The Curse of Finiteness: Undecidability of 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 triad 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

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 [24].

Two major conceptual contributions of database theory can be identified: query answering as the central reasoning problem and finite-model semantics.

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

* This is a shortened version of a paper accepted at KR'16 [25], appended to this submission. Definitions common in the DL community and proofs have been omitted due to space reasons.

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 [12] and *unions of conjunctive queries* (UCQs). Answering conjunctive queries over DL knowledge bases was first mentioned as a topic in the 1990s [17] 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 $\mathcal{ALCHOIQ}b$ [26] and Horn-$\mathcal{SROIQ}$ [20].

In the context of semi-structured databases, other query formalisms have been developed which allow for expressing information needs related to reachability, so-called *path queries* or *navigational queries* [8]. Beyond expressing more elaborate information needs, such queries can also be used to internalize ontological knowledge into the query to a certain degree [3]. Over the past decade, a variety of results regarding answering of (diverse variants of) path queries over DL knowledge bases have been established [10, 11, 2] 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-$\mathcal{SROIQ}$ [20]. Current research progresses to even more expressive query languages most of which can be seen as fragments of Datalog [27, 7].

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 (all sublogics of $\mathcal{SROI}$) 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 [9] but spawned only little follow-up work [18, 14]. It was only in 2008 when finite satisfiability for $\mathcal{SROIQ}$ (and all its sublogics) was shown to be decidable [15], exploiting a result on the finite satisfiability for the counting two-variable fragment of first-order logic [22].

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 [1] which extend previous work on finite controllability in databases under the open-world assumption [23] entail that for $\mathcal{ALCHOI}b$ (and all its sublogics), 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 $\mathcal{ALCHOIQ}b$ [26]. 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 $\mathcal{ALCHOI}b$ for which finite UCQ entailment is known to be decidable is Horn-$\mathcal{ALCFI}$ [14].

[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.

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.

Preliminaries

Finite Model Reasoning Beyond the standard semantics for DLs, this paper also addresses reasoning under the finite-model semantics, a prominent (or even the standard) setting in database theory.

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 $\mathcal{K} \models_{\mathrm{fin}} q$ ($\mathcal{K} \models_{\mathrm{fin}} Q$), if $\mathcal{I} \models q$ ($\mathcal{I} \models Q$) holds for every interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ with $\mathcal{I} \models \mathcal{K}$ and finite $\Delta^{\mathcal{I}}$.*

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 1. Consider the knowledge base $\mathcal{K}_1$ consisting of the axioms $\mathsf{Fun}(r^-)$, $\top \sqsubseteq \exists r.\top$, and $\{a\} \sqsubseteq \neg\exists r^-.\top$. We find that $\mathcal{K}_1$ is satisfiable (witnessed by the interpretation $(\mathbb{N}, \{a \mapsto 0, r \mapsto \mathit{succ}\})$) but not finitely satisfiable (since the sum of r-indegrees and the sum of r-outdegrees cannot match in a finite model).

In a similar way, the $\mathcal{SHOIF}$ knowledge base $\mathcal{K}_2$ containing the axioms $\mathsf{Fun}(r^-)$, $\top \sqsubseteq \exists r.\top$, $r \sqsubseteq r'$, and $\mathsf{Trans}(r')$ does not entail the CQ $\{r'(x,x)\}$ (witnessed by the interpretation $(\mathbb{N}, \{r \mapsto \mathit{succ}, r' \mapsto <\})$), but $\mathcal{K}_2 \models_{\mathrm{fin}} \{r'(x,x)\}$.

The Post Correspondence Problem We will establish our undecidability result by a reduction from the well-known Post Correspondence Problem [21] defined as follows:

Definition 2 (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 2. Let $\mathbb{P} = \{(g_1, g'_1), (g_2, g'_2), (g_3, g'_3)\}$ where $g_1 = b$, $g'_1 = bbb$, $g_2 = ab$, $g'_2 = a$, $g_3 = bbba$, and $g'_3 = a$. Then $2, 1, 1, 3$ is a solution sequence due to the fact that $g_2\ g_1\ g_1\ g_3 = (ab)(b)(b)(bbba) = abbbbbba = (a)(bbb)(bbb)(a) = g'_2\ g'_1\ g'_1\ g'_3$. Therefore the answer to the PCP instance $\mathbb{P}$ is "yes".

