Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning
Main Track

Finite Entailment of UCRPQs over $\mathcal{ALC}$ Ontologies

Víctor Gutiérrez-Basulto[1] , **Albert Gutowski**[2] , **Yazmín Ibáñez-García**[1] , **Filip Murlak**[2]
[1]Cardiff University, UK
[2]University of Warsaw, Poland
{a.gutowski, f.murlak}@mimuw.edu.pl, {gutierrezbasultov, ibanezgarciay}@cardiff.ac.uk

Abstract

We investigate the problem of *finite* entailment of ontology-mediated queries. We consider the expressive query language, unions of conjunctive regular path queries (UCRPQs), extending the well-known class of union of conjunctive queries, with regular expressions over roles. We look at ontologies formulated using the description logic $\mathcal{ALC}$, and show a tight 2EXPTIME upper bound for entailment of UCRPQs. At the core of our decision procedure, there is a novel automata-based technique introducing a stratification of interpretations induced by the deterministic finite automaton underlying the input UCRPQ.

1 Introduction

At the intersection of knowledge representation and database theory lies the fundamental problem of *ontology-mediated query entailment (OMQE)*, where the background knowledge provided by an ontology is used to enrich the answers to queries posed to databases. In this context, description logics (DLs) are a widely accepted family of logics used to formulate ontologies. By now, the OMQE problem under the unrestricted semantics (reasoning over arbitrary models) is well understood for various query languages and DLs (Schneider and Simkus 2020). In contrast, for the *finite* OMQE problem, where one is interested in reasoning over finite models only, the overall landscape is rather incomplete. However, in recent years, the study of finite OMQE has been gaining traction, considering both lightweight and expressive DLs and (mostly) unions of conjunctive queries (Rosati 2008; Ibáñez-García, Lutz, and Schneider 2014; Rudolph 2016; Gogacz, Ibáñez-García, and Murlak 2018; Gogacz et al. 2019; Danielski and Kieronski 2019; Gogacz et al. 2020; Bednarczyk and Kieroński 2022).

In this paper we consider the problem of finite OMQE with unions of conjunctive regular path queries (UCRPQs) as the query language. UCRPQs (Florescu, Levy, and Suciu 1998; Calvanese et al. 2000) are a powerful navigational query language for graph databases in which one can express that two entities are related by a path of edges that can be specified by a regular language over binary relations. So, UCRPQs extend unions of conjunctive queries (UCQs) with atoms that might contain regular expressions that traverse the edges of the database. Indeed, path navigation is included in the query language XPath 2.0 for XML data, and it is also present in the SPARQL 1.1 query language for RDF data through the property path feature. Given the resemblance of instance data stored in ABoxes in DLs to graph-like data, several investigations on unrestricted entailment of various types of navigational query languages mediated by DL ontologies have been carried out (Stefanoni et al. 2014; Calvanese, Eiter, and Ortiz 2014; Bienvenu, Ortiz, and Simkus 2015; Gutiérrez-Basulto, Ibáñez-García, and Jung 2018; Gogacz et al. 2019; Bednarczyk and Rudolph 2019), yielding algorithmic approaches and optimal complexity bounds. For finite entailment of regular path queries mediated by DL ontologies, there are only undecidability results available (Rudolph 2016). The most relevant positive news are the decidability and computational complexity results by Danielski and Kieronski (2019) and Gogacz et al. (2020) on finite entailment of conjuctive queries with transitive closure over roles mediated by expressive DL ontologies.

We focus on ontologies formulated using the description logic $\mathcal{ALC}$. Note that entailment of UCRPQs over $\mathcal{ALC}$ ontologies is not *finitely controllable*, i.e. finite and unrestricted entailment do not coincide as it is *not* the case that for any $\mathcal{ALC}$ knowledge base $\mathcal{K}$ and any UCRPQ φ, it holds that $\mathcal{K}$ entails φ over all (unrestricted) models iff $\mathcal{K}$ entails φ over all finite models. By assuming that the represented world is finite, we can therefore not reuse existing complexity bounds or algorithmic approaches to UCRPQ entailment. From a usability perspective, the suitability of this assumption depends on the potential applications. A particular interest for navigational queries comes from bioinformatics and cheminformatics (Lysenko et al. 2016; Cook et al. 2016; Galgonek et al. 2016; Hu, Qiu, and Dumontier 2015; Rajabi and Sanchez-Alonso 2021; Chen et al. 2020). For instance, experts often need to find associations between entities in protein, cellular, drug, and disease networks (represented as graph databases), so that e.g. gene-disease-drug associations (corresponding to paths in the database) can be discovered for developing new treatment methods. In this type of applications, databases and the models they represent are clearly meant to be finite. Importantly, biochemical networks contain complex motifs involving e.g. *cycles* or cliques. This type of patterns can be described using UCRPQs, however, without the finiteness assumption these patterns could be disregarded as the associated query might not be entailed when reasoning over all models (including infinite ones).

1.1 Contribution

The main technical contribution of our investigation is the development of a dedicated automata-based method for entailment of UCRPQs over $\mathcal{ALC}$ ontologies, providing an optimal upper bound. More precisely, we obtain the following result, where the matching lower bound is inherited from (Ortiz and Simkus 2014).

Theorem 1. *Finite entailment of UCRPQs over $\mathcal{ALC}$ ontologies is* 2EXPTIME*-complete.*

In prior work, Rudolph (2016) showed that finite entailment of 2RPQs in $\mathcal{ALCIOF}$ is undecidable. Theorem 1 thus provides a key step towards delimiting the decidability boundary of finite OMQE with navigational queries.

At the heart of our approach to finite entailment of UCRPQs in $\mathcal{ALC}$ there is a stratification of interpretations induced by the deterministic finite automaton underlying the UCRPQ. This stratification builds upon the so-called *tape construction*, previously used to efficiently evaluate queries in the extension of XPath 1.0 where arbitrary regular expressions may appear as path expressions (Bojańczyk and Parys 2011). To realize the tape construction, our method represents UCRPQs by means of a semiautomaton $\mathcal{B}$ (Ginzburg 1968) and defines an expansion of $\mathcal{B}$, allowing to trace runs of $\mathcal{B}$ that begin in all possible states, on all infixes of the input word. We make interpretations $\mathcal{I}$ knowledgeable of the expansion by enriching paths of $\mathcal{I}$ with its possible runs and by associating edges of $\mathcal{I}$ with levels ℓ induced by the transitions of the expansion. In a similar fashion we also make CRPQs sensible of levels. With this at hand, we tackle finite entailment by eliminating the lowest level from a query and from an interpretation, and then recursively solving the simpler problem. At each step of this process, we should be able to arrange solutions to simpler problems in a hierarchical way so that we can reason over them. To this aim, we consider a variant of entailment that includes an *environment*, which will provide the necessary information to position the arranged solutions to simpler problems in the context of larger interpretations. To better keep track of the complexity of our recursive method, we introduce a modification of the entailment problem modulo environment in which we look at a particular type of finite models: (ℓ, ℓ')-*models*, which are models with edges of levels ℓ or higher that are 'consistent' w.r.t. queries referring to edges of level ℓ' or higher. We solve the problem of finding (ℓ, ℓ')-models recursively by increasing ℓ and ℓ' in an alternating way, until both reach the maximum level $n + 1$, with n the number of states of $\mathcal{B}$. This will mean solving finite entailment modulo environment, and thus standard finite entailment as well.

Missing proofs can be found in the technical report available at https://arxiv.org/pdf/2204.14261.pdf.

1.2 Related Work

We next discuss some existing work relevant to our study.

OMQE of Navigational Queries. As previously discussed, there exist various works on unrestricted entailment of navigational query languages mediated by DL ontologies. Most of them concentrate on extensions of regular path queries (RPQs), such as UCRPQs, and consider both Horn (Bienvenu, Ortiz, and Simkus 2015) and expressive DLs (Calvanese, Eiter, and Ortiz 2014; Gutiérrez-Basulto, Ibáñez-García, and Jung 2018; Gogacz et al. 2019; Bednarczyk and Rudolph 2019). There have been also some studies on entailment of graph XPath queries (Stefanoni et al. 2014; Bienvenu et al. 2014; Kostylev, Reutter, and Vrgoc 2014).

Finite OMQE. There exist various decidability results and optimal complexity bounds for finite entailment of union of conjunctive queries in Horn DLs (Rosati 2008; Ibáñez-García, Lutz, and Schneider 2014) and in expressive DLs from the $\mathcal{S}$ family (Gogacz, Ibáñez-García, and Murlak 2018; Gogacz et al. 2019; Danielski and Kieronski 2019). In most cases, the computational complexity coincides with that of the unrestricted case, but the algorithmic approaches are completely different. On the negative side, undecidability of finite entailment of UCQs in the more expressive DL $\mathcal{SHOIF}$ was shown by (Rudolph 2016), as well as the undecidability result for finite entailment of 2RPQs in $\mathcal{ALCIOF}$. Closer to our work are the positive results on finite entailment of UCQs with transitive closure over roles in expressive DLs allowing for transitivity or transitive closure over roles (Danielski and Kieronski 2019; Gogacz et al. 2020). These results close the distance to the undecidability frontier for finite entailment from a different angle by considering ontology languages more expressive than $\mathcal{ALC}$, but a subclass of UCRPQs as query language. In the context of database theory research, finite OMQE (also called open-world query entailment) has also been investigated; for instance, Amarilli and Benedikt (2020) study finite OMQE for inclusion dependencies and functional dependencies over relations of arbitrary arity, and Pratt-Hartmann (2009) looks at finite OMQE in the two-variable fragment of FOL with counting quantifiers.

