A Journey into Ontology Approximation: From Non-Horn to Horn

Anneke Haga[1] , **Carsten Lutz**[1] , **Johannes Marti**[2] and **Frank Wolter**[3]

[1]Universität Bremen, Germany
[2]Universiteit van Amsterdam, Netherlands
[3]University of Liverpool, UK

{anneke, clu}@uni-bremen.de, johannes.marti@gmail.com, wolter@liverpool.ac.uk

Abstract

We study complete approximations of an ontology formulated in a non-Horn description logic (DL) such as $\mathcal{ALC}$ in a Horn DL such as $\mathcal{EL}$. We provide concrete approximation schemes that are necessarily infinite and observe that in the $\mathcal{ELU}$-to-$\mathcal{EL}$ case finite approximations tend to exist in practice and are guaranteed to exist when the source ontology is acyclic. In contrast, neither of these are the case for $\mathcal{ELU}_\perp$-to-$\mathcal{EL}_\perp$ and for $\mathcal{ALC}$-to-$\mathcal{EL}_\perp$ approximations. We also define a notion of approximation tailored towards ontology-mediated querying, connect it to subsumption-based approximations, and identify a case where finite approximations are guaranteed to exist.

1 Introduction

Despite prominent standardization efforts such as OWL, a large variety of description logics (DLs) continues to be used as ontology languages. In fact, ontology designers choose a DL suitable for their purposes based on many factors including expressive power, computational properties, and tool support [Baader *et al.*, 2017]. Since ontology engineering frequently involves (partial) reuse of existing ontologies, this raises the problem of converting an ontology written in some source DL $\mathcal{L}_S$ into a desired target DL $\mathcal{L}_T$. A particularly important case is ontology approximation where $\mathcal{L}_T$ is a fragment of $\mathcal{L}_S$, studied for example in [Pan and Thomas, 2007; Ren *et al.*, 2010; Botoeva *et al.*, 2010; Carral *et al.*, 2014; Zhou *et al.*, 2015; Bötcher *et al.*, 2019].

In practice, ontology approximation is often done in an ad hoc way by dropping all statements from the source ontology $\mathcal{O}_S$ that are not expressible in $\mathcal{L}_T$, or at least the inexpressible parts of such statements. It is well-known that this results in incomplete approximations, that is, there will be knowledge in $\mathcal{O}_S$ that could be expressed in $\mathcal{L}_T$, but is not contained in the resulting approximated ontology. The degree and nature of the resulting incompleteness is typically neither understood nor analyzed. One reason for this unsatisfactory situation might be the fact that it is by no means easy to construct complete approximations and, even worse, finite complete approximations are not guaranteed to exist. This was studied in depth in [Bötcher *et al.*, 2019] where ontologies formulated

in expressive Horn DLs such as Horn-$\mathcal{SHIF}$ and $\mathcal{ELI}$ are approximated in tractable Horn DLs such as $\mathcal{EL}$. For example, it is shown there that finite complete $\mathcal{ELI}$-to-$\mathcal{EL}$ approximations do not exist even in extremely simple cases including those occurring in practice. The authors then lay out a new research program for ontology approximation that consists in mapping out the structure of complete (infinite) ontology approximations as a tool for guiding informed decisions when constructing incomplete (finite) approximations in practice, and also to enable a better understanding of the degree and nature of incompleteness.

In this paper, we consider $\mathcal{L}_S$-to-$\mathcal{L}_T$ ontology approximation where $\mathcal{L}_S$ is a non-Horn DL such as $\mathcal{ALC}$ and $\mathcal{L}_T$ is a tractable Horn DL such as $\mathcal{EL}$. Arguably, these are extremely natural cases of ontology approximation given that Horn vs. non-Horn is nowadays the most important classification criterion for DLs [Baader *et al.*, 2017]. Non-Horn DLs include expressive features such as negation and disjunction and require 'reasoning by cases' which is computationally costly, but also have considerably higher expressive power than Horn DLs. Horn DLs, in contrast, enjoy favourable properties such as the existence of universal models and of 'consequence-based' reasoning algorithms that avoid reasoning by cases [Cucala *et al.*, 2019]. Despite being natural, however, non-Horn-to-Horn approximation turns out to be a challenging endeavour.

We start with the fundamental case of $\mathcal{ELU}$-to-$\mathcal{EL}$ approximation. Given an $\mathcal{ELU}$ ontology $\mathcal{O}_S$, we aim to find a (potentially infinite) $\mathcal{EL}$ ontology $\mathcal{O}_T$ such that for all $\mathcal{EL}$ concepts C, D in the signature of $\mathcal{O}_S$, $\mathcal{O}_S \models C \sqsubseteq D$ iff $\mathcal{O}_T \models C \sqsubseteq D$.

Example 1. *Consider the $\mathcal{ELU}$ ontology*

$$\mathcal{O}_S = \{ \qquad \texttt{Job} \sqsubseteq \texttt{MainJob} \sqcup \texttt{SideJob} \\ \exists \texttt{job}.\texttt{SideJob} \sqsubseteq \exists \texttt{job}.(\texttt{MainJob} \sqcap \texttt{PartTime}) \}.$$

Then the following is an $\mathcal{EL}$ approximation of $\mathcal{O}_S$:

$$\mathcal{O}_T = \{ \quad \exists \texttt{job}.\texttt{SideJob} \sqsubseteq \exists \texttt{job}.(\texttt{MainJob} \sqcap \texttt{PartTime}) \\ \exists \texttt{job}.\texttt{Job} \sqsubseteq \exists \texttt{job}.\texttt{MainJob} \\ \exists \texttt{job}.(\texttt{Job} \sqcap \texttt{PartTime}) \sqsubseteq \exists \texttt{job}.(\texttt{MainJob} \sqcap \texttt{PartTime}) \}.$$

The last two lines of $\mathcal{O}_T$ illustrate that $\mathcal{EL}$ consequences of $\mathcal{ELU}$ ontologies can be rather non-obvious.

We first prove that finite approximations need not exist in the $\mathcal{ELU}$-to-$\mathcal{EL}$ case and that depth bounded approximations

may be non-elementary in size. Our main result is then a concrete approximation scheme that makes explicit the structure of complete infinite approximations and aims to keep as much structure of the source ontology as possible. An interesting and, given the results in [Bötcher *et al.*, 2019], surprising feature of our scheme is that it can be expected to often deliver *finite* approximations in practical cases. We perform a case study based on the Manchester ontology corpus that confirm this expectation. We also show that if $\mathcal{O}_S$ is an acyclic $\mathcal{ELU}$ ontology, then a finite $\mathcal{EL}$ approximation always exists (though it need not be acyclic). The finite approximations that we obtain are too large to be directly used in practice. Nevertheless, we view our results as positive and believe that in practice approximations of reasonable size often exist, as in Example 1. A 'push button technology' for constructing them, however, is outside of the scope of this paper.

We then proceed to the cases of $\mathcal{ELU}_\perp$-to-$\mathcal{EL}_\perp$ and $\mathcal{ALC}$-to-$\mathcal{EL}_\perp$ approximations which turn out to be closely related to each other. They also turn out to be significantly different from the $\mathcal{ELU}$-to-$\mathcal{EL}$ case in that finite approximations do not exist in extremely simple (and practical) cases, much like in the Horn approximation cases studied in [Bötcher *et al.*, 2019]. Also, finite approximations of acyclic ontologies are no longer guaranteed to exist. While this is not good news, it is remarkable that the addition of the $\perp$ symbol has such a dramatic effect. We again provide an (infinite) approximation scheme.