Theorem 1 (Post, 1946). *The Post Correspondence Problem is undecidable.*

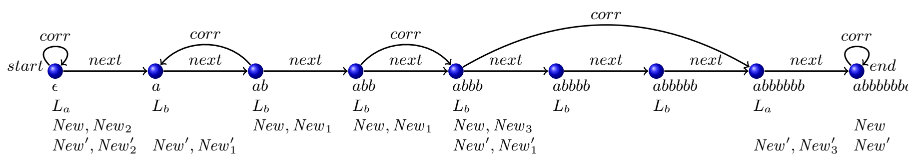

Fig. 1. Solution model for the PCP instance and solution sequence described in Example 2.

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 construct 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 3 (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:*

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

Thereby, $start$ *and* end *are two individual names,* L_a*,* L_b*,* New*,* New'*,* New_1*,* New'_1*,* $\ldots$ 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 2.

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.

$$\neg\{end\} \equiv \exists next.\top \quad (2) \qquad \neg\{start\} \equiv \exists next^{-}.\top \quad (3)$$

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

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

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

$$\neg\{end\} \equiv L_a \sqcup L_b \quad (6) \qquad L_a \sqcap L_b \sqsubseteq \bot \quad (7)$$

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 $I'_g := L_{c_1} \sqcap \exists next.(\neg New' \sqcap L_{c_2} \sqcap \exists next.(\neg New' \sqcap \ldots L_{c_\ell} \sqcap \exists next.New' \ldots))$.

$$\{start\} \sqsubseteq New \sqcap New' \quad (8)$$

$$New \sqcap \neg\{end\} \equiv New_1 \sqcup \ldots \sqcup New_\mu \quad (9) \qquad New_i \sqcap New_j \sqsubseteq \bot \quad (10)$$

$$New' \sqcap \neg\{end\} \equiv New'_1 \sqcup \ldots \sqcup New'_\mu \quad (11) \qquad New'_i \sqcap New'_j \sqsubseteq \bot \quad (12)$$

$$New_k \sqsubseteq I_{g_k} \quad (13) \qquad New'_k \sqsubseteq I'_{g_k} \quad (14)$$

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.

$$New \equiv \exists corr.\top \quad (15) \qquad New' \equiv \exists corr^{-}.\top \quad (16)$$

$$\mathsf{Fun}(corr) \quad (17) \qquad \mathsf{Fun}(corr^{-}) \quad (18)$$

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

$$New_k \sqsubseteq \exists corr.New'_k \quad (19) \qquad New'_k \sqsubseteq \exists corr^{-}.New_k \quad (20)$$

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.

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

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 $\mathsf{q}_1 = \{word(x,x)\}$ is supposed to detect looping $next$-chains (which must not exist in a solution model) and

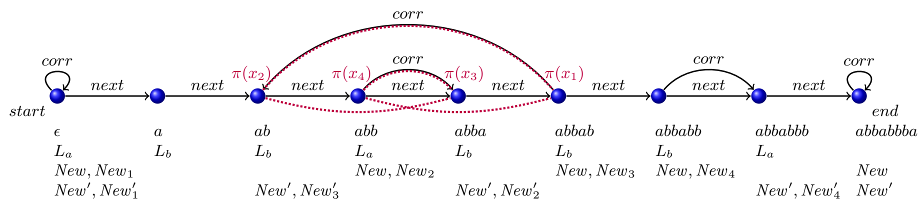

Fig. 2. Model for the knowledge base $\mathcal{K}_{\mathbb{P}'}$ derived from the PCP instance $\mathbb{P}'$ described in Example 3. For better readability, the $word$ role has not been drawn, it is defined to hold between any two individuals connected by a directed chain of $next$ roles. Note that this 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 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 1. *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 $\mathcal{I}' \not\models \{\mathfrak{q}_1, \mathfrak{q}_2\}$.*

Lemma 2. *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 $\mathbb{P}$.*

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

Example 3. Consider $\mathbb{P}' = \{(g_1, g_1'), (g_2, g_2'), (g_3, g_3'), (g_4, g_4')\}$ with $g_1 = abb$, $g_1' = ab$, $g_2 = ab$, $g_2' = bbb$, $g_3 = b$, $g_3' = ba$, $g_4 = ba$, 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 2. *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 $\mathcal{K}_{\mathbb{P}} \not\models_{\text{fin}} \{\mathfrak{q}_1, \mathfrak{q}_2\}$.*

Corollary 1. *Finite entailment of unions of conjunctive queries from $\mathcal{SHOIF}$ knowledge bases is undecidable.*[2]

[2] Briefly before submitting the camera ready version of this paper, the author was made aware by Carsten Lutz that this corollary can indeed be strengthened to plain conjunctive queries by a refinement of the construction used in the proof. For didactic and space reasons, we decided against including the not too difficult, yet somewhat unwieldy argument.