Finite Controllability. There have been also a few works on finite controllability in the context of DLs. For instance, Bednarczyk and Kieroński (2022) recently showed that the $\mathcal{ZOI}$ and $\mathcal{ZOQ}$ members of the $\mathcal{Z}$ family are finitely controllable for UCQs. Beyond DLs, there have been several works on UCQ-finite controllability: for the guarded fragment of FOL (Bárány, Gottlob, and Otto 2014) or for various fragments of existential rules (Civili and Rosati 2012; Gogacz and Marcinkowski 2013; Baget et al. 2011; Amendola, Leone, and Manna 2018; Gottlob, Manna, and Pieris 2018). Closer to our study, is the work by Figueira, Figueira, and Baque (2020) on the classification of finitely and non-finitely controllable subclasses of CRPQs over ontologies formulated in the guarded-negation fragment of FOL or in the frontier fragment of existential rules. However, no complexity results or algorithms for finite entailment are provided for the non-finitely controllable cases.

2 Preliminaries

2.1 Description Logics

We consider a vocabulary consisting of countably infinite disjoint sets of *concept names* $\mathsf{N_C}$, *role names* $\mathsf{N_R}$, and *individual names* $\mathsf{N_I}$. *$\mathcal{ALC}$-concepts* C, D are defined by the grammar

$$C, D ::= A \mid \neg C \mid C \sqcap D \mid \exists r.C$$

where $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. We use standard abbreviations $\bot$, $\top$, $C \sqcup D$ and $\forall r.C$.

An *$\mathcal{ALC}$-TBox $\mathcal{T}$* is a finite set of *concept inclusions (CIs)* $C \sqsubseteq D$, where C, D are $\mathcal{ALC}$-concepts. An *ABox $\mathcal{A}$* is a finite non-empty set of *concept* and *role assertions* of the form $A(a)$, $r(a, b)$, where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$ and $\{a, b\} \subseteq \mathsf{N_I}$. We write $\mathsf{ind}(\mathcal{A})$ for the *set of individual names* occurring in $\mathcal{A}$. A *knowledge base (KB)* is a pair $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. We write $\mathsf{CN}(\mathcal{K})$ and $\mathsf{rol}(\mathcal{K})$ for the *sets of all concept and role names* occurring in $\mathcal{K}$. We let $\|\mathcal{K}\|$ be the total size of the representation of $\mathcal{K}$.

Without loss of generality, we assume throughout the paper that all CIs are in one of the following *normal forms*:

$$\bigsqcap_i A_i \sqsubseteq \bigsqcup_j B_j, \quad A \sqsubseteq \exists r.B, \quad A \sqsubseteq \forall r.B,$$

where $A, A_i, B, B_j \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and empty disjunction and conjunction are equivalent to $\bot$ and $\top$, respectively. Additionally, for each $A \in \mathsf{CN}(\mathcal{K})$ there is a complementary $\bar{A} \in \mathsf{CN}(\mathcal{K})$ axiomatized with $\top \sqsubseteq A \sqcup \bar{A}$ and $A \sqcap \bar{A} \sqsubseteq \bot$.

2.2 Interpretations

The semantics is given as usual via *interpretations* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ consisting of a non-empty *domain* $\Delta^\mathcal{I}$ and an *interpretation function* $\cdot^\mathcal{I}$ mapping concept names to subsets of the domain and role names to binary relations over the domain, and individual names to elements of the domain. The interpretation of complex concepts C is defined in the usual way (Baader et al. 2017). An interpretation $\mathcal{I}$ is a *model of a TBox $\mathcal{T}$*, written $\mathcal{I} \models \mathcal{T}$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$ for all CIs $C \sqsubseteq D \in \mathcal{T}$. It is a *model of an ABox $\mathcal{A}$*, written $\mathcal{I} \models \mathcal{A}$, if $\mathsf{ind}(\mathcal{A}) \subseteq \Delta^\mathcal{I}$, $a^\mathcal{I} = a$ for each $a \in \mathsf{ind}(\mathcal{A})$, $(a, b) \in r^\mathcal{I}$ for all $r(a, b) \in \mathcal{A}$, and $a \in A^\mathcal{I}$ for all $A(a) \in \mathcal{A}$. The first two conditions constitute the so-called *standard name assumption*. Finally, $\mathcal{I}$ is a *model of a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$*, written $\mathcal{I} \models \mathcal{K}$, if $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I} \models \mathcal{A}$.

An interpretation $\mathcal{I}$ is *finite* if $\Delta^\mathcal{I}$ is finite. An interpretation $\mathcal{I}'$ is a *sub-interpretation* of $\mathcal{I}$, written as $\mathcal{I}' \subseteq \mathcal{I}$, if $\Delta^{\mathcal{I}'} \subseteq \Delta^\mathcal{I}$, $A^{\mathcal{I}'} \subseteq A^\mathcal{I}$, and $r^{\mathcal{I}'} \subseteq r^\mathcal{I}$ for all $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. For $\Sigma \subseteq \mathsf{N_C} \cup \mathsf{N_R}$, $\mathcal{I}$ is an interpretation *over signature* Σ if $A^\mathcal{I} = \emptyset$ and $r^\mathcal{I} = \emptyset$ for all $A \in \mathsf{N_C} \setminus \Sigma$ and $r \in \mathsf{N_R} \setminus \Sigma$. The union $\mathcal{I} \cup \mathcal{J}$ of $\mathcal{I}$ and $\mathcal{J}$ is an interpretation such that $\Delta^{\mathcal{I} \cup \mathcal{J}} = \Delta^\mathcal{I} \cup \Delta^\mathcal{J}$, $A^{\mathcal{I} \cup \mathcal{J}} = A^\mathcal{I} \cup A^\mathcal{J}$, and $r^{\mathcal{I} \cup \mathcal{J}} = r^\mathcal{I} \cup r^\mathcal{J}$ for all $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$.

A *unary $\mathcal{K}$-type* is a subset of $\mathsf{CN}(\mathcal{K})$ including either A or $\bar{A}$ for each $A \in \mathsf{CN}(\mathcal{K})$. For an interpretation $\mathcal{I}$ and an element $d \in \Delta^\mathcal{I}$, the *unary $\mathcal{K}$-type of d in $\mathcal{I}$* is $\mathsf{tp}^\mathcal{I}(d) = \{A \in \mathsf{CN}(\mathcal{K}) \mid d \in A^\mathcal{I}\}$. We say that $\mathcal{I}$ *realizes* a unary $\mathcal{K}$-type τ if $\tau = \mathsf{tp}^\mathcal{I}(d)$ for some $d \in \Delta^\mathcal{I}$.

2.3 Queries and Finite Entailment

We next introduce the query language. We concentrate on Boolean queries, that is, queries without answer variables. The extension to queries with answer variables is standard; see, for example, (Glimm et al. 2008). A *conjunctive regular path query (CRPQ)* is a first-order formula

$$\varphi = \exists \mathbf{x}\, \psi(\mathbf{x})$$

such that $\psi(\mathbf{x})$ is constructed using $\wedge$ over atoms of the form $A(t)$ or $\mathcal{E}(t, t')$ where $A \in \mathsf{N_C}$, t, t' are variables from $\mathbf{x}$ or individual names from $\mathsf{N_I}$, and $\mathcal{E}$ is a *path expression* defined by the grammar

$$\mathcal{E}, \mathcal{E}' ::= r \mid \mathcal{E}^* \mid \mathcal{E} \cup \mathcal{E}' \mid \mathcal{E} \circ \mathcal{E}'$$

where $r \in \mathsf{N_R}$. Thus, $\mathcal{E}$ is essentially a regular expression over the (infinite) alphabet $\{r \mid r \in \mathsf{N_R}\}$. The set of individual names in φ is denoted with $\mathsf{ind}(\varphi)$. A *conjunctive query (CQ)* is a CRPQ that does not use the operators $*, \cup$ and $\circ$ in path expressions, and a *regular path query (RPQ)* consists of a single atom of the form $\mathcal{E}(t, t')$.

