Proceedings, Fifteenth International Conference on
Principles of Knowledge Representation and Reasoning (KR 2016)

Undecidability Results for Database-Inspired Reasoning Problems in Very Expressive Description Logics

Sebastian Rudolph
TU Dresden
Germany
sebastian.rudolph@tu-dresden.de

Abstract

Recently, the field of knowledge representation is drawing a lot of inspiration from database theory. In particular, in the area of description logics and ontology languages, interest has shifted from satisfiability checking to query answering, with various query notions adopted from databases, like (unions of) conjunctive queries or different kinds of path queries. Likewise, the finite model semantics is being established as a viable and interesting alternative to the traditional semantics based on unrestricted models.

In this paper, we investigate diverse database-inspired reasoning problems for very expressive description logics (all featuring the worrisome trias of inverses, counting, and nominals) which have in common that role paths of unbounded length can be described (in the knowledge base or of the query), leading to a certain non-locality of the reasoning problem. We show that for all the cases considered, undecidability can be established by very similar means.

Most notably, we show undecidability of finite entailment of unions of conjunctive queries for a fragment of $\mathcal{SHOIQ}$ (the logic underlying the OWL DL ontology language), and undecidability of finite entailment of conjunctive queries for a fragment of $\mathcal{SROIQ}$ (the logical basis of the more recent and popular OWL 2 DL standard).

Introduction

Logic-based knowledge representation and reasoning is a prospering discipline that currently sees a lot of practical uptake in areas where intelligent information processing is key. One of the major transition paths to practice is through ontological specifications, using well-established and widely accepted standardized ontology languages like OWL (McGuinness and van Harmelen 10 February 2004) or its newer, refined version OWL 2 (OWL Working Group 27 October 2009). Logically, these languages are based on very expressive members of the family of *description logics* (DLs) (Baader et al. 2007; Rudolph 2011; Krötzsch, Simančík, and Horrocks 2012).

In the spirit of mathematical logic, the central reasoning problem traditionally considered in DL research was satisfiability of knowledge bases (and other problems straightforwardly reducible to it – so called *standard reasoning tasks*), employing the common model-theoretic semantics adopted from first-order logic. Over the past two decades, fostered by the growing practical impact of DL research, the scope of interest has widened to include new types of reasoning problems. Thereby, not very surprisingly, the area of databases has been an important source of inspiration. In fact, the fields of logic-based knowledge representation and reasoning have been significantly converging over the past years and seen a lot of cross-fertilization (cf. Rudolph 2014).

On the formal side, two major conceptual contributions of database theory can be identified: First, instead of focusing on satisfiability checks, the central reasoning problem in databases is query answering. Thereby the formalism to express queries may be different from the language used to specify the queried knowledge, such that an immediate reduction to satisfiability checking is not possible. Second, since databases are necessarily finite, the semantics commonly employed in database theory is based on finite models only. As we will see, this change of semantics may make a big difference regarding satisfiability and query answering.

Query Answering As opposed to satisfiability checking, evaluating queries in the presence of a background knowledge base (referred to as *ontology-based query answering*) allows us to express more complex information needs. A very basic, yet prominent query formalism often encountered in databases and nowadays in description logics is that of *conjunctive queries* (CQs) corresponding to the SELECT-PROJECT-JOIN fragment of SQL (Chandra and Merlin 1977) and *unions of conjunctive queries* (UCQs). Answering conjunctive queries over DL knowledge bases has first been mentioned as a topic in the 1990s (Levy and Rousset 1996) and since then investigated for a great variety of description logic languages. The most expressive DLs with inverses, counting, and nominals where CQ and UCQ entailment[1] are known to be decidable are $\mathcal{ALCHOIQb}$ (Rudolph and Glimm 2010) and Horn-$\mathcal{SROIQ}$ (Ortiz, Rudolph, and Simkus 2011).

In the context of semi-structured databases, other query formalisms have been developed which allow to express information needs related to reachability, so-called *path*

[1]The computation problem of query answering is polynomially reducible to the decision problem of (Boolean) query entailment, so we focus on the latter in the following.

queries or *navigational queries* (Buneman 1997). Beyond expressing more elaborate information needs, such queries can also be used to internalize ontological knowledge into the query to a certain degree (Bischoff et al. 2014). Over the past decade, a variety of results regarding answering of (diverse variants of) path queries over DL knowledge bases have been established (Calvanese, Eiter, and Ortiz 2007; 2009; Bienvenu et al. 2014) the most popular classes of queries currently considered are *two-way regular path queries* (2RPQs) and *(unions of) conjunctive two-way regular path queries* ((U)C2RPQs). The most expressive DL fragment with inverses, counting, and nominals combined where a UC2RPQs answering is known to be decidable is again Horn-$\mathcal{SROIQ}$ (Ortiz, Rudolph, and Simkus 2011).

Current research progresses to even more expressive query languages most of which can be seen as fragments of Datalog (Rudolph and Krötzsch 2013; Bourhis, Krötzsch, and Rudolph 2014).

Finite Satisfiability As stated above, the finite model semantics, while very popular in the database domain, has historically received little attention from DL researchers. This may be partially due to the fact, that many of the less expressive DLs (up to $\mathcal{SROI}$) have the *finite model property*, where the two satisfiability notions (for finite vs. arbitrary models) coincide. This property, however is lost as soon as inverses and counting are involved. First investigations into finite satisfiability of such DLs go back to the last millenium (Calvanese 1996) but spawned only little follow-up work (Lutz, Sattler, and Tendera 2005; Ibáñez-García, Lutz, and Schneider 2014). It was only in 2008 when finite satisfiability for $\mathcal{SROIQ}$ (and all its sublogics) was shown to be decidable (Kazakov 2008), exploiting a result on the finite satisfiability for the counting two-variable fragment of first-order logic (Pratt-Hartmann 2005).

Finite Query Entailment Query entailment under the finite model semantics (short: finite query entailment) has so far received very little attention from the DL community. Note that the finite model property does not help here. The equivalent notion, holding when query entailment and finite query entailment coincide, is called *finite controllability*. Luckily, very recent results on the guarded fragment of first order logic (Bárány, Gottlob, and Otto 2014) which extend previous work on finite controllability in databases under the open-world assumption (Rosati 2011) entail that for DLs up to $\mathcal{ALCHOIb}$, answering CQs and UCQs is finitely controllable, therefore for all those logics, decidability of finite (U)CQ entailment follows from decidability of (U)CQ entailment of the more expressive $\mathcal{ALCHOIQb}$ (Rudolph and Glimm 2010). For the case where the underlying logic has counting, or role chains can be described in the knowledge base or the query, results on finite query entailment are very scarce, the only DL not subsumed by $\mathcal{ALCHOIb}$ for which finite UCQ entailment is known to be decidable is Horn-$\mathcal{ALCFI}$ (Ibáñez-García, Lutz, and Schneider 2014).

The contribution of this paper consists in a sequence of undecidability results regarding database-inspired reasoning problems which are established by very similar constructions encoding the classical undecidable Post Correspondence Problem. In particular, we prove undecidability of

1. finite UCQ entailment from $\mathcal{SHOIF}$ KBs,
2. finite CQ entailment from $\mathcal{SROIF}^-$ KBs,
3. finite 2RPQ entailment from $\mathcal{ALCOIF}$ KBs,
4. 2RPQ entailment from $\mathcal{ALCOIF}_{\mathrm{reg}}$ KBs,
5. satisfiability of $\mathcal{ALCOIF}_{\omega\mathrm{reg}}$ KBs, and
6. 2ωRPQ entailment from $\mathcal{ALCOIF}$ KBs.

The last two reasoning problems feature two-way ω-regular path expressions (in the logic vs. in the query language) used to describe infinite paths. We will draw connections from this novel descriptive feature to existing logics.

We will treat the first reasoning problem in great detail, with necessary preliminaries, examples and full proofs. For the later problems, we will introduce preliminaries in the place needed and sketch the necessary changes that need to be made to the construction.

Preliminaries

The Description Logic $\mathcal{SHOIF}$

We now introduce the description logic $\mathcal{SHOIF}$, a sublogic of the prominent description logics $\mathcal{SHOIQ}$ (Horrocks and Sattler 2007) and $\mathcal{SROIQ}$ (Horrocks, Kutz, and Sattler 2006) underlying the OWL DL and the OWL 2 DL standards, respectively (McGuinness and van Harmelen 10 February 2004; OWL Working Group 27 October 2009).

As signature of $\mathcal{SHOIF}$ we have countably infinite disjoint sets N_C, N_R and N_I of *concept names*, *role names* and *individual names* respectively. Further the set N_R is partitioned into two sets namely, $\mathbf{R_s}$ and $\mathbf{R_n}$ of *simple* and *non-simple* roles respectively. The set $\mathbf{R}$ of $\mathcal{SHOIF}$ *roles* contains r and r^- (the *inverse* of r) for every $r \in N_R$.

