasets (ABoxes), which are then accessible via queries. Ontology materialization is a reasoning task that computes logical consequences of the dataset w.r.t. the TBox and it is the most important task in some languages, e.g. OWL 2 RL. In other languages, e.g., those that allow existential quantification, materialization is a stepping stone for query answering via rewriting (Kontchakov et al. 2011).

To make ontology materialization useful in practice, especially for large datasets, scalable materialization is of great importance. Several approaches have been proposed to achieve this goal. The RDFox (Motik et al. 2014) and WebPIE (Urbani et al. 2012) systems operate on the entire dataset and utilize parallel computing to perform a rule-based materialization for OWL 2 RL. Other approaches try to reduce the size of the dataset. Modules or so-called 'individual islands' (Wandelt and Möller 2012) are used for reducing the set of ABox assertions to those that are sufficient for computing the entailed assertions for a given individual. The SHER system (Dolby et al. 2009) improves consistency checking of a large ABox by computing a so-called 'summary ABox' in which several original individuals are merged into one. If the resulting ABox is found consistent, then so is the original ABox. If not, then explanations (Kalyanpur et al. 2007) are used to pinpoint the contradictory axioms or relax the summary to avoid inconsistency. Similar to SHER, in the abstraction refinement method for Horn $\mathcal{ALCHOI}$ (Glimm et al. 2014b) several individuals of the original ABox can be represented by one individual in a corresponding 'abstract ABox'. In contrast to the summary ABox, the abstract ABox provides an under-approximation rather than the over-approximation of entailments (e.g., if the abstraction is inconsistent then so is the original ABox but not necessarily vice versa). To ensure completeness of the method, the so-called refinement step is used that recomputes the abstraction based on new (sound) entailments obtained from a previous abstraction. This has the added benefit that not only consistency but also the full materialization of the ABox can be computed without (rather expensive) explanation computations or repeated consistency checks. This paper significantly advances the abstraction refinement method in several directions:

1. We extend the approach to guarantee completeness in the presence of transitive and functional roles, thus fully supporting Horn $\mathcal{SHOIF}$ ontologies. Reasoning with nominals, inverse roles, and functionality is known to be challenging due to the loss of the tree-model property and the existence of implicitly cardinality constrained concepts (implicit nominals).
2. We materialize not only concept assertions, but also role and equality assertions. In $\mathcal{ALCHOI}$, role and equality assertions can be computed by expanding the role hierarchy and analyzing assertions of nominals. In $\mathcal{SHOIF}$ special techniques are needed to properly handle functionality and the consequences of implicit nominals.
3. We present a new set of materialization rules, which loosely decouple the ABox from the TBox. Although we use them only for proving completeness of the method, these rules can be of interest on its own, e.g., as a basis of an efficient implementation for ABox reasoning. This provides a fresh view of the approach as the completeness proofs are principally different from the proofs by Glimm et al. (2014b) and the method can potentially be extended to other languages having similar rules.

	Syntax	Semantics
Concepts:		
atomic concept	A	$A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
nominal	o	$o^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}, \lVert o^{\mathcal{I}} \rVert = 1$
top	$\top$	$\Delta^{\mathcal{I}}$
bottom	$\bot$	$\emptyset$
negation	$\neg C$	$\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
disjunction	$C \sqcup D$	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$
existential restriction	$\exists R.C$	$\{d \mid \exists e \in C^{\mathcal{I}} : \langle d, e\rangle \in R^{\mathcal{I}}\}$
universal restriction	$\forall R.C$	$\{d \mid \langle d, e\rangle \in R^{\mathcal{I}} \rightarrow e \in C^{\mathcal{I}}\}$
Axioms:		
concept inclusion	$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
role inclusion	$R \sqsubseteq S$	$R^{\mathcal{I}} \subseteq S^{\mathcal{I}}$
role transitivity	$\mathsf{tran}(R)$	$R^{\mathcal{I}} \circ R^{\mathcal{I}} \subseteq R^{\mathcal{I}}$
role functionality	$\mathsf{func}(R)$	$\langle d, e\rangle, \langle d, e'\rangle \in R^{\mathcal{I}} \rightarrow e = e'$
concept assertion	$C(a)$	$a^{\mathcal{I}} \in C^{\mathcal{I}}$
role assertion	$R(a, b)$	$\langle a^{\mathcal{I}}, b^{\mathcal{I}}\rangle \in R^{\mathcal{I}}$
equality assertion	$a \approx b$	$a^{\mathcal{I}} = b^{\mathcal{I}}$

Table 1: The syntax and semantics of the DL $\mathcal{SHOIF}$

4. We evaluate our approach on several real life and benchmark ontologies. The abstractions are often significantly smaller than the original ontologies (by orders of magnitude) and the materialization can be computed efficiently.

Preliminaries

The syntax of $\mathcal{SHOIF}$ is defined using a vocabulary consisting of countably infinite disjoint sets $\mathsf{N_C}$ of *atomic concepts*, $\mathsf{N_O}$ of *nominals*, $\mathsf{N_R}$ of *atomic roles*, and $\mathsf{N_I}$ of *individuals*. A role is either atomic or an *inverse role* r^-, $r \in \mathsf{N_R}$. We define $R^- := r^-$ if $R = r$ and $R^- := r$ if $R = r^-$. Complex *concepts* and *axioms* are defined in Table 1. An *ontology* $\mathcal{O}$ is a finite set of axioms, written as $\mathcal{O} = \mathcal{A} \cup \mathcal{T}$, where $\mathcal{A}$ is an *ABox* consisting of the concept, role, and equality assertions in $\mathcal{O}$ and $\mathcal{T}$ a *TBox* consisting of the concept and role inclusion, transitivity, and functionality axioms in $\mathcal{O}$. To simplify presentation, we do not distinguish between $R(a, b)$, $a \approx b$, $R \sqsubseteq S$ and $\mathsf{tran}(R)$ and, respectively, $R^-(b, a)$, $b \approx a$, $R^- \sqsubseteq S^-$ and $\mathsf{tran}(R^-)$. We use $\mathsf{con}(\mathcal{O})$, $\mathsf{rol}(\mathcal{O})$, $\mathsf{ind}(\mathcal{O})$, and $\mathsf{nom}(\mathcal{O})$ for the sets of atomic concepts, atomic roles, individuals, and nominals occurring in $\mathcal{O}$, respectively.

An *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consists of a non-empty set $\Delta^{\mathcal{I}}$, the *domain* of $\mathcal{I}$, and an *interpretation function* $\cdot^{\mathcal{I}}$, that assigns to each $A \in \mathsf{N_C}$ a subset $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, to each $o \in \mathsf{N_O}$ a singleton subset $o^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, $\lVert o^{\mathcal{I}} \rVert = 1$, to each $R \in \mathsf{N_R}$ a binary relation $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$, and to each $a \in \mathsf{N_I}$ an element $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$. This assignment is extended to roles by $(r^-)^{\mathcal{I}} = \{\langle e, d\rangle \mid \langle d, e\rangle \in r^{\mathcal{I}}\}$ and to complex concepts as shown in Table 1. $\mathcal{I}$ *satisfies* an axiom α (written $\mathcal{I} \models \alpha$) if the corresponding condition in Table 1 holds. Given an ontology $\mathcal{O}$, $\mathcal{I}$ is a *model* of $\mathcal{O}$ (written $\mathcal{I} \models \mathcal{O}$) if $\mathcal{I} \models \alpha$ for all axioms $\alpha \in \mathcal{O}$; $\mathcal{O}$ *is consistent* if $\mathcal{O}$ has a model; and $\mathcal{O}$ *entails* an axiom α (written $\mathcal{O} \models \alpha$), if every model of $\mathcal{O}$ satisfies α. A role R is *functional* (in $\mathcal{O}$) if $\mathsf{func}(R) \in \mathcal{O}$ and *transitive* (in $\mathcal{O}$) if $\mathsf{tran}(R) \in \mathcal{O}$. For an ontology $\mathcal{O}$, let $\sqsubseteq^*_{\mathcal{H}}$ be the reflexive transitive closure of the *role hierarchy* $\mathcal{H} = \{R \sqsubseteq S \in \mathcal{O}\}$. If $R \sqsubseteq^*_{\mathcal{H}} S$, then we say that R is a *sub-role* of S and S is a *super-role* of R; a role R is *simple* (in $\mathcal{O}$) if it has no transitive sub-roles. If $\mathsf{func}(R) \in \mathcal{O}$, then R must be simple.

A $\mathcal{SHOIF}$ ontology $\mathcal{O}$ is *Horn* (Krötzsch, Rudolph, and Hitzler 2013), if for every $D(a) \in \mathcal{O}$ and $C \sqsubseteq D \in \mathcal{O}$, the concepts C and D satisfy the following grammar definitions:

$$C_{(i)} ::= \top \mid \bot \mid A \mid o \mid C_1 \sqcap C_2 \mid C_1 \sqcup C_2 \mid \exists R.C \quad (1)$$

$$D_{(i)} ::= \top \mid \bot \mid A \mid o \mid D_1 \sqcap D_2 \mid \exists R.D \mid \forall R.D \mid \neg C \quad (2)$$

(Horn) $\mathcal{ALCHOI}$ is the fragment of (Horn) $\mathcal{SHOIF}$ in which functionality and transitivity are disallowed.

A Horn $\mathcal{SHOIF}$ ontology $\mathcal{O}$ is in normalized form, if (1) for every $C(a) \in \mathcal{O}$, C is an atomic concept; (2) for every $C \sqsubseteq D \in \mathcal{O}$, the concepts C and D satisfy the following grammar:

$$C_{(i)} ::= \top \mid \bot \mid A \mid o \mid C_1 \sqcap C_2 \mid C_1 \sqcup C_2 \quad (3)$$

$$D_{(i)} ::= \top \mid \bot \mid A \mid o \mid \exists R.A \mid \forall R.A \mid \neg C \quad (4)$$

and (3) for every $C \sqsubseteq \forall R.A \in \mathcal{O}$ and every transitive sub-role T of R, there exists an atomic concept B that occurs only in $\{C \sqsubseteq \forall T.B, B \sqsubseteq \forall T.B, B \sqsubseteq A\} \subseteq \mathcal{O}$ and not in C or A. The normalized form can be obtained for any Horn $\mathcal{SHOIF}$ ontology using structural transformation and the known technique for eliminating transitivity (see, e.g., (Kazakov 2009)). We assume all Horn $\mathcal{SHOIF}$ ontologies to be normalized, which is w.l.o.g.

For an ontology $\mathcal{O}$, we say that $\mathcal{O}$ is *concept-materialized* if $\mathcal{O} \models A(a)$ implies $A(a) \in \mathcal{O}$ for each $A \in \mathsf{con}(\mathcal{O})$ and $a \in \mathsf{ind}(\mathcal{O})$; $\mathcal{O}$ is *role-materialized* if $\mathcal{O} \models r(a, b)$ implies $r(a, b) \in \mathcal{O}$ for each $r \in \mathsf{rol}(\mathcal{O})$ and $a, b \in \mathsf{ind}(\mathcal{O})$; $\mathcal{O}$ is *equality-materialized* if $\mathcal{O} \models a \approx b$ implies $a \approx b \in \mathcal{O}$ for each $a, b \in \mathsf{ind}(\mathcal{O})$; $\mathcal{O}$ is (fully) *materialized* if it is concept-, role-, and equality-materialized. Given an ontology $\mathcal{O}$, the concept-, role-, equality-, and/or (full) materialization of $\mathcal{O}$ is the smallest super-set of $\mathcal{O}$ that is concept-, role-, equality-, and/or fully materialized respectively.

Computing Materialization by Abstraction

The main idea of the abstraction refinement method is to materialize an ontology $\mathcal{O} = \mathcal{A} \cup \mathcal{T}$ with a large ABox $\mathcal{A}$ by constructing a smaller ABox $\mathcal{B}$ such that the materialization of $\mathcal{O}$ is obtained from the materialization of $\mathcal{B} \cup \mathcal{T}$ by transferring entailments to $\mathcal{O}$. The ABox $\mathcal{B}$ is usually called the *abstraction* of the *original* ABox $\mathcal{A}$ (or just the *abstract ABox*), and the individuals in $\mathcal{B}$ are called *representatives* of the *original* individuals in $\mathcal{A}$. All results in this section apply to any DL with (classical) set-theoretic semantics, e.g., $\mathcal{SROIQ}$ (Horrocks, Kutz, and Sattler 2006).

Definition 1. Let $\mathcal{A}$ and $\mathcal{B}$ be ABoxes. A mapping $h : \mathsf{ind}(\mathcal{B}) \rightarrow \mathsf{ind}(\mathcal{A})$ is called a *homomorphism (from $\mathcal{B}$ to $\mathcal{A}$)* if, for every assertion $\alpha \in \mathcal{B}$, we have $h(\alpha) \in \mathcal{A}$, where $h(C(a)) := C(h(a))$, $h(R(a, b)) := R(h(a), h(b))$, and $h(a \approx b) := h(a) \approx h(b)$. We say an individual $b \in \mathsf{ind}(\mathcal{B})$ is a *representative* of an individual $a \in \mathsf{ind}(\mathcal{A})$ if there exists a homomorphism $h : \mathsf{ind}(\mathcal{B}) \rightarrow \mathsf{ind}(\mathcal{A})$ such that $h(b) = a$.