Finally, we propose a notion of approximation that is tailored towards applications in ontology-mediated querying [Calvanese *et al.*, 2009] and show that it is intimately related to the subsumption-based approximations that we had studied before. Remarkably, if we concentrate on atomic queries (AQs), then we obtain finite approximations even in the $\mathcal{ALC}$-to-$\mathcal{EL}_\perp$ case. Compared to the related work presented in [Kaminski *et al.*, 2016], we do not require the preservation of all query answers, but only of a maximal subset thereof, and our method is applicable to all ontologies formulated in the source DL chosen rather than to a syntactically restricted class. We also observe an interesting application to the rewritability of ontology-mediated queries.

All proofs are deferred to the appendix [Haga *et al.*, 2020].

2 Preliminaries

Let N_C and N_R be disjoint and countably infinite sets of *concept names* and *role names*. In the description logic $\mathcal{ALC}$, *concepts* C, D are built according to the syntax rule

$$C, D ::= \top \mid \perp \mid A \mid \neg C \mid C \sqcap D \mid C \sqcup D \mid \exists r.C \mid \forall r.C$$

where A ranges over N_C and r over N_R. The *depth* of a concept is the nesting depth of the constructors $\exists r$ and $\forall r$ in it. For example, the concept $\exists r.B \sqcap \exists r.\exists s.A$ is of depth 2. We introduce other DLs as fragments of $\mathcal{ALC}$. An $\mathcal{ELU}_\perp$ *concept* is an $\mathcal{ALC}$ concept that does not contain negations $\neg C$ and value restrictions $\forall r.C$. An $\mathcal{EL}_\perp$ *concept* is an $\mathcal{ELU}_\perp$ concept that does not contain disjunctions $C \sqcup D$. $\mathcal{ELU}$ concepts and $\mathcal{EL}$ concepts are defined likewise, but additionally forbid the use of the bottom concept $\perp$.

For any of these DLs $\mathcal{L}$, an $\mathcal{L}$ *ontology* is a set of *concept inclusions (CIs)* $C \sqsubseteq D$ where C and D are $\mathcal{L}$ concepts. While

ontologies used in practice have to be finite, we frequently consider also infinite ontologies. W.l.o.g., we assume that all occurrences of $\perp$ in $\mathcal{ELU}_\perp$ ontologies are in CIs of the form $C \sqsubseteq \perp$, where C does not contain $\perp$. An *acyclic ontology* $\mathcal{O}$ is a set of concept inclusions $A \sqsubseteq C$ and *concept equivalences* $A \equiv C$ where A is a concept name (that is, it is not a compound concept), the left-hand sides are unique, and $\mathcal{O}$ does not contain a definitiorial cycle $A_0 \bowtie_1 C_0, \dots, A_n \bowtie_n C_n, \bowtie_i \in \{\sqsubseteq, \equiv\}$, where C_i contains $A_{i+1 \bmod n+1}$ for all $i \le n$. An equivalence $A \equiv C$ can be viewed as two CIs $A \sqsubseteq C$ and $C \sqsubseteq A$ and thus every acyclic ontology is an ontology in the original sense.

A *signature* Σ is a set of concept and role names, uniformly referred to as *symbols*. We use $\mathtt{sig}(X)$ to denote the set of symbols used in any syntactic object X such as a concept or an ontology. If $\mathtt{sig}(X) \subseteq \Sigma$, we also say that X is *over* Σ. The *size* of a (finite) syntactic object X, denoted $||X||$, is the number of symbols needed to write it, with every occurrence of a concept and role name contributing one.

The semantics of concepts and ontologies is defined in terms of *interpretations* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ as usual, see [Baader *et al.*, 2017]. An interpretation $\mathcal{I}$ *satisfies* a CI $C \sqsubseteq D$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$, an equivalence $A \equiv C$ if $A^\mathcal{I} = C^\mathcal{I}$, and it is a *model* of an ontology $\mathcal{O}$ if it satisfies all CIs in $\mathcal{O}$. Concept C is *subsumed* by concept D w.r.t. ontology $\mathcal{O}$, written $\mathcal{O} \models C \sqsubseteq D$, if every model $\mathcal{I}$ of $\mathcal{O}$ satisfies the CI $C \sqsubseteq D$; we then also say that the CI is a *consequence* of $\mathcal{O}$. Subsumption can be decided in polynomial time in $\mathcal{EL}_\perp$ and is EXPTIME-complete between $\mathcal{ELU}$ and $\mathcal{ALC}$ [Baader *et al.*, 2017]. We now give our main definition of approximation. With concept of depth bounded by ω, we mean concepts of unrestricted depth.

Definition 1. *Let $\mathcal{O}_S$ be an $\mathcal{ALC}$ ontology, $sig(\mathcal{O}_S) = \Sigma$, $\mathcal{L}_T$ any of the DLs introduced above, and $\ell \in \mathbb{N} \cup \{\omega\}$. A (potentially infinite) $\mathcal{L}_T$ ontology $\mathcal{O}_T$ is an ℓ-bounded $\mathcal{L}_T$ approximation of $\mathcal{O}_S$ if*

$$\mathcal{O}_S \models C \sqsubseteq D \text{ iff } \mathcal{O}_T \models C \sqsubseteq D$$

for all $\mathcal{L}_T$ concepts C, D over Σ of depth bounded by ℓ. $\mathcal{O}_T$ is non-projective if $\mathtt{sig}(\mathcal{O}_T) \subseteq \Sigma$ and projective otherwise. We refer to ω-bounded $\mathcal{L}_T$ approximations as $\mathcal{L}_T$ approximations.

We refer to the "if" direction of the biimplication in Definition 1 as *soundness* of the approximation and to the "only if" direction as *completeness*. Infinite approximations always exist: take as $\mathcal{O}_T$ the set of all $\mathcal{L}$ CIs $C \sqsubseteq D$ with C, D over Σ and $\mathcal{O}_S \models C \sqsubseteq D$. In the same way, finite (non-projective) depth-bounded approximations always exist. With $\mathcal{L}_S$-to-$\mathcal{L}_T$ *approximation*, $\mathcal{L}_S$ a DL and $\mathcal{L}_T$ a fragment of $\mathcal{L}_S$, we mean the task to approximate an $\mathcal{L}_S$ ontology in $\mathcal{L}_T$, possibly using an infinite ontology.

3 $\mathcal{ELU}$-to-$\mathcal{EL}$ Approximation

We consider $\mathcal{ELU}$-to-$\mathcal{EL}$ approximation as the simplest case of approximating non-Horn ontologies in a Horn DL.

Fundamentals. We start with observing that projective approximations are more powerful than non-projective ones.