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 $r_1, \ldots, r_n, r \in \mathbf{R}$.

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.

$$word \sqsubseteq badcycle \quad (23) \qquad\qquad corr \circ word \circ corr^- \circ word \sqsubseteq badcycle \quad (24)$$

Note that these axioms are in accordance with the simplicity and regularity constraints commonly imposed on DLs with role chain axioms. 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 3. *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 $\mathcal{K}'_{\mathbb{P}} \not\models_{\mathrm{fin}} \{badcycle(x,x)\}$.*

Corollary 2. *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 (2RPQs). We assume the reader to be familiar with these queries as well as the underlying notion of *two-way regular path expressions* (2RPEs). 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 [15, 13, 20]. 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.

[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 4. *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 $\mathcal{K}''_{\mathbb{P}} \not\models_{\mathrm{fin}} (next)^+ \cup corr{\cdot}(next)^+{\cdot}corr^-{\cdot}(next)^+(x,x)$.*

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

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

The description logic $\mathcal{ALCOIF}_{\mathrm{reg}}$ 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 $\mathcal{ALCOIF}_{\mathrm{reg}}$ is quite a significant extension. Most notably, unlike most mainstream description logics, $\mathcal{ALCOIF}_{\mathrm{reg}}$ 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 $start^{\mathcal{I}}$.

$$\top \sqsubseteq \exists next^*.\{end\} \quad (25) \qquad\qquad \top \sqsubseteq \exists (next^-)^*.\{start\} \quad (26)$$

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

Theorem 5. *Let $\mathbb{P}$ be a PCP instance and let $\mathcal{K}'''_{\mathbb{P}}$ be the $\mathcal{ALCOIF}_{\mathrm{reg}}$ 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 4. *Entailment of two-way regular path queries from $\mathcal{ALCOIF}_{\mathrm{reg}}$ knowledge bases is undecidable.*

It might be worth noting that dropping one of the three constructs of 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 [11].

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 $\mathcal{ALCOIF}_{\mathrm{reg}}$ 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 [19]. 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 5. *Entailment of TC-CQs from $\mathcal{ALCOIF}^*$ knowledge bases is undecidable.*

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

The DL $\mathcal{ALCOIF}_{\mathrm{reg}}$ 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 4 **(ω-Regular Expressions, 2ωRPQs).** *For an alphabet $\mathcal{A}$, the set OEX of ω-regular expressions over $\mathcal{A}$ is $OEX ::= NEX^\omega \mid OEX \cup OEX \mid EX \cdot OEX$, where NEX are the regular expressions not matching the empty word, and EX are all regular expressions. We associate with each ω-regular expression exp over $\mathcal{A}$ a set of infinite words over $\mathcal{A}$, denoted by $[exp]$, in the straightforward way (where exp^ω denotes indefinite repetition of words matching 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 $\mathcal{ALCOIF}_{\omega\mathrm{reg}}$ 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 $r_1r_2\cdots$ 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 $(\delta_i, \delta_{i+1}) \in r_i^\mathcal{I}$ holds for every $i \in \mathbb{N}$.

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.

$$\exists next^\omega.\infty \sqsubseteq \bot \quad (27) \qquad\qquad \exists(corr \cdot next^+ \cdot corr^- \cdot next^+)^\omega.\infty \sqsubseteq \bot \quad (28)$$

Theorem 6. *Let $\mathbb{P}$ be a PCP instance and let $\mathcal{K}_\mathbb{P}''''$ be the $\mathcal{ALCOIF}_{\mathrm{reg}}$ 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.*

Corollary 6. *Satisfiability of $\mathcal{ALCOIF}_{\omega\mathrm{reg}}$ knowledge bases is undecidable.*

The DL $\mathcal{ALCOIF}_{\omega\mathrm{reg}}$ 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 DL version $\mu\mathcal{ALCIO}_f$ [4, 6, 5]. We will not go into details 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 $\nu X.\exists corr.\exists next.\mu Y.\big((\exists next.Y) \sqcup \exists.corr^-.\exists next.\mu Z.(\exists next.Z) \sqcup X\big)$. These concept expressions correspond to the so-called *aconjunctive fragment* of the μ-calculus [16] which, roughly speaking, only allows one to describe situations which are essentially linear. We let $\mu\mathcal{ALCIO}_f^{\mathrm{acon}}$ denote $\mu\mathcal{ALCIO}_f$ where fixpoint expressions are in aconjunctive form. Then the following corollary improves on a previous undecidability result for $\mu\mathcal{ALCIO}_f$ [4] (the proof of which hinges upon the use of non-aconjunctive fixpoint expressions).