The semantics of CRPQs is defined via matches. Let us fix a CRPQ $\varphi = \exists \mathbf{x}\, \psi(\mathbf{x})$ and an interpretation $\mathcal{I}$. A *match for φ in $\mathcal{I}$* is a function

$$\pi : \mathbf{x} \cup \mathsf{ind}(\varphi) \to \Delta^\mathcal{I}$$

such that $\pi(a) = a$, for all $a \in \mathsf{ind}(\varphi)$, and $\mathcal{I}, \pi \models \psi(\mathbf{x})$ under the standard semantics of first-order logic extended with a rule for atoms of the form $\mathcal{E}(t, t')$. More formally, we define:

- $\mathcal{I}, \pi \models \psi_1 \wedge \psi_2$ iff $\mathcal{I}, \pi \models \psi_1$ and $\mathcal{I}, \pi \models \psi_2$;
- $\mathcal{I}, \pi \models A(t)$ iff $\pi(t) \in A^\mathcal{I}$;
- $\mathcal{I}, \pi \models \mathcal{E}(t, t')$ iff $(\pi(t), \pi(t')) \in \mathcal{E}^\mathcal{I}$, where $\mathcal{E}^\mathcal{I}$ is defined inductively as $(\mathcal{E}^*)^\mathcal{I} = (\mathcal{E}^\mathcal{I})^*$, $(\mathcal{E}_1 \cup \mathcal{E}_2)^\mathcal{I} = \mathcal{E}_1^\mathcal{I} \cup \mathcal{E}_2^\mathcal{I}$, $(\mathcal{E}_1 \circ \mathcal{E}_2)^\mathcal{I} = \mathcal{E}_1^\mathcal{I} \circ \mathcal{E}_2^\mathcal{I}$.

An interpretation $\mathcal{I}$ *satisfies* φ, written $\mathcal{I} \models \varphi$, if there exists a match for φ in $\mathcal{I}$. A *union of CRPQs (UCRPQ)* is a finite set of CRPQs and a *union of CQs (UCQ)* is a finite set of CQs. An interpretation $\mathcal{I}$ satisfies an UCRPQ Φ, written as $\mathcal{I} \models \Phi$, if $\mathcal{I} \models \varphi$ for some $\varphi \in \Phi$. We say that $\mathcal{K}$ *finitely entails* Φ, written $\mathcal{K} \models_{\mathsf{fin}} \Phi$, if each finite model of $\mathcal{K}$ satisfies Φ. A model of $\mathcal{K}$ that does not satisfy Φ is a *counter-model*. The *finite entailment problem* asks if a given KB $\mathcal{K}$ finitely entails a given query Φ.

2.4 UCRPQs via Semiautomata

We work with UCRPQs represented by means of a *semiautomaton* (Ginzburg 1968) $\mathcal{B} = (Q, \Gamma, \delta)$ where Q is a finite set of states, $\Gamma \subseteq \{r \mid r \in \mathsf{N_R}\}$ is a finite alphabet—throughout the paper we assume $\Gamma = \mathsf{rol}(\mathcal{K})$, and $\delta : Q \times \Gamma \to Q$ is the transition function. A semiautomaton is essentially a deterministic finite automaton without initial and final states; a run of a semiautomaton $\mathcal{B}$ over a word w is defined just like for a finite automaton, except that it can begin in any state and there is no notion of accepting runs. Under this representation, an RPQ is an atom over a binary predicate of the form $\mathcal{B}_{q,q'}$ where $q, q' \in Q$ are states of $\mathcal{B}$. We let $\mathcal{I}, \pi \models \mathcal{B}_{q,q'}(t, t')$ iff $(\pi(t), \pi(t')) \in \mathcal{B}^\mathcal{I}_{q,q'}$ where $\mathcal{B}^\mathcal{I}_{q,q'}$ is the set of pairs (e, e') such that for some $n \in \mathbb{N}$ there exist $r_1, \ldots, r_n \in \Gamma$ and $e_0, \ldots, e_n \in \Delta^\mathcal{I}$ such that

- $e_0 = e$ and $e_n = e'$;

- $(e_{i-1}, e_i) \in (r_i)^\mathcal{I}$ for all $i \in \{1, \ldots, n\}$;
- there exists a run of $\mathcal{B}$ on the word $r_1 \ldots r_n$ that begins in state q and ends in state q'.

We also allow *edge atoms* of the form $r(x, x')$ for $r \in \Gamma$.

Each UCRPQ Φ can be effectively rewritten into a UCRPQ Φ' expressed by means of a semiautomaton $\mathcal{B}$ of size $k \cdot 2^{O(m)}$ where k is the number of path expressions in Φ and m is their maximal size. The size of CRPQs in Φ' is bounded by the size of CRPQs in Φ and $|\Phi'| = 2^{\text{poly}\|\Phi\|}$, where $\|\Phi\|$ is the total size of Φ.

For simplicity we work with KBs $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ where the ABox $\mathcal{A}$ is *trivial*; that is, $\mathsf{ind}(\mathcal{A}) = \{a\}$ for some $a \in \mathsf{N_I}$ and $\mathcal{A}$ contains only concept assertions. The general finite entailment problem can be reduced to this special case using the following lemma.

Lemma 1. *Given an oracle for finite entailment for trivial ABoxes, the general finite entailment* $(\mathcal{T}, \mathcal{A}) \models_{\mathsf{fin}} \Phi$ *can be decided in time* $2^{\text{poly}(\|(\mathcal{T},\mathcal{A})\|)\cdot 2^{\text{poly}(\|\Phi\|)}}$ *using calls to the oracle for* $\mathcal{K}' = (\mathcal{T}, \mathcal{A}')$ *and* Φ' *consisting of* $2^{\text{poly}(\|\Phi\|)}$ *CRPQs of linear size over the same semiautomaton as* Φ.

2.5 Entailment Modulo Environment

We solve the entailment problem using a divide-and-conquer approach in which counter-models are decomposed into simpler ones, whose existence is easier to decide. Each level of this recursive procedure will involve certain modifications to the TBox. For complexity reasons we need to pay close attention to these changes, making sure that no blow-up is involved. To make it easier, we generalize the entailment problem by turning the modifications into a separate part of the input, which allows fixing the TBox for the duration of the whole procedure. At every level of the recursion, we will need to reason 'externally' about the way simpler pieces are put together to form the larger counter-model, and 'internally' about how to specify the required properties of a piece depending on what is happening outside. We will think of the models as induced subinterpretations of a larger interpretation. Dually, the remaining part of the larger interpretation can be seen as an external context, in which our models live. The relevant features of this context will be represented by environments, which we now define.

An *environment* $\mathcal{E} = (\Theta, \varepsilon)$ consists of a set Θ of unary types and a function $\varepsilon : \Theta \to 2^{\mathsf{rol}(\mathcal{K}) \times \mathsf{CN}(\mathcal{K})}$. The intended meaning is that only types from Θ are allowed and each element of an allowed unary type τ has an r-edge to an element in the extension of B in the external context for each $(r, B) \in \varepsilon(\tau)$. Accordingly, we say that $\mathcal{I}$ is a *model of* $\mathcal{K}$ *modulo* $\mathcal{E}$ and write $\mathcal{I} \models^{\mathcal{E}} \mathcal{K}$ if $\mathcal{I}$ realizes only unary types from Θ and it is a model of $\mathcal{K}$ under the following *relaxed semantics of existential restrictions*:

- for every existential restriction $\exists r.B$ in $\mathcal{K}$ and every element $d \in \Delta^\mathcal{I}$, $d \in (\exists r.B)^\mathcal{I}$ iff either there is an r-edge in $\mathcal{I}$ from d to an element $e \in B^\mathcal{I}$ or $(r, B) \in \varepsilon\big(\mathsf{tp}^\mathcal{I}(d)\big)$.

(The semantics of universal restrictions is not altered and it is the environment's reponsibility to account for them.) Correspondingly, a query Φ is *finitely entailed by* $\mathcal{K}$ *modulo* $\mathcal{E}$, written $\mathcal{K} \models^{\mathcal{E}}_{\mathsf{fin}} \Phi$, if for each finite interpretation $\mathcal{I}$, if $\mathcal{I} \models^{\mathcal{E}} \mathcal{K}$ then $\mathcal{I} \models \Phi$. The problem of *finite entailment modulo environment* is to decide for a given KB $\mathcal{K}$, environment $\mathcal{E}$, and query Φ if $\mathcal{K} \models^{\mathcal{E}}_{\mathsf{fin}} \Phi$.

Note that finite entailment modulo environment and ordinary finite entailment are interreducible. In one direction, it is enough to take the set of all unary $\mathcal{K}$-types for Θ and set $\varepsilon(\tau) = \emptyset$ for all $\tau \in \Theta$. In the other direction, the conditions imposed on unary types and the relaxed semantics of existential restrictions can be expressed easily in the TBox. The latter reduction, however, might significantly increase the size of the TBox. It is easier to control the size of the input at different levels of the recursion when these conditions are explicitly represented in the environment.

3 Expansion and Decorations

In order to handle UCRPQs expressed by means of a semiautomaton $\mathcal{B}$ we need to be able to trace runs of $\mathcal{B}$ that begin in all possible states, on all infixes of the input word. We achieve this using the following construction.