Example 1. Consider the ABoxes $\mathcal{A} = \{A(a),\ A(b)\}$ and $\mathcal{B} = \{A(u)\}$. Then the individual u of $\mathcal{B}$ is a representative for both individuals a and b.

The following property of homomorphisms allows transferring entailments from abstractions to original ABoxes.

Lemma 1. *Let $h\colon \mathsf{ind}(\mathcal{B}) \to \mathsf{ind}(\mathcal{A})$ be a homomorphism between the ABoxes $\mathcal{B}$ and $\mathcal{A}$. Then, for every TBox $\mathcal{T}$ and every axiom β, $\mathcal{B} \cup \mathcal{T} \models \beta$ implies $\mathcal{A} \cup \mathcal{T} \models h(\beta)$.*

Corollary 2. *If an individual $u \in \mathsf{ind}(\mathcal{B})$ is a representative for an individual $a \in \mathsf{ind}(\mathcal{A})$, then, for every TBox $\mathcal{T}$ and concept C, if $\mathcal{B} \cup \mathcal{T} \models C(u)$, then $\mathcal{A} \cup \mathcal{T} \models C(a)$.*

According to Corollary 2, in Example 1 one can *transfer* any entailed concept assertion for u to the corresponding assertions for a and b. In fact, in this particular case, all entailed concept assertions for $\mathcal{A}$ can be computed this way because there is also a homomorphism from $\mathcal{A}$ to $\mathcal{B}$.

Example 2 (Example 1 continued)**.** Consider the homomorphism $h : \mathsf{ind}(\mathcal{A}) \to \mathsf{ind}(\mathcal{B})$ defined by $h = \{a \mapsto u,\ b \mapsto u\}$. Then by Lemma 1, for every TBox $\mathcal{T}$ and concept C if $\mathcal{A} \cup \mathcal{T} \models C(a)$ or $\mathcal{A} \cup \mathcal{T} \models C(b)$ then $\mathcal{B} \cup \mathcal{T} \models C(u)$.

In practice, computing a sufficiently small abstraction $\mathcal{B}$ of $\mathcal{A}$ such that there are homomorphisms in both directions is rarely possible, so the set of concept assertions transferred using Corollary 2 is usually incomplete. To ensure completeness, one can employ further *refinement* steps that recompute the abstraction based on the new information derived. This method was shown to be complete for concept materialization of Horn $\mathcal{ALCHOI}$ ontologies (Glimm et al. 2014b). The aim of this paper is to extend this approach to Horn $\mathcal{SHOIF}$. Before we go into further details of our extension, we first describe challenges that the new functionality and transitivity axioms pose for ontology materialization.

Full Materialization for Horn $\mathcal{SHOIF}$

It is easy to show using model-theoretic arguments that an $\mathcal{ALCHOI}$ ontology $\mathcal{O}$ without equality assertions entails an equality between individuals $a \approx b$ iff either $a = b$ or both a and b are instances of some nominal concept o occurring in $\mathcal{O}$. To compute such entailed equality assertions, it is sufficient to compute instances of nominals, which can be accomplished by introducing an axiom $o \sqsubseteq A_o$ with a fresh concept A_o for each nominal o and computing instances of A_o. If $\mathcal{O}$ contains equality assertions, one needs to additionally perform the transitive symmetric closure of the resulting equality assertions. For role-materialization, similarly, one can show that if $\mathcal{O} \models R(a, b)$ then either (i) there exists $R'(a', b') \in \mathcal{O}$ such that $\mathcal{O} \models a \approx a'$, $\mathcal{O} \models b \approx b'$, and $R' \sqsubseteq^*_{\mathcal{H}} R$, or (ii) a is an instance of $\exists R.o$ and b is an instance of o for some nominal o, or (iii) a is an instance of o and b is an instance of $\exists R^-.o$ for some nominal o. All these conditions can be checked by introducing fresh concepts and computing the concept-materialization.

That is, (full) materialization of Horn $\mathcal{ALCHOI}$ ontologies can be reduced to concept-materialization by syntactic transformations. The following examples illustrate that for Horn $\mathcal{SHOIF}$ ontologies such reductions do not work.

Example 3. Consider the ontology $\mathcal{O} = \mathcal{A} \cup \mathcal{T}$ with $\mathcal{A} = \{A(a),\ A(b)\}$ and $\mathcal{T} = \{A \sqsubseteq \exists F^-.o, \mathsf{func}(F)\}$. Then $\mathcal{O} \models a \approx b$ but neither a nor b are instances of the nominal o.

As Example 3 illustrates, equality testing in (Horn) $\mathcal{SHOIF}$ becomes less trivial; the main reason is that using a combination of functional roles, inverse roles, and nominals one can express *entailed nominal concepts* such as A in Example 3, which can be interpreted by at most one element.

In the following example, we demonstrate how functional roles can also result in some non-trivial entailments of role assertions, even if no equality or nominals are used.

Example 4. Consider the ontology $\mathcal{O} = \mathcal{A} \cup \mathcal{T}$ with $\mathcal{A} = \{A(a), R(a, b)\}$ and $\mathcal{T} = \{A \sqsubseteq \exists S.\top,\ R \sqsubseteq F,\ S \sqsubseteq F,\ \mathsf{func}(F)\}$. Then $\mathcal{O} \models S(a, b)$, but $\mathcal{O} \not\models R \sqsubseteq S$.

As can be seen from Examples 3 and 4, the computation of equality- and role-materialization becomes a non-trivial problem for Horn $\mathcal{SHOIF}$ ontologies. Fortunately, using the following corollary of Lemma 1, one can extend the main idea behind concept-materialization described in the previous section to equality- and role-materialization.

Corollary 3. *Let $h : \mathsf{ind}(\mathcal{B}) \to \mathsf{ind}(\mathcal{A})$ be a homomorphism between the ABoxes $\mathcal{B}$ and $\mathcal{A}$, $u, v \in \mathsf{ind}(\mathcal{B})$, $a = h(u), b = h(v) \in \mathsf{ind}(\mathcal{A})$, and $\mathcal{T}$ a $\mathcal{SROIQ}$ TBox. Then $\mathcal{B} \cup \mathcal{T} \models u \approx v$ implies $\mathcal{A} \cup \mathcal{T} \models a \approx b$, and $\mathcal{B} \cup \mathcal{T} \models R(u, v)$ implies $\mathcal{A} \cup \mathcal{T} \models R(a, b)$, for every role R.*

Unfortunately, the abstract ABoxes that are sufficient to guarantee completeness of concept-materialization are not sufficient to guarantee completeness of equality- and role-materialization as demonstrated in the following example.

Example 5. Consider the ABox $\mathcal{A}$ and its abstraction $\mathcal{B}$ in Example 1. As stated in Example 2, for any TBox $\mathcal{T}$ all entailed concept assertions of $\mathcal{T} \cup \mathcal{A}$ can be obtained using Corollary 2 for the abstraction $\mathcal{B}$. However, the abstraction $\mathcal{B}$ may be insufficient for computing all entailed role or equality assertions using Corollary 3. Indeed, consider $\mathcal{T} = \{A \sqsubseteq o\}$. Then $\mathcal{A} \cup \mathcal{T} \models a \approx b$, but, clearly, there is no homomorphism $h : \mathsf{ind}(\mathcal{B}) \to \mathsf{ind}(\mathcal{A})$ such that $h(u) = a$ and $h(u) = b$ required to derive this assertion using Corollary 3. Similarly, it is possible that $\mathcal{A} \cup \mathcal{T} \models R(a, b)$ for some role R, but we are not able to derive this assertion using Corollary 3, for example, for $\mathcal{T} = \{A \sqsubseteq \exists T.o,\ A \sqsubseteq \exists T^-.o,\ \mathsf{tran}(T),\ T \sqsubseteq R\}$.

Abstraction Refinement for Horn $\mathcal{SHOIF}$

The general algorithm for ontology reasoning using the abstraction refinement method can be summarized as follows:

1. Build a suitable abstraction of the original ontology;
2. Compute the entailments from the abstraction using a reasoner and transfer them to the original ontology using homomorphisms (Lemma 1);
3. Compute the deductive closure of the original ontology using some (light-weight) rules;
4. Repeat from Step 1 until no new entailments can be added to the original ontology.

The efficiency and theoretical properties of this method depend on the choices of how the abstraction is computed

in Step 1, which entailments are transferred in Step 2, and which rules are used to compute the deductive closure in Step 3. In the following we detail these choices.

To compute the abstraction of the original ABox (Step 1), we define *types* of individuals based on their assertions.

Definition 2. Let $\mathcal{A}$ be an ABox and a an individual. The *concept type* of a is a set of concepts $\tau_C(a) = \{C \mid C(a) \in \mathcal{A}\}$. The *role type* of a is a set of roles $\tau_R(a) = \{R \mid \exists b : R(a,b) \in \mathcal{A}\}$. The *(combined) type* of a is the pair $\tau(a) = \langle \tau_C(a), \tau_R(a) \rangle$ where $\tau_C(a)$ is the concept type of a and $\tau_R(a)$ is the role type of a.

Note that we assume w.l.o.g. that, for every individual a, $\top(a)$ is included in the ontology and, therefore, $\top$ occurs in every (concept) type. To simplify the presentation, however, we usually omit $\top$ in the remainder.

Example 6. Let $\mathcal{A} = \{A(a), A(b), R(a,b)\}$. Then $\tau_C(a) = \tau_C(b) = \tau_1 = \{A\}$, $\tau_R(a) = \{R\}$, $\tau_R(b) = \{R^-\}$, $\tau(a) = \tau_2 = \langle \{A\}, \{R\} \rangle$, and $\tau(b) = \tau_3 = \langle \{A\}, \{R^-\} \rangle$.

The abstract ABox is then constructed by choosing one representative for each type with the respective assertions.

Definition 3. The *abstraction* of an ABox $\mathcal{A}$ is an ABox $\mathcal{B} = \bigcup_{a \in \mathsf{ind}(\mathcal{A})} (\mathcal{B}_{\tau_C(a)} \cup \mathcal{B}_{\tau(a)})$, where:

- for each concept type τ_C, $\mathcal{B}_{\tau_C} = \{C(u_{\tau_C}) \mid C \in \tau_C\}$,
- for each combined type $\tau = \langle \tau_C, \tau_R \rangle$, $\mathcal{B}_\tau = \{C(v_\tau) \mid C \in \tau_C\} \cup \{R(v_\tau, w_\tau^R) \mid R \in \tau_R\}$,

where u_{τ_C}, v_τ, and w_τ^R are distinguished *abstract individuals* for each concept type τ_C and each combined type τ.

Example 7. The abstraction for $\mathcal{A}$ in Example 6 is the ABox $\mathcal{B} = \mathcal{B}_{\tau_C(a)} \cup \mathcal{B}_{\tau_C(b)} \cup \mathcal{B}_{\tau(a)} \cup \mathcal{B}_{\tau(b)}$, where $\mathcal{B}_{\tau_C(a)} = \mathcal{B}_{\tau_C(b)} = \mathcal{B}_{\tau_1} = \{A(u_{\tau_1})\}$, $\mathcal{B}_{\tau(a)} = \mathcal{B}_{\tau_2} = \{A(v_{\tau_2}), R(v_{\tau_2}, w_{\tau_2}^R)\}$, $\mathcal{B}_{\tau(b)} = \mathcal{B}_{\tau_3} = \{A(v_{\tau_3}), R^-(v_{\tau_3}, w_{\tau_3}^{R^-})\}$.

Intuitively, the abstraction is a disjoint union of ABoxes simulating concept and combined types. Note that each mapping $h : \mathsf{ind}(\mathcal{B}) \to \mathsf{ind}(\mathcal{A})$ such that:

$$h(u_{\tau_C}) \in \{a \in \mathsf{ind}(\mathcal{A}) \mid \tau_C(a) = \tau_C\}, \tag{5}$$

$$h(v_\tau) \in \{a \in \mathsf{ind}(\mathcal{A}) \mid \tau(a) = \tau\}, \tag{6}$$

$$h(w_\tau^R) \in \{b \in \mathsf{ind}(\mathcal{A}) \mid R(h(v_\tau), b) \in \mathcal{A}\}, \tag{7}$$

is a homomorphism from $\mathcal{B}$ to $\mathcal{A}$. This allows us to transfer entailments back to the original ABox using Corollaries 2 and 3. Note that each original individual a in $\mathcal{A}$ has at least two representatives in $\mathcal{B}$: $u_{\tau_C(a)}$, which has exactly the same concept assertions as a, and $v_{\tau(a)}$, which additionally has assertions with the same roles. The use of two representatives distinguishes the abstractions from the previously introduced ones (Glimm et al. 2014b) and solves the problem with role and equality assertions in Example 5.