Further, we define a function Inv on roles such that $\mathsf{Inv}(r) = r^-$ if r is a role name and $\mathsf{Inv}(r) = s$ if $r = s^-$.

The set of $\mathcal{SHOIF}$ concepts (or simply concepts) is the smallest set satisfying the following properties:

- every concept name $A \in N_C$ is a concept;
- if C, D are concepts, r is a role, $a_1, \ldots, a_n$ are individual names and n is a non-negative integer, then the following are concepts:

$\top$	(top concept)
$\bot$	(bottom concept)
$\neg C$	(negation)
$C \sqcap D$	(intersection)
$C \sqcup D$	(union)
$\forall r.C$	(universal quantification)
$\exists r.C$	(existential quantification)
$\{a_1, \ldots, a_n\}$	(nominals / one-of)

A $\mathcal{SHOIF}$ axiom is an expression of one the following forms:

1. $C \sqsubseteq D$, where C and D are $\mathcal{SHOIF}$ concepts,
2. $s \sqsubseteq r$, where s and r are $\mathcal{SHOIF}$ roles, and if $s \in \mathbf{R_n}$ then also $r \in \mathbf{R_n}$

Axiom α	$\mathcal{I} \models \alpha$, if
$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
$s \sqsubseteq r$	$s^{\mathcal{I}} \subseteq r^{\mathcal{I}}$
$\mathsf{Fun}(s)$	$(\delta, \delta_1), (\delta, \delta_2) \in s^{\mathcal{I}}$ implies $\delta_1 = \delta_2$ for all $\delta, \delta_1, \delta_2 \in \Delta^{\mathcal{I}}$
$\mathsf{Trans}(r)$	$(\delta_1, \delta_2), (\delta_2, \delta_3) \in r^{\mathcal{I}}$ implies $(\delta_1, \delta_3) \in r^{\mathcal{I}}$ for all $\delta_1, \delta_2, \delta_3 \in \Delta^{\mathcal{I}}$

Table 1: Semantics of $\mathcal{SHOIF}$ axioms

3. $\mathsf{Fun}(s)$ (functionality) for some $s \in \mathbf{R_s}$,
4. $\mathsf{Trans}(r)$ (transitivity) for some $r \in \mathbf{R_n}$.

Axioms of the first form are called *general concept inclusion axioms* (GCIs), axioms of the second form are called *role inclusions*, axioms of the third type are called *functionality axioms*, and axioms of the third form are called *transitivity axioms*. As usual, we write $C \equiv D$ to assert both $C \sqsubseteq D$ and $D \sqsubseteq C$. A $\mathcal{SHOIF}$ knowledge base is a set of $\mathcal{SHOIF}$ axioms.[2]

The semantics of $\mathcal{SHOIF}$ is defined in the standard model-theoretic way. A $\mathcal{SHOIF}$ interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ is composed of a non-empty set $\Delta^{\mathcal{I}}$, called the *domain of* $\mathcal{I}$ and a *mapping function* $\cdot^{\mathcal{I}}$ such that:

- $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ for every concept name A;
- $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ for every role name $r \in N_R$;
- $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ for every individual name a.

The mapping $\cdot^{\mathcal{I}}$ is extended to roles and concepts as follows:

$$\begin{aligned}
(r^-)^{\mathcal{I}} &= \{(\delta, \delta') \mid (\delta', \delta) \in r^{\mathcal{I}}\} \\
\top^{\mathcal{I}} &= \Delta^{\mathcal{I}} \\
\bot^{\mathcal{I}} &= \emptyset \\
(\neg C)^{\mathcal{I}} &= \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}} \\
(C \sqcap D)^{\mathcal{I}} &= C^{\mathcal{I}} \cap D^{\mathcal{I}} \\
(C \sqcup D)^{\mathcal{I}} &= C^{\mathcal{I}} \cup D^{\mathcal{I}} \\
(\forall r.C)^{\mathcal{I}} &= \{\delta \in \Delta^{\mathcal{I}} \mid \forall \delta'.(\delta, \delta') \in r^{\mathcal{I}} \rightarrow y \in C^{\mathcal{I}}\} \\
(\exists r.C)^{\mathcal{I}} &= \{\delta \in \Delta^{\mathcal{I}} \mid \exists \delta'.(\delta, \delta') \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\} \\
\{a_1, \ldots, a_n\}^{\mathcal{I}} &= \{a_1^{\mathcal{I}}, \ldots, a_n^{\mathcal{I}}\}
\end{aligned}$$

where C, D are concepts and r is a role.

Given an axiom α, we say the an interpretation $\mathcal{I}$ *satisfies* α, written $\mathcal{I} \models \alpha$, if it satisfies the corresponding condition given in Table 1. Similarly, $\mathcal{I}$ satisfies a $\mathcal{SHOIF}$ knowledge base $\mathcal{K}$, written $\mathcal{I} \models \mathcal{K}$, if it satisfies all the axioms in $\mathcal{K}$; we then call $\mathcal{I}$ a *model* of $\mathcal{K}$. A knowledge base is said to be *consistent* or *satisfiable* if it has a model.

(Unions of) Conjunctive Queries

We next formally introduce (unions of) conjunctive queries. Assume a $\mathcal{SHOIF}$ signature as above and let N_V be a countably infinite set of variables disjoint from N_C, N_R, and N_I. A *term* t is an element from $N_V \cup N_I$. Let $A \in N_C$ be a concept name, $r \in N_R$ a role name, and t, t' terms. An *atom* is an expression $A(t)$ or $r(t, t')$ and we refer to these two types of atoms as *concept atoms* and *role atoms* respectively. A *conjunctive query* (CQ) is a non-empty set of atoms.

Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ be an interpretation. A total function π from the terms of a CQ q to $\Delta^{\mathcal{I}}$ is an *evaluation* if $\pi(a) = a^{\mathcal{I}}$ for each individual name a occurring in q. For $A(t), r(t, t')$ atoms, we write

- $\mathcal{I} \models^{\pi} A(t)$ if $\pi(t) \in A^{\mathcal{I}}$;
- $\mathcal{I} \models^{\pi} r(t, t')$ if $(\pi(t), \pi(t')) \in r^{\mathcal{I}}$.

If, for an evaluation π, $\mathcal{I} \models^{\pi} At$ for all atoms $At \in q$, we write $\mathcal{I} \models^{\pi} q$. We say that $\mathcal{I}$ *satisfies* q and write $\mathcal{I} \models q$ if there exists an evaluation π such that $\mathcal{I} \models^{\pi} q$. We call such a π a *match* for q in $\mathcal{I}$.

Let $\mathcal{K}$ be a knowledge base and q a conjunctive query. If $\mathcal{I} \models \mathcal{K}$ implies $\mathcal{I} \models q$, we say that $\mathcal{K}$ *entails* q and write $\mathcal{K} \models q$.

The *query entailment problem* is defined as follows: given a knowledge base $\mathcal{K}$ and a query q, decide whether $\mathcal{K} \models q$.

A *union of conjunctive queries* (UCQ) is a finite set $Q = \{q_1, \ldots, q_n\}$ of CQs. Some interpretation $\mathcal{I}$ *satisfies* Q (written: $\mathcal{I} \models Q$) if $\mathcal{I}$ it satisfies one of $q_1, \ldots, q_n$. We say that some knowledge base $\mathcal{K}$ *entails* Q and write $\mathcal{K} \models Q$ if $\mathcal{I} \models \mathcal{K}$ implies $\mathcal{I} \models Q$.

Finite Model Reasoning

Above we introduced the standard semantics for satisfiability and entailment of (unions of) conjunctive queries in description logic. This paper, however, also addresses reasoning under the finite-model semantics, which is a prominent (or even the standard) setting in database theory. Given the current convergence of the fields of knowledge representation and database theory, research into finite-model reasoning in description logics has intensified lately.

Definition 1 (Finite Model Semantics)**.** *A knowledge base $\mathcal{K}$ is said to be* finitely satisfiable *if it has a finite model, i.e., there exists an interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ with $\mathcal{I} \models \mathcal{K}$ and $\Delta^{\mathcal{I}}$ finite. Likewise we say $\mathcal{K}$* finitely entails *a conjunctive query q (or a union of conjunctive queries Q) and write $\mathcal{K} \models_{\text{fin}} q$ ($\mathcal{K} \models_{\text{fin}} Q$), if for every interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ with $\mathcal{I} \models \mathcal{K}$ and finite $\Delta^{\mathcal{I}}$ holds $\mathcal{I} \models q$ ($\mathcal{I} \models Q$).*

It is obvious that finite satisfiability implies satisfiability, while the other direction holds only if the underlying logic has the finite model property. Likewise, entailment implies finite entailment but not vice versa.