Let us fix an arbitrary linear order on the set Q of the states of $\mathcal{B}$. The *expansion* of $\mathcal{B}$ is a semiautomaton $\widehat{\mathcal{B}}$ whose set of states is the set $\widehat{Q}$ of all permutations of Q. Thus, an element of $\widehat{Q}$ can be seen as a tuple $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ such that p_i is the image of the ith state of $\mathcal{B}$ under the respective permutation. We refer to positions in this tuple as *levels*. In particular, the *level of* $q \in Q$ *in* $\mathbf{p}$ is the unique i such that $q = p_i$. Assuming $\delta : Q \times \mathsf{rol}(\mathcal{K}) \to Q$ is the transition function of $\mathcal{B}$, we define the transition function

$$\hat{\delta} : \widehat{Q} \times \mathsf{rol}(\mathcal{K}) \to \widehat{Q}$$

of $\widehat{\mathcal{B}}$ by letting $\hat{\delta}\big(\mathbf{p}, r\big)$ be the permutation $\mathbf{p}'$ obtained by listing all states appearing in the sequence

$$\delta(\mathbf{p}, r) = \big(\delta(p_1, r), \delta(p_2, r), \ldots, \delta(p_n, r)\big)$$

in the order of their first appearances, followed by all remaining states of $\mathcal{B}$ ordered as in Q. Note that the level of $\delta(p_i, r)$ in $\mathbf{p}'$ is at most i. Consider the set $P \subseteq \{1, 2, \ldots, n\}$ of levels i such that the level of $\delta(p_i, r)$ in $\mathbf{p}'$ is equal to i. It follows from the definition of $\mathbf{p}'$ that $P = \{1, 2, \ldots, \ell\}$ for some $\ell \in \{1, 2, \ldots, n\}$. We call this number ℓ the *level of transition* $\mathbf{p} \xrightarrow{r} \mathbf{p}'$.

From each run of $\widehat{\mathcal{B}}$ on a word w we can reconstruct all runs of $\mathcal{B}$ on w. Let $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_m$ be a run of $\widehat{\mathcal{B}}$ on w. Consider a run $q_0, q_1, \ldots, q_m$ of $\mathcal{B}$ on w. For $i = 0, 1, \ldots, m$, let ℓ_i be the level of q_i in $\mathbf{p}_i$. Any sequence $\ell_0, \ell_1, \ldots, \ell_m$ associated like this with a run of $\mathcal{B}$ will be called a *thread* in the run of $\widehat{\mathcal{B}}$ (see Fig. 1). Notice that two threads that begin at different levels can meet at the same level somewhere along the run; if this happens they remain equal until the end of the run. Also, threads can be born in the middle of a run of $\widehat{\mathcal{B}}$, but they never disappear. A crucial property of threads is that they are non-increasing sequences: the level of q_{i+1} in $\mathbf{p}_{i+1}$ is bounded by the level of q_i in $\mathbf{p}_i$.

Lemma 2. *Let* $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_m$ *be a run of* $\widehat{\mathcal{B}}$ *on* w*, and let* q, q' *be states of* $\mathcal{B}$*. There is a run of* $\mathcal{B}$ *on* w *from* q *to* q' *iff there exist positions* $0 \leq j_1 < j_2 < \cdots < j_k = m$*, levels* $n \geq \ell_1 > \ell_2 > \cdots > \ell_k \geq 1$*, and states* $q_0, q_1, \ldots, q_k$ *with* $1 \leq k \leq n$ *such that* $q_0 = q$*,* $q_k = q'$*, and*

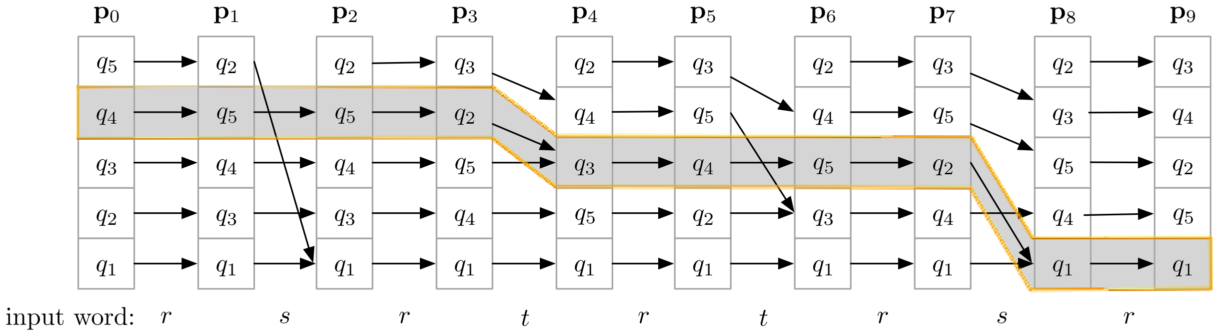

Figure 1: A thread in a run of the expansion of a semiautomaton.

- *the level of q_0 in $\mathbf{p}_0$ is ℓ_1 and the level of q_k in $\mathbf{p}_m$ is ℓ_k;*
- *for all $i \in \{1, 2, \dots, k-1\}$, the level of q_i in $\mathbf{p}_{j_i}$ is ℓ_i and the level of $\delta\big(q_i, w[j_i+1]\big)$ in $\mathbf{p}_{j_i+1}$ is ℓ_{i+1};*
- *for all $i \in \{1, 2, \dots, k\}$, each transition taken in the segment of the run from $\mathbf{p}_{j_{i-1}+1}$ (or $\mathbf{p}_0$ for $i = 1$) to $\mathbf{p}_{j_i}$ has level at least ℓ_i.*

As an illustration of Lemma 2, consider the run of the expanded semiautomaton shown in Fig. 1. Tracing the run of the original semiautomaton on the same word, starting in state q_4, we discover the positions $j_1 = 3$ and $j_2 = 7$ where the corresponding thread drops to a lower level. Between these positions, the thread stays at the same level, beginning with $\ell_1 = 4$ (taking transitions of levels $5, 4, 5 \geq \ell_1$), followed by $\ell_2 = 3$ (taking transitions of levels $5, 3, 5 \geq \ell_2$), and $\ell_3 = 1$ (taking a transition of level $5 \geq \ell_3$).

The next step is to account for the possible runs of $\widehat{\mathcal{B}}$ over paths in the interpretation. Towards this goal, we decorate elements of the interpretation with states of $\widehat{\mathcal{B}}$. To avoid additional blow-up, we represent states of $\widehat{\mathcal{B}}$ using combinations of fresh concept names $C_{q,\ell}$ where q is a state of $\mathcal{B}$ and $\ell \in \{1, 2, \dots, n\}$ is a level; we write $\mathsf{CN}(\widehat{\mathcal{B}})$ for the set of all $C_{q,\ell}$. For a state $\mathbf{p} = (p_1, p_2, \dots, p_n)$ of $\widehat{\mathcal{B}}$, by $C_\mathbf{p}$ we mean the concept $C_{p_1,1} \sqcap C_{p_2,2} \sqcap \dots \sqcap C_{p_n,n}$. We say that an element $e \in \Delta^\mathcal{I}$ is *decorated* with state $\mathbf{p}$ if $e \in C^\mathcal{I}_\mathbf{p}$. An interpretation $\mathcal{I}$ is *$\widehat{\mathcal{B}}$-decorated* if no element has incoming edges over different roles from $\mathsf{rol}(\mathcal{K})$ and $\mathcal{I}$ satisfies the CIs

$$C_\mathbf{p} \sqsubseteq \forall r.C_{\hat{\delta}(\mathbf{p},r)}\,, \quad C_\mathbf{p} \sqcap C_{\mathbf{p}'} \sqsubseteq \bot\,, \quad \top \sqsubseteq \bigsqcup_{\mathbf{p} \in \widehat{Q}} C_\mathbf{p}$$

for all states $\mathbf{p}, \mathbf{p}'$ of $\widehat{\mathcal{B}}$ such that $\mathbf{p} \neq \mathbf{p}'$. The axiomatization above is exponential in the size of $\mathcal{B}$, but we can do better.

Lemma 3. *Given $\mathcal{B}$ one can compute in polynomial time a TBox $\widehat{\mathcal{T}}_\mathcal{B}$ such that $\mathcal{I} \models \widehat{\mathcal{T}}_\mathcal{B}$ iff $\mathcal{I}$ is $\widehat{\mathcal{B}}$-decorated.*

To every edge in a $\widehat{\mathcal{B}}$-decorated interpretation $\mathcal{I}$ we can assign a level as follows. Consider elements $e, e' \in \Delta^\mathcal{I}$ such that $(e, e') \in r^\mathcal{I}$ for some $r \in \mathsf{rol}(\mathcal{K})$. Note that $(e, e') \notin s^\mathcal{I}$ for every $s \in \mathsf{rol}(\mathcal{K}) \setminus \{r\}$. Let $\mathbf{p}$ and $\mathbf{p}'$ be the states decorating e and e', respectively. It holds that $\mathbf{p} \xrightarrow{r} \mathbf{p}'$. By *the level of the edge* (e, e') we shall understand the level of this transition. A *level-ℓ interpretation* is a $\widehat{\mathcal{B}}$-decorated interpretation that does not contain edges of level strictly below ℓ; if $\ell > n$, this means that there are no edges at all. The following lemma is the key to our algorithm.