Example 8. Consider the ABox $\mathcal{A} = \{A(a), A(b)\}$ and TBox $\mathcal{T} = \{A \sqsubseteq o\}$ mentioned in Example 5. We have $\tau_C(a) = \tau_C(b) = \tau_1 = \{A\}$, and $\tau(a) = \tau(b) = \tau_2 = \langle \{A\}, \emptyset \rangle$. The abstraction of $\mathcal{A}$ is defined as $\mathcal{B} = \mathcal{B}_{\tau_1} \cup \mathcal{B}_{\tau_2}$ with $\mathcal{B}_{\tau_1} = \{A(u_{\tau_1})\}$, $\mathcal{B}_{\tau_2} = \{A(v_{\tau_2})\}$. Since $\mathcal{B} \cup \mathcal{T} \models u_{\tau_1} \approx v_{\tau_2}$, and $h = \{u_{\tau_1} \mapsto a,\ v_{\tau_2} \mapsto b\}$ is a homomorphism from $\mathcal{B}$ to $\mathcal{A}$, using Corollary 3 we obtain $\mathcal{A} \cup \mathcal{T} \models a \approx b$.

Next, we detail which entailed assertions are transferred from $\mathcal{B}$ to $\mathcal{A}$ in Step 2 of the algorithm. To achieve completeness it is not necessary to transfer all of them.

Definition 4. Let $\mathcal{B}$ be the abstraction of an ABox $\mathcal{A}$ (by Definition 3), and $\Delta\mathcal{B}$ a set of assertions. The *update of $\mathcal{A}$ (using $\Delta\mathcal{B}$)* is the smallest set of assertions $\Delta\mathcal{A}$ such that:

$$C(v_{\tau(a)}) \in \Delta\mathcal{B} \Rightarrow C(a) \in \Delta\mathcal{A}, \tag{8}$$

$$C(w_{\tau(a)}^R) \in \Delta\mathcal{B},\ R(a,b) \in \mathcal{A} \Rightarrow C(b) \in \Delta\mathcal{A}, \tag{9}$$

$$S(v_{\tau(a)}, w_{\tau(a)}^R) \in \Delta\mathcal{B}, R(a,b) \in \mathcal{A} \Rightarrow S(a,b) \in \Delta\mathcal{A}, \tag{10}$$

$$S(v_{\tau(a)}, v_{\tau(a)}) \in \Delta\mathcal{B} \Rightarrow S(a,a) \in \Delta\mathcal{A} \tag{11}$$

$$u_{\tau_C(a)} \approx v_{\tau(b)} \in \Delta\mathcal{B} \Rightarrow a \approx b \in \Delta\mathcal{A}, \tag{12}$$

$$R(u_{\tau_C(a)}, v_{\tau(b)}) \in \Delta\mathcal{B} \Rightarrow R(a,b) \in \Delta\mathcal{A}. \tag{13}$$

The following lemma can be established using homomorphisms from $\mathcal{B}$ to $\mathcal{A}$ satisfying conditions (5)–(7).

Lemma 4. *Let $\mathcal{B}$ be the abstraction of $\mathcal{A}$, $\Delta\mathcal{A}$ an update for $\Delta\mathcal{B}$, and $\mathcal{T}$ a TBox. Then $\mathcal{B} \cup \mathcal{T} \models \Delta\mathcal{B}$ implies $\mathcal{A} \cup \mathcal{T} \models \Delta\mathcal{A}$.*

Proof. It is easy to see that for each $\alpha \in \Delta\mathcal{A}$ and $\beta \in \Delta\mathcal{B}$ in cases (8)–(13), there is a homomorphism h from $\mathcal{B}$ to $\mathcal{A}$ satisfying conditions (5)–(7) such that $h(\beta) = \alpha$. Hence, by Lemma 1, $\mathcal{B} \cup \mathcal{T} \models \beta$ implies $\mathcal{A} \cup \mathcal{T} \models h(\beta) = \alpha$. □

After transferring the entailed assertions according to Definition 4, in Step 3 we compute the closure of the ABox $\mathcal{A}$ under equality, transitivity, and functionality.

Definition 5. We say that an ABox $\mathcal{A}$ is *equality-closed* if:

$$a \in \mathsf{ind}(\mathcal{A}) \text{ implies } a \approx a \in \mathcal{A} \tag{14}$$

$$\{a \approx b, b \approx c\} \subseteq \mathcal{A} \text{ implies } a \approx c \in \mathcal{A}, \tag{15}$$

$$\{a \approx b, A(a)\} \subseteq \mathcal{A} \text{ implies } A(b) \in \mathcal{A}, \tag{16}$$

$$\{a \approx b, R(a,c)\} \subseteq \mathcal{A} \text{ implies } R(b,c) \in \mathcal{A}, \tag{17}$$

$\mathcal{A}$ is *closed under the axiom* $\mathsf{tran}(T)$ if:

$$\{T(a,b), T(b,c)\} \subseteq \mathcal{A} \text{ implies } T(a,c) \in \mathcal{A}. \tag{18}$$

$\mathcal{A}$ is *closed under the axiom* $\mathsf{func}(F)$ if:

$$\{F(a,b), F(a,c)\} \subseteq \mathcal{A} \text{ implies } b \approx c \in \mathcal{A}. \tag{19}$$

The *closure of $\mathcal{A}$ (w.r.t. a TBox $\mathcal{T}$) under equality, transitivity, and/or functionality* is the smallest super-set of $\mathcal{A}$ that is closed under equality, for each $\mathsf{tran}(T) \in \mathcal{T}$ and/or each $\mathsf{func}(F) \in \mathcal{T}$, respectively.

Computing the closure of an ABox under equality, functionality, and transitivity is a relatively lightweight operation that does not require using a DL reasoner. Note that all these assertions must be derived in order to compute the full materialization. The previous method (Glimm et al. 2014b) does not involve the computation of the closure as in Step 3. One can easily check that this and the lack of the additional individuals for the concept types results in incompleteness even for concept-materialization of Horn $\mathcal{SHOIF}$ ontologies.

Steps 2 and 3 can only extend the original ABox with entailed atomic assertions, the procedure is, therefore, *sound*.

Since the number of such assertions is bounded by the size of the materialization, the procedure eventually terminates and the number of repeat loops is polynomial in the size of the ontology. We next show that the procedure is *complete*.

A Criterion for Ontology Materialization

To prove completeness of the abstraction refinement procedure in the case of Horn $\mathcal{SHOIF}$, we characterize when such ontologies are fully materialized by means of closure of the ABox assertions under certain rules. The rules are similar to the rules for reasoning in Horn $\mathcal{SHOIQ}$ (Ortiz, Rudolph, and Simkus 2010) in the sense that they derive logical consequences of axioms. Since we are only interested in ABox consequences and not going to use these rules for computing the materialization (but merely for proving completeness of the algorithm), however, we will not derive TBox axioms explicitly, but use their entailments in side conditions of the rules.

Recall from the discussion after Example 3, that in $\mathcal{SHOIF}$ one can express some non-trivial *nominal concepts*. We extend the language with new axiom types.

Definition 6. A *concept cardinality restriction* is an axiom of the form $|C| \geq n$ or $|C| = n$ with C a concept and $n \in \mathbb{N}$. An interpretation $\mathcal{I}$ satisfies $|C| \geq n$ ($|C| = n$), written $\mathcal{I} \models |C| \geq n$ ($\mathcal{I} \models |C| = n$), iff $|C^{\mathcal{I}}| \geq n$ ($|C^{\mathcal{I}}| = n$).

We also use *role conjunctions* $R \sqcap S$, interpreted by $(R \sqcap S)^{\mathcal{I}} = R^{\mathcal{I}} \cap S^{\mathcal{I}}$. The new constructors and axioms are used only in the conditions of rules and not in the ontology.

In the following, we denote by N and M conjunctions of atomic concepts and/or nominals and by H conjunctions of roles (possibly with superscripts). If we write $C \in N$, we treat N as the corresponding set of conjuncts, where $N = \top$ denotes the empty conjunction. For a role conjunction H we denote by H^- the conjunction of inverses of roles occurring in H. We write $N(a) \in \mathcal{A}$ if $C(a) \in \mathcal{A}$ for every $C \in N$. Note that if $N(a) \in \mathcal{A}$ then $\mathcal{A} \models |N| \geq 1$.

Definition 7. Let $\mathcal{A}$ be an ABox. Then $\mathfrak{N}(\mathcal{A}) = \{|N| \geq 1 \mid N(a) \in \mathcal{A}\}$ is the *set of cardinality axioms induced by* $\mathcal{A}$.

The materialization rules are presented in Table 2. These rules are complete for ontology materialization.

Theorem 5. *Let $\mathcal{O} = \mathcal{A} \cup \mathcal{T}$ be a normalized Horn $\mathcal{SHOIF}$ ontology and $\mathcal{T}' = \mathcal{T} \cup \mathfrak{N}(\mathcal{A})$. Then $\mathcal{A}$ is closed under the rules in Table 2 w.r.t. $\mathcal{T}$ iff $\mathcal{O}$ is fully materialized.*

Proof. The *if* direction is trivial since all rules derive logical consequences of the axioms in premises and side conditions. For the *only if* direction, if $\mathcal{T}'$ is inconsistent then we can derive all assertions using rules $\mathbf{R}^2_{\sqsubseteq}$, $\mathbf{R}_{\exists}$, and $\mathbf{R}_{||}$. And if $\mathsf{ind}(\mathcal{O}) = \emptyset$, then Theorem 5 trivially holds. From now on we assume $\mathsf{ind}(\mathcal{O}) \neq \emptyset$ and $\mathcal{T}'$ is consistent. We then construct a model $\mathcal{J}$ of $\mathcal{O}$ such that $\mathcal{J} \models \alpha$ implies $\alpha \in \mathcal{A}$, for every atomic assertion α. Then, $\mathcal{O} \models \alpha$ implies $\mathcal{J} \models \alpha$, which implies $\alpha \in \mathcal{A}$, i.e., $\mathcal{O}$ is materialized. Such model $\mathcal{J}$ is obtained in two steps: 1) construct an interpretation $\mathcal{I}$ satisfying all but transitivity axioms in $\mathcal{O}$; 2) obtain $\mathcal{J}$ from $\mathcal{I}$ by extending the interpretation of non-simple roles to satisfy transitivity axioms.

Intuitively, $\mathcal{I}$ has a forest-like structure consisting of a graph part and tree parts. The graph part contains domain elements for individuals and concepts that have exactly one instance. The tree parts grow from the graph part to satisfy entailed existential axioms. To construct $\mathcal{I}$, we use a *chase-like* technique (Abiteboul, Hull, and Vianu 1995) in which new domain elements are introduced to satisfy axioms of the form $M \sqsubseteq \exists H.N$ entailed by $\mathcal{O}$ such that H and N are maximal (w.r.t M), i.e., there is no $M \sqsubseteq \exists H'.N'$ entailed by $\mathcal{O}$ and $H \cup N \subsetneq H' \cup N'$. This guarantees that new domain elements (and their respective assertions) also satisfy universal restrictions and role inclusions. Additionally, a new domain element is introduced only when an axiom cannot be satisfied by the existing domain elements. This ensures all the existential axioms are satisfied while the functionality axioms are not violated.

For the conjunction of atomic concepts and/or nominals N, we say N is *singleton* w.r.t. $\mathcal{T}'$ if $\mathcal{T}' \models |N| = 1$. A singleton N is *maximal* if $N \subsetneq N'$ for no other singleton N'. A maximal singleton N is *anonymous* if there exists no M s.t. $\mathcal{T}' \models M \sqsubseteq N$ and $|M| \geq 1 \in \mathfrak{N}(\mathcal{A})$. Let $[a] = \{b \mid a \approx b \in \mathcal{A}\}$, we define domain elements of the graph part as the set $\mathbb{G}(\mathcal{O}) = \{[a] \mid a \in \mathsf{ind}(\mathcal{O})\} \cup \{N \mid N$ is an anonymous maximal singleton w.r.t. $\mathcal{T}'\}$.

Let Γ be the set of all role conjunctions, $\Sigma = \{N \mid \mathcal{T}' \models |N| \geq 1\}$, and, for each word $w \in \mathbb{G}(\mathcal{O}) \times (\Gamma \times \Sigma)^*$, we denote the *conjunction type* of w as $\mathbf{ct}(w) = N$ if $w = [a], N(a) \in \mathcal{A}$ and N is maximal; $w = N \in \mathbb{G}(\mathcal{O})$; or $w = v.H.N$. Then, $\Delta^{\mathcal{I}}$ is defined constructively s.t. $\mathbb{G}(\mathcal{O}) \subseteq \Delta^{\mathcal{I}}$ and $w.H.N \in \Delta^{\mathcal{I}}$, if $w \in \Delta^{\mathcal{I}}, \mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \exists H.N$ with $H \in \Gamma$, $N \in \Sigma$ and the following conditions hold:

1. N is not singleton,
2. there exist no a and b s.t. $w = [a], N(b) \in \mathcal{A}$, and $R(a, b) \in \mathcal{A}$ for every $R \in H$,
3. if $\mathbf{ct}(w)$ is singleton, then there exist no $M \in \Sigma$ and H' s.t. $H^- \subseteq H'$ and $\mathcal{T}' \models \{M \sqsubseteq \exists H'.\mathbf{ct}(w), M \sqsubseteq N\}$,
4. there exist no $v \in \Delta^{\mathcal{I}}$, H', and N' s.t. $w = v.H'.N'$, $H^- \subseteq H'$, and $\mathcal{T}' \models \{\mathbf{ct}(v) \sqsubseteq N\}$,
5. there exist no H' and N' s.t. $H \cup N \subsetneq H' \cup N'$ and $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \exists H'.N'$.