Proposition 1. *The $\mathcal{ELU}$ ontology*

$$\mathcal{O}_S = \{ \quad A \quad \sqsubseteq \quad B_1 \sqcup B_2,$$
$$\exists r.B_i \quad \sqsubseteq \quad B_i, \qquad \text{for } i \in \{1, 2\}$$
$$B_i \sqcap A' \quad \sqsubseteq \quad M \qquad \text{for } i \in \{1, 2\} \}.$$

$$
\begin{aligned}
C &\sqsubseteq X_C \\
X_{D_1} \sqcap C &\sqsubseteq X_{D_2} && \text{if } \mathcal{O}_S \models D_1 \sqcap C \sqsubseteq D_2 \\
X_{D_1} \sqcap X_{D_2} &\sqsubseteq X_{D_3} && \text{if } \mathcal{O}_S \models D_1 \sqcap D_2 \sqsubseteq D_3 \\
\exists r.X_{D_1} &\sqsubseteq X_{D_2} && \text{if } \mathcal{O}_S \models \exists r.D_1 \sqsubseteq D_2 \\
X_{D_1} &\sqsubseteq \exists r.X_{D_2} && \text{if } \mathcal{O}_S \models D_1 \sqsubseteq \exists r.D_2 \\
X_{D_1} &\sqsubseteq C && \text{if } \mathcal{O}_S \models D_1 \sqsubseteq C
\end{aligned}
$$

Figure 1: Candidate $\mathcal{EL}$ approximation $\mathcal{O}_T$.

has a finite projective $\mathcal{EL}$ approximation, but every non-projective $\mathcal{EL}$ approximation is infinite.

In fact, a finite projective $\mathcal{EL}$ approximation $\mathcal{O}_T$ of the ontology $\mathcal{O}_S$ from Proposition 1 is obtained from $\mathcal{O}_S$ by replacing the CI in the first line with

$$A \sqsubseteq X_{B_1 \sqcup B_2}, \ \exists r.X_{B_1 \sqcup B_2} \sqsubseteq X_{B_1 \sqcup B_2}, \ X_{B_1 \sqcup B_2} \sqcap A' \sqsubseteq M.$$

The intuitive reason for why $\mathcal{O}_S$ has no finite non-projective $\mathcal{EL}$ approximation is that $\mathcal{O}_S \models A' \sqcap \exists r^n.A \sqsubseteq M$ for all $n \geq 0$. Proposition 1 indicates that projective approximations are preferable. Since they also seem perfectly acceptable from an application viewpoint, we concentrate on the projective case and from now on mean projective approximations whenever we speak of approximations.

To illustrate the challenges of $\mathcal{ELU}$-to-$\mathcal{EL}$ approximation, it is instructive to consider a candidate approximation scheme that might be suggested by Proposition 1. We use $\mathtt{sub}(\mathcal{O}_S)$ to denote the set of all subconcepts of (concepts in) the ontology $\mathcal{O}_S$ and $\mathtt{sub}^-(\mathcal{O}_S)$ to denote the restriction of $\mathtt{sub}(\mathcal{O}_S)$ to concept names and existential restrictions $\exists r.C$. We use $\mathtt{Con}(\mathcal{O}_S)$ to denote the set of all non-empty conjunctions of concepts from $\mathtt{sub}^-(\mathcal{O}_S)$ without repetitions and $\mathtt{Dis}(\mathcal{O}_S)$ to mean the set of all disjunctions of concepts from $\mathtt{Con}(\mathcal{O}_S)$ without repetitions. Now, a (finite projective) candidate $\mathcal{EL}$ approximation scheme is given in Figure 1 where C ranges over $\mathtt{sub}(\mathcal{O}_S)$ and D_1, D_2, D_3 range over $\mathtt{Dis}(\mathcal{O}_S)$. It indeed yields an approximation when applied to the ontology $\mathcal{O}_S$ in Proposition 1. There are, however, two major problems. First, the syntactic structure of $\mathcal{O}_S$ is lost completely, which is undesirable in practice where ontologies are the result of a careful modeling effort. We could include all $\mathcal{EL}$ concept inclusions from $\mathcal{O}_S$ in the approximation, but this would be purely cosmetic since all such CIs are already implied. Second, the approximation is incomplete in general. In fact, finite approximations need not exist also in the projective case while the approximation scheme in Figure 1 is always finite.

Proposition 2. *The $\mathcal{ELU}$ ontology*

$$
\begin{aligned}
\mathcal{O}_S = \{ \quad A &\sqsubseteq B_1 \sqcup B_2, \\
\exists r.B_2 &\sqsubseteq \exists r.(B_1 \sqcap L), \\
L &\sqsubseteq \exists s.L \quad \}
\end{aligned}
$$

has no finite $\mathcal{EL}$ approximation.

The intuitive reason for why $\mathcal{O}_S$ has no finite $\mathcal{EL}$ approximation is that $\mathcal{O}_S \models \exists r.(A \sqcap \exists s^n.\top) \sqsubseteq \exists r.(B_1 \sqcap \exists s^n.\top)$ for all $n \geq 0$.

The ontology in Proposition 2 can be varied to show that even bounded depth approximations can get very large. The function $\mathtt{tower} : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ is defined as $\mathtt{tower}(0, n) := n$ and $\mathtt{tower}(k+1, n) := 2^{\mathtt{tower}(k,n)}$.

$$
\begin{aligned}
C &\sqsubseteq \mathtt{DNF}(E)^{\uparrow} && \text{if } C \sqsubseteq E \in \mathcal{O}_S \\
X_D \sqcap D_1^{\uparrow} &\sqsubseteq D_2^{\uparrow} && \text{if } \mathcal{O}_S \models D \sqcap D_1 \sqsubseteq D_2 \\
\exists r.X_D &\sqsubseteq D_1^{\uparrow} && \text{if } \mathcal{O}_S \models \exists r.D \sqsubseteq D_1 \\
F^{\uparrow} &\sqsubseteq \exists r.G && \text{if } \mathcal{O}_S \models F \sqsubseteq \exists r.G
\end{aligned}
$$

where in the last line

- F is an $\mathcal{EL}$ concept over $\mathtt{sig}(\mathcal{O}_S)$ decorated with disjunctions from $\mathtt{Dis}(\mathcal{O}_S)$ at leaves and

- G is an $\mathcal{O}_S$-generatable $\mathcal{EL}$ concept over $\mathtt{sig}(\mathcal{O}_S)$

such that $\mathtt{depth}(F) \leq \mathtt{depth}(G) < \ell$.

Figure 2: ℓ-bounded $\mathcal{EL}$ approximation $\mathcal{O}_T^{\ell}$.

Proposition 3. *Let $\mathcal{O}_S^n$ be obtained from the ontology $\mathcal{O}_S$ in Proposition 2 by replacing the bottommost CI with*

$$L \sqsubseteq A_1 \sqcap \hat{A}_1 \sqcap \cdots \sqcap A_n \sqcap \hat{A}_n \sqcap \exists r_1.L \sqcap \exists r_2.L$$

Then for all $n, \ell \geq 1$ and any ℓ-bounded $\mathcal{EL}$ approximation $\mathcal{O}_T$ of $\mathcal{O}_S^n$, $||\mathcal{O}_T|| \geq \mathtt{tower}(\ell, n)$.