Lemma 4. *Consider a level-ℓ interpretation $\mathcal{I}$ and elements $e \in C^\mathcal{I}_{q,\ell}$ and $e' \in C^\mathcal{I}_{q',\ell}$. Then, $(e, e') \in \mathcal{B}^\mathcal{I}_{q,q'}$ iff there is a path from e to e' in $\mathcal{I}$.*

We make use of Lemma 4 by decomposing RPQs into segments corresponding to different levels, as was done for the runs of $\widehat{\mathcal{B}}$ in Lemma 2. To facilitate this, we make our queries aware of levels. A *$\widehat{\mathcal{B}}$-decorated CRPQ* is a CRPQ φ represented by means of semiautomaton $\mathcal{B}$ that contains exactly one atom of the form $C_{q,\ell}(x)$ and exactly one atom of the form $C_{q',\ell'}(x')$ for each atom $\mathcal{B}_{q,q'}(x, x')$ in φ. We call ℓ and ℓ' the *begin level* and the *end level* of atom $\mathcal{B}_{q,q'}(x, x')$ in φ, respectively. Because levels never increase in a thread of a run of $\widehat{\mathcal{B}}$, we can assume without loss of generality that $\ell \geq \ell'$ always holds. A *level-ℓ CRPQ* is a $\widehat{\mathcal{B}}$-decorated CRPQ that contains no RPQ atoms of end level strictly below ℓ. As all end levels are at most n, a level-ℓ CRPQ for $\ell > n$ contains no RPQ atoms; that is, it is a CQ. To *complete* a CRPQ φ means to turn it into a $\widehat{\mathcal{B}}$-decorated CRPQ φ' by adding unary atoms over concepts $C_{q,\ell}$ in an arbitrary minimal way. Each resulting φ' is called a *completion* of φ. Over $\widehat{\mathcal{B}}$-decorated interpretations, φ is equivalent to the union of its completions. The *completion of a UCRPQ* Φ is the union of all completions of all CRPQs in Φ.

We conclude this section by showing how to turn any counterexample to $\mathcal{K} \models^\mathcal{E}_\mathsf{fin} \Phi$ into a $\widehat{\mathcal{B}}$-decorated one. Let $\mathcal{I}$ be an interpretation over $\mathsf{CN}(\mathcal{K}) \cup \mathsf{rol}(\mathcal{K})$. The *product* of $\mathcal{I}$ and $\widehat{\mathcal{B}}$ is the interpretation $\mathcal{I} \times \widehat{\mathcal{B}}$ over $\mathsf{CN}(\widehat{\mathcal{B}}) \cup \mathsf{CN}(\mathcal{K}) \cup \mathsf{rol}(\mathcal{K})$ such that

- $\Delta^{\mathcal{I} \times \widehat{\mathcal{B}}} = \Delta^\mathcal{I} \times \mathsf{rol}(\mathcal{K}) \times \widehat{Q}$,
- $C^{\mathcal{I} \times \widehat{\mathcal{B}}} = C^\mathcal{I} \times \mathsf{rol}(\mathcal{K}) \times \widehat{Q}$ for all $C \in \mathsf{CN}(\mathcal{K})$,
- $C^{\mathcal{I} \times \widehat{\mathcal{B}}}_{q,\ell} = \Delta^\mathcal{I} \times \mathsf{rol}(\mathcal{K}) \times \{(p_1, p_2, \dots, p_n) \in \widehat{Q} : p_\ell = q\}$ for all $q \in Q$ and $\ell \in \{1, 2, \dots, n\}$,
- $r^{\mathcal{I} \times \widehat{\mathcal{B}}} = \big\{\big((e, s, \mathbf{p}), (e', r, \mathbf{p}')\big) : (e, e') \in r^\mathcal{I}, \mathbf{p} \xrightarrow{r} \mathbf{p}',$ $s \in \mathsf{rol}(\mathcal{K})\big\}$ for $r \in \mathsf{rol}(\mathcal{K})$.

Note that if $\mathcal{I}$ is finite, so is $\mathcal{I} \times \widehat{\mathcal{B}}$.

Lemma 5. *Let Φ be a UCRPQ, $\mathcal{K}$ an $\mathcal{ALC}$ KB with a trivial ABox, and $\mathcal{E}$ an environment.*

- *$\mathcal{I} \times \widehat{\mathcal{B}}$ is a $\widehat{\mathcal{B}}$-decorated interpretation.*
- *If $\mathcal{I} \not\models \Phi$ then $\mathcal{I} \times \widehat{\mathcal{B}} \not\models \Phi$.*
- *If $\mathcal{I} \models^{\mathcal{E}} \mathcal{K}$ then $\mathcal{I} \times \widehat{\mathcal{B}} \models^{\mathcal{E}} \mathcal{K}$ up to identifying the unique individual a in $\mathcal{K}$ with some $(a, r, \mathbf{p}) \in \Delta^{\mathcal{I} \times \widehat{\mathcal{B}}}$.*

4 Core Computational Problem

To solve the entailment problem we eliminate the lowest level from the query and from the interpretation, and solve the problem with fewer levels recursively. Eliminating each level will involve interpretations built from pieces that are solutions for the simplified problem. Evaluating CRPQs over such interpretations requires breaking them down into fragments and it must accommodate single RPQs witnessed across multiple pieces.

For a UCRPQ Φ let $\tilde{\Phi}$ be the completion of an equivalent UCRPQ represented by means of a semiautomaton $\mathcal{B}$. A *fragment* of $\varphi \in \tilde{\Phi}$ is either of the following:

- a $\widehat{\mathcal{B}}$-decorated CRPQ of the form $C_{q_1,\ell_1}(y_1) \wedge \mathcal{B}_{q_1,q_2}(y_1, y_2) \wedge C_{q_2,\ell_2}(y_2)$ or $C_{q_1,\ell_1}(y_1) \wedge \mathcal{B}_{q_1,q_2}(y_1, y_2) \wedge C_{q_2,\ell_2}(y_2) \wedge r(y_2, y_3) \wedge C_{q_3,\ell_3}(y_3)$ where y_1, y_2, y_3 are fresh variables and $r \in \mathsf{rol}(\mathcal{K})$,
- a connected $\widehat{\mathcal{B}}$-decorated CRPQ that can be obtained from φ by dropping selected atoms, replacing selected RPQ atoms $\mathcal{B}_{q,q'}(x, x')$ by a subset of $\mathcal{B}_{q,q_1}(x, y_1)$, $r(y_1, y_2)$, $\mathcal{B}_{q_3,q'}(y_3, x')$ for some fresh variables y_1, y_2, y_3 and $r \in \mathsf{rol}(\mathcal{K})$, and completing the resulting CRPQ.

A fragment of Φ is a fragment of any of the CRPQs in $\tilde{\Phi}$. Importantly, a fragment of a fragment of Φ is also a fragment of Φ, and each $\varphi \in \tilde{\Phi}$ is a fragment of Φ. Up to renaming fresh variables, Φ has $2^{\mathrm{poly}(\|\Phi\|)}$ different fragments, despite $\mathcal{B}$ being exponential in $\|\Phi\|$.

We now enrich interpretations again by including information about matched fragments of Φ. For each fragment φ of Φ and each $\emptyset \neq V \subseteq \mathit{var}(\varphi)$ we choose a fresh concept name $A_{\varphi,V}$. We call an interpretation $\mathcal{I}$ *correct (wrt. Φ)* if $e \in A^{\mathcal{I}}_{\varphi,V}$ iff $\pi(V) = \{e\}$ for some match π for φ in $\mathcal{I}$. Assuming $\mathcal{I}$ is correct, $\mathcal{I} \models \Phi$ iff $A^{\mathcal{I}}_{\varphi,V} \neq \emptyset$ for some $\varphi \in \tilde{\Phi}$ and $\emptyset \neq V \subseteq \mathit{var}(\varphi)$. Correctness is not compositional: the union of two correct interpretations sharing a single element need not be correct. As our method of eliminating levels relies on such decompositions of interpretations, we replace correctness with a notion that is weaker, but compositional.