Since $\mathsf{ind}(\mathcal{O}) \neq \emptyset$, we have $\Delta^{\mathcal{I}} \neq \emptyset$. The interpretation function $\cdot^{\mathcal{I}}$ is defined as follows:

$$
\begin{aligned}
a^{\mathcal{I}} &= [a], \text{ for } a \in \mathsf{ind}(\mathcal{O}) \\
C^{\mathcal{I}} &= \{w \in \Delta^{\mathcal{I}} \mid \mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq C\}, \\
&\quad \text{for } C \in \mathsf{con}(\mathcal{O}) \cup \mathsf{nom}(\mathcal{O}) \\
r^{\mathcal{I}} &= \{([a], [b]) \mid r(a, b) \in \mathcal{A}\} \ (i.1) \\
&\quad \cup \{(w, v) \mid \mathcal{T}' \models \{\mathbf{ct}(w) \sqsubseteq \exists r.\mathbf{ct}(v), |\mathbf{ct}(v)| = 1\}\} \ (i.2) \\
&\quad \cup \{(v, w) \mid \mathcal{T}' \models \{\mathbf{ct}(w) \sqsubseteq \exists r^-.\mathbf{ct}(v), |\mathbf{ct}(v)| = 1\}\} (i.3) \\
&\quad \cup \{(w, w.H.N) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \mid r \in H\} \ (i.4) \\
&\quad \cup \{(w.H.N, w) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \mid r^- \in H\} \ (i.5) \\
&\quad \text{for } r \in \mathsf{rol}(\mathcal{O}), \text{ where } w, v \in \Delta^{\mathcal{I}}
\end{aligned}
$$

Our goal is to construct the interpretation $\mathcal{I}$ that entails all except transitivity axioms in $\mathcal{A} \cup \mathcal{T}$. To show $\mathcal{I}$ is such an

$$\mathbf{R}^1_{\approx}\ \frac{}{a \approx a} : a \in \mathsf{ind}(\mathcal{A}) \qquad \mathbf{R}^2_{\approx}\ \frac{a \approx b \quad b \approx c}{a \approx c} \qquad \mathbf{R}^3_{\approx}\ \frac{a \approx b \quad R(a,c)}{R(b,c)} \qquad \mathbf{R}^4_{\approx}\ \frac{a \approx b \quad A(a)}{A(b)}$$

$$\mathbf{R}^1_t\ \frac{T(a,b) \quad T(b,c)}{T(a,c)} : \mathsf{tran}(T) \in \mathcal{T} \quad \mathbf{R}^2_t\ \frac{N(a) \quad M(b)}{T(a,b)} : \mathcal{T}' \models \{N \sqsubseteq \exists T.N', M \sqsubseteq \exists T^-.N', |N'| = 1\}, \mathsf{tran}(T) \in \mathcal{T}$$

$$\mathbf{R}^3_t\ \frac{N(a)}{T(a,a)} : \mathcal{T}' \models N \sqsubseteq \exists (T \sqcap T^-).\top, \mathsf{tran}(T) \in \mathcal{T} \qquad \mathbf{R}_{\exists}\ \frac{N(a) \quad M(b)}{R(a,b)} : \mathcal{T}' \models \{N \sqsubseteq \exists R.M,\ |M| = 1\}$$

$$\mathbf{R}^1_{\leq}\ \frac{F(a,b) \quad F(a,c)}{b \approx c} : \mathsf{func}(F) \in \mathcal{T} \quad \mathbf{R}_{||}\ \frac{N(a) \quad N(b)}{a \approx b} : \mathcal{T}' \models |N| = 1 \quad \mathbf{R}_{\forall}\ \frac{N(a) \quad R(a,b)}{B(b)} : \mathcal{T}' \models N \sqsubseteq \forall R.B$$

$$\mathbf{R}^2_{\leq}\ \frac{M(a) \quad F(a,b)}{S(a,b) \quad A(b)} : \mathcal{T}' \models M \sqsubseteq \exists (F \sqcap S).A, \mathsf{func}(F) \in \mathcal{T} \quad \mathbf{R}^1_{\sqsubseteq}\ \frac{R(a,b)}{S(a,b)} : R \sqsubseteq S \in \mathcal{T} \quad \mathbf{R}^2_{\sqsubseteq}\ \frac{N(a)}{B(a)} : \mathcal{T}' \models N \sqsubseteq B$$

Table 2: Materialization rules

interpretation, we first prove some auxiliary lemmas. The following lemma follows directly from the definition of $\mathcal{I}$ and the assumption that $\mathcal{A}$ is closed under rules $\mathbf{R}^i_{\approx}, 1 \leq i \leq 4$.

Lemma 6. *Let N be a set of atomic concepts and/or nominals, R a role, and a, b individuals. Then, we have:*

1. $[a] \in N^{\mathcal{I}}$ *iff* $N(a) \subseteq \mathcal{A}$ *and*
2. $([a],[b]) \in R^{\mathcal{I}}$ *iff* $R(a,b) \in \mathcal{A}$.

Under the interpretation $\mathcal{I}$, only non-empty concepts, e.g. N s.t. $\mathcal{T}' \models |N| \geq 1$, have instances.

Lemma 7. *Let N be a set of atomic concepts and/or nominals, and d a domain element. Then, $d \in N^{\mathcal{I}}$ implies $\mathcal{T}' \models |N| \geq 1$.*

Proof. We consider all forms of d in $\Delta^{\mathcal{I}}$.

Case $d = [a]$. By Lemma 6, we have $N(a) \in \mathcal{A}$, which implies $\mathcal{T}' \models |N| \geq 1$.

Case $d = N'$ for some anonymous maximal singleton N'. By definition of $\mathcal{I}$, $\mathcal{T}' \models N' \sqsubseteq N$. Therefore $\mathcal{T}' \models |N| \geq 1$.

Case $d = w.H.N'$. Since $w.H.N' \in \Delta^{\mathcal{I}}$, we have $N' \in \Sigma$, which implies $\mathcal{T}' \models |N'| \geq 1$.

□

We prove the following properties of concept C, D in (3), (4) respectively. These properties allow us to show $\mathcal{I} \models C \sqsubseteq D$ for every axiom $C \sqsubseteq D \in \mathcal{O}$.

Lemma 8. *If $d \in \Delta^{\mathcal{I}}$ and C is a concept of the form* (3) *then $d \in C^{\mathcal{I}}$ iff $\models \mathbf{ct}(d) \sqsubseteq C$.*

Proof. Case $C = \top$ is trivial and cases $C = A$, $C = o$ directly follow from the definition of $\mathcal{I}$. Case $C = \bot$ holds as there is no such $d \in \Delta^{\mathcal{I}}$ (by Lemma 7). For the case $C = C_1 \sqcap C_2$. We have $d \in (C_1 \sqcap C_2)^{\mathcal{I}}$ iff $d \in C_1^{\mathcal{I}}$ and $w \in C_2^{\mathcal{I}}$. By induction hypothesis, this is the case iff $\mathcal{T}' \models \mathbf{ct}(d) \sqsubseteq C_1, \mathbf{ct}(d) \sqsubseteq C_2$, i.e. $\mathcal{T}' \models \mathbf{ct}(d) \models C_1 \sqcap C_2$. Case $C = C_1 \sqcup C_2$ is analogous. □

Lemma 9. *If $d \in \Delta^{\mathcal{I}}$, D is a concept of the form* (4)*, and $\mathcal{T}' \models \mathbf{ct}(d) \sqsubseteq D$ then $d \in D^{\mathcal{I}}$.*

Proof. Case $D = \top$ is trivial and cases $D = A$, $D = o$ directly follow from the definition of $\mathcal{I}$. Case $D = \bot$ holds as there is no such $d \in \Delta^{\mathcal{I}}$ (by Lemma 7).

Case $D = \exists R.A$. We consider all forms of d.

- $d = [a]$. If A is singleton and $A(b) \in \mathcal{A}$ for some individual b, then $R(a,b) \in \mathcal{A}$ as $\mathcal{A}$ is closed under $\mathbf{R}_{\exists}$; therefore $([a],[b]) \in R^{\mathcal{I}}$. Otherwise, the definition of $\mathcal{I}$ ensures that there exists a pair $([a], v)$ in either $(i.2)$, $(i.3)$, or $(i.4)$ such that $([a], v) \in R^{\mathcal{I}}$ and $v \in A^{\mathcal{I}}$.
- We argue analogously for the case $d = N$, where N is anonymous maximal singleton , and for the case $d = w.H.N$.

Case $D = \forall R.A$ with R is a role name; the case of inverse role name is analogous. We examine all cases for the pair $(d,e) \in R^{\mathcal{I}}$ and show that $e \in A^{\mathcal{I}}$.

- Case $(i.1)$: $(d,e) = ([a],[b])$. By Lemma 6, we have $R(a,b) \in \mathcal{A}$ and $M(a) \in \mathcal{A}$, with $M = \mathbf{ct}(d)$. As $\mathcal{A}$ is closed under the rule $\mathbf{R}_{\forall}$, we obtain $A(b) \in \mathcal{A}$, which implies $[b] \in A^{\mathcal{I}}$, i.e. $e \in A^{\mathcal{I}}$.
- Case $(i.2)$: $(d,e) = (w,v)$ with $\mathcal{T}' \models \{\mathbf{ct}(w) \sqsubseteq \exists R.\mathbf{ct}(v), |\mathbf{ct}(v)| = 1\}$. Since $\mathcal{T}' \models |\mathbf{ct}(w)| \geq 1$ and $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \forall R.A$, we obtain $\mathcal{T}' \models \mathbf{ct}(v) \sqsubseteq A$. By induction hypothesis, we have $v \in A^{\mathcal{I}}$, i.e. $e \in A^{\mathcal{I}}$.
- Case $(i.3)$: $(d,e) = (v,w)$ with $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \exists R^-.\mathbf{ct}(v)$. Since $\mathcal{T}' \models \mathbf{ct}(v) \sqsubseteq \forall R.A$, we obtain

$\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq A$. By induction hypothesis, we have $w \in A^{\mathcal{I}}$.

- Case $(i.4)$: $(d, e) = (w, w.H.N)$ with $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \exists H.N$ and $R \in H$. Since $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \forall R.A$, we obtain $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \exists H.(N \sqcap A)$. By maximality of N, we have $A \in N$, which implies $\mathcal{T}' \models \mathbf{ct}(w.H.N) \sqsubseteq A$. By induction hypothesis we obtain $w.H.N \in A^{\mathcal{I}}$, i.e. $e \in A^{\mathcal{I}}$.
- Case $(i.5)$: $(d, e) = (w.H.N, w)$ with $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \exists H.N$ and $R^- \in H$. Since $\mathcal{T}' \models N \sqsubseteq \forall R.A$, we obtain $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq A$. By induction hypothesis, we have $w \in A^{\mathcal{I}}$.

Case $D = \neg C$. By Lemma 8, we have $d \notin C^{\mathcal{I}}$, which implies $d \in (\neg C)^{\mathcal{I}}$.

□

We are now ready to show that $\mathcal{I}$ satisfies all but transitivity axioms in $\mathcal{O}$.

Lemma 10. *Let $\mathcal{T}^*$ be a TBox obtained from $\mathcal{T}$ by removing all transitivity axioms. Then, $\mathcal{I} \models \mathcal{A} \cup \mathcal{T}^*$.*

Proof. We show $\mathcal{I} \models \alpha$, for every axiom $\alpha \in \mathcal{A} \cup \mathcal{T}^*$.

Case $\alpha = A(a)$, $\alpha = R(a, b)$, and $\alpha = a \approx b$. $\mathcal{I} \models \alpha$ by the definition of $\mathcal{I}$.

Case $\alpha = C \sqsubseteq D$. Since $\mathcal{T}$ is normalized, C is of the form (3) and D is of the form (4). Let $w \in \Delta^{\mathcal{I}}$ such that $w \in C^{\mathcal{I}}$. Then by Lemma 8 we have $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq C$. Since $C \sqsubseteq D \in \mathcal{T} \subseteq \mathcal{T}'$, we have $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq D$. Hence by Lemma 9, $w \in D^{\mathcal{I}}$. Since $w \in \Delta^{\mathcal{I}}$ is arbitrary, we have proved that $\mathcal{I} \models C \sqsubseteq D$.

Case $\alpha = R \sqsubseteq S$, where R is a role name; the case of inverse of a role name is analogous. Let $(d, e) \in R^{\mathcal{I}}$, we show $(d, e) \in S^{\mathcal{I}}$.