A Complete Approximation. We present a more careful approximation scheme that aims to preserve the structure of $\mathcal{O}_S$, is complete, and yields a finite approximation in many practical cases. Let $\mathcal{O}_S$ be an $\mathcal{ELU}$ ontology to be approximated. As a preliminary, we assume that for all CIs $C \sqsubseteq D \in \mathcal{O}_S$, C is an $\mathcal{EL}$ concept. If this is not the case, then we can rewrite $\mathcal{O}_S$ by exhaustively replacing every disjunction $C \sqcup D$ that occurs (possibly as a subconcept) on the left-hand side of a concept inclusion in $\mathcal{O}_S$ with a fresh concept name $X_{C \sqcup D}$ and adding the inclusions $C \sqsubseteq X_{C \sqcup D}$ and $D \sqsubseteq X_{C \sqcup D}$. It is not hard to see that the resulting ontology $\mathcal{O}_S'$ is of size polynomial in $||\mathcal{O}_S||$ and a conservative extension of $\mathcal{O}_S$ in the sense that $\mathcal{O}_S \models C \sqsubseteq D$ iff $\mathcal{O}_S' \models C \sqsubseteq D$ for all $\mathcal{ELU}$ concepts C, D over $\mathtt{sig}(\mathcal{O}_S)$. Consequently, every $\mathcal{EL}$ approximation of $\mathcal{O}_S'$ is also a projective $\mathcal{EL}$ approximation of $\mathcal{O}_S$ and we can work with $\mathcal{O}_S'$ in place of $\mathcal{O}_S$.

Let $\ell \in \mathbb{N} \cup \{\omega\}$. The proposed $\mathcal{EL}$ approximation $\mathcal{O}_T^{\ell}$ of $\mathcal{O}_S$ is given in Figure 2 where D_1, D_2 range over $\mathtt{Dis}(\mathcal{O}_S)$ and D ranges over $\mathtt{Dis}^-(\mathcal{O}_S)$, the set of all disjunctions in $\mathtt{Dis}(\mathcal{O}_S)$ that have at least two disjuncts. We still have to define the notation and terminology used in the figure. For an $\mathcal{ELU}$ concept C such that all disjunctions in C are from $\mathtt{Dis}(\mathcal{O}_S)$, we use $C^{\uparrow}$ to denote the $\mathcal{EL}$ concept obtained from C by replacing every outermost $D \in \mathtt{Dis}^-(\mathcal{O}_S)$ with a fresh concept name X_D. Set $\mathtt{DNF}(C) = C$ if C is a concept name or of the form $\exists r.D$, $\mathtt{DNF}(C_1 \sqcap C_2) = \mathtt{DNF}(C_1) \sqcap \mathtt{DNF}(C_2)$, and define $\mathtt{DNF}(C_1 \sqcup C_2)$ to be the $\mathcal{ELU}$-concept obtained by converting $C_1 \sqcup C_2$ into disjunctive normal form (DNF), treating existential restrictions $\exists r.D$ as atomic concepts, that is, the argument D is not modified. Note that while $||\mathtt{DNF}(C)||$ may be exponential in $||C||$, we have $||\mathtt{DNF}(C)^{\uparrow}|| \leq ||C||$. By *decorating an $\mathcal{EL}$ concept C with disjunctions from $\mathtt{Dis}(\mathcal{O}_S)$ at leaves*, we mean to replace subconcepts $\exists r.E$ of C with E of depth 0 by $\exists r.(E \sqcap D)$, $D \in \mathtt{Dis}(\mathcal{O}_S)$. As a special case, we can replace C with $C \sqcap D$, $D \in \mathtt{Dis}(\mathcal{O}_S)$, if C is of depth 0.

Definition 2. *An $\mathcal{EL}$ concept C is $\mathcal{O}_S$-generatable if there is an $\exists r.D \in \mathtt{sub}(\mathcal{O}_S)$ that occurs on the right-hand side of a CI in $\mathcal{O}_S$ and satisfies $\mathcal{O}_S \models D \sqsubseteq C$.*

Let us explain the proposed approximation. The first three lines of Figure 2 can be viewed as a more careful version of the first four lines of Figure 1. In the first line, we preserve the structure of $\mathcal{O}_S$ as long as it lies outside the scope of a disjunction operator, thanks to the careful definition of $\mathrm{DNF}(C)$. This is not cosmetic as in the candidate approximation in Figure 1: since we introduce the concept names X_D only when a disjunction is 'derived' (first line) and only for disjunctions $D \in \mathrm{Dis}^-(\mathcal{O}_S)$, $\mathcal{O}_T^\ell$ is no longer guaranteed to be an approximation when the first line in Figure 2 is dropped. The last line of the approximation addresses the effect illustrated by Proposition 2. It is strong enough so that a counterpart of the second last line in Figure 1 is not needed. An example application of our approximation scheme is given in [Haga *et al.*, 2020].

An interesting aspect of our approximation is that it turns out to be finite in many practical cases. In fact, it is easy to see that $\mathcal{O}_T^\ell$ is finite for all $\ell < \omega$ and that $\mathcal{O}_T^\omega$ is finite if and only if there are only finitely many $\mathcal{EL}$ concepts that are $\mathcal{O}_S$-generatable, up to logical equivalence; we then say that $\mathcal{O}_S$ is *finitely generating*. Since ontologies from practice tend to have a simple structure, one might expect that they often enjoy this property. Below, we report about a case study that confirms this expectation.

How does the approximation scheme in Figure 2 relate to the examples given above? For the ontologies $\mathcal{O}_S$ in Example 1 and in Proposition 1, our approximation $\mathcal{O}_T^\omega$ contains all CIs in the approximation $\mathcal{O}_T$ given in place. Of course, $\mathcal{O}_T^\omega$ also contains a lot of additional CIs that, however, do not result in any new consequences $C \sqsubseteq D$ with C, D $\mathcal{EL}$ concepts over $\mathrm{sig}(\mathcal{O}_S)$. It seems very difficult to identify up front those CIs that are really needed. We can remove them after constructing $\mathcal{O}_T^\omega$ by repeatedly deciding conservative extensions [Lutz and Wolter, 2010], but this is not practical given the size of $\mathcal{O}_T^\omega$. Nevertheless, both ontologies $\mathcal{O}_S$ are finitely generating and thus in both cases $\mathcal{O}_T^\omega$ is finite. In Example 1, the $\mathcal{O}_S$-generatable concepts are $\top$, `MainJob`, `PartTime`, and `MainJob` $\sqcap$ `PartTime` (up to logical equivalence) while there are no $\mathcal{O}_S$-generatable concepts for Proposition 1. For Proposition 2, there are infinitely many $\mathcal{O}_S$-generatable concepts such as $\exists s^n.\top$ for all $n \geq 0$.

Case Study. We have considered the seven non-trivial $\mathcal{EL}\mathcal{U}$ ontologies that are part of the Manchester OWL corpus.[1] The size of the ontologies ranges from 113 to 813 concept inclusions and equalities. All ontologies use disjunction on the right-hand side of CIs (thus in a non-trivial way) and none of them is acyclic. We have been able to prove that all these ontologies are finitely generating and thus the approximation $\mathcal{O}_T^\omega$ is finite. Our proof relies on the following observation.

Lemma 1. $\mathcal{O}_S$ *is not finitely generating iff for every* $n \geq 0$, *there is an* $\exists r.D \in \mathrm{sub}(\mathcal{O}_S)$ *that occurs on the right-hand side of a CI and a sequence* $r_1, \ldots, r_n$ *of role names from* $\mathcal{O}_S$ *such that* $\mathcal{O}_S \models D \sqsubseteq \exists r_1. \cdots . \exists r_n.\top$.