We first abstract the decomposition of a $\widehat{\mathcal{B}}$-decorated CRPQ induced by a match in a union of disjoint 'peripheric' interpretations, each sharing a single element with a single 'core' interpretation (Fig. 2 shows three 'peripheric' interpretations connected to the 'core' by single edges, included in the 'peripheric' interpretations). A *partition* of a $\widehat{\mathcal{B}}$-decorated CRPQ φ into $\varphi', \varphi_1, \dots, \varphi_k$ is obtained as follows. Choose $X', X_1, \dots X_k \subseteq \mathit{var}(\varphi)$ such that

- $X_i \cap X_j = \emptyset$ for all $i \neq j$;

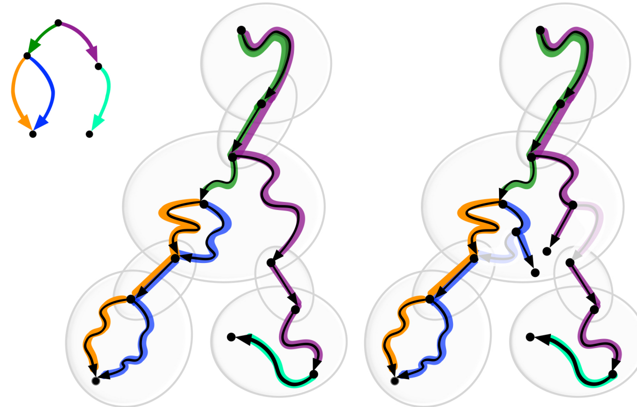

Figure 2: CRPQ φ is distributed over the bags constituting $\mathcal{I}$.

- for each atom of the form $r(x, x')$ in φ there exists i such that either $\{x, x'\} \subseteq X_i$ or $\{x, x'\} \subseteq X'$.

Based on $X', X_1, \dots X_k$ define $\varphi', \varphi_1, \dots, \varphi_k$ as follows. For each atom of the form $r(x, x')$ in φ choose i such that $\{x, x'\} \subseteq X_i$ and add $r(x, x')$ to φ_i or add $r(x, x')$ to φ' provided that $\{x, x'\} \subseteq X'$. For each RPQ atom $\mathcal{B}_{q,q'}(x, x')$ of begin level ℓ and end level ℓ' in φ do one of the following:

- provided that $\{x, x'\} \subseteq X'$, add $\mathcal{B}_{q,q'}(x, x')$ to φ';
- choose i such that $\{x, x'\} \subseteq X_i$ but $\{x, x'\} \not\subseteq X'$, and add $\mathcal{B}_{q,q'}(x, x')$ to φ_i (light green RPQ in Fig. 2);
- choose i such that $x \in X' \setminus X_i$ and $x' \in X_i \setminus X'$, a level m such that $\ell \geq m \geq \ell'$, a state p of $\mathcal{B}$, and a fresh variable y, and add $\mathcal{B}_{q,p}(x, y) \wedge C_{p,m}(y)$ to φ' and $C_{p,m}(y) \wedge \mathcal{B}_{p,q'}(y, x')$ to φ_i (blue and orange in Fig. 2);
- choose i such that $x \in X_i \setminus X'$ and $x' \in X' \setminus X_i$, a level m such that $\ell \geq m \geq \ell'$, a state p of $\mathcal{B}$, and a fresh variable y, and add $\mathcal{B}_{q,p}(x, y) \wedge C_{p,m}(y)$ to φ_i and $C_{p,m}(y) \wedge \mathcal{B}_{p,q'}(y, x')$ to φ' (dark green in Fig. 2);
- choose $i \neq j$ such that $x \in X_i \setminus X'$ and $x' \in X_j \setminus X'$, levels m, m' such that $\ell \geq m \geq m' \geq \ell'$, states p, p' of $\mathcal{B}$, and fresh variables y, y', add $\mathcal{B}_{q,p}(x, y) \wedge C_{p,m}(y)$ to φ_i, $C_{p,m}(y) \wedge \mathcal{B}_{p,p'}(y, y') \wedge C_{p',m'}(y')$ to φ', and $C_{p',m'}(y') \wedge \mathcal{B}_{p',q'}(y', x')$ to φ_j (purple in Fig. 2).

Note that for each $\mathcal{B}_{q,q'}(x, x')$ exactly one of the above actions can be performed and the choice of i and j is unique. To complete the construction, add to φ' all unary atoms of φ over variables already used in φ', and similarly for each φ_i. Observe that for each $X' \subseteq \mathit{var}(\varphi)$ there is exactly one choice of $X_1, X_2, \dots, X_k$ (up to a permutation) such that the resulting $\varphi_1, \varphi_2, \dots, \varphi_k$ are connected (regardless of the choice of p, p' and m, m'). Assuming that φ is a fragment of Φ, it then holds that so are $\varphi_1, \varphi_2, \dots, \varphi_k$.

We call $\mathcal{I}$ *consistent (wrt. Φ)* if for each partition of a fragment φ of Φ into a CRPQ φ' and fragments $\varphi_1, \varphi_2, \dots, \varphi_k$ with $\mathit{var}(\varphi_i) \cap \mathit{var}(\varphi_j) = \emptyset$ for $i \neq j$, $V_i = \mathit{var}(\varphi_i) \cap \mathit{var}(\varphi')$, and $\emptyset \neq V \subseteq \mathit{var}(\varphi) \cap \mathit{var}(\varphi')$, there is no match π for φ' in $\mathcal{I}$ such that $\pi(V_i) = \{e_i\} \subseteq \left(A_{\varphi_i,V_i}\right)^{\mathcal{I}}$ for all i but $\pi(V) = \{e\} \not\subseteq \left(A_{\varphi,V}\right)^{\mathcal{I}}$. Clearly, all correct interpretations are consistent. The converse is not true in general, but the following key property is preserved.

Lemma 6. *For every UCRPQ Φ and every consistent $\widehat{\mathcal{B}}$-decorated interpretation $\mathcal{I}$, if $A^{\mathcal{I}}_{\varphi,V} = \emptyset$ for each $\varphi \in \tilde{\Phi}$ and $\emptyset \neq V \subseteq var(\varphi)$, then $\mathcal{I} \not\models \Phi$.*

Consistency is sufficient to express entailment, but it does not yield well to the recursive elimination of levels. We generalize it by refining the information about matched fragments of Φ. We introduce fresh concepts $A^{\kappa}_{\varphi,V}$ where φ is a fragment of Φ, $\emptyset \neq V \subseteq var(\varphi)$,

$$\kappa : var(\varphi) \to \{1, 2, \dots, \ell\},$$

and $\kappa(V) = \{\ell\}$ for some $\ell \in \{1, 2, \dots, n+1\}$. We write $\mathsf{CN}^{\Phi}_{\ell}$ for the set of $A^{\kappa}_{\psi,V}$ such that $\kappa(V) = \{\ell\}$. Intuitively, κ is a synopsis of when specific fragments of ψ were matched during the recursive search for the model. Specifically, $\kappa(x) = \ell$ indicates that x was matched after all levels strictly below ℓ had been eliminated from the query, but while level ℓ was still present. Accordingly, ℓ-consistency, defined below, ensures that the synopses built so far are consistently updated while level ℓ is being handled.

We call $\mathcal{I}$ *ℓ-consistent (wrt. Φ)* if for each partition of a fragment φ of Φ into a CRPQ φ' of level ℓ and fragments $\varphi_1, \varphi_2, \dots, \varphi_k$ with $var(\varphi_i) \cap var(\varphi_j) = \emptyset$ for $i \neq j$, $V_i = var(\varphi_i) \cap var(\varphi')$, and $\emptyset \neq V \subseteq var(\varphi) \cap var(\varphi')$, there is no match π for φ' in $\mathcal{I}$ such that $\pi(V_i) = \{e_i\} \subseteq \left(A^{\kappa_i}_{\varphi_i,V_i}\right)^{\mathcal{I}}$ for all i but $\pi(V) = \{e\} \not\subseteq (A^{\kappa}_{\varphi,V})^{\mathcal{I}}$ where

- $\kappa_i(x) \leq \ell$ for all $x \in var(\varphi_i)$,
- $\kappa(x) = \kappa_i(x)$ for all $x \in var(\varphi_i) \setminus V_i$,
- $\kappa(x) = \ell$ for all $x \in var(\varphi) \cap var(\varphi')$.

We stress that while φ' has level ℓ, fragments $\varphi, \varphi_1, \dots, \varphi_k$ can have any level. Note also that ℓ-consistency speaks only of concept names in $\mathsf{CN}^{\Phi}_1 \cup \mathsf{CN}^{\Phi}_2 \cup \dots \cup \mathsf{CN}^{\Phi}_{\ell}$. Identifying $A_{\varphi,V}$ with $A^{\kappa}_{\varphi,V}$ for κ constantly equal to 1, we get that consistency and 1-consistency are equivalent.

In what follows, by an *(ℓ, ℓ')-interpretation* we mean an ℓ'-consistent level-ℓ interpretation. By an *(ℓ, ℓ')-model* of $\mathcal{K}$ modulo $\mathcal{E}$ we mean an (ℓ, ℓ')-interpretation that is model of $\mathcal{K}$ modulo $\mathcal{E}$. The actual problem we will be solving is the following *(ℓ, ℓ')-model problem* for $\ell \leq \ell'$: Given a KB $\mathcal{K}$ with a trivial ABox, an environment $\mathcal{E}$, and a UCRPQ Φ decide if there exists a finite (ℓ, ℓ')-model of $\mathcal{K}$ modulo $\mathcal{E}$.