- Case $(i.1)$: $(d, e) = ([a], [b])$. By Lemma 6 we have $R(a, b) \in \mathcal{A}$. As $\mathcal{A}$ is closed under $\mathbf{R}^1_{\sqsubseteq}$, we obtain $S(a, b) \in \mathcal{A}$, which implies $([a], [b]) \in S^{\mathcal{I}}$.
- Cases $(i.2)$: $(d, e) = (w, v)$ with $\mathcal{T}' \models \{\mathbf{ct}(w) \models \exists R.\mathbf{ct}(v), |\mathbf{ct}(v)| = 1\}$. Since $R \sqsubseteq S \in \mathcal{T}$, we have $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \exists S.\mathbf{ct}(v)$, which implies $(w, s) \in S^{\mathcal{I}}$.
- Case $(i.3)$ is analogous to Case $(i.2)$.
- Case $(i.4)$: $(d, e) = (w, w.H.N)$ with $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \exists H.N, R \in H$. Since $R \sqsubseteq S \in \mathcal{T}'$, we have $\mathcal{T} \models \mathbf{ct}(w) \sqsubseteq \exists (H \sqcap S).N$. By maximality of H, we have $S \in H$, which implies $(w, w.H.N) \in S^{\mathcal{I}}$.
- Case $(i.5)$ is similar to Case $(i.4)$.

Case $\alpha = \mathsf{func}(F)$ with F is a role name; the case of inverse role name is analogous. We consider all possible combination of $p_1 = (d, e), p_2 = (d, e') \in F^{\mathcal{I}}$ and show $e = e'$.

- Case $(i.1)$ and $(i.1)$: $p_1 = ([a], [b])$, $p_2 = ([a], [c])$. By Lemma 6 we have $F(a, b), F(a, c) \in \mathcal{A}$. As $\mathcal{A}$ is closed under $\mathbf{R}^1_{\leq}$, we have $b \approx c \in \mathcal{A}$, which implies $[b] = [c]$.
- Case $(i.1)$ and $(i.2)$: $p_1 = ([a], [b]), p_2 = ([a], v)$ with $\mathcal{T}' \models \{\mathbf{ct}(a) \sqsubseteq \exists F.\mathbf{ct}(v), |\mathbf{ct}(v)| = 1\}$. By Lemma 6 and by definition of $\mathcal{I}$ we have $F(a, b) \in \mathcal{A}$, $M(a) \subseteq \mathcal{A}$ with $M = \mathbf{ct}([a])$. Since $\mathcal{A}$ is closed under $\mathbf{R}^2_{\leq}$, we have $N(b) \in \mathcal{A}$, with $N = \mathbf{ct}(v)$. Since $\mathcal{T}' \models |N| = 1$, we have $v = [b]$.
- Case $(i.1)$ and $(i.3)$: $p_1 = ([a], [b]), p_2 = ([a], w)$ with $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \exists F^-.\mathbf{ct}(a), |\mathbf{ct}(a)| = 1$. Since $\mathsf{func}(F) \in \mathcal{T}'$, we obtain $\mathcal{T}' \models |\mathbf{ct}(w)| = 1$, which implies $\mathcal{T}' \models \mathbf{ct}(a) \sqsubseteq \exists F.\mathbf{ct}(w)$. Similar to case $(i.1)$ and $(i.2)$, we obtain $w = [b]$.
- Case $(i.1)$ and $(i.4)$: $p_1 = ([a], [b])$, $p_2 = ([a], [a].H.N)$ with $F(a, b) \in \mathcal{A}$ and $\mathcal{T}' \models \mathbf{ct}([a]) \sqsubseteq \exists H.N, F \in H$. Since $\mathcal{A}$ is closed under $\mathbf{R}^2_{\leq}$, we have $N(b) \in \mathcal{A}$. This violates the condition for $[a].H.N$ being introduced. Therefore, the combination of $(i.1)$ and $(i.4)$ cannot occur.
- Case $(i.1)$ and $(i.5)$, which is analogous to the case $(i.1)$ and $(i.4)$, cannot occur.
- Case $(i.2)$ and $(i.2)$: $p_1 = (w, v)$, $p_2 = (w, v')$, with $\mathcal{T}' \models \{\mathbf{ct}(w) \sqsubseteq \exists F.\mathbf{ct}(v), |\mathbf{ct}(v)| = 1\}$ and $\mathcal{T}' \models \{\mathbf{ct}(w) \sqsubseteq \exists F.\mathbf{ct}(v'), |\mathbf{ct}(v')| = 1\}$. Clearly $\mathcal{T}' \models \{\mathbf{ct}(v) \sqsubseteq \mathbf{ct}(v'), \mathbf{ct}(v') \sqsubseteq \mathbf{ct}(v)\}$. Since $\mathcal{T}' \models \{|\mathbf{ct}(v)| = 1, |\mathbf{ct}(v')| = 1\}$, by the definition of $\Delta^{\mathcal{I}}$, either $\mathbf{ct}(v)$ and $\mathbf{ct}(v')$ are anonymous maximal singletons or $v = [a], v' = [b]$ for some individuals a, b. In the former case, by maximality of $\mathbf{ct}(v)$ and $\mathbf{ct}(v')$, we have $v = v'$. In the latter case, since $\mathcal{A}$ is closed under $\mathbf{R}^2_{\sqsubseteq}$, we have $[a] = [b]$, i.e. $v = v'$.
- Case $(i.2)$ and $(i.3)$: $p_1 = (w, v), p_2 = (w, v')$, with $\mathcal{T}' \models \{\mathbf{ct}(w) \sqsubseteq \exists F.\mathbf{ct}(v), |\mathbf{ct}(v)| = 1\}$ and $\mathcal{T}' \models \{\mathbf{ct}(v') \sqsubseteq \exists F^-.\mathbf{ct}(w), |\mathbf{ct}(w)| = 1\}$. This case is analogous to the case $(i.1)$ and $(i.3)$.
- Case $(i.2)$ and $(i.4)$: $p_1 = (w, v), p_2 = (w, w.H.N)$ with $\mathcal{T}' \models \{\mathbf{ct}(w) \sqsubseteq \exists F.\mathbf{ct}(v), |\mathbf{ct}(v)| = 1\}$ and $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \exists H.N, F \in H$. Since $\mathsf{func}(F) \in \mathcal{T}'$, we have $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \exists H.N \sqcap \mathbf{ct}(v)$, which violates the conditions for $w.H.N$ being introduced.
- Case $(i.2)$ and $(i.5)$: $p_1 = (w.S.N, v)$, $p_2 = (w.S.N, w)$ with $\mathcal{T}' \models \{N \sqsubseteq \exists F.\mathbf{ct}(v), |\mathbf{ct}(v)| = 1\}$ and $w.H.N \in \Delta^{\mathcal{I}}, F^- \in H$. Since $w.H.N \in \Delta^{\mathcal{I}}$, we have $\mathbf{ct}(w) \models \exists H.N$, which, together with $\mathsf{func}(F) \in \mathcal{T}'$, implies $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \mathbf{ct}(v)$. Since $\mathcal{T}' \models |\mathbf{ct}(v)| = 1$ and $\mathcal{T}' \models |\mathbf{ct}(w)| \geq 1$ (by Lemma 7), we also obtain $\mathcal{T}' \models \{\mathbf{ct}(v) \sqsubseteq \mathbf{ct}(w), |\mathbf{ct}(w)| = 1\}$. Using arguments as in the case $(i.2)$ and $(i.2)$, we obtain $w = v$.
- Case $(i.3)$ and $(i.3)$: $p_1 = (v, w)$, $p_2 = (v, w')$ with $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \exists F^-.\ \mathbf{ct}(v)$, $\mathbf{ct}(w') \sqsubseteq \exists F^-.\mathbf{ct}(v), |\mathbf{ct}(v)| = 1$. Since $\mathsf{func}(F) \in \mathcal{T}'$, we have $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \mathbf{ct}(w'), \mathbf{ct}(w') \sqsubseteq \mathbf{ct}(w), |\mathbf{ct}(w)| = 1, |\mathbf{ct}(w')| = 1$. Using arguments as in the case $(i.2)$ and $(i.2)$, we obtain $w = w'$.
- Case $(i.3)$ and $(i.4)$: $p_1 = (w, v), p_2 = (w, w.H.N)$ with $\mathcal{T}' \models \{\mathbf{ct}(v) \sqsubseteq \exists F^-.\mathbf{ct}(w), |\mathbf{ct}(w)| = 1\}$ and $w.H.N \in \Delta^{\mathcal{I}}, F \in H$. From $\mathcal{T}' \models \{\mathbf{ct}(v) \sqsubseteq \exists F^-.\mathbf{ct}(w), |\mathbf{ct}(w)| = 1\}$ and $\mathsf{func}(F) \in \mathcal{T}'$, we obtain $\mathcal{T}' \models \mathbf{ct}(w) \models \exists F.\mathbf{ct}(v), |\mathbf{ct}(v)| = 1\}$. Similar

to the case $(i.2)$ and $(i.4)$, the combination of $(i.3)$ and $(i.4)$ cannot occur.

- Case $(i.3)$ and $(i.5)$ cannot occur.
- Case $(i.4)$ and $(i.4)$: $p_1 = (w, w.H.N)$, $p_2 = (w, w.H'.N')$. Since $p_1, p_2 \in F^{\mathcal{I}}$, we have $\mathcal{T}' \models \{M \sqsubseteq \exists H.N, M \sqsubseteq \exists H'.N'\}$, $F \in H$, and $F \in H'$. Since $\mathsf{func}(F) \in \mathcal{T}$, we obtain $\mathcal{T}' \models M \sqsubseteq \exists(H \sqcap H').(N \sqcap N')$. By the maximality of $H \sqcap H'$ and $N \sqcap N'$, we have $H = H \sqcap H' = H'$ and $N = N \sqcap N' = N'$.
- Case $(i.4)$ and $(i.5)$: $p_1 = (w.H.N, w.H.N.H'.N')$, $p_2 = (w.H.N, w)$ with $\mathcal{T}' \models N \sqsubseteq \exists H'.N'$, $F \in H'$ and $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq \exists H.N$, $F^- \in H$. Clearly, $\mathcal{T}' \models \mathbf{ct}(w) \sqsubseteq N', \mathbf{ct}(w) \sqsubseteq \exists(H \sqcap H'^-).N$. This violates the conditions for $w.H.N.H'.N'$ being in $\Delta^{\mathcal{I}}$. Thus, the combination of $(i.4)$ and $(i.5)$ cannot occur.
- Case $(i.5)$ and $(i.5)$ can occur only if $p_1 = p_2$, which is trivial.

This finishes the proof of $\mathcal{I} \models \mathcal{A} \cup \mathcal{T}^*$. □

To obtain a model $\mathcal{J}$ from $\mathcal{I}$ that also satisfies the transitivity axioms in $\mathcal{A} \cup \mathcal{T}$, we define $\mathcal{J}$ as identical to $\mathcal{I}$ apart from the interpretations of roles with transitive sub-roles. We set, for each role name r:

$$r^{\mathcal{J}} = r^{\mathcal{I}} \cup \{(d_0, d_n) \mid d_0, \ldots, d_n \in \Delta^{\mathcal{I}}, (d_{i-1}, d_i) \in T^{\mathcal{I}}, 1 \leq i \leq n, \mathcal{T} \models T \sqsubseteq r, \mathsf{tran}(T) \in \mathcal{T}\} \tag{20}$$

Lemma 11. *$\mathcal{J} \models \mathcal{A} \cup \mathcal{T}$ and, for every $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and $a, b \in \mathsf{N_I}$, we have:*

1. *$\mathcal{J} \models A(a)$ iff $A(a) \in \mathcal{A}$,*
2. *$\mathcal{J} \models a \approx b$ iff $a \approx b \in \mathcal{A}$, and*
3. *$\mathcal{J} \models r(a, b)$ iff $r(a, b) \in \mathcal{A}$.*

Proof. We first show $\mathcal{J} \models \mathcal{A} \cup \mathcal{T}$. Clearly, $\mathcal{J}$ entails each assertion and role inclusion in $\mathcal{A} \cup \mathcal{T}$. Since functional roles must be simple and the interpretation of simple roles remains the same, $\mathcal{J}$ satisfies functionality axioms, and, clearly, the extension of roles in (20) ensures that $\mathcal{J}$ satisfies transitivity axioms. For concept inclusions in $\mathcal{T}$, the only non-trivial case to be revised is $C \sqsubseteq \forall R.A$, where R has a transitive sub-role T. Let $d_0 \in C^{\mathcal{J}}$ and (d_0, d_n) is newly added to $R^{\mathcal{J}}$ in (20). Since $\mathcal{O}$ is normalized, there exists B such that $d_i \in B^{\mathcal{I}} \subseteq A^{\mathcal{I}} = A^{\mathcal{J}}, 1 \leq i \leq n$, hence, $d_0 \in (\forall R.A)^{\mathcal{J}}$.

Claims (1)–(2) follow directly from the definitions of $\mathcal{I}$ and $\mathcal{J}$. To show Claim (3), we show that for every transitive role T and individuals a and b, if role extension in (20) results in $([a], [b]) \in T^{\mathcal{J}}$, then $T(a, b) \in \mathcal{A}$. Intuitively, there are three types of domain elements in $\Delta^{\mathcal{I}}$: domain elements for individuals, for anonymous maximal singletons, and for the tree parts. With the observation that d_i $(1 \leq i \leq n-1)$ in (20) belongs to one of those three types of domain elements, we use three rules $\mathbf{R}_t^1, \mathbf{R}_t^2$, and $\mathbf{R}_t^3$, which cover all possible ways to connect $[a]$ and $[b]$ via a chain of transitive role (T) assertions, to derive $T(a, b) \in \mathcal{A}$.