In our implementation, we use role inclusions to avoid going through all of the exponentially many sequences $r_1, \ldots, r_n$. Lemma 1 can also be used to show the following.

---

[1] http://owl.cs.manchester.ac.uk/publications/supporting-material/owlcorpus/

Theorem 1. *It is decidable whether a given* $\mathcal{EL}\mathcal{U}$-*ontology* $\mathcal{O}_S$ *is finitely generating.*

By what was said above, this implies that it is decidable whether the approximation $\mathcal{O}_T^\omega$ from Figure 2 is finite.

Soundness and Completeness. We now establish soundness and completeness of the proposed approximation, the main result in this section.

Theorem 2. *For every* $\ell \in \mathbb{N} \cup \{\omega\}$, $\mathcal{O}_T^\ell$ *is an* ℓ-*bounded* $\mathcal{EL}$ *approximation of* $\mathcal{O}_S$.

While soundness is easy to show, completeness is remarkably subtle to prove. It is stated by the following lemma which shows that our approximation $\mathcal{O}_T^\ell$ is actually stronger than required in that it preserves all $\mathcal{EL}$ subsumptions $C \sqsubseteq D$ with D of depth bounded by ℓ and C of unrestricted depth.

Lemma 2. *Let* $\ell \in \mathbb{N} \cup \{\omega\}$. *Then* $\mathcal{O}_S \models C_0 \sqsubseteq D_0$ *implies* $\mathcal{O}_T^\ell \models C_0 \sqsubseteq D_0$ *for all* $\mathcal{EL}$ *concepts* C_0, D_0 *over* $\mathrm{sig}(\mathcal{O}_S)$ *such that the role depth of* D_0 *is bounded by* ℓ.

The proof of Lemma 2 is the most substantial one in this paper. It uses a chase procedure for $\mathcal{EL}\mathcal{U}$ ontologies that is specifically tailored towards proving completeness in that it is deterministic rather than disjunctive and mimics the concept inclusions in Figure 2. Showing that this chase is complete is far from trivial.

Fewer Symbols. The number of fresh concept names X_D in $\mathcal{O}_T^\ell$ is double exponential in $||\mathcal{O}_S||$ since the number of disjunctions in $\mathrm{Dis}^-(\mathcal{O})$ is. However, $\mathcal{O}_T^\ell$ can be rewritten into an ontology $\widehat{\mathcal{O}}_T^\ell$ that uses only single exponentially many fresh concept names and is still an ℓ-bounded approximation of $\mathcal{O}_S$. The idea is to transition from disjunctive normal form to conjunctive normal form, that is, to replace each concept name X_D, $D \in \mathrm{Dis}^-(\mathcal{O})$, with a conjunction of concept names $Y_{D'}$ where D' is a disjunction of concepts from $\mathrm{sub}^-(\mathcal{O})$, rather than conjunctions thereof. Details are in [Haga *et al.*, 2020].

Theorem 3. *For every* $\ell \in \mathbb{N} \cup \{\omega\}$, $\widehat{\mathcal{O}}_T^\ell$ *is an* ℓ-*bounded* $\mathcal{EL}$ *approximation of* $\mathcal{O}_S$.

Acyclic Ontologies. Using Lemma 1, one can show that $\mathcal{O}_T^\omega$ is finite whenever $\mathcal{O}_S$ is an acyclic $\mathcal{EL}\mathcal{U}$ ontology. In fact, the length n of role sequences with the properties stated in the lemma is bounded by $||\mathcal{O}_S||$ if $\mathcal{O}_S$ is acyclic.

Theorem 4. *Every acyclic* $\mathcal{EL}\mathcal{U}$ *ontology has a finite* $\mathcal{EL}$ *approximation.*

There is, however, more that we can say about acyclic ontologies. We first observe that there are acyclic $\mathcal{EL}\mathcal{U}$ ontologies that have finite $\mathcal{EL}$ approximations, but no $\mathcal{EL}$ approximation that is an acyclic ontology.

Example 2. *Consider the acyclic* $\mathcal{EL}\mathcal{U}$ *ontology*

$$\mathcal{O}_S = \{A \equiv (B_1 \sqcap B_2) \sqcup (B_1 \sqcap B_3)\}.$$

Then $\mathcal{O}_T = \{B_1 \sqcap B_2 \sqsubseteq A, B_1 \sqcap B_3 \sqsubseteq A, A \sqsubseteq B_1\}$ *is an* $\mathcal{EL}$ *approximation of* $\mathcal{O}_S$, *but* $\mathcal{O}_S$ *has no* $\mathcal{EL}$ *approximation that is an acyclic ontology, finite or infinite.*

Further, our approximations $\mathcal{O}_T^\ell$ can be simplified for acyclic $\mathcal{EL}\mathcal{U}$ ontologies $\mathcal{O}_S$. Let $\widetilde{\mathcal{O}}_T^\ell$ be defined like $\mathcal{O}_T^\ell$

in Figure 2, except that in the last line, F ranges only over concept names (not decorated with disjunctions) rather than over compound concepts, a significant simplification.

Theorem 5. *Let $\ell \in \mathbb{N} \cup \{\omega\}$ and let $\mathcal{O}_S$ be an acyclic $\mathcal{ELU}$ ontology. Then $\widetilde{\mathcal{O}}_T^{\ell}$ is an ℓ-bounded $\mathcal{EL}$ approximation of $\mathcal{O}_S$.*

Based on this observation, constructing finite $\mathcal{EL}$ approximations of acyclic $\mathcal{ELU}$ ontologies does not seem infeasible in practice.

4 $\mathcal{ALC}$-to-$\mathcal{EL}_{\perp}$ Approximation

We consider $\mathcal{ELU}_{\perp}$-to-$\mathcal{EL}_{\perp}$ and $\mathcal{ALC}$-to-$\mathcal{EL}_{\perp}$ approximation which turn out to be closely related to each other and significantly different from $\mathcal{ELU}$-to-$\mathcal{EL}$ approximation.

It immediately follows from the results in Section 3 that finite approximations are guaranteed to exist neither in the $\mathcal{ELU}_{\perp}$-to-$\mathcal{EL}_{\perp}$ nor in the $\mathcal{ALC}$-to-$\mathcal{EL}_{\perp}$ case. However, while we have argued that finite $\mathcal{ELU}$-to-$\mathcal{EL}$ approximations can be expected to exist in many practical cases, this does not appear to be true for $\mathcal{ELU}_{\perp}$-to-$\mathcal{EL}_{\perp}$ and $\mathcal{ALC}$-to-$\mathcal{EL}_{\perp}$. The following example illustrates the problem.

Example 3. *Consider the $\mathcal{ELU}_{\perp}$ ontology*

$$\mathcal{O}_S = \{ \qquad \begin{aligned} A_1 &\sqsubseteq M \sqcup N_1, \\ A_2 &\sqsubseteq M \sqcup N_2, \\ \exists r.N_1 \sqcap \exists r.N_2 &\sqsubseteq \perp \end{aligned} \qquad \}.$$

There are no $\mathcal{O}_S$-generatable $\mathcal{EL}$ concepts. Yet, there is no finite $\mathcal{EL}_{\perp}$ approximation of $\mathcal{O}_S$. Informally, this is because

$$\mathcal{O}_S \models \exists r.(A_1 \sqcap \exists r^n.\top) \sqcap \exists r.(A_2 \sqcap \exists r^n.\top) \sqsubseteq \exists r.(M \sqcap \exists r^n.\top)$$

for all $n \geq 1$.[2]

While the above example is for $\mathcal{ELU}_{\perp}$-to-$\mathcal{EL}_{\perp}$, there is an additional effect in $\mathcal{ALC}$-to-$\mathcal{EL}_{\perp}$ that already occurs for very simple ontologies $\mathcal{O}_S$.