By Lemma 5, entailment modulo environment (with trivial ABox) can be reduced to the $(1, 1)$-model problem by modifying the environment to forbid all unary types containing $A^{\kappa}_{\varphi,V}$ for any $\varphi \in \tilde{\Phi}$, $\emptyset \neq V \subseteq var(\varphi)$, and κ constantly equal 1. Note that the reduction does not affect the query Φ, nor the KB $\mathcal{K}$. However, it introduces up to $2^{\text{poly}(\|\Phi\|)}$ new concept names $A_{\varphi,V}$ and $C_{q,\ell}$. Consequently, the number of unary types is at most $2^{|\mathsf{CN}(\mathcal{K})| + 2^{\text{poly}(\|\Phi\|)}}$. It follows that the size of the environment is bounded by $2^{\|\mathcal{K}\| + 2^{\text{poly}(\|\Phi\|)}}$.

To solve the $(1, 1)$-model problem we will proceed recursively, incrementing ℓ and ℓ' in an alternating fashion, until $\ell = \ell' = n + 1$. At each level of the recursion we will be making multiple recursive calls. During the recursion the UCRPQ Φ and the TBox $\mathcal{T}$ will remain unchanged, but the ABox and the environment will evolve. Importantly, we will not introduce any new concepts, so the size of the environment will always be bounded by $2^{\|\mathcal{K}\| + 2^{\text{poly}(\|\Phi\|)}}$. The size of the ABox will be bounded by $\|\mathcal{K}\| + 2^{\text{poly}(\|\Phi\|)}$ and the number of individuals will never grow. In consequence, the total cost of the algorithm can be computed as the cost of a single recursion step times the number of steps. In the following sections we will show that each recursion step can be carried out in time $2^{O(\|\mathcal{K}\|) + 2^{\text{poly}(\|\Phi\|)}}$, excluding the cost of the recursive calls. The depth of the recursion is $O(n) = 2^{\text{poly}(\|\Phi\|)}$. The number of recursive calls within a single recursion step is also bounded by $2^{O(\|\mathcal{K}\|) + 2^{\text{poly}(\|\Phi\|)}}$. This means that the total number of recursion steps is $2^{\|\mathcal{K}\| \cdot 2^{\text{poly}(\|\Phi\|)}}$ and so is the overall complexity of the recursive algorithm for the $(1, 1)$-model problem.

5 Incrementing the Level of Queries

The main goal of this section is to solve the (ℓ, ℓ)-model problem by reduction to multiple instances of the $(\ell, \ell+1)$-model problem for $\ell \leq n$. The $(n+1, n+1)$-model problem is discussed briefly at the end of the section.

As a first step, we observe that it is enough to consider interpretations whose DAG of strongly connected components is a tree. For this purpose we define *tree-like* interpretations as those that can be decomposed into multiple finite subinterpretations, called *bags*, arranged into a (possibly infinite) tree such that: (1) all bags are pairwise disjoint; (2) between each parent and child bag there is a single edge, pointing from an element of the parent bag to an element of the child bag; (3) all other edges are between elements of the same bag. We think of edges between bags as 2-element interpretations, called *edge-bags*, sharing the origin with the parent bag and the target with the child bag. A tree-like interpretation is then a union of all its bags and edge-bags. Fig. 2 shows a tree-like interpretation with 4 bags and 3 edge-bags. In tree-like interpretations ℓ-consistency is a local property.

Lemma 7. *A tree-like interpretation is ℓ-consistent iff each of its bags and edge-bags is ℓ-consistent.*

Lemma 8. *There is a finite (ℓ, ℓ')-model of $\mathcal{K}$ modulo $\mathcal{E}$ iff there is a finite tree-like (ℓ, ℓ')-model of $\mathcal{K}$ modulo $\mathcal{E}$ whose bags are strongly connected.*

The next step is to eliminate the lowest level from the queries. An *$(\ell + 1)$-reduct* of a level-ℓ CRPQ φ is any CRPQ that can be obtained from φ by first splitting each RPQ atom $\mathcal{B}_{q_1,q_2}(x_1, x_2)$ of begin level $\ell_1 > \ell$ and end level ℓ into $\mathcal{B}_{q_1,q'_1}(x_1, x'_1) \wedge C_{q'_1,\ell'_1}(x'_1) \wedge r(x'_1, x'_2) \wedge C_{q'_2,\ell}(x'_2) \wedge \mathcal{B}_{q'_2,q_2}(x'_2, x_2)$ where $\ell_1 \geq \ell'_1 \geq \ell + 1$, and then dropping from the resulting CRPQ all atoms whose begin and end level is ℓ (all unary atoms are kept). Note that each $(\ell+1)$-reduct φ' of φ is a conjunction of at most $|\varphi|$ disjoint fragments of φ and that $var(\varphi) \subseteq var(\varphi')$.

Lemma 9. *Over $\widehat{\mathcal{B}}$-decorated interpretations, each level-ℓ CRPQ implies the union of its $(\ell+1)$-reducts. Over strongly-connected level-ℓ interpretations, each level-ℓ CRPQ is equivalent to the union of its $(\ell + 1)$-reducts.*

Because Lemma 8 guarantees tree-like solutions with strongly connected bags, we can replace ℓ-consistency with *strong ℓ-consistency*: the only difference is that π ranges over matches of all possible $(\ell+1)$-reducts of φ', rather than over matches of φ' itself. We restate Lemma 8 as follows.

Lemma 10. *There is a finite (ℓ,ℓ)-model of $\mathcal{K}$ modulo $\mathcal{E}$ iff there is a finite tree-like level-ℓ model of $\mathcal{K}$ modulo $\mathcal{E}$ whose edge bags are ℓ-consistent and bags strongly ℓ-consistent.*

It remains to show how to find models of the latter form. Let us first see how to find one consisting of a single bag; that is, how to find a finite strongly ℓ-consistent level-ℓ model of $\mathcal{K}$ modulo $\mathcal{E}$. We will show that this amounts to finding a finite $(\ell,\ell+1)$-model of $\mathcal{K}$ modulo $\mathcal{E}'$ for one of the $(\ell+1)$-reducts $\mathcal{E}'$ of $\mathcal{E}$ described below.

Consider a fragment φ, a non-empty set $V \subseteq \mathit{var}(\varphi)$, a partition of φ into a CRPQ φ' of level ℓ, and fragments $\varphi_1,\varphi_2,\ldots,\varphi_k$, as in the definition of (strong) ℓ-consistency. Let $\kappa : \mathit{var}(\varphi) \to \{1,2,\ldots,\ell\}$ be such that $\kappa\big(\mathit{var}(\varphi)\cap \mathit{var}(\varphi')\big) = \{\ell\}$. Let ψ' be an $(\ell+1)$-reduct of φ'. Consider CRPQs ψ with $\mathit{var}(\varphi) \subseteq \mathit{var}(\psi)$ that can be partitioned into ψ' and $\varphi_1,\varphi_2,\ldots,\varphi_k$. Choose the one with minimal $\mathit{var}(\psi)$. This amounts to merging back all RPQ atoms split during the partition of φ, provided that their segments were not affected by replacing φ' with ψ'. The CRPQ ψ is not a fragment, because it need not be connected: Figure 2 right illustrates passing from φ to ψ consisting of two disconnected fragments. Let $\psi_1,\psi_2,\ldots,\psi_m$ be the fragments constituting ψ and let $U_i = V \cap \mathit{var}(\psi_i)$. An $(\ell+1)$-*reduct* $\mathcal{E}'$ of $\mathcal{E}$ is constructed by iterating over all possible choices of φ, V, φ', $\varphi_1,\varphi_2,\ldots,\varphi_k$, ψ', κ, as above, and pruning $\mathcal{E}$ for each choice in one of the following ways:

- either pick i such that $U_i = \emptyset$ and remove all unary types that contain any $A^{\kappa_i}_{\psi_i,W_i}$ with $W_i \subseteq \mathit{var}(\psi_i)\cap \mathit{var}(\psi')$, $\kappa_i\big(\mathit{var}(\psi_i)\cap \mathit{var}(\psi')\big) = \{\ell+1\}$, and $\kappa_i(x) = \kappa(x)$ for all $x \in \mathit{var}(\psi_i)\setminus \mathit{var}(\psi')$;
- or remove all unary types that contain some $A^{\kappa_i}_{\psi_i,U_i}$ with $\kappa_i\big(\mathit{var}(\psi_i)\cap \mathit{var}(\psi')\big) = \{\ell+1\}$ and $\kappa(x) = \kappa_i(x)$ for all $x \in \mathit{var}(\psi_i)\setminus \mathit{var}(\psi')$, for each i such that $U_i \neq \emptyset$, but do not contain $A^{\kappa}_{\varphi,V}$.