Example 2. *Consider the knowledge base $\mathcal{K}_1$ consisting of the following axioms:*

$$\begin{aligned}
&\mathsf{Fun}(r^-) \\
\top &\sqsubseteq \exists r.\top \\
\{a\} &\sqsubseteq \neg \exists r^-.\top.
\end{aligned}$$

We find that $\mathcal{K}_1$ is satisfiable (witnessed by the interpretation $(\mathbb{N}, \{a \mapsto 0, r \mapsto succ\})$) but not finitely satisfiable (since the sum of r-indegrees and the sum of r-outdegrees cannot match in a finite model).

[2]We omit the typical subdivision of knowledge bases into ABox, TBox, and RBox, as it is not needed for our line of argumentation.

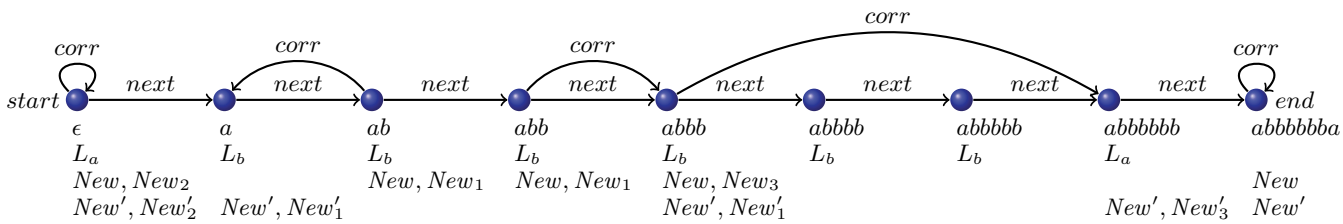

Figure 1: Solution model for the PCP instance and solution sequence described in Example 4.

In a similar way, the $\mathcal{SHOIF}$ knowledge base $\mathcal{K}_2$ containing the axioms

$$\begin{array}{rcl} \top & \sqsubseteq & \exists r.\top \\ r & \sqsubseteq & r' \\ & & \mathsf{Trans}(r') \end{array}$$

does not entail the CQ $\{r'(x,x)\}$ (witnessed by the interpretation $(\mathbb{N}, \{r \mapsto succ, r' \mapsto <\})$), but $\mathcal{K}_2 \models_{\text{fin}} \{r'(x,x)\}$.

The Post Correspondence Problem

We will establish our undecidability result by a reduction from the well-known Post Correspondence Problem (Post 1946) defined as follows:

Definition 3 (Post Correspondence Problem)**.** *Let $\mathbb{P} = \{(g_1, g'_1), \ldots, (g_\mu, g'_\mu)\}$ be an arbitrary finite set of pairs of non-empty strings over the alphabet $\{a, b\}$. A nonempty finite sequence $i_1, \ldots, i_n$ of natural numbers from $\{1, \ldots, \mu\}$ is called a* solution sequence *of $\mathbb{P}$ if $g_{i_1} \cdots g_{i_n} = g'_{i_1} \cdots g'_{i_n}$. The* Post Correspondence Problem *(short: PCP) requires to determine if there exists a solution sequence for a given $\mathbb{P}$.*

Example 4. *Let $\mathbb{P} = \{(g_1, g'_1), (g_2, g'_2), (g_3, g'_3)\}$ with*

- $g_1 = b$ *and* $g'_1 = bbb$,
- $g_2 = ab$ *and* $g'_2 = a$,
- $g_3 = bbba$ *and* $g'_3 = a$.

Then $2, 1, 1, 3$ is a solution sequence since

$$\begin{array}{rc} & g_2\ g_1\ g_1\ g_3 \\ = & (ab)(b)(b)(bbba) \\ = & a\ b\ b\ b\ b\ b\ b\ a \\ = & (a)(bbb)(bbb)(a) \\ = & g'_2\ g'_1\ g'_1\ g'_3 \end{array}$$

Therefore the answer to the PCP instance $\mathbb{P}$ is "yes".

Theorem 5 (Post, 1946)**.** *The Post Correspondence Problem is undecidable.*

Undecidability of finite UCQ Entailment in $\mathcal{SHOIF}$

We are now ready to establish our first undecidability result. To this end, we will for a given instance of the PCP establish a $\mathcal{SHOIF}$ knowledge base and a union of conjunctive queries such that every model of the knowledge base *not* satisfying the UCQ (also called a *counter-model*) encodes a solution to the problem instance, and, conversely, every solution to the problem instance gives rise to such a counter-model.

Solution Models

We first formally define in which way the counter-models are supposed to encode solutions to the provided PCP instance.

Definition 6 (Solution Model)**.** *Given a PCP instance $\mathbb{P} = \{(g_1, g'_1), \ldots, (g_\mu, g'_\mu)\}$, an interpretation $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ is called a* solution model *for $\mathbb{P}$ if there is a solution sequence $i_1, \ldots, i_n$ of $\mathbb{P}$ such that for $w = g_{i_1} \cdots g_{i_n} = g'_{i_1} \cdots g'_{i_n}$, the following hold:*

- $\Delta^\mathcal{I} = \text{Prefixes}(w) = \{v \mid w = vv', v' \in \{a, b\}^*\}$
- $start^\mathcal{I} = \epsilon$
- $end^\mathcal{I} = w$
- $L_a^\mathcal{I} = \{v \mid va \in \Delta^\mathcal{I}\}$
- $L_b^\mathcal{I} = \{v \mid vb \in \Delta^\mathcal{I}\}$
- $New^\mathcal{I} = \{\epsilon\} \cup \{g_{i_1} \cdots g_{i_\ell} \mid 1 \leq \ell \leq n\}$
- $New'^\mathcal{I} = \{\epsilon\} \cup \{g'_{i_1} \cdots g'_{i_\ell} \mid 1 \leq \ell \leq n\}$
- ${New_k}^\mathcal{I} = \{g_{i_1} \cdots g_{i_{\ell-1}} \mid i_\ell = k, 1 \leq \ell \leq n\}$
- ${New'_k}^\mathcal{I} = \{g'_{i_1} \cdots g'_{i_{\ell-1}} \mid i_\ell = k, 1 \leq \ell \leq n\}$
- $next^\mathcal{I} = \{(v, vc) \mid c \in \{a, b\}, v, vc \in \Delta^\mathcal{I}\}$
- $corr^\mathcal{I} = \{(\epsilon, \epsilon)\} \cup \{(g_{i_1} \cdots g_{i_\ell}, g'_{i_1} \cdots g'_{i_\ell}) \mid 1 \leq \ell \leq n\}$

Thereby, start and end are two individual names, L_a, L_b, New, New', New_1, New'_1, … New_μ, New'_μ, are concept names and next and corr are role names.

Figure 1 displays a solution model for the PCP instance $\mathbb{P}$ and solution sequence presented in Example 4.

Axiomatization of Solution Models

The purpose of the subsequently defined knowledge base $\mathcal{K}_\mathbb{P}$ is to enforce that all its finite models that do not satisfy a certain UCQ must be isomorphic to some solution model of $\mathbb{P}$. We now introduce the axioms bit by bit and explain their purpose. First, we stipulate that the starting and the ending element do not coincide (and thereby the word encoded by the solution model is nonempty).

$$\{start\} \sqcap \{end\} \sqsubseteq \bot \tag{1}$$

Next, we enforce that every but the ending element has an outgoing *next* role, and that every but the starting element has an incoming such role.

$$\neg\{end\} \equiv \exists next.\top \tag{2}$$

$$\neg\{start\} \equiv \exists next^{-}.\top \tag{3}$$

Also, we make sure that there is no more than one outgoing and no more than one incoming $next$ role for every element.

$$\mathsf{Fun}(next) \tag{4}$$
$$\mathsf{Fun}(next^-) \tag{5}$$

Now we ensure that every domain element except $end^{\mathcal{I}}$ is labeled with exactly one of L_a or L_b.

$$\neg\{end\} \equiv L_a \sqcup L_b \tag{6}$$
$$L_a \sqcap L_b \sqsubseteq \bot \tag{7}$$

Next, we describe "marker concepts" for the elements at the boundaries of the concatenated words (two versions for the two different concatenations). Also, we make sure that at each such boundary that is not the ending element, a choice is made regarding which of the μ possible words comes next, and we implement this choice. Thereby, for a word $g = c_1 \cdots c_\ell$ we let $I_g := L_{c_1} \sqcap \exists next.(\neg New \sqcap L_{c_2} \sqcap \exists next.(\neg New \sqcap \ldots L_{c_\ell} \sqcap \exists next.New \ldots))$ and $I'_g := L_{c_1} \sqcap \exists next.(\neg New' \sqcap L_{c_2} \sqcap \exists next.(\neg New' \sqcap \ldots L_{c_\ell} \sqcap \exists next.New' \ldots))$.