We first show some auxiliary lemmas covering cases in which two domain elements (not necessary for individuals) are connected via a chain of transitive role assertions.

Lemma 12. *Let T be a transitive role, $n \geq 1$, and $d_0 \in \mathbb{G}(\mathcal{O})$, $d_1, \ldots, d_n \in \Delta^{\mathcal{I}} \setminus \mathbb{G}(\mathcal{O})$. If $(d_{i-1}, d_i) \in T^{\mathcal{I}}, 1 \leq i \leq n$, then $\mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq \exists T.\mathbf{ct}(d_n)$.*

Proof. The proof is by induction on n.

Case $n = 1$. Since $(d_0, d_1) \in T^{\mathcal{I}}$, $d_0 \in \mathbb{G}(\mathcal{O})$, and $d_1 \notin \mathbb{G}(\mathcal{O})$, from the definition of $\mathcal{I}$ we have either $\mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq \exists H.\mathbf{ct}(d_1), T \in H$ or $\mathcal{T}' \models \{\mathbf{ct}(d_1) \sqsubseteq \exists T^-.\mathbf{ct}(d_0), |\mathbf{ct}(d_0)| = 1\}$. In both cases, we obtain $\mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq \exists T.\mathbf{ct}(d_1)$.

Case $n \geq 2$. If there exists $1 \leq i < j \leq n$ such that $d_i = d_j$, then using induction hypothesis on $m < n$ elements: $d_0, \ldots, d_i, d_{j+1}, \ldots, d_n$, we have $\mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq \exists T.\mathbf{ct}(d_n)$. From now on, we consider the case that no such i and j exist. Let $\mathbf{len}(w)$ denote the length of the word w, i.e. $\mathbf{len}(w) = 1$ if $w \in \mathbb{G}(\mathcal{O})$; and $\mathbf{len}(w) = \mathbf{len}(v) + 2$ if $w = v.H.N$. And let $1 \leq l \leq n$ s.t. $\mathbf{len}(d_l)$ is maximal. We consider three sub-cases:

- Case $l = 1$. From the definition of $\mathcal{I}$ we have either $\mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq \exists H.\mathbf{ct}(d_1)$, $T \in H$ or $\mathcal{T}' \models \{\mathbf{ct}(d_1) \sqsubseteq \exists T^-.\mathbf{ct}(d_0), |\mathbf{ct}(d_0)| = 1\}$. The former cannot occur as it will lead to $d_2 = d_0 \in \mathbb{G}(\mathcal{O})$, which is not the case. Therefore, we only consider the later. From the definition of $\mathcal{I}$, for each pair $(d_{i-1}, d_i) \in T^{\mathcal{I}}$, $1 \leq i \leq n-1$, there are two possibilities: either $d_{i-1} = w, d_i = w.H.N$ or $d_{i-1} = w.H.N, d_i = w$ for some w, H, N. Since $\mathbf{len}(d_1)$ is maximal and no $1 \leq i < j \leq n$ such that $d_i = d_j$, we have $\mathcal{T}' \models \{\mathbf{ct}(d_n) \sqsubseteq \exists H_n.\mathbf{ct}(d_{n-1}), \ldots, \mathbf{ct}(d_2) \sqsubseteq \exists H_2.\mathbf{ct}(d_1)$, with $T^- \in H_i, 2 \leq i \leq n$. As T is transitive and $\mathcal{T}' \models \mathbf{ct}(d_1) \sqsubseteq \exists T^-.\mathbf{ct}(d_0)$, we obtain $\mathcal{T}' \models \mathbf{ct}(d_n) \sqsubseteq \exists T^-.\mathbf{ct}(d_0)$. Additionally, since $\mathcal{T}' \models |\mathbf{ct}(d_0)| = 1$, we have $\mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq \exists T.\mathbf{ct}(d_n)$.
- Case $1 < l < n$. Since $(d_{l-1}, d_l) \in T^{\mathcal{I}}$, and $\mathbf{len}(d_l)$ is maximal, from the definition of $\mathcal{I}$ we have $d_{l-1} = w, d_l = w.H.N$ for some H, N, and $w \in \Delta^{\mathcal{I}}$. Similarly, we have $d_{l+1} = w$. Therefore, $d_{l-1} = d_{l+1}$, which cannot be the case.
- Case $l = n$. Since $(d_{n-1}, d_n) \in T^{\mathcal{I}}$ and $\mathbf{len}(d_n)$ is maximal, from the definition of $\mathcal{I}$ we have $d_n = d_{n-1}.H.\mathbf{ct}(d_n)$ for some H s.t. $\mathcal{T}' \models \mathbf{ct}(d_{n-1}) \sqsubseteq \exists H.\mathbf{ct}(d_n), T \in H$. Additionally, by induction hypothesis, we have $\mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq \exists T.\mathbf{ct}(d_{n-1})$. Since T is transitive, we obtain $\mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq \exists T.\mathbf{ct}(d_n)$.

□

In the following lemma, we show that for a chain of transitive role assertions in the tree parts, all elements have the same prefix.

Lemma 13. *Let T be a role, $n \geq 1$, and $d_1, \ldots, d_n \in \Delta^{\mathcal{I}} \setminus \mathbb{G}(\mathcal{O})$. If $d_1 = [a].H.N$ for some $H.N \in \Gamma \times \Sigma$ and $(d_{i-1}, d_i) \in T^{\mathcal{I}}, 2 \leq i \leq n$, then $d_n = [a].H.N.v$, where $v \in (\Gamma \times \Sigma)^*$.*

Proof. The proof is by induction on n. For $n = 1$, the lemma trivially holds. For $n \geq 2$, by induction hypothesis, either $d_{n-1} = [a].H.N$ or d_{n-1} is of the form $[a].H.N.u.H'.N'$ for some (possibly empty) word u.

- Case: $d_{n-1} = [a].H.N$. Since $(d_{n-1}, d_n) \in T^{\mathcal{I}}$, from the definition of $\mathcal{I}$, cases $(i.4)$ and $(i.5)$, we obtain either $d_n = [a].H.N.H''.N''$ for some H'', N'', or $d_n = [a]$. The latter cannot occur as $d_n \notin \mathbb{G}(\mathcal{O})$.
- Case: $d_{n-1} = [a].H.N.u.H'.N'$. Since $(d_{n-1}, d_n) \in T^{\mathcal{I}}$, from the definition of $\mathcal{I}$, cases $(i.4)$ and $(i.5)$, we obtain either $d_n = [a].H.N.u.H'.N'.H''.N''$ for some H'', N'', or $d_n = [a].H.N.u$.

In all cases, we have d_n is of the form $[a].H.N.v$, where $v \in (\Gamma \times \Sigma)^*$ □

Using Lemma 12 and Lemma 13 we are able to claim that if two domain elements in the graph part, e.g. d_0 and d_n, are connected via a chain of transitive role assertions in the tree parts, then $(d_0, d_n) \in T^{\mathcal{I}}$ and this can be infered from $\mathcal{T}'$.

Lemma 14. *Let T be a transitive role, $n \geq 1$, and $d_0, d_n \in \mathbb{G}(\mathcal{O})$, $d_1, \ldots, d_{n-1} \in \Delta^{\mathcal{I}} \setminus \mathbb{G}(\mathcal{O})$. If $(d_{i-1}, d_i) \in T^{\mathcal{I}}, 1 \leq i \leq n$, then one of the following holds:*

1. $\mathcal{T}' \models \{\mathbf{ct}(d_0) \sqsubseteq \exists T.\mathbf{ct}(d_n), |\mathbf{ct}(d_n)| = 1\}$,
2. $\mathcal{T}' \models \{\mathbf{ct}(d_n) \sqsubseteq \exists T^-.\mathbf{ct}(d_0), |\mathbf{ct}(d_0)| = 1\}$, *or*
3. $d_0 = d_n, \mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq \exists (T \sqcap T^-).\top$.

Proof. We distinguish cases in definition of $\mathcal{I}$ that lead to $(d_0, d_1) \in T^{\mathcal{I}}$:

$$\mathcal{T}' \models \mathbf{ct}(d_1) \sqsubseteq \exists T^-.\mathbf{ct}(d_0), |\mathbf{ct}(d_0)| = 1 \quad (21)$$

$$\mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq \exists H.\mathbf{ct}(d_1), T \in H \quad (22)$$

and those that lead to $(d_{n-1}, d_n) \in T^{\mathcal{I}}$:

$$\mathcal{T}' \models \mathbf{ct}(d_{n-1}) \sqsubseteq \exists T.\mathbf{ct}(d_n), |\mathbf{ct}(d_n)| = 1 \quad (23)$$

$$\mathcal{T}' \models \mathbf{ct}(d_n) \sqsubseteq \exists H'.\mathbf{ct}(d_{n-1}), T^- \in H' \quad (24)$$

We examine combinations of (21), (22) with (23), (24).

- Cases (21) + (23), and (22) + (23). By Lemma 12, we have $\mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq \exists T.\mathbf{ct}(d_{n-1})$. From (23) and T is transitive, we obtain Claim (1).
- Case (21) + (24). By Lemma 12, we have $\mathcal{T}' \models \mathbf{ct}(d_n) \sqsubseteq \exists T^-.\mathbf{ct}(d_1)$. From (21) and T is transitive, we obtain Claim (2).
- Case (22) + (24). From (22), we have $d_1 = d_0.H.N$ for some H, N. By Lemma 13, we have $d_{n-1} = d_0.H.N.v$. On the other hand, from (24), we have $d_{n-1} = d_n.H'.N'$. Therefore, $d_0 = d_n, H = H', N = N'$, and v is empty. Hence, $\mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq (T \sqcap T^-).N$, which implies Claim (3).

□

We present another case where two domain elements in the graph part are connected via a chain of transitive role assertions that contain no domain elements for individuals.

Lemma 15. *Let T be a transitive role, $n \geq 1$, $d_0, \ldots, d_n \in \Delta^{\mathcal{I}}$ s.t. $d_0 \in \mathbb{G}(\mathcal{O})$, $d_n = N$ for some anonymous maximal singleton N, and there exists no $1 \leq i \leq n-1$ s.t. $d_i = [a]$ for some individual a. If $(d_{i-1}, d_i) \in T^{\mathcal{I}}, 1 \leq i \leq n$, then $\mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq \exists T.N$.*

Proof. The proof is by induction on n. For $n = 1$, we have $(d_0, d_1) \in T^{\mathcal{I}}$. By the definition of $\mathcal{I}$, this is the case only when $\mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq \exists T.N$. For $n \geq 2$, we distinguish two cases:

- There exists $1 \leq i \leq n-1$ s.t. $d_i = M$ for some anonymous maximal singleton M. By induction hypothesis, we have $\mathcal{T}' \models \{\mathbf{ct}(d_0) \sqsubseteq \exists T.M, M \sqsubseteq \exists T.N\}$. Since T is transitive, $\mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq \exists T.N$.
- There exists no $1 \leq i \leq n-1$ s.t. $d_i = M$ for some anonymous maximal singleton M. From the assumption of the lemma that there exists no $1 \leq i \leq n-1$ s.t. $d_i = [a]$ for some individual a, we have $d_i \notin \mathbb{G}(\mathcal{O})$ for every $1 \leq i \leq n-1$. By Lemma 14, we have either $\mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq \exists T.N$, or $\mathcal{T}' \models \{N \sqsubseteq \exists T^-.\mathbf{ct}(d_0), |\mathbf{ct}(d_0)| = 1\}$, or $\mathbf{ct}(d_0) = N, N \sqsubseteq \exists (T \sqcap T^-).\top$. In all cases, we obtain $\mathcal{T}' \models \mathbf{ct}(d_0) \sqsubseteq \exists T.N$.

□

After inspecting all possible ways that role extension in (20) results in new pairs in the interpretation of transitive roles. We are now ready to show that $T(a, b) \in \mathcal{A}$ if $([a], [b]) \in T^{\mathcal{J}}$ results from the extension of T.