Example 4. *The $\mathcal{ALC}$ ontology $\mathcal{O}_S = \{A \sqsubseteq \forall r.B\}$ has no finite $\mathcal{EL}_{\perp}$ approximation. This is shown in [Bötcher et al., 2019] for the equivalent $\mathcal{ELI}$ ontology $\{\exists r^-.A \sqsubseteq B\}$. Informally, this is because $\mathcal{O}_S \models A \sqcap \exists r^{n+1}.\top \sqsubseteq \exists r.(B \sqcap \exists r^n.\top)$ for all $n \geq 1$.*

Note that the ontology $\mathcal{O}_S$ in Example 4 is acyclic and thus in contrast to the $\mathcal{ELU}$-to-$\mathcal{EL}$ case, finite $\mathcal{EL}_{\perp}$ approximations of acyclic $\mathcal{ALC}$ ontologies need not exist. In a sense, Example 3 shows the same negative result for the $\mathcal{ELU}_{\perp}$-to-$\mathcal{EL}_{\perp}$ case. While the ontology used there is not strictly acyclic, acyclic ontologies do not make much sense in the case of $\mathcal{ELU}_{\perp}$ and additionally admitting CIs $C_1 \sqcap C_2 \sqsubseteq \perp$ as used in Example 3 seems to be the most modest extension possible that incorporates $\perp$ in a meaningful way.

Despite these additional challenges, we can extend the approximation given in Section 3 to $\mathcal{ELU}_{\perp}$-to-$\mathcal{EL}_{\perp}$ and to $\mathcal{ALC}$-to-$\mathcal{EL}_{\perp}$ when we are willing to drop $\mathcal{O}_S$-generatability and, as a consequence, accept the fact that approximations are infinite unless they are depth bounded. Note that the latter is also the case in $\mathcal{L}$-to-$\mathcal{EL}$ approximation where $\mathcal{L}$ is an expressive Horn DL such as $\mathcal{ELI}$ [Bötcher et al., 2019].

---

[2]A formal proof is analogous to that of Proposition 2.

$$\begin{array}{rcll} C &\sqsubseteq& \mathtt{DNF}(E)^{\uparrow} & \text{if } C \sqsubseteq E \in \mathcal{O}_S \\ X_D \sqcap D_1^{\uparrow} &\sqsubseteq& D_2^{\uparrow} & \text{if } \mathcal{O}_S \models D \sqcap D_1 \sqsubseteq D_2 \\ \exists r.X_D &\sqsubseteq& D_1^{\uparrow} & \text{if } \mathcal{O}_S \models \exists r.D \sqsubseteq D_1 \\ F^{\uparrow} &\sqsubseteq& \exists r.G & \text{if } \mathcal{O}_S \models F \sqsubseteq \exists r.G \end{array}$$

where in the last line F is an $\mathcal{EL}$ concept over $\mathtt{sig}(\mathcal{O}_S)$ decorated with disjunctions from $\mathtt{Dis}(\mathcal{O}_S)$ at leaves and G is an $\mathcal{EL}$ concept over $\mathtt{sig}(\mathcal{O}_S)$ such that

1. F has no top-level conjunct $\exists r.F'$ s.t. $\mathcal{O}_S \models F' \sqsubseteq G$;
2. $\mathtt{depth}(F) \leq \mathtt{depth}(G) < \ell$.

Figure 3: ℓ-bounded $\mathcal{EL}_{\perp}$ approximation $\mathcal{O}_T^{\ell}$.

We first reduce $\mathcal{ALC}$-to-$\mathcal{EL}_{\perp}$ approximations to $\mathcal{ELU}_{\perp}$-to-$\mathcal{EL}_{\perp}$ approximations. Let $\mathcal{O}_S$ be an $\mathcal{ALC}$ ontology. We can transform $\mathcal{O}_S$ into an $\mathcal{ELU}_{\perp}$ ontology as follows:

1. replace each subconcept $\forall r.C$ with $\neg \exists r.\neg C$;
2. select a concept $\neg C$ such that C contains no negation, replace all occurrences of $\neg C$ with the fresh concept name $A_{\neg C}$, and add the CIs $\top \sqsubseteq C \sqcup A_{\neg C}$ and $C \sqcap A_{\neg C} \sqsubseteq \perp$; repeat until no longer possible.

The resulting ontology $\mathcal{O}_S'$ is of size polynomial in $\|\mathcal{O}_S\|$ and a conservative extension of $\mathcal{O}_S$ in the sense that $\mathcal{O}_S \models C \sqsubseteq D$ iff $\mathcal{O}_S' \models C \sqsubseteq D$ for all $\mathcal{ALC}$ concepts C, D over $\mathtt{sig}(\mathcal{O}_S)$. Consequently, every $\mathcal{EL}_{\perp}$ approximation of $\mathcal{O}_S'$ is also a (projective) $\mathcal{EL}_{\perp}$ approximation of $\mathcal{O}_S$.

It thus suffices to consider $\mathcal{ELU}_{\perp}$-to-$\mathcal{EL}_{\perp}$ approximations. Thus let $\mathcal{O}_S$ be an $\mathcal{ELU}_{\perp}$ ontology. For each $\ell \in \mathbb{N} \cup \{\omega\}$, the $\mathcal{EL}_{\perp}$ approximation $\mathcal{O}_T^{\ell}$ of $\mathcal{O}_S$ is given in Figure 2 where again D ranges over $\mathtt{Dis}^-(\mathcal{O}_S)$ and D_1, D_2 range over $\mathtt{Dis}(\mathcal{O}_S)$; both $\mathtt{Dis}(\mathcal{O}_S)$ and $\mathtt{Dis}^-(\mathcal{O}_S)$ are defined exactly as for $\mathcal{ELU}$ ontologies and in $\mathtt{DNF}(C)$ we drop all disjuncts that contain $\perp$ as a conjunct, possibly resulting in the empty disjunction (which represents $\perp$). Point 1 can be viewed as an optimization that sometimes helps to avoid the expensive last line. There, a *top-level conjunct* means a concept F_i if F takes the form $F_1 \sqcap \cdots \sqcap F_n$, $n \geq 1$. In [Haga et al., 2020], we point out another non-trivial such optimization.

Theorem 6. *$\mathcal{O}_T^{\ell}$ is an ℓ-bounded $\mathcal{EL}_{\perp}$ approximation of $\mathcal{O}_S$.*

The proof of Theorems 2 and 6 also establishes another result that will turn out to be interesting in the context of ontology-mediated queries in Section 5. We use $\mathcal{O}_T^-$ to denote the restriction of $\mathcal{O}_T^{\omega}$ to the (instantiations) of the first three lines in Figure 3 (equivalently: Figure 2). Clearly, $\mathcal{O}_T^-$ is always finite.