$$\{start\} \sqsubseteq New \sqcap New' \tag{8}$$
$$New \sqcap \neg\{end\} \equiv New_1 \sqcup \ldots \sqcup New_\mu \tag{9}$$
$$New_i \sqcap New_j \sqsubseteq \bot \qquad 1 \leq i < j \leq \mu \tag{10}$$
$$New' \sqcap \neg\{end\} \equiv New'_1 \sqcup \ldots \sqcup New'_\mu \tag{11}$$
$$New'_i \sqcap New'_j \sqsubseteq \bot \qquad 1 \leq i < j \leq \mu \tag{12}$$
$$New_k \sqsubseteq I_{g_k} \qquad 1 \leq k \leq \mu \tag{13}$$
$$New'_k \sqsubseteq I'_{g_k} \qquad 1 \leq k \leq \mu \tag{14}$$

We now turn to the $corr$ role which is supposed to help synchronizing the two concatenation schemes. To this end, $corr$ is supposed to connect corresponding word boundaries of one scheme with those of the other. We let $corr$ connect exactly the New elements with New' elements and make sure that this connection is a bijection.

$$New \equiv \exists corr.\top \tag{15}$$
$$New' \equiv \exists corr^-.\top \tag{16}$$
$$\mathsf{Fun}(corr) \tag{17}$$
$$\mathsf{Fun}(corr^-) \tag{18}$$

Also, we require that at corresponding word boundaries of the two schemes, the corresponding words are to be chosen.

$$New_k \sqsubseteq \exists corr.New'_k \tag{19}$$
$$New'_k \sqsubseteq \exists corr^-.New_k \tag{20}$$

Last, we use a role inclusion and a transitivity axiom to introduce and describe an auxiliary role: the $word$ role spans over chains of consecutive $next$ roles.

$$next \sqsubseteq word \tag{21}$$
$$\mathsf{Trans}(word) \tag{22}$$

Lastly but importantly, we define conjunctive queries which are supposed to detect "errors" in a model of the knowledge base defined so far. The CQ $\mathfrak{q}_1 = \{word(x, x)\}$ is supposed to detect looping $next$-chains (which must not exist in a solution model) and the CQ $\mathfrak{q}_2 = \{corr(x_1, x_2), word(x_2, x_3), corr(x_4, x_3), word(x_4, x_1)\}$ intuitively encodes the phenomenon of two "crossing" $corr$ relationships, which also are not allowed to occur in a solution model.

Correctness of the Reduction

After presenting the knowledge base and the queries, we will now formally prove the correspondence between the PCS and non-entailment. Thereby, the introduced notion of solution models will come in handy.

Lemma 7. *Let $\mathbb{P}$ be a PCP instance, and let $\mathcal{I}$ be a corresponding solution model. Then $\mathcal{I}$ can be extended into a model $\mathcal{I}'$ of $\mathcal{K}_\mathbb{P}$ such that $\mathcal{I}' \not\models \{\mathfrak{q}_1, \mathfrak{q}_2\}$.*

Proof. Let $\mathcal{I}'$ be defined by extending $\mathcal{I}$ from Definition 6 by letting $word^{\mathcal{I}'} = \{(v, v') \mid v' = vv'', v'' \in \{a, b\}^*, v, v' \in \Delta^{\mathcal{I}}\}$. Then it is straightforward to check that $\mathcal{I}'$ is a model of $\mathcal{K}_\mathbb{P}$ and does not satisfy $\mathfrak{q}_1$ nor $\mathfrak{q}_2$. □

Lemma 8. *Let $\mathbb{P}$ be a PCP. Then every finite model $\mathcal{I}$ of $\mathcal{K}_\mathbb{P}$ with $\mathcal{I} \not\models \{\mathfrak{q}_1, \mathfrak{q}_2\}$ is isomorphic to a solution model of $\mathbb{P}$.*

Proof. Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ with $\Delta^{\mathcal{I}}$ finite as well as $\mathcal{I} \models \mathcal{K}_\mathbb{P}$ and $\mathcal{I} \not\models \{\mathfrak{q}_1, \mathfrak{q}_2\}$.

We first show that every such model must be a finite chain of elements connected by $next$ roles starting with $start^{\mathcal{I}}$ and ending with $end^{\mathcal{I}}$.

First we label some of the elements of $\Delta^{\mathcal{I}}$ by natural numbers as follows: we label $start^{\mathcal{I}}$ with 0, the (thanks to Axiom 2 existing and thanks to Axiom 4 uniquely determined) $next$-successor of $start^{\mathcal{I}}$ with 1. In case the 1-labelled element is not $end^{\mathcal{I}}$, we label the (again existing and uniquely determined) $next$-successor of 1 by 2 and so forth. Due to Axiom 5, no element can get assigned two different numbers. The only way the labeling procedure can stop is when $end^{\mathcal{I}}$ is labeled. The procedure has to stop since the labeling is injective and the domain finite. Therefore there must be a chain of $next$ roles connecting $start^{\mathcal{I}}$ with $end^{\mathcal{I}}$. We now show that no elements outside this chain exists. Toward a contradiction, let $\mathcal{I}^-$ be obtained from $\mathcal{I}$ by removing all labeled elements. If we see $\mathcal{I}^-$ as finite directed graph with edge relation $next^{\mathcal{I}}$, we find that every vertex has indegree and outdegree 1. Therefore every element δ of $\mathcal{I}^-$ must lie on a directed circle, causing $\mathfrak{q}_1$ to be satisfied, contradicting our assumption.

Moreover, every element on that chain except $end^{\mathcal{I}}$ is either in $L_a^{\mathcal{I}}$ or in $L_b^{\mathcal{I}}$.

This finding allows us to rename the elements of the interpretation into words from $\{a, b\}^*$: to find the word for some domain element δ we start from $start^{\mathcal{I}}$ and follow the $next$ chain and concatenate all letters c found in the L_c labels of the traversed elements for all elements before δ (but excluding δ itself). After this renaming, the considered interpretation is identical to a solution model, as we continue to show now.

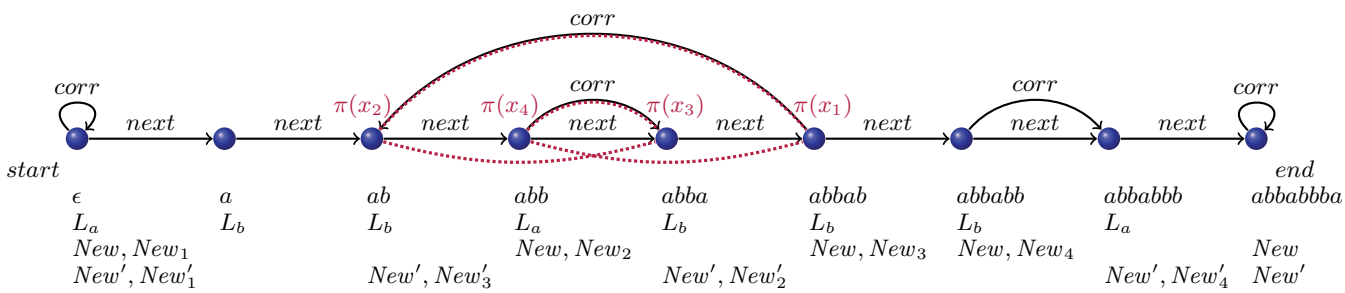

Figure 2: Model for the knowledge base $\mathcal{K}_{\mathbb{P}'}$ derived from the PCP instance $\mathbb{P}'$ described in Example 9. For better readability, the $word$ role has not been drawn, it is defined to hold between any two individuals connected by a directed chain of $next$ roles. Note that this model is not a solution model. The provided evaluation π witnesses that the query $\mathfrak{q}_2 = \{corr(x_1, x_2), word(x_2, x_3), corr(x_4, x_3), word(x_4, x_1)\}$ is satisfied in that model.

The prefix-order on the elements induces a linear order on both $New^{\mathcal{I}}$ and $New'^{\mathcal{I}}$. Moreover $New^{\mathcal{I}} \setminus \{end^{\mathcal{I}}\}$ is partitioned into the sets ${New_1}^{\mathcal{I}}, \dots, {New_\mu}^{\mathcal{I}}$ due to Axioms 9 and 10 and $New'^{\mathcal{I}} \setminus \{end^{\mathcal{I}}\}$ is partitioned into the sets ${New'_1}^{\mathcal{I}}, \dots, {New'_\mu}^{\mathcal{I}}$ due to Axioms 11 and 12. Thanks to Axioms 15 – 18, $corr$ is a bijection between $New^{\mathcal{I}}$ and $New'^{\mathcal{I}}$. Moreover, Axioms 19 and 20 make sure that $corr$ only connects corresponding partitions.