Corollary 7. *Satisfiability of $\mu\mathcal{ALCIO}_f^{\mathrm{acon}}$ 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 [5].

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 5 (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 arity one, 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 $\mathcal{ALCOIF}_{\omega\mathrm{reg}}$ 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 7. *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 8. *Entailment of two-way ω-regular path queries from $\mathcal{ALCOIF}$ knowledge bases is undecidable.*

Conclusion and Future Work

We have established undecidability for database-inspired reasoning problems regarding very expressive description logics that allow for inverses, counting and nominals coupled with expressive means for describing role chains of unbounded or even infinite length. Focusing on query answering and the finite model semantics, we showed that for several 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, many 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 a very prominent longstanding 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 $\mathcal{ALCOIF}_{\mathrm{reg}}$ KBs
- P7 Satisfiability of $\mathcal{ALCOIF}_{\mathrm{reg}}$ KBs. Decidability of this problem entails decidability of P6, since model-finiteness can be axiomatized in $\mathcal{ALCOIF}_{\mathrm{reg}}$.
- P8 Finite CQ entailment from $\mathcal{ALCOIF}_{\mathrm{reg}}$ KBs. Clearly, decidability of this problem entails decidability of P6.
- P9 CQ entailment from $\mathcal{ALCOIF}_{\mathrm{reg}}$ 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 [11]. 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 [20].[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-$\mathcal{ALCOIF}_{\mathrm{reg}}$ is not more expressive than Horn-$\mathcal{ALCOIF}$.

References

1. Vince Bárány, Georg Gottlob, and Martin Otto. Querying the guarded fragment. *Logical Methods in Computer Science*, 10(2), 2014.
2. Meghyn Bienvenu, Diego Calvanese, Magdalena Ortiz, and Mantas Simkus. Nested regular path queries in description logics. In Chitta Baral, Giuseppe De Giacomo, and Thomas Eiter, editors, *Proc. 14th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'14)*. AAAI Press, 2014.
3. Stefan Bischoff, Markus Krötzsch, Axel Polleres, and Sebastian Rudolph. Schema-agnostic query rewriting for SPARQL 1.1. In Peter Mika, Tania Tudorache, Abraham Bernstein, Chris Welty, Craig A. Knoblock, Denny Vrandečić, Paul T. Groth, Natasha F. Noy, Krzysztof Janowicz, and Carole A. Goble, editors, *Proc. 13th Int. Semantic Web Conf. (ISWC'14)*, volume 8796 of *LNCS*, pages 584–600. Springer, 2014.
4. Piero A. Bonatti. On the undecidability of description and dynamic logics with recursion and counting. In Georg Gottlob and Toby Walsh, editors, *Proc. 18th Int. Joint Conf. on Artificial Intelligence (IJCAI'03)*, pages 331–336. Morgan Kaufmann, 2003.
5. Piero A. Bonatti, Carsten Lutz, Aniello Murano, and Moshe Y. Vardi. The complexity of enriched mu-calculi. *Logical Methods in Computer Science*, 4(3), 2008.
6. Piero A. Bonatti and Adriano Peron. On the undecidability of logics with converse, nominals, recursion and counting. *Artificial Intelligence*, 158(1):75–96, 2004.
7. Pierre Bourhis, Markus Krötzsch, and Sebastian Rudolph. How to best nest regular path queries. In Meghyn Bienvenu, Magdalena Ortiz, Riccardo Rosati, and Mantas Simkus, editors, *Informal Proc. 27th Int. Workshop on Description Logics (DL'04)*, volume 1193 of *CEUR Workshop Proceedings*, pages 404–415. CEUR-WS.org, 2014.
8. Peter Buneman. Semistructured data. In Alberto O. Mendelzon and Z. Meral Özsoyoglu, editors, *Proc. 16th Symposium on Principles of Database Systems (PODS'97)*, pages 117–121. ACM Press, 1997.
9. Diego Calvanese. Finite model reasoning in description logics. In Luigia Carlucci Aiello, Jon Doyle, and Stuart C. Shapiro, editors, *Proc. 5th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'96)*, pages 292–303. Morgan Kaufmann, 1996.
10. Diego Calvanese, Thomas Eiter, and Magdalena Ortiz. Answering regular path queries in expressive description logics: An automata-theoretic approach. In *Proc. 22nd AAAI Conf. on Artificial Intelligence (AAAI'07)*, pages 391–396. AAAI Press, 2007.
11. Diego Calvanese, Thomas Eiter, and Magdalena Ortiz. Regular path queries in expressive description logics with nominals. In Craig Boutilier, editor, *Proc. 21st Int. Joint Conf. on Artificial Intelligence (IJCAI'09)*, pages 714–720. IJCAI, 2009.
12. Ashok K. Chandra and Philip M. Merlin. Optimal implementation of conjunctive queries in relational data bases. pages 77–90, 1977.
13. Stéphane Demri and Hans Nivelle. Deciding regular grammar logics with converse through first-order logic. *J. of Logic, Language and Information*, 14(3):289–329, 2005.
14. Yazmin Angélica Ibáñez-García, Carsten Lutz, and Thomas Schneider. Finite model reasoning in Horn description logics. In Chitta Baral, Giuseppe De Giacomo, and Thomas Eiter, editors, *Proc. 14th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'14)*. AAAI Press, 2014.
15. Yevgeny Kazakov. $\mathcal{RIQ}$ and $\mathcal{SROIQ}$ are harder than $\mathcal{SHOIQ}$. In Gerhard Brewka and Jérôme Lang, editors, *Proc. 11th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'08)*, pages 274–284. AAAI Press, 2008.
16. Dexter Kozen. Results on the propositional mu-calculus. *Theoretical Computer Science*, 27:333–354, 1983.