Proposition 16. *Let T be a transitive role, a, b individuals, $n \geq 1$, and $d_0, \ldots, d_n \in \Delta^{\mathcal{I}}$, with $d_0 = [a]$, $d_n = [b]$. If $(d_{i-1}, d_i) \in T^{\mathcal{I}}, 1 \leq i \leq n$, then $T(a, b) \in \mathcal{A}$.*

Proof. The proof is by induction on n. For $n = 1$, we have $([a], [b]) \in T^{\mathcal{I}}$. From the definition of $\mathcal{I}$, we obtain $T(a, b) \in \mathcal{A}$. For $n \geq 2$, we distinguish three cases:

- There exists $1 \leq i \leq n-1$ s.t. $d_i = [c]$ for some individual c. By induction hypothesis we have $T(a, c), T(c, b) \in \mathcal{A}$. Since $\mathcal{A}$ is closed under $\mathbf{R}_t^1$, we have $T(a, b) \in \mathcal{A}$.
- There exists no $1 \leq i \leq n-1$ s.t. $d_i = [c]$ for some individual c, but there exists $1 \leq j \leq n-1$ s.t. $\mathcal{T}' \models |\mathbf{ct}(d_j)| = 1$. By Lemma 15, we have $\mathcal{T}' \models \{\mathbf{ct}([a]) \sqsubseteq \exists T.\mathbf{ct}(d_j), \mathbf{ct}([b]) \sqsubseteq \exists T^-.\mathbf{ct}(d_j)\}$. Since $\mathcal{A}$ is closed under $\mathbf{R}_t^2$, we have $T(a, b) \in \mathcal{A}$.
- The complement of the previous cases, i.e. for every $1 \leq i \leq n-1$, neither $d_i = [c]$ for some individual c nor $\mathcal{T}' \models |\mathbf{ct}(d_i)| = 1$. This implies that $d_i \in \Delta^{\mathcal{I}} \setminus \mathbb{G}(\mathcal{O}), 1 \leq i \leq n-1$. By Lemma 14, one of the following cases holds.
  - $\mathcal{T}' \models \{\mathbf{ct}([a]) \sqsubseteq \exists T.\mathbf{ct}([b]), |\mathbf{ct}([b])| = 1\}$. Since $\mathcal{A}$ is closed under $\mathbf{R}_\exists$, we have $T(a, b) \in \mathcal{A}$.
  - $\mathcal{T}' \models \{\mathbf{ct}([b]) \sqsubseteq \exists T^-.\mathbf{ct}([a]), |\mathbf{ct}([a])| = 1\}$. Since $\mathcal{A}$ is closed under $\mathbf{R}_\exists, \mathbf{R}_{inv}$, we have $T(a, b) \in \mathcal{A}$.
  - $a = b$ and $\mathcal{T}' \models \mathbf{ct}([a]) \sqsubseteq \exists (T \sqcap T^-).\top$. Since $\mathcal{A}$ is closed under $\mathbf{R}_t^3$, we have $T(a, a) \in \mathcal{A}$.

In all cases, we have $T(a, b) \in \mathcal{A}$, which is required to show. □

Let $([a], [b]) \in r^{\mathcal{J}}$. If $([a], [b]) \in r^{\mathcal{I}}$, then from the definition of $\mathcal{I}$ we have $r(a, b) \in \mathcal{A}$. Otherwise, $([a], [b]) \in r^{\mathcal{J}}$ results from (20). By Proposition 16 and the condition that $\mathcal{A}$ is closed under $\mathbf{R}_\sqsubseteq^1$, we have $r(a, b) \in \mathcal{A}$. This finishes the proof of Lemma 11. □

We have shown that there exists a model $\mathcal{J}$ of $\mathcal{O}$ as required. This completes the proof of Theorem 5. □

Completeness

Once the abstraction refinement procedure terminates, we claim that the ontology is fully materialized by showing that it is closed under the rules in Table 2.

Lemma 17. *Let $\mathcal{A} \cup \mathcal{T}$ be an ontology s.t. $\mathcal{A}$ is equality-, transitivity-, and functionality-closed, $\mathcal{B}$ the abstraction of $\mathcal{A}$, $\mathcal{B}'$ an ABox s.t. $\mathcal{B} \subseteq \mathcal{B}'$, $\mathfrak{N}(\mathcal{B}') = \mathfrak{N}(\mathcal{A})$ and $\Delta\mathcal{A} \subseteq \mathcal{A}$ with $\Delta\mathcal{A}$ the update of $\mathcal{A}$ using $\mathcal{B}' \setminus \mathcal{B}$. Then, $\mathcal{B}'$ is closed under the rules in Table 2 w.r.t. $\mathcal{T}$ implies that $\mathcal{A}$ is also closed under the rules w.r.t. $\mathcal{T}$.*

Proof. Since $\mathfrak{N}(\mathcal{B}') = \mathfrak{N}(\mathcal{A})$, the side condition of each rule holds for $\mathcal{B}' \cup \mathcal{T}$ iff it holds for $\mathcal{A} \cup \mathcal{T}$. We examine each rule in Table 2 and show that $\mathcal{A}$ is closed under that rule. Clearly, $\mathcal{A}$ is closed under $\mathbf{R}^1_{\approx}$–$\mathbf{R}^4_{\approx}$, $\mathbf{R}^1_{\leq}, \mathbf{R}^1_t$. For the other rules, the intuition is: if the premises of a rule $\mathbf{R}$ hold for some assertions γ in $\mathcal{A}$, then the premises of $\mathbf{R}$ also hold for the corresponding abstract assertions γ' in $\mathcal{B}$, and in $\mathcal{B}'$ consequently. Since $\mathcal{B}'$ is closed under $\mathbf{R}$, the conclusion κ' of γ' w.r.t. $\mathbf{R}$ is already in $\mathcal{B}'$. Then, the condition $\Delta\mathcal{A} \subseteq \mathcal{A}$ guarantees that the conclusion κ of γ w.r.t. $\mathbf{R}$ is already in $\mathcal{A}$, which implies that $\mathcal{A}$ is closed under $\mathbf{R}$. Let $\Delta\mathcal{B} = \mathcal{B}' \setminus \mathcal{B}$, we now consider each rule in detail.

- For $\mathbf{R}^2_t$, let $\{N(a), M(b)\} \subseteq \mathcal{A}$, and the side condition holds. We have $\{N(v_{\tau(a)}), M(v_{\tau(b)})\} \subseteq \mathcal{B} \subseteq \mathcal{B}'$. Since $\mathcal{B}'$ is closed under $\mathbf{R}^2_t$, we have $T(v_{\tau(a)}, v_{\tau(b)}) \in \mathcal{B}'$. If $v_{\tau(a)} = v_{\tau(b)} = v_\tau$, i.e. a and b have the same type τ, we are not able to obtain $T(a, b)$ to add to $\mathcal{A}$. That explains why concept types are required. Since $\{N(u_{\tau_C}), M(v_\tau)\} \subseteq \mathcal{B} \subseteq \mathcal{B}'$ and $\mathcal{B}'$ is closed under the $\mathbf{R}^2_t$, we have $T(u_{\tau_C}, v_\tau) \in \mathcal{B}'$, and consequently $T(u_{\tau_C}, v_\tau) \in \Delta\mathcal{B}$. By Definition 4, we have $T(a, b) \in \Delta\mathcal{A} \subseteq \mathcal{A}$. Therefore, $\mathcal{A}$ is closed under $\mathbf{R}^2_t$.
- For $\mathbf{R}^3_t$, let $N(a) \in \mathcal{A}$, and the side condition holds. We have $N(v_{\tau(a)}) \in \mathcal{B} \subseteq \mathcal{B}'$. Since $\mathcal{B}'$ is closed under $\mathbf{R}^3_t$, we have $T(v_{\tau(a)}, v_{\tau(a)}) \in \mathcal{B}'$, and consequently $T(v_{\tau(a)}, v_{\tau(a)}) \in \Delta\mathcal{B}$. By Definition 4, we have $T(a, a) \in \Delta\mathcal{A} \subseteq \mathcal{A}$. Therefore, $\mathcal{A}$ is closed under $\mathbf{R}^3_t$.
- For $\mathbf{R}_{||}$, let $\{N(a), N(b)\} \subseteq \mathcal{A}$ and the side condition holds. We have $\{N(v_{\tau(a)}), N(v_{\tau(b)})\} \subseteq \mathcal{B} \subseteq \mathcal{B}'$. Since $\mathcal{B}'$ is closed under $\mathbf{R}_{||}$, we have $v_{\tau(a)} \approx v_{\tau(b)} \in \mathcal{B}$. If $v_{\tau(a)} \neq v_{\tau(b)}$, then it is sound to derive $a \approx b$. However, in case $v_{\tau(a)} = v_{\tau(b)} = v_\tau$, i.e. a and b have the same type τ, it is unsound to derive $a \approx b$ in general. That explains why concept types are required and in this case, we use equality of u_{τ_C}. Since $\{N(u_{\tau_C}), N(v_\tau)\} \subseteq \mathcal{B} \subseteq \mathcal{B}'$ and $\mathcal{B}'$ is closed under $\mathbf{R}_{||}$, we have $u_{\tau_C} \approx v_\tau \in \Delta\mathcal{B}$. By Definition 4, we have $a \approx b \in \Delta\mathcal{A} \subseteq \mathcal{A}$. Therefore, $\mathcal{A}$ is closed under $\mathbf{R}_{||}$.
- For $\mathbf{R}_\forall$, let $\{N(a), R(a, b)\} \subseteq \mathcal{A}$ and the side condition holds. Since $\{N(v_{\tau(a)}), R(v_{\tau(a)}, w^R_{\tau(a)})\} \subseteq \mathcal{B} \subseteq \mathcal{B}'$ and $\mathcal{B}'$ is closed under the rule, we obtain $B(w^R_{\tau(a)}) \in \Delta\mathcal{B}$. By Definition 4, $B(b) \in \Delta\mathcal{A} \subseteq \mathcal{A}$. Therefore, $\mathcal{A}$ is closed under $\mathbf{R}_\forall$.
- For $\mathbf{R}_\exists$, let $\{N(a), M(b)\} \subseteq \mathcal{A}$, and the side condition holds. We have $\{N(u_{\tau_C(a)}), M(v_{\tau(b)})\} \subseteq \mathcal{B} \subseteq \mathcal{B}'$. Since $\mathcal{B}'$ is closed under $\mathbf{R}_\exists$, we have $R(u_{\tau_C(a)}, v_{\tau(b)}) \in \mathcal{B}'$, and consequently $R(u_{\tau_C(a)}, v_{\tau(b)}) \in \Delta\mathcal{B}$. By Definition 4, we have $R(a, b) \in \Delta\mathcal{A} \subseteq \mathcal{A}$. Therefore, $\mathcal{A}$ is closed under $\mathbf{R}_\exists$.
- For $\mathbf{R}^1_\sqsubseteq$, let $R(a, b) \in \mathcal{A}$, and the side condition holds. We have $R(v_{\tau(a)}, w^R_{\tau(a)}) \in \mathcal{B} \subseteq \mathcal{B}'$. Since $\mathcal{B}'$ is closed under $\mathbf{R}^1_\sqsubseteq$, we have $S(v_{\tau(a)}, w^R_{\tau(a)}) \in \mathcal{B}'$, and consequently $S(v_{\tau(a)}, w^R_{\tau(a)}) \in \Delta\mathcal{B}$. By Definition 4, we have $S(a, b) \in \Delta\mathcal{A} \subseteq \mathcal{A}$. Therefore, $\mathcal{A}$ is closed under $\mathbf{R}^1_\sqsubseteq$.
- For $\mathbf{R}^2_\sqsubseteq$, we proceed analogously.
- For $\mathbf{R}^2_\leq$, let $\{M(a), F(a, b)\} \subseteq \mathcal{A}$ and the side condition holds. We have $\{M(v_{\tau(a)}), F(v_{\tau(a)}, w^F_{\tau(a)})\} \subseteq \mathcal{B} \subseteq \mathcal{B}'$. Since $\mathcal{B}'$ is closed under $\mathbf{R}^2_\leq$, $\{H(v_{\tau(a)}, w^F_{\tau(a)}), N(w^F_{\tau(a)})\} \subseteq \mathcal{B}'$, and consequently $\{H(v_{\tau(a)}, w^F_{\tau(a)}), N(w^F_{\tau(a)})\} \subseteq \Delta\mathcal{B}$. By Definition 4, we have $\{H(a, b), N(b)\} \subseteq \Delta\mathcal{A} \subseteq \mathcal{A}$. Therefore, $\mathcal{A}$ is closed under $\mathbf{R}^2_\leq$.

We have shown that $\mathcal{A}$ is closed under all rules in Table 2 w.r.t. $\mathcal{T}$. □

Using Lemma 17, we show that the procedure is complete.

Theorem 18. *The ontology $\mathcal{O} = \mathcal{A} \cup \mathcal{T}$ obtained from the abstraction refinement procedure is fully materialized.*

Proof. Let $\mathcal{B}$ be the abstraction of $\mathcal{A}$, $\mathcal{B}' \cup \mathcal{T}$ the materialization of $\mathcal{B} \cup \mathcal{T}$, $\Delta\mathcal{B} = \mathcal{B}' \setminus \mathcal{B}$, and $\Delta\mathcal{A}$ the update of $\mathcal{A}$. For every $A(a) \in \mathcal{A}$, we have $A(v_{\tau(a)}) \in \mathcal{B} \subseteq \mathcal{B}'$, which implies $\mathfrak{N}(\mathcal{A}) \subseteq \mathfrak{N}(\mathcal{B}')(\star)$. For every $A(u_{\tau_C(a)}) \in \mathcal{B}'$, we have $A(v_{\tau(a)}) \in \mathcal{B}'$. Since the procedure terminated, i.e. $\Delta\mathcal{A} \subseteq \mathcal{A}$, for every $A(v_{\tau(a)}) \in \mathcal{B}'$ or $A(w^R_{\tau(a)}) \in \mathcal{B}'$, we have $A(a) \in \mathcal{A}$ for some $a \in \mathsf{ind}(\mathcal{A})$. Hence, we obtain $\mathfrak{N}(\mathcal{B}') \subseteq \mathfrak{N}(\mathcal{A})(\dagger)$. From $(\star)$ and $(\dagger)$ we have $\mathfrak{N}(\mathcal{A}) = \mathfrak{N}(\mathcal{B}')$, which allows us to apply Lemma 17. By Theorem 5, $\mathcal{B}'$ is closed under the rules in Table 2. Therefore, by Lemma 17, $\mathcal{A}$ is closed under those rules. Consequently, by Theorem 5, $\mathcal{O}$ is materialized. □

Implementation and Evaluation

We implemented a prototype system X for full materialization of Horn $\mathcal{SHOIF}$ ontologies, evaluated X on popular ontologies, and compared it with other reasoners. The purpose of our experiments was to measure the reduction in the sizes of the ontologies and the subsequent improvements of the reasoning times on the reduced ontologies. For this reason, the computation of the abstractions and of the deductive closure by the rules is kept simple and is not (yet) very optimized. In particular, after each refinement step, the abstraction is computed from scratch.