Furthermore, Axiom 13 ensures that every element marked with some New_k starts a chain of $next$s encoding the word g_k such that New holds for the last element of that chain but for none of the intermediate elements. Similarly, due to Axiom 13, every New'_k element starts such a chain corresponding to g'_k.

Thereby, we have established that the word encoded by the chain from $start^{\mathcal{I}}$ to $end^{\mathcal{I}}$ can be represented as concatenation of words from $g_1, \dots, g_\mu$ but also of words from $g'_1, \dots, g'_\mu$. With $corr$ being a partition-faithful bijection, we even know that the numbers of words used for the two concatenation schemes must be the same, and, more specifically that for every $i \in \{1, \dots \mu\}$ the number of occurrences of g_i in the first concatenation must be equal to the number of occurrences of g'_i in the second concatenation.

To really ensure that the concatenations are synchronous in the sense required to constitute a solution model, we finally need to show that the $corr$ role indeed connects the first New-element (regarding the order introduced above) with the first New'-element, the second with the second, and so forth. Toward a contradiction, suppose it does not, i.e., there are $\kappa \neq \kappa'$ such that $corr$ connects the κth New-element (let us call it δ_1) with the κ'th New'-element (denoted δ'_1). Since $corr$ is a bijection between $New^{\mathcal{I}}$ and $New'^{\mathcal{I}}$, there must be numbers $\iota \neq \iota'$ where $corr$ connects the ιth New-element (denoted δ_2) with the ι'th New'-element (denoted δ'_2) such that one of the two holds: ether $\iota > \kappa$ and $\iota' < \kappa'$ or $\iota < \kappa$ and $\iota' > \kappa'$. W.l.o.g., we assume the first case (otherwise just swap δ_1 with δ_2 and δ'_1 with δ'_2). Then there exists a path of $next$ and $next$ roles from δ_1 to δ_2 and likewise another such path from δ'_2 to δ'_1. This means that $(\delta_1, \delta_2), (\delta'_2, \delta'_1) \in word^{\mathcal{I}}$. On the other hand, we already know that $(\delta_1, \delta'_1), (\delta_2, \delta'_2) \in corr^{\mathcal{I}}$. But then, $\mathcal{I} \models \mathfrak{q}_2$, witnessed by the match $\pi = \{x_1 \mapsto \delta_2, x_2 \mapsto \delta'_2, x_3 \mapsto \delta'_1, x_4 \mapsto \delta_1\}$, contradicting our assumption. □

To illustrate the idea behind the construction and the proof, we will provide an example with an "out of sync" pseudo-solution and show how the query $\mathfrak{q}_2$ catches this problem.

Example 9. *Consider* $\mathbb{P}' = \{(g_1, g'_1), (g_2, g'_2), (g_3, g'_3), (g_4, g'_4)\}$ *with*

- $g_1 = abb$ *and* $g'_1 = ab$,
- $g_2 = ab$ *and* $g'_2 = bbb$,
- $g_3 = b$ *and* $g'_3 = ba$,
- $g_4 = ba$ *and* $g'_4 = a$.

Then, the interpretation depicted in Fig. 2 is a model of $\mathcal{K}_{\mathbb{P}'}$ *but not a solution model, as witnessed by* $\mathfrak{q}_2$ *being satisfied.*

The two lemmas together now give rise to the following theorem linking the PCP with finite UCQ entailment in $\mathcal{SHOIF}$.

Theorem 10. *Let* $\mathbb{P}$ *be a PCP instance and let* $\mathcal{K}_{\mathbb{P}}$ *be the* $\mathcal{SHOIF}$ *knowledge base consisting of Axioms 1–22. Then the answer to* $\mathbb{P}$ *is "yes" if and only if* $\mathcal{K}_{\mathbb{P}} \not\models_{\text{fin}} \{\mathfrak{q}_1, \mathfrak{q}_2\}$.

Proof. For the "only if" direction, we can invoke Lemma 7 to show that every solution sequence for $\mathbb{P}$ gives rise to a solution model which is a model of $\mathcal{K}_{\mathbb{P}}$ but does not satisfy $\{\mathfrak{q}_1, \mathfrak{q}_2\}$. For the "if" direction, Lemma 8 ensures that every model witnessing the finite non-entailment is isomorphic to a solution model, from which, by definition, a solution sequence can be extracted. □

Corollary 11. *Finite entailment of unions of conjunctive queries from* $\mathcal{SHOIF}$ *knowledge bases is undecidable.*

Related Undecidability Results

The construction used to establish the above undecidability result can be modified to show undecidability of other reasoning problems where nominals, counting, inverses and path expressions are involved. In the following we will introduce the logics and queries considered and describe how the reasoning problem needs to be adapted

Finite CQ Entailment in $\mathcal{SROIF}^-$

The description logic $\mathcal{SROIF}^-$ is obtained from $\mathcal{SHOIF}$ by allowing so called *complex role inclusion axioms*[3] of the form $r_1 \circ \ldots \circ r_n \sqsubseteq r$ for $r_1, \ldots, r_n, r \in \mathbf{R}$. Semantically, such an axiom is satisfied in an interpretation $\mathcal{I}$ if $r_1^{\mathcal{I}} \circ \ldots \circ r_n^{\mathcal{I}} \sqsubseteq r^{\mathcal{I}}$, where $\circ$ denotes the relational product (or, in database terms: join). Obviously, role inclusions known from $\mathcal{SHOIF}$ are a special form of such axioms (for $n = 1$) and also transitivity axioms can be expressed ($\mathsf{Trans}(r)$ can be written as $r \circ r \sqsubseteq r$). As complex role inclusions are very powerful constructs that immediately lead to undecidability when used freely, one has to control their usage by imposing so-called global restrictions, known as simplicity and regularity constraints. The simplicity constraint requires that, given a role inclusion axiom $r_1 \circ \ldots \circ r_n \sqsubseteq r$, r must be from $\mathbf{R_n}$ if $n > 1$ or if $n = 1$ and $r_1 \in \mathbf{R_n}$. The regularity constraint requires that there be a strict (irreflexive) partial order $\prec$ on $\mathbf{R}$ such that

- for $r \in \{s, \mathsf{Inv}(s)\}$, we have that $s \prec r$ iff $\mathsf{Inv}(s) \prec r$ and
- every role incusion axiom is of the form $r \circ r \sqsubseteq r$, $\mathsf{Inv}(r) \sqsubseteq r$, $r_1 \circ \ldots r_n \sqsubseteq r$, $r \circ r_1 \circ \cdots \circ r_n \sqsubseteq r$ or $r_1 \circ \cdots \circ r_n \circ r \sqsubseteq r$ where $r, r_1 \ldots, r_n \in \mathbf{R}$ and $r_i \prec r$ for $1 \leq i \leq n$.

We now show how the added expressive power of complex role inclusions can be used to incorporate the error detection previously carried out by two CQs into just one CQ. The basic idea is that both CQs are supposed to detect cycles of a certain kind. So we can define a new role $badcycle$ that spans role chains which, if we identified their first and their last elements would lead to $\mathfrak{q}_1$ or $\mathfrak{q}_2$ being satisfied.

$$word \sqsubseteq badcycle \quad (23)$$

$$corr \circ word \circ corr^- \circ word \sqsubseteq badcycle \quad (24)$$

Note that these axioms are in accordance with the mentioned global constraints. Obviously, in order to ensure that an interpretation matches neither $\mathfrak{q}_1$ nor $\mathfrak{q}_2$, we just have to forbid $badcycle$-loops, i.e., we must require that the one-atom CQ $\{badcycle(x, x)\}$ is not satisfied.

Theorem 12. *Let $\mathbb{P}$ be a PCP instance and let $\mathcal{K}'_{\mathbb{P}}$ be the $\mathcal{SROIF}^-$ KB consisting of Axioms 1–24. Then the answer to $\mathbb{P}$ is "yes" if and only if $\mathcal{K}'_{\mathbb{P}} \not\models_{\text{fin}} \{badcycle(x, x)\}$.*

Corollary 13. *Finite conjunctive query entailment from $\mathcal{SROIF}^-$ knowledge bases is undecidable.*

Finite 2RPQ Entailment from $\mathcal{ALCOIF}$ KBs

We next show undecidability of a problem involving two-way regular path queries, which we first will formally define.