17. Alon Y. Levy and Marie-Christine Rousset. CARIN: A representation language combining Horn rules and description logics. In Wolfgang Wahlster, editor, *Proc. 12th European Conf. on Artificial Intelligence (ECAI'96)*, pages 323–327. John Wiley and Sons, 1996.
18. Carsten Lutz, Ulrike Sattler, and Lidia Tendera. The complexity of finite model reasoning in description logics. *Information and Computation*, 199(1-2):132–171, 2005.
19. Magdalena Ortiz, Sebastian Rudolph, and Mantas Simkus. Query answering is undecidable in DLs with regular expressions, inverses, nominals, and counting. INFSYS Research Report 1843-10-03, TU Vienna, 2010.
20. Magdalena Ortiz, Sebastian Rudolph, and Mantas Simkus. Query answering in the Horn fragments of the description logics SHOIQ and SROIQ. In Toby Walsh, editor, *Proc. 22nd Int. Joint Conf. on Artificial Intelligence (IJCAI'11)*, pages 1039–1044. IJCAI/AAAI, 2011.
21. Emil L. Post. A variant of a recursively unsolvable problem. *Bulletin of the American Mathematical Society*, 52, 1946.
22. Ian Pratt-Hartmann. Complexity of the two-variable fragment with counting quantifiers. *J. of Logic, Language and Information*, 14:369–395, 2005.
23. Riccardo Rosati. On the finite controllability of conjunctive query answering in databases under open-world assumption. *J. of Computer and System Sciences*, 77(3):572–594, 2011.
24. Sebastian Rudolph. The two views on ontological query answering. In Georg Gottlob and Jorge Pérez, editors, *Proc. 8th Alberto Mendelzon Workshop on Foundations of Data Management (AMW'14)*, volume 1189 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.
25. Sebastian Rudolph. Undecidability results for database-inspired reasoning problems in very expressive description logics. In *Proc. 15th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'15)*. AAAI Press, 2016. to appear.
26. Sebastian Rudolph and Birte Glimm. Nominals, inverses, counting, and conjunctive queries or: Why infinity is your friend! *J. of Artificial Intelligence Research*, 39:429–481, 2010.
27. Sebastian Rudolph and Markus Krötzsch. Flag & check: Data access with monadically defined queries. In Richard Hull and Wenfei Fan, editors, *Proc. 32nd Symp. on Principles of Database Systems (PODS'13)*, pages 151–162. ACM, 2013.