Ontology	# ax.	# con.	# rol.	# ind.	# ast.	# mater. ast.
NPD	773	331	143	264 081	911 517	1 728 168
DBPedia$^+$	1 762	442	766	3 790 936	25 276 255	28 452 224
IMDb$^+$	134	82	41	6 502 964	27 682 276	83 787 170
L 10	96	43	25	207 426	850 433	1 322 755
L 50	96	43	25	1 082 818	4 445 949	6 914 186
L 100	96	43	25	2 179 766	8 954 615	13 924 116
L 500	96	43	25	10 847 183	44 573 624	69 315 045
U 10	157	67	35	242 491	1 926 941	3 922 810
U 50	157	67	35	1 227 123	9 751 725	19 864 408
U 100	157	67	35	2 461 347	19 571 799	39 857 382
U 500	157	67	35	12 375 804	98 374 736	200 296 996

Table 3: Test ontologies with the number of TBox axioms (# ax.), atomic concepts (# con.), roles (# rol.), individuals (# ind.), concept and role assertions (# ast.) in the original ABox and in the materialization (# mater. ast.)

Ontology	1st abstraction				2nd abstraction			
	# types	# indiv.	# ast.	% ast.	# types	# indiv.	# ast.	% ast.
NPD	1 102	18 565	21 263	2.333	–	–	–	–
DBPedia$^+$	187 968	1 576 392	3 239 328	12.816	186 924	1 568 633	3 914 593	15.487
IMDb$^+$	366	1 739	2 160	0.008	–	–	–	–
L10	29	223	227	0.027	29	223	321	0.038
L50	27	213	217	0.005	27	213	309	0.007
L100	27	213	217	0.002	27	213	309	0.003
L500	27	213	217	0.000	27	213	309	0.001
U10	4 477	68 007	91 889	4.769	9 204	150 203	369 277	19.164
U50	8 205	126 499	172 627	1.770	18 674	308 808	780 930	8.008
U100	10 220	158 146	216 750	1.107	24 362	405 697	1 036 017	5.293
U500	15 521	243 273	335 809	0.341	41 814	702 892	1 836 102	1.866

Table 4: Number of types, abstract individuals (# indiv.), sizes of the abstract ABoxes and the comparison with sizes of the original ABoxes

Table 3 presents detailed information about the test ontologies. NPD is an ontology about petroleum activities, DBPedia$^+$ is an extension of the DBPedia ontology, and IMDb$^+$ consists of the Movie ontology and the dataset from the IMDb website. For the popular benchmarks LUBM (Guo, Pan, and Heflin 2005) and UOBM (Ma et al. 2006), we use Ln and Un to denote the datasets for n universities respectively. LUBM is in Horn $\mathcal{SHI}$ and IMDb$^+$ is in Horn $\mathcal{SHIF}$. For other ontologies, we extracted the relevant Horn fragment by removing axioms with disjunctions; NPD is in Horn $\mathcal{SHIF}$, DBPedia$^+$ and UOBM are in Horn $\mathcal{SHOIF}$. Test ontologies and our system are available online.[1]

For each ontology, X reached the fixpoint after at most two abstraction steps. As seen in Table 4, the sizes of the abstractions are significantly smaller than the sizes of the original ABoxes. For LUBM and IMDb$^+$, the reduction is over four orders of magnitude. LUBM is also known to be easily handled by other related approaches such as SHER (Fokoue et al. 2006). For NPD the reduction is also significant: the abstraction is just 2% of the size of the original ABox. UOBM and DBPedia$^+$ are more challenging as the ontologies are in full Horn $\mathcal{SHOIF}$. For U10, the size of the abstract ABox is approximately 19% of the size of the original one. Unsurprisingly, the size reductions are better for larger datasets, e.g., the abstraction for U500 is less than 2% in the size of the original ABox. For DBPedia$^+$, the abstract ABox is merely 15% of the size of the original one. This is because individuals in this ontology are more diverse due to relatively large number of atomic concepts and roles.

Table 5 provides a comparison of reasoning times with and without abstraction. We limit this comparison to well-known reasoners Konclude 0.6.2 (Steigmiller, Liebig, and Glimm 2014), PAGOdA 2.0 (Zhou et al. 2015), HermiT 1.3.8 (Glimm et al. 2014a), and Pellet 2.3.6 (Sirin et al. 2007). HermiT managed to compute the full materialization of L10 and timed out for the other ontologies. Pellet performed relatively well on NPD but took significantly more time to materialize DBPedia$^+$, IMDb$^+$, and L10; it timed out for other ontologies. For Konclude and PAGOdA we had difficulties in computing full materialization. The results of PAGOdA (P_{reas}) do not include equality materialization. Konclude can only be controlled via OWLLink API (Liebig et al. 2011), which causes a significant communication overhead due to the large number of individuals. In particular, Konclude timed out for all our test ontologies. For reference, we provide the (incomplete) results of Konclude for just con-

[1] https://sites.google.com/site/subm1566

Ontology	X_{load}	$\mathsf{K}^{\text{abst}}_{\text{reas}}$	X_{reas}	K_{load}	$\mathsf{K}^{\text{CLI}}_{\text{reas}}$	$\mathsf{K}^{\text{link}}_{\text{reas}}$	P_{load}	P_{reas}	HermiT$_{\text{reas}}$	Pellet$_{\text{reas}}$
NPD	12	4	6	9	5	35	21	49	*to*	94
DBP$^+$	470	297	412	240	145	335	101	167	*to*	13 958
IMDb$^+$	171	1	173	280	172	254	36	209	*to*	1 866
L10	6	2	7	10	5	15	3	8	4 222	10 776
L50	33	2	27	55	31	66	20	48	*to*	*to*
L100	68	2	61	102	67	296	43	97	*to*	*to*
L500	433	2	304	592	351	*err*	221	621	*to*	*to*
U10	35	35	54	23	25	36	7	25	*to*	*to*
U50	129	69	155	120	144	164	43	138	*to*	*to*
U100	199	95	267	254	315	444	89	298	*to*	*to*
U500	1 152	143	1 213	1 440	2107	*err*	509	1 591	*to*	*to*

Table 5: The reasoning times of our system (X_{reas}), command-line Konclude ($\mathsf{K}^{\text{CLI}}_{\text{reas}}$), Konclude using OWLLink ($\mathsf{K}^{\text{link}}_{\text{reas}}$), PAGOdA ($\mathsf{P}_{\text{reas}}$), HermiT (HermiT$_{\text{reas}}$) and Pellet (Pellet$_{\text{reas}}$), excluding the respective ontology loading times (X_{load}), (K_{load}), and (P_{load}); HermiT and Pellet have the same loading times as of X; *err* and *to* stand for "error" and "timeout" respectively

cept materialization using the command-line client ($\mathsf{K}^{\text{CLI}}_{\text{reas}}$), and using OWLLink ($\mathsf{K}^{\text{link}}_{\text{reas}}$). All results were obtained using a compute server with 40 Intel Xeon E5-2660V3 processors and 512 GB RAM and a timeout of 5 hours.

We present the full reasoning times of X (X_{reas}) excluding loading time (X_{load}) and including reasoning time of Konclude via OWLLink on the abstract ABoxes ($\mathsf{K}^{\text{abst}}_{\text{reas}}$). As mentioned, our computation of the abstractions is not very optimized; currently the total reasoning time is dominated by copying the entailed assertions from the abstract ABoxes to the original ABox. This step could be avoided by recomputing the abstractions directly. In any case, the purpose of our evaluation was not to show the superiority of X, but to see if our approach can improve the performance of existing reasoners on large data sets. As the abstract ABoxes are often significantly smaller than the original ones, directly implementing the technique in existing reasoners could bring even better performance improvements.

Discussion and Future Work

The presented approach for *full* materialization of Horn $\mathcal{SHOIF}$ ontologies results in abstractions that are significantly smaller than the original ontologies (by orders of magnitude) and can be computed efficiently. This is despite the more complex structure of the abstractions and the additionally required closure rules to achieve completeness in the presence of nominals, inverse roles, and functionality.

A remaining challenge is the extension to non-Horn ontologies, which is non-trivial since reasoners do not communicate non-deterministic information.

References

Abiteboul, S.; Hull, R.; and Vianu, V. 1995. *Foundations of Databases*. Addison-Wesley.

Dolby, J.; Fokoue, A.; Kalyanpur, A.; Schonberg, E.; and Srinivas, K. 2009. Scalable highly expressive reasoner (SHER). *J. of Web Semantics* 7(4):357–361.

Fokoue, A.; Kershenbaum, A.; Ma, L.; Schonberg, E.; and Srinivas, K. 2006. The summary ABox: Cutting ontologies down to size. In *Proc. of the 5th Int. Semantic Web Conference (ISWC 2006)*, volume 4273 of *LNCS*, 343–356. Springer.

Glimm, B.; Horrocks, I.; Motik, B.; Stoilos, G.; and Wang, Z. 2014a. Hermit: An OWL 2 reasoner. *J. Autom. Reasoning* 53(3):245–269.

Glimm, B.; Kazakov, Y.; Liebig, T.; Tran, T.; and Vialard, V. 2014b. Abstraction refinement for ontology materialization. In *Proc. of the 13th Int. Semantic Web Conference (ISWC 2014)*, 180–195. Springer.

Guo, Y.; Pan, Z.; and Heflin, J. 2005. LUBM: A benchmark for OWL knowledge base systems. *J. Web Sem.* 3(2-3):158–182.

Horrocks, I.; Kutz, O.; and Sattler, U. 2006. The even more irresistible SROIQ. In *Proc. of 10th International Conference on Principles of Knowledge Representation and Reasoning (KR 2006)*, 57–67.

Kalyanpur, A.; Parsia, B.; Horridge, M.; and Sirin, E. 2007. Finding all justifications of OWL DL entailments. In *Proc. of the 6th Int. Semantic Web Conf. (ISWC 2007)*, 267–280.

Kazakov, Y. 2009. Consequence-driven reasoning for Horn $\mathcal{SHIQ}$ ontologies. In *Proc. of the 21st Int. Joint Conf. on Artificial Intelligence (IJCAI 2009)*, 2040–2045.

Kontchakov, R.; Lutz, C.; Toman, D.; Wolter, F.; and Zakharyaschev, M. 2011. The combined approach to ontology-based data access. In *Proc. of the 22nd International Joint Conference on Artificial Intelligence, IJCAI 2011*, 2656–2661.

Krötzsch, M.; Rudolph, S.; and Hitzler, P. 2013. Complexities of horn description logics. *ACM Trans. Comput. Log.* 14(1):2.

Liebig, T.; Luther, M.; Noppens, O.; and Wessel, M. 2011. Owllink. *Semantic Web* 2(1):23–32.

Ma, L.; Yang, Y.; Qiu, Z.; Xie, G. T.; Pan, Y.; and Liu, S. 2006. Towards a complete OWL ontology benchmark. In *Proc. of European Semantic Web Conference, ESWC 2006*, 125–139.

Motik, B.; Nenov, Y.; Piro, R.; Horrocks, I.; and Olteanu, D. 2014. Parallel materialisation of datalog programs in centralised, main-memory RDF systems. In *Proc. of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 129–137.

Ortiz, M.; Rudolph, S.; and Simkus, M. 2010. Worst-case optimal reasoning for the horn-dl fragments of OWL 1 and 2. In *Proc. of the Twelfth International Conference on Principles of Knowledge Representation, KR 2010*.

Sirin, E.; Parsia, B.; Grau, B. C.; Kalyanpur, A.; and Katz, Y. 2007. Pellet: A practical OWL-DL reasoner. *J. Web Sem.* 5(2):51–53.

Steigmiller, A.; Liebig, T.; and Glimm, B. 2014. Konclude: System description. *J. Web Sem.* 27:78–85.

Urbani, J.; Kotoulas, S.; Maassen, J.; van Harmelen, F.; and Bal, H. E. 2012. WebPIE: A web-scale parallel inference engine using MapReduce. *J. Web Semantics* 10:59–75.

Wandelt, S., and Möller, R. 2012. Towards ABox modularization of semi-expressive description logics. *J. of Applied Ontology* 7(2):133–167.

Zhou, Y.; Grau, B. C.; Nenov, Y.; Kaminski, M.; and Horrocks, I. 2015. Pagoda: Pay-as-you-go ontology query answering using a datalog reasoner. *J. Artif. Intell. Res. (JAIR)* 54:309–367.