Definition 14 (Two-way Regular Path Queries)**.** *A two-way regular path expression (2RPE) is a regular expression over the alphabet $\mathbf{R}$ consisting of role names and their inverses. Given an interpretation $\mathcal{I}$, the semantics of a 2RPE exp is a binary relation such that $exp^{\mathcal{I}}$ contains all pairs $(\delta, \delta') \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ for which there is a word $r_1 \ldots r_n$ of roles matching exp such that there exist domain elements $\delta_0 \ldots \delta_n$ with $\delta_0 = \delta$ and $\delta_n = \delta'$ and $(\delta_{i-1}, \delta_i) \in r_i^{\mathcal{I}}$ for $1 \leq i \leq n$. A 2-way regular path query (2RPQ) is one atom $exp(t, t')$ where exp is a 2RPE and t, t' are terms. Evaluation, satisfaction and entailment for 2RPQs are then defined in the same way as for conjunctive queries.*

Furthermore, we recall that an $\mathcal{ALCOIF}$ knowledge base is a $\mathcal{SHOIF}$ knowledge base that does not have role inclusions nor transitivity axioms.

It has been established that the problem of CQ entailment from $\mathcal{SROIQ}$ KBs can be reduced to the problem of conjunctive 2RPQ entailment from $\mathcal{ALCHOIQ}$ KBs using automata-theoretic methods for modifying the knowledge base and rewriting the query (Kazakov 2008; Demri and Nivelle 2005; Ortiz, Rudolph, and Simkus 2011). As this technique is modular with respect to most used modeling features and preserves (cardinality of) models, it can be used to transform the problem of (finite) entailment of one-atom-CQ from $\mathcal{SROIF}^-$ KBs to the problem of (finite) 2RPQ entailment from $\mathcal{ALCOIF}$ KBs. In particular, this reduction can be used to establish the following result.

Theorem 15. *Let $\mathbb{P}$ be a PCP instance and let $\mathcal{K}''_{\mathbb{P}}$ be the $\mathcal{ALCOIF}$ knowledge base consisting of Axioms 1–20. Then the answer to $\mathbb{P}$ is "yes" if and only if $\mathcal{K}''_{\mathbb{P}} \not\models_{\text{fin}} (next)^+ \cup corr \cdot (next)^+ \cdot corr^- \cdot (next)^+(x, x)$.*

Note that, instead of employing the transformation sketched above, this theorem can also be directly proven very much along the lines of the previous proof with only very minor modifications.

Corollary 16. *Finite entailment of two-way regular path queries from $\mathcal{ALCOIF}$ knowledge bases is undecidable.*

2RPQ Entailment from $\mathcal{ALCOIF}_{\text{reg}}$ KBs

The description logic $\mathcal{ALCOIF}_{\text{reg}}$ is obtained from $\mathcal{ALCOIF}$ by allowing concept expressions of the form $\exists exp.C$ where exp is a 2RPE and C is a concept expression. The semantics of such concept expressions is defined in the straightforward way, based on semantics of 2RPEs introduced above.

Note that progressing from $\mathcal{ALCOIF}$ to $\mathcal{ALCOIF}_{\text{reg}}$ is quite a significant extension. Most notably, unlike most mainstream description logics, $\mathcal{ALCOIF}_{\text{reg}}$ is not a fragment of first-order logic, as it for instance allows for expressing reachability.

In our case, we can use the new type of expressions to axiomatically enforce that each model must be a finite chain of $next$s leading from $start^{\mathcal{I}}$ to $end^{\mathcal{I}}$ without "externally" imposing the finite model assumption. We simply state that every domain element starts a path of $next$s ending in $end^{\mathcal{I}}$ and a path of $next^-$s ending in $start^{\mathcal{I}}$.

$$\top \sqsubseteq \exists next^*.\{end\} \quad (25)$$

$$\top \sqsubseteq \exists (next^-)^*.\{start\} \quad (26)$$

With this additional axioms at hand, we can now easily establish the next theorem.

[3] We denote this description logic by $\mathcal{SROIF}^-$, since according to the common nomenclature, $\mathcal{SROIF}$ would contain more modeling features such as self-loops, the universal role, and role disjointness.

Theorem 17. *Let $\mathbb{P}$ be a PCP instance and let $\mathcal{K}_{\mathbb{P}}'''$ be the $\mathcal{ALCOIF}_{\text{reg}}$ knowledge base consisting of Axioms 1–20 and Axioms 25 and 26. Then the answer to $\mathbb{P}$ is "yes" if and only if $\mathcal{K}_{\mathbb{P}}'' \not\models corr \cdot (next)^+ \cdot corr^- \cdot (next)^+(x,x)$.*

Note that here, the query does not need to detect looping $next$ chains since their existence is already prevented by Axioms 25 and 26 together with Axioms 1–5.

Corollary 18. *Entailment of two-way regular path queries from $\mathcal{ALCOIF}_{\text{reg}}$ knowledge bases is undecidable.*

It might be worth noting that dropping one of the three constructs inverses, functionality or nominals from the logic makes the problem decidable again, even if further modeling features are added and positive 2RPQs (i.e., arbitrary Boolean combinations of 2RPQs) are considered (Calvanese, Eiter, and Ortiz 2009).

Note that the above finding can be turned into a slight generalization of an already known result: Let $\mathcal{ALCOIF}^*$ be the restriction of the description logic $\mathcal{ALCOIF}_{\text{reg}}$ where all regular expressions are of the form r^* for $r \in \mathbf{R}$. A *transitive closure-enhanced conjunctive query* (TC-CQ) is a conjunctive query allowing for atoms of the form $r^*(t_1, t_2)$ for $r \in \mathbf{R}$. Satisfaction and entailment of such queries are defined in the straightforward way. It was shown that entailment of unions of TC-CQs from $\mathcal{ALCOIF}^*$ knowledge bases is undecidable (Ortiz, Rudolph, and Simkus 2010). By using the above construction and noting that the 2RPQ $corr \cdot (next)^+ \cdot corr^- \cdot (next)^+(x,x)$ is (with respect to entailment) equivalent to the TC-CQ $\{corr(x_1,x_2),\ next(x_2,x_3),\ next^*(x_3,x_4),\ corr(x_5,x_4),\ next(x_5,x_6),\ next^*(x_6,x_1)\}$, we can establish the following corollary slightly strengthening the previous result.

Corollary 19. *Entailment of TC-CQs from $\mathcal{ALCOIF}^*$ knowledge bases is undecidable.*

Satisfiability of $\mathcal{ALCOIF}_{\omega\text{reg}}$ KBs

The DL $\mathcal{ALCOIF}_{\text{reg}}$ introduced in the previous section featured the possibility to describe unbounded, yet finite chains of roles. Opposed to this, it might also be desirable to describe infinite chains of roles. In fact, this is a feature not uncommon in temporal variants of modal logics and can, e.g., be used to express liveness properties. While regular expressions are used to characterize finite role chains, the appropriate notion for infinite role chains would be ω-regular expressions.

Definition 20 (ω-Regular Expressions, 2ωRPQs)**.** *Given an alphabet $\mathcal{A}$, an ω-regular expression is defined as follows:*

- *if exp is a regular expression not matching ϵ, then exp^ω is an ω-regular expression,*
- *if exp_1 and exp_2 are two ω-regular expressions, then $exp_1 \cup exp_2$ is an ω-regular expression,*
- *for every regular expression exp and every ω-regular expressions exp', $exp \cdot exp'$ is an ω-regular expression.*

We associate with each ω-regular expression exp over $\mathcal{A}$ a set of infinite words over $\mathcal{A}$, denoted by $[exp]$, inductively as follows:

- *if exp is a regular expression not matching ϵ, then $[exp^\omega] = \{v_1 v_2 \cdots \mid \forall i \in \mathbb{N} : v_i$ matches $exp\}$,*
- *if exp_1 and exp_2 are two ω-regular expressions, then $[exp_1 \cup exp_2] = [exp_1] \cup [exp_2]$,*
- *for every regular expression exp and every ω-regular expressions exp', $[exp \cdot exp'] = \{vv' \mid v$ matches $exp,\ v' \in [exp']\}$.*

If for an ω-regular expression exp, an infinite word v satisfies $v \in [exp]$, we also say v matches *exp.*

*Given a set **R** of roles (i.e., role names and their inverses), a* two-way ω-regular path expression *(2ωRPE) is a ω-regular expression over the alphabet **R**.*

We now let $\mathcal{ALCOIF}_{\omega\text{reg}}$ denote the description logic $\mathcal{ALCOIF}$ extended by concept expressions of the form $\exists exp.\infty$ with exp an 2ωRPE. The semantics of these expressions, which we call *ω-concepts*, is defined as follows $(\exists exp.\infty)^{\mathcal{I}}$ consists of those $\delta \in \Delta^{\mathcal{I}}$ for which there exist an infinite word $r_1 r_2 \cdots$ over role names and their inverses matching exp and an infinite sequence $\delta_0, \delta_1, \ldots$ of elements from $\Delta^{\mathcal{I}}$ such that $\delta = \delta_0$ and for every $i \in \mathbb{N}$ holds $(\delta_i, \delta_{i+1}) \in r_i^{\mathcal{I}}$.