Theorem 7. *Let C_0, D_0 be $\mathcal{EL}_{\perp}$ concepts with $D_0 \in \mathtt{sub}(\mathcal{O}_S)$. Then $\mathcal{O}_S \models C_0 \sqsubseteq D_0$ iff $\mathcal{O}_T^- \models C_0 \sqsubseteq D_0$.*

5 Approximations and Query Evaluation

The notion of approximations given in Section 2 is tailored towards preserving subsumptions. In ontology-mediated querying, in contrast, the main aim of approximation is to preserve as many query answers as possible. We propose a suitable notion of approximation and show that the results obtained in the previous sections have interesting applications also in ontology-mediated querying.

Let N_I be a countably infinite set of *individual names* disjoint from N_C and N_R. An *ABox* is a finite set of *concept assertions* $A(a)$ and *role assertions* $r(a,b)$ where $A \in N_C$, $r \in N_R$, and $a, b \in N_I$. We use $\mathrm{Ind}(\mathcal{A})$ to denote the set of individual names in the ABox $\mathcal{A}$. An interpretation $\mathcal{I}$ *satisfies* a concept assertion $A(a)$ if $a \in A^{\mathcal{I}}$ and a role assertion $r(a,b)$ if $(a,b) \in r^{\mathcal{I}}$. It is a *model* of an ABox if it satisfies all assertions in it. A Σ-*ABox* is an ABox $\mathcal{A}$ with $\mathrm{sig}(\mathcal{A}) = \Sigma$.

An *ontology-mediated query (OMQ)* is a triple $Q = (\mathcal{O}, \Sigma, q)$ with $\mathcal{O}$ an ontology, $\Sigma \subseteq \mathrm{sig}(\mathcal{O}) \cup \mathrm{sig}(q)$ an ABox signature, and q an actual query. While conjunctive queries (CQs) and unions of CQs are a popular choice for formulating q and our central Definition 3 below makes sense also for these richer query languages, for simplicity we concentrate on *atomic queries (AQs)* $A(x)$ where A is a concept name and on $\mathcal{EL}$ *queries (ELQs)* $C(x)$ where C an $\mathcal{EL}$ concept. We also mention $\mathcal{ALC}$ *queries (ALCQs)* $C(x)$ where C is an $\mathcal{ALC}$ concept. Note that all such queries are unary. We use $\mathrm{ELQ}(\Sigma)$ to denote the language of all ELQs that use only symbols from signature Σ. Let $(\mathcal{L}, \mathcal{Q})$ denote the *OMQ language* that contains all OMQs Q in which $\mathcal{O}$ is formulated in DL $\mathcal{L}$ and q in query language $\mathcal{Q}$, such as in $(\mathcal{EL}, \mathrm{AQ})$.

Let $Q = (\mathcal{O}, \Sigma, C(x))$ be an OMQ and $\mathcal{A}$ a Σ-ABox. Then $a \in \mathrm{Ind}(\mathcal{A})$ is an *answer* to Q on $\mathcal{A}$, written $\mathcal{A} \models Q(a)$, if $a \in C^{\mathcal{I}}$ for all models $\mathcal{I}$ of $\mathcal{O}$ and $\mathcal{A}$. For OMQs Q_1 and Q_2, $Q_i = (\mathcal{O}_i, \Sigma, q_i)$, we say that Q_1 is *contained* in Q_2 and write $Q_1 \subseteq Q_2$ if for every Σ-ABox $\mathcal{A}$ and $a \in \mathrm{Ind}(\mathcal{A})$, $\mathcal{A} \models Q_1(a)$ implies $\mathcal{A} \models Q_2(a)$. We say that Q_1 is *equivalent* to Q_2 and write $Q_1 \equiv Q_2$ if $Q_1 \subseteq Q_2$ and $Q_2 \subseteq Q_1$.

A natural definition of ontology approximation in the context of OMQs is as follows.

Definition 3. *Let $\mathcal{O}_S$ be an $\mathcal{ALC}$ ontology, $\mathcal{L}_T$ one of the DLs from Section 2, and $\mathcal{Q}$ a query language. An $\mathcal{L}_T$ ontology $\mathcal{O}_T$ is an $\mathcal{L}_T$ approximation of $\mathcal{O}_S$ w.r.t $\mathcal{Q}$ if for all queries $q \in \mathcal{Q}$ and all signatures Σ with $\Sigma \cap \mathrm{sig}(\mathcal{O}_T) \subseteq \mathrm{sig}(\mathcal{O}_S)$,*

1. *$(\mathcal{O}_S, \Sigma, q) \supseteq (\mathcal{O}_T, \Sigma, q)$ and*

2. *$(\mathcal{O}_S, \Sigma, q) \supseteq Q$ implies $(\mathcal{O}_T, \Sigma, q) \supseteq Q$ for all OMQs $Q = (\mathcal{O}'_T, \Sigma, q)$ with $\mathcal{O}'_T \in \mathcal{L}_T$.*

$\mathcal{O}_T$ might use fresh symbols and thus approximations are projective. Informally, Point 1 is a soundness condition and Point 2 formalizes 'to preserve as many query answers as possible'. It is not guaranteed that the OMQs $(\mathcal{O}_S, \Sigma, q)$ and $(\mathcal{O}_T, \Sigma, q)$ are equivalent for all relevant queries q and signatures Σ, and the following example shows that this is in fact impossible to achieve.

Example 5. *Let $\mathcal{O}_S$ be the $\mathcal{ELU}$ ontology*

$$\mathcal{O}_S = \{\top \sqsubseteq B_1 \sqcup B_2\} \cup \{B_i \sqcap \exists r.B_i \sqsubseteq A \mid i \in \{1,2\}\}$$

Then an $\mathcal{EL}$ approximation of $\mathcal{O}_S$ w.r.t. ELQ is

$$\mathcal{O}_T = \{B_1 \sqcap B_2 \sqcap \exists r.\top \sqsubseteq A, \ \exists r.(B_1 \sqcap B_2) \sqsubseteq A\}$$
$$\cup \{B_i \sqcap \exists r.B_i \sqsubseteq A \mid i \in \{1,2\}\}.$$

However, there is no OMQ in $(\mathcal{EL}, \mathrm{ELQ})$ that is equivalent to $(\mathcal{O}_S, \{r\}, A(x))$ since it would have to return a as an answer on the ABox $\{r(a,a)\}$, but not on the ABox $\{r(a,b), r(b,a)\}$. No OMQ from $(\mathcal{EL}, \mathrm{ELQ})$ has this property.

It turns out that the approximations from Sections 3 and 4 are also useful in the context of Definition 3 when we choose ELQ or AQ as the query language. In particular, it follows from Theorem 7 that every $\mathcal{ALC}$ ontology $\mathcal{O}_S$ has a finite $\mathcal{EL}_\perp$ approximation w.r.t. AQ.

Theorem 8. *Let $\mathcal{O}_S$ be an $\mathcal{ALC}$ ontology, $\mathrm{sig}(\mathcal{O}_S) = \Sigma$. Then*

1. *the ontology $\mathcal{O}_T^\omega$ from Section 4 is an $\mathcal{EL}_\perp$ approximation of $\mathcal{O}_S$ w.r.t. $\mathrm{ELQ}(\Sigma)$;*

2. *the ontology $\mathcal{O}_T^-$ from Section 4 is a (finite) $\mathcal{EL}_\perp$ approximation of $\mathcal{O}_S$ w.r.t. AQ;*

3. *if $\mathcal{O}_S$ falls within $\mathcal{ELU}$, then the ontology $\mathcal{O}_T^\omega$ from Section 3 is an $\mathcal{EL}$ approximation of $\mathcal{O}_S$ w.r.t. $\mathrm{ELQ}(\Sigma)$.*

Point 2 also implies that $\mathcal{O}_T^-$ is an $\mathcal{EL}$ approximation of $\mathcal{O}_S$ w.r.t. AQ whenever $\mathcal{O}_S$ is an $\mathcal{ELU}$ ontology. We close with an interesting application of Theorem 8.