Intuitively, we will use the new expressivity provided by ω-concepts to prevent the existence of infinite paths of certain shapes. In particular, we prevent infinite $next$-paths as well as paths of infinitely repeated $corr \cdot next^n \cdot corr^- \cdot next^m$-sequences.

$$\exists next^\omega.\infty \quad \sqsubseteq \quad \bot \qquad (27)$$

$$\exists (corr \cdot next^+ \cdot corr^- \cdot next^+)^\omega.\infty \quad \sqsubseteq \quad \bot \qquad (28)$$

Theorem 21. *Let $\mathbb{P}$ be a PCP instance and let $\mathcal{K}_{\mathbb{P}}''''$ be the $\mathcal{ALCOIF}_{\text{reg}}$ knowledge base consisting of Axioms 1–20 and Axioms 27 and 28. Then the answer to $\mathbb{P}$ is "yes" if and only if $\mathcal{K}_{\mathbb{P}}''''$ is satisfiable.*

Proof. The proof is very similar to that of Theorem 10, in turn referring to Lemmas 7 and 8, with the following notable modifications: First, Axiom 27 is the one to ensure that every model of $\mathcal{K}_{\mathbb{P}}''''$ is a finite $next$-chain starting from $start^{\mathcal{I}}$ and ending in $end^{\mathcal{I}}$. Second, if $corr$ connects non-corresponding word boundaries, we find a looping $corr \cdot next^n \cdot corr^- \cdot next^m$-chain as argued in the proof of Lemma 8, therefore $(\exists (corr \cdot next^+ \cdot corr^- \cdot next^+)^\omega.\infty)^{\mathcal{I}}$ is non-empty; a contradiction. Third, it is easy to check that in any solution model, $(\exists (corr \cdot next^+ \cdot corr^- \cdot next^+)^\omega.\infty)^{\mathcal{I}}$ is necessarily empty. □

Corollary 22. *Satisfiability of $\mathcal{ALCOIF}_{\omega\text{reg}}$ knowledge bases is undecidable.*

The description logic $\mathcal{ALCOIF}_{\omega\text{reg}}$ might seem a bit contrived at the first glance. It should however be noted that it constitutes a fragment of the so-called *fully enriched μ-calculus* and its description logic version $\mu\mathcal{ALCIO}_f$ (Bonatti 2003; Bonatti and Peron 2004; Bonatti et al. 2008). We will not go into details about this logic here, we just note that in particular, $\exists next^\omega.\infty$ can be expressed in $\mu\mathcal{ALCIO}_f$ as $\nu X.\exists next.X$ and $\exists (corr \cdot next^+ \cdot corr^- \cdot next^+)^\omega.\infty$ can be expressed by $\nu X.\exists corr.\exists next.\mu Y.\big((\exists next.Y) \sqcup \exists.corr^-.\exists next.\mu Z.(\exists next.Z) \sqcup X\big)$.

We note that these concept expressions correspond to the so-called *aconjunctive fragment* of the μ-calculus (Kozen 1983) which, roughly speaking, only allows one to describe situations which are essentially linear. We let $\mu\mathcal{ALCIO}_f^{\text{acon}}$ denote $\mu\mathcal{ALCIO}_f$ where fixpoint expressions must be in aconjunctive form. Then the following corollary improves on a previous undecidability result for $\mu\mathcal{ALCIO}_f$ (Bonatti 2003) (the proof of which hinges upon the use of non-aconjunctive fixpoint expressions).

Corollary 23. *Satisfiability of $\mu\mathcal{ALCIO}_f^{\text{acon}}$ knowledge bases is undecidable.*

Again it is noteworthy that removing any of the three modeling features inverses, functionality, or nominals (in μ-calculus terminology: the features of being full, graded, or hybrid), makes the problem decidable again (Bonatti et al. 2008).

ω2RPQ Entailment from $\mathcal{ALCOIF}$ KBs

The last reasoning problem considered here is very close to the previous one, the difference being that we allow ω-regular expressions in the query language rather than in the logic itself.

Definition 24 (Two-way ω-Regular Path Queries)**.** A two-way ω-regular path query *(2ωRPQ) is an atom of the shape $exp(t)$ where exp is a 2ωRPE and t is a term. For an interpretation $\mathcal{I}$ and an evaluation π, we define that $\mathcal{I} \models_\pi exp(t)$ holds iff there exist an infinite word $r_1 r_2 \cdots$ over role names and their inverses matching exp and an infinite sequence $\delta_0, \delta_1, \ldots$ of elements from $\Delta^\mathcal{I}$ such that $\pi(t) = \delta_0$ and for every $i \in \mathbb{N}$ holds $(\delta_i, \delta_{i+1}) \in r_i^\mathcal{I}$. Entailment of 2$\omega$RPQs from knowledge bases is defined in the straightforward way.*

Note that the query atom must be of unary arity, since an infinite chain of roles has only a defined starting but no ending point. As it turns out, the previous undecidability result concerning satisfiability of $\mathcal{ALCOIF}_{\omega\text{reg}}$ KBs can be directly transformed into one regarding ω2RPQ entailment from $\mathcal{ALCOIF}$ KBs, since in the former, ω-concepts were only used to detect and exclude problematic situations. This allows us to effortlessly rephrase the construction into a query entailment problem.

Theorem 25. *Let $\mathbb{P}$ be a PCP instance and, as before, let $\mathcal{K}''_\mathbb{P}$ be the $\mathcal{ALCOIF}$ knowledge base consisting of Axioms 1–20. Then the answer to $\mathbb{P}$ is "yes" if and only if $\mathcal{K}''_\mathbb{P} \not\models next^\omega \cup (corr \cdot next^+ \cdot corr^- \cdot next^+)^\omega(x)$.*

Corollary 26. *Entailment of two-way ω-regular path queries from $\mathcal{ALCOIF}$ knowledge bases is undecidable.*

Conclusion and Future Work

In this paper, we have approached the decidability boundary from above for database-inspired reasoning problems for very expressive description logics that allow for inverses, counting and nominals, a combination that is known for causing complications when it comes to reasoning tasks, in particular when coupled with expressive means for describing role chains of unbounded or even infinite length. We have focused on query answering and the finite model semantics and showed that for a bunch of reasoning problems from that realm, a reduction of the Post Correspondence Problem can be achieved through slight modifications of one generic construction.

These findings clarify the decidability status of interesting reasoning problems around very expressive DLs, some of which are complemented by decidability results for sublogics with just one modeling feature removed. Still, there are numerous related reasoning problems whose decidability status remains open. In particular, decidability is unknown for the following problems (with some dependencies between them as stated below):

- P1 (U)CQ entailment from $\mathcal{SHOIF}$ KBs. A version of *the* long-standing open problem. For UCQs, the finite-model version has been settled (negatively) in this paper, but there is little hope that this will provide insights toward a solution of the unrestricted model case.
- P2 Finite CQ entailment from $\mathcal{SHOIF}$ KBs.
- P3 (U)CQ entailment from $\mathcal{SROIF}$ KBs. Decidability of this problem would entail decidability of P1 and essentially boil down to decidability of conjunctive query answering in OWL 2 DL.
- P4 2RPQ entailment from $\mathcal{ALCOIF}$ KBs. Note that the case is open only for "looping" 2RPQs, where the two terms in the atom are the same variable. For all other 2RPQs, the problem is decidable by a reduction to (un)satisfiability of $\mathcal{ALCOIF}$. The finite entailment case was settled (negatively) in this paper.
- P5 (Unions of) Conjunctive 2RPQ entailment from $\mathcal{ALCOIF}$ KBs. This problem is equivalent to P3 and its decidability would entail decidability of P4 and P1.
- P6 Finite satisfiability of $\mathcal{ALCOIF}_{\text{reg}}$ KBs
- P7 Satisfiability of $\mathcal{ALCOIF}_{\text{reg}}$ KBs. Decidability of this problem entails decidability of P6, since model-finiteness can be axiomatized in $\mathcal{ALCOIF}_{\text{reg}}$.
- P8 Finite CQ entailment from $\mathcal{ALCOIF}_{\text{reg}}$ KBs. Clearly, decidability of this problem entails decidability of P6.
- P9 CQ entailment from $\mathcal{ALCOIF}_{\text{reg}}$ KBs. For the aforementioned reasons, decidability of this problem would entail decidability of all P8, P7, and P6.