The topic of rewriting an OMQ into a simpler query language has received a lot of interest in the literature, see for example [Calvanese *et al.*, 2007; Gottlob *et al.*, 2014; Kaminski *et al.*, 2016; Feier *et al.*, 2019]. An OMQ Q is $(\mathcal{L}, \mathcal{Q})$-*rewritable* if there is an OMQ Q' in the OMQ language $(\mathcal{L}, \mathcal{Q})$ such that $Q \equiv Q'$.

By virtue of Theorem 8, we can decide whether an OMQ $Q = (\mathcal{O}, \Sigma, A(x))$ from $(\mathcal{ALC}, \mathrm{AQ})$ is $(\mathcal{EL}_\perp, \mathrm{AQ})$-rewritable. It can be seen that this is the case if and only if Q is equivalent to an OMQ $Q' \in (\mathcal{EL}_\perp, \mathrm{AQ})$ of the form $(\mathcal{O}', \Sigma, A(x))$. By Condition 2 of Definition 3, it thus suffices to construct the finite $\mathcal{EL}_\perp$ approximation $\mathcal{O}_T^-$ of $\mathcal{O}$ w.r.t. AQ from Theorem 8 and check whether $Q \equiv (\mathcal{O}_T^-, \Sigma, A(x))$, which is decidable [Bienvenu *et al.*, 2014]. This result extends to $(\mathcal{ALC}, \mathrm{ALCQ})$ since every OMQ from this language is equivalent to one from $(\mathcal{ALC}, \mathrm{AQ})$. Via the results in [Feier *et al.*, 2018], this can be lifted further to a certain class of conjunctive queries.

Theorem 9. *Given an OMQ $Q \in (\mathcal{ALC}, \mathrm{ALCQ})$, it is decidable whether Q is $(\mathcal{EL}_\perp, \mathrm{AQ})$-rewritable.*

6 Conclusion

We have investigated the structure and finiteness of ontology approximations when transitioning from non-Horn DLs to Horn DLs. We believe that our results shed significant light on the situation. It remains, however, an important and challenging topic for future work to push our techniques further towards practical applicability. Also, there are many other relevant cases of approximation. As a first step, one might think about extending the DLs considered in this paper with role inclusions. It might further be interesting to study the problem to decide whether a given (finite) candidate is an approximation of a given ontology. We expect this to be quite non-trivial. A related result in [Lutz *et al.*, 2012] states that it is between EXPTIME and 2EXPTIME to decide whether a given $\mathcal{ELU}$ ontology $\mathcal{O}_S$ of a restricted syntactic form has a finite complete $\mathcal{EL}$ approximation. Without the restriction, even decidability is open.

References

[Baader *et al.*, 2017] Franz Baader, Ian Horrocks, Carsten Lutz, and Ulrike Sattler. *An Introduction to Description Logic*. Cambridge University Press, 2017.

[Bienvenu *et al.*, 2014] Meghyn Bienvenu, Balder ten Cate, Carsten Lutz, and Frank Wolter. Ontology-based data access: A study through disjunctive datalog, CSP, and MMSNP. *ACM Trans. Database Syst.*, 39(4):33:1–33:44, 2014.

[Bötcher *et al.*, 2019] Anneke Bötcher, Carsten Lutz, and Frank Wolter. Ontology approximation in Horn description logics. In *Proc. of IJCAI*, pages 1574–1580. ijcai.org, 2019.

[Botoeva *et al.*, 2010] Elena Botoeva, Diego Calvanese, and Mariano Rodriguez-Muro. Expressive approximations in *DL-Lite* ontologies. In *Proc. of AIMSA*, volume 6304 of *LNCS*, pages 21–31. Springer, 2010.

[Calvanese *et al.*, 2007] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. Autom. Reasoning*, 39(3):385–429, 2007.

[Calvanese *et al.*, 2009] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Antonella Poggi, Mariano Rodriguez-Muro, and Riccardo Rosati. Ontologies and databases: The DL-Lite approach. In *Reasoning Web*, volume 5689 of *LNCS*, pages 255–356. Springer, 2009.

[Carral *et al.*, 2014] David Carral, Cristina Feier, Bernardo Cuenca Grau, Pascal Hitzler, and Ian Horrocks. $\mathcal{EL}$-ifying ontologies. In *Proc. of IJCAR*, pages 464–479, 2014.

[Cucala *et al.*, 2019] David Tena Cucala, Bernardo Cuenca Grau, and Ian Horrocks. 15 years of consequence-based reasoning. In *Description Logic, Theory Combination, and All That - Essays Dedicated to Franz Baader on the Occasion of His 60th Birthday*, volume 11560 of *LNCS*, pages 573–587. Springer, 2019.

[Feier *et al.*, 2018] Cristina Feier, Carsten Lutz, and Frank Wolter. From conjunctive queries to instance queries in ontology-mediated querying. In *Proc. of IJCAI*, pages 1810–1816. ijcai.org, 2018.

[Feier *et al.*, 2019] Cristina Feier, Antti Kuusisto, and Carsten Lutz. Rewritability in monadic disjunctive datalog, MMSNP, and expressive description logics. *Logical Methods in Computer Science*, 15(2), 2019.

[Gottlob *et al.*, 2014] Georg Gottlob, Stanislav Kikot, Roman Kontchakov, Vladimir V. Podolskii, Thomas Schwentick, and Michael Zakharyaschev. The price of query rewriting in ontology-based data access. *Artif. Intell.*, 213:42–59, 2014.

[Haga *et al.*, 2020] Anneke Haga, Carsten Lutz, Johannes Marti, and Frank Wolter. A journey into ontology approximation: From Non-Horn to Horn. *CoRR*, abs/2001.07754, 2020.

[Kaminski *et al.*, 2016] Mark Kaminski, Yavor Nenov, and Bernardo Cuenca Grau. Datalog rewritability of disjunctive datalog programs and non-Horn ontologies. *Artif. Intell.*, 236:90–118, 2016.

[Lutz and Wolter, 2010] Carsten Lutz and Frank Wolter. Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$. *J. Symb. Comput.*, 45(2):194–228, 2010.

[Lutz *et al.*, 2012] Carsten Lutz, Inanç Seylan, and Frank Wolter. An automata-theoretic approach to uniform interpolation and approximation in the description logic $\mathcal{EL}$. In *Proc. of KR*. AAAI Press, 2012.

[Pan and Thomas, 2007] Jeff Z. Pan and Edward Thomas. Approximating *OWL-DL* ontologies. In *AAAI*, pages 1434–1439, 2007.

[Ren *et al.*, 2010] Yuan Ren, Jeff Z. Pan, and Yuting Zhao. Soundness preserving approximation for tbox reasoning. In *Proc. of AAAI*. AAAI Press, 2010.

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