It should be noted that for many of the problems, removing one of the features inverses, nominals, or functionality would make the problem decidable. This is the case for P1, P3, P4, P5, P7, and P9 as can be inferred from decidability of positive two-way relational path query (P2RPQ) entailment from the extremely expressive DLs $\mathcal{ZIQ}$, $\mathcal{ZOQ}$, and $\mathcal{ZOI}$ knowledge bases (Calvanese, Eiter, and Ortiz 2009).

On another note, the same subset of the problems are known to be decidable when just the Horn fragment of the underlying description logic is considered, following from the decidability of entailment of unions of conjunctive 2RPQs from Horn-$\mathcal{SROIQ}$ KBs (Ortiz, Rudolph, and Simkus 2011).[4]

[4]Regarding P7 and P9, to be fair, one should state that going to the Horn fragment essentially disables the interesting uses of regular expressions, i.e., Horn-$\mathcal{ALCOIF}_{\text{reg}}$ is not more expressive than Horn-$\mathcal{ALCOIF}$.

Acknowledgements The author sincerely thanks the three reviewers for their valuable and detailed comments. On a more casual note, he is commemorating selected female beings for epitomizing the notion of undecidability.

References

Baader, F.; Calvanese, D.; McGuinness, D.; Nardi, D.; and Patel-Schneider, P., eds. 2007. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, second edition.

Bárány, V.; Gottlob, G.; and Otto, M. 2014. Querying the guarded fragment. *Logical Methods in Computer Science* 10(2).

Bienvenu, M.; Calvanese, D.; Ortiz, M.; and Simkus, M. 2014. Nested regular path queries in description logics. In Baral, C.; Giacomo, G. D.; and Eiter, T., eds., *Proc. 14th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'14)*. AAAI Press.

Bischoff, S.; Krötzsch, M.; Polleres, A.; and Rudolph, S. 2014. Schema-agnostic query rewriting for SPARQL 1.1. In Mika, P.; Tudorache, T.; Bernstein, A.; Welty, C.; Knoblock, C. A.; Vrandečić, D.; Groth, P. T.; Noy, N. F.; Janowicz, K.; and Goble, C. A., eds., *Proc. 13th Int. Semantic Web Conf. (ISWC'14)*, volume 8796 of *LNCS*, 584–600. Springer.

Bonatti, P. A., and Peron, A. 2004. On the undecidability of logics with converse, nominals, recursion and counting. *Artificial Intelligence* 158(1):75–96.

Bonatti, P. A.; Lutz, C.; Murano, A.; and Vardi, M. Y. 2008. The complexity of enriched mu-calculi. *Logical Methods in Computer Science* 4(3).

Bonatti, P. A. 2003. On the undecidability of description and dynamic logics with recursion and counting. In Gottlob, G., and Walsh, T., eds., *Proc. 18th Int. Joint Conf. on Artificial Intelligence (IJCAI'03)*, 331–336. Morgan Kaufmann.

Bourhis, P.; Krötzsch, M.; and Rudolph, S. 2014. How to best nest regular path queries. In Bienvenu, M.; Ortiz, M.; Rosati, R.; and Simkus, M., eds., *Informal Proc. 27th Int. Workshop on Description Logics (DL'04)*, volume 1193 of *CEUR Workshop Proceedings*, 404–415. CEUR-WS.org.

Buneman, P. 1997. Semistructured data. In Mendelzon, A. O., and Özsoyoglu, Z. M., eds., *Proc. 16th Symposium on Principles of Database Systems (PODS'97)*, 117–121. ACM Press.

Calvanese, D.; Eiter, T.; and Ortiz, M. 2007. Answering regular path queries in expressive description logics: An automata-theoretic approach. In *Proc. 22nd AAAI Conf. on Artificial Intelligence (AAAI'07)*, 391–396. AAAI Press.

Calvanese, D.; Eiter, T.; and Ortiz, M. 2009. Regular path queries in expressive description logics with nominals. In Boutilier, C., ed., *Proc. 21st Int. Joint Conf. on Artificial Intelligence (IJCAI'09)*, 714–720. IJCAI.

Calvanese, D. 1996. Finite model reasoning in description logics. In Aiello, L. C.; Doyle, J.; and Shapiro, S. C., eds., *Proc. 5th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'96)*, 292–303. Morgan Kaufmann.

Chandra, A. K., and Merlin, P. M. 1977. Optimal implementation of conjunctive queries in relational data bases. 77–90.

Demri, S., and Nivelle, H. 2005. Deciding regular grammar logics with converse through first-order logic. *J. of Logic, Language and Information* 14(3):289–329.

Horrocks, I., and Sattler, U. 2007. A tableau decision procedure for *SHOIQ*. *J. Automated Reasoning* 39(3):249–276.

Horrocks, I.; Kutz, O.; and Sattler, U. 2006. The even more irresistible $\mathcal{SROIQ}$. In Doherty, P.; Mylopoulos, J.; and Welty, C. A., eds., *Proc. 10th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'06)*, 57–67. AAAI Press.

Ibáñez-García, Y. A.; Lutz, C.; and Schneider, T. 2014. Finite model reasoning in Horn description logics. In Baral, C.; Giacomo, G. D.; and Eiter, T., eds., *Proc. 14th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'14)*. AAAI Press.

Kazakov, Y. 2008. $\mathcal{RIQ}$ and $\mathcal{SROIQ}$ are harder than $\mathcal{SHOIQ}$. In Brewka, G., and Lang, J., eds., *Proc. 11th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'08)*, 274–284. AAAI Press.

Kozen, D. 1983. Results on the propositional mu-calculus. *Theoretical Computer Science* 27:333–354.

Krötzsch, M.; Simančík, F.; and Horrocks, I. 2012. A description logic primer. *CoRR* abs/1201.4089.

Levy, A. Y., and Rousset, M. 1996. CARIN: A representation language combining Horn rules and description logics. In Wahlster, W., ed., *Proc. 12th European Conf. on Artificial Intelligence (ECAI'96)*, 323–327. John Wiley and Sons.

Lutz, C.; Sattler, U.; and Tendera, L. 2005. The complexity of finite model reasoning in description logics. *Information and Computation* 199(1-2):132–171.

McGuinness, D. L., and van Harmelen, F., eds. 10 February 2004. *OWL Web Ontology Language Overview*. W3C Recommendation. Available at http://www.w3.org/TR/owl-features/.

Ortiz, M.; Rudolph, S.; and Simkus, M. 2010. Query answering is undecidable in DLs with regular expressions, inverses, nominals, and counting. INFSYS Research Report 1843-10-03, TU Vienna.

Ortiz, M.; Rudolph, S.; and Simkus, M. 2011. Query answering in the Horn fragments of the description logics SHOIQ and SROIQ. In Walsh, T., ed., *Proc. 22nd Int. Joint Conf. on Artificial Intelligence (IJCAI'11)*, 1039–1044. IJCAI/AAAI.

OWL Working Group, W. 27 October 2009. *OWL 2 Web Ontology Language: Document Overview*. W3C Recommendation. Available at http://www.w3.org/TR/owl2-overview/.

Post, E. L. 1946. A variant of a recursively unsolvable problem. *Bulletin of the American Mathematical Society* 52.

Pratt-Hartmann, I. 2005. Complexity of the two-variable fragment with counting quantifiers. *J. of Logic, Language and Information* 14:369–395.

Rosati, R. 2011. On the finite controllability of conjunctive query answering in databases under open-world assumption. *J. of Computer and System Sciences* 77(3):572–594.

Rudolph, S., and Glimm, B. 2010. Nominals, inverses, counting, and conjunctive queries or: Why infinity is your friend! *J. of Artificial Intelligence Research* 39:429–481.

Rudolph, S., and Krötzsch, M. 2013. Flag & check: Data access with monadically defined queries. In Hull, R., and Fan, W., eds., *Proc. 32nd Symp. on Principles of Database Systems (PODS'13)*, 151–162. ACM.

Rudolph, S. 2011. Foundations of description logics. In Polleres, A.; d'Amato, C.; Arenas, M.; Handschuh, S.; Kroner, P.; Ossowski, S.; and Patel-Schneider, P. F., eds., *Reasoning Web. Semantic Technologies for the Web of Data – 7th International Summer School 2011*, volume 6848 of *LNCS*. Springer. 76–136.

Rudolph, S. 2014. The two views on ontological query answering. In Gottlob, G., and Pérez, J., eds., *Proc. 8th Alberto Mendelzon Workshop on Foundations of Data Management (AMW'14)*, volume 1189 of *CEUR Workshop Proceedings*. CEUR-WS.org.