Lemma 11. *$\mathcal{I}$ is a strongly ℓ-consistent level-ℓ model of $\mathcal{K}$ modulo $\mathcal{E}$ iff some interpretation that agrees with $\mathcal{I}$ over all role names and all concept names except $\mathsf{CN}^{\Phi}_{\ell+1}$ is an $(\ell+1)$-consistent level-ℓ model of $\mathcal{K}$ modulo $\mathcal{E}'$ for some $(\ell+1)$-reduct $\mathcal{E}'$ of $\mathcal{E}$.*

Finite models consisting of multiple bags can be constructed bottom-up by a least fixed point procedure, using Lemma 11 to find each bag.

Lemma 12. *The (ℓ,ℓ)-model problem for an $\mathcal{ALC}$ KB $\mathcal{K}$, a UCRPQ Φ, and an environment $\mathcal{E}$ can be solved in time*

$$2^{O(\|\mathcal{K}\|)+2^{\mathrm{poly}(\|\Phi\|)}}$$

given an oracle for the $(\ell,\ell+1)$-model problem (with the same UCRPQ and TBox).

At the bottom of the recursion we need to check if there exists a $(n+1,n+1)$-model for $\mathcal{K}$ modulo $\mathcal{E}$. Now, a $\widehat{\mathcal{B}}$-decorated interpretation is level-$(n+1)$ iff it is *discrete*; that is, it has no edges at all. This allows solving the problem by a direct inspection. Because the ABox is trivial and ℓ-consistency is preserved under restrictions of the domain, it is enough to go through all singleton interpretations.

Lemma 13. *The $(n+1,n+1)$-model problem for an $\mathcal{ALC}$ KB $\mathcal{K}$, a UCRPQ Φ, and an environment $\mathcal{E}$ can be solved in time $2^{O(\|\mathcal{K}\|)+2^{\mathrm{poly}(\|\Phi\|)}}$.*

6 Incrementing the Level of Models

In this section we solve the (ℓ,ℓ')-model problem by reduction to multiple instances of the (ℓ',ℓ')-model problem for $\ell<\ell'$; that is, we eliminate level-ℓ edges from the interpretations. Like in Section 5, we rely on tree-like models of a special form; this time, however, they may be infinite and an additional step is needed to turn them into finite ones.

A $\widehat{\mathcal{B}}$-decorated interpretation is *ℓ'-flat* if it is a tree-like interpretation where all edges between bags have level strictly below ℓ', whereas all edges inside bags have level at least ℓ'.

Lemma 14. *If there exists a finite (ℓ,ℓ')-model of $\mathcal{K}$ modulo $\mathcal{E}$ then there exists an ℓ'-flat (ℓ,ℓ')-model of $\mathcal{K}$ modulo $\mathcal{E}$ with bounded degree and bag size.*

In contrast to Lemma 8, the above only shows that the reformulated condition is necessary. We show that it is sufficient, by turning an arbitrary ℓ'-flat (ℓ,ℓ')-model of $\mathcal{K}$ modulo $\mathcal{E}$ with bounded degree and bag size into a finite (ℓ,ℓ')-model of $\mathcal{K}$ modulo $\mathcal{E}$. For this we use *coloured blocking*. For $d \in \Delta^{\mathcal{I}}$, the *$m$-neighbourhood $N^{\mathcal{I}}_m(d)$ of d* is the interpretation obtained by restricting $\mathcal{I}$ to elements $e \in \Delta^{\mathcal{I}}$ within distance m from d in $\mathcal{I}$, enriched with a fresh concept interpreted as $\{d\}$. A *colouring of $\mathcal{I}$ with k colours* is an extension $\mathcal{I}'$ of $\mathcal{I}$ to k fresh concept names $B_1,\ldots,B_k$ such that $B_1^{\mathcal{I}'},\ldots,B_k^{\mathcal{I}'}$ is a partition of $\Delta^{\mathcal{I}'} = \Delta^{\mathcal{I}}$. We say that $d \in B_i^{\mathcal{I}'}$ has colour B_i. We call $\mathcal{I}'$ *m-proper* if for each $d \in \Delta^{\mathcal{I}'}$ all elements of $N^{\mathcal{I}'}_m(d)$ have different colours.

Fact 1 (Gogacz, Ibáñez-García, and Murlak 2018)**.** *If $\mathcal{I}$ has bounded degree, then for all $m \geq 0$ there exists an m-proper colouring $\mathcal{I}'$ of $\mathcal{I}$ with finitely many colours. Consider interpretation $\mathcal{J}$ obtained from $\mathcal{I}'$ by redirecting some edges such that the old target and the new target have isomorphic m-neighbourhoods in $\mathcal{I}'$. Then, for each conjunctive query φ with at most $\sqrt{m}$ binary atoms, if $\mathcal{I} \not\models \varphi$, then $\mathcal{J} \not\models \varphi$.*

Let $\mathcal{I}$ be an ℓ'-flat (ℓ,ℓ')-model of $\mathcal{K}$ modulo $\mathcal{E}$ of bounded degree with bags of size at most M. In order to make Fact 1 applicable, we need to express the ℓ'-consistency condition over $\mathcal{I}$ by means of a finite set of conjunctive queries, rather than CRPQs. Towards this end, we show that over $\mathcal{I}$ each level-ℓ' CRPQ is equivalent to a UCQ. We rely on the following observation.

Lemma 15. *In a match of a $\widehat{\mathcal{B}}$-decorated CRPQ in a $\widehat{\mathcal{B}}$-decorated interpretation, each path witnessing an RPQ atom of end level at least ℓ' uses at most $n-\ell'$ edges of level strictly below ℓ'.*

We say that a CRPQ φ is *bounded by* K over an interpretation $\mathcal{J}$ if for each match of φ in $\mathcal{J}$ each RPQ atom of φ can be witnessed by a path of length at most K.

Lemma 16. *Let $\mathcal{J}$ be $\widehat{\mathcal{B}}$-decorated interpretation made up of disjoint level-ℓ' interpretations of size at most M connected by edges of level strictly below ℓ'. Assuming $\ell' \leq n$, each level-ℓ' CRPQ is bounded by $M(n-\ell'+1)^2$ over $\mathcal{J}$.*

For a $\widehat{\mathcal{B}}$-decorated CRPQ φ, let $\varphi^{(K)}$ be the UCQ obtained by taking the union of all CQs that can be obtained from φ by eliminating each RPQ atom $\mathcal{B}_{q,q'}(x,x')$ in one of the following ways: either remove the atom and equate variables x and x', or replace the atom with a CQ of the form

$$r_1(x,y_1) \wedge r_2(y_1,y_2) \wedge \cdots \wedge r_N(y_{N-1},x')$$

where $N \leq K$, $y_1,\ldots,y_{N-1}$ are fresh variables, and there is a run of $\mathcal{B}$ on $r_1 \ldots r_N$ that begins in q and ends in q'.

Fact 2. *If a $\widehat{\mathcal{B}}$-decorated CRPQ φ is bounded by K on an interpretation $\mathcal{J}$, then $\mathcal{J} \models \varphi$ iff $\mathcal{J} \models \varphi^{(K)}$.*

The final step before we can apply Fact 1 is to express ℓ'-consistency as query evaluation. Consider a partition of a fragment φ of Φ into a CRPQ φ' of level ℓ' and fragments $\varphi_1,\varphi_2,\ldots,\varphi_k$ with $\mathit{var}(\varphi_i)\cap \mathit{var}(\varphi_j)=\emptyset$ for $i\neq j$, $V_i = \mathit{var}(\varphi_i)\cap\mathit{var}(\varphi')$, and $\emptyset \neq V \subseteq \mathit{var}(\varphi)\cap\mathit{var}(\varphi')$. Let ψ be the CRPQ obtained from φ' as follows. Begin from a copy of φ'. For each $i \in \{1,\ldots,k\}$, add to ψ an atom $A^{\kappa_i}_{\varphi_i,V_i}(u)$ for some κ_i satisfying $\kappa_i(x) \leq \ell$ for all $x \in \mathit{var}(\varphi_i)$ and $\kappa_i(x)=\ell$ for all $x\in V_i$, and some variable u in V_i (V_i is nonempty, because φ is connected), and equate all variables in V_i. Similarly, add to ψ the atom $\bar{A}_{\varphi',V}(u)$ for some κ satisfying $\kappa(x)=\ell$ for all $x\in \mathit{var}(\varphi)\cap\mathit{var}(\varphi')$ and $\kappa(x)=\kappa_i(x)$ for all $x\in\mathit{var}(\varphi)\cap\mathit{var}(\varphi_i)$, and some $u\in V$, and equate all variables in V. Let $\Phi_{\ell'}$ be the union of all CRPQs ψ obtained as above for different choices of φ, φ', $\varphi_1,\varphi_2,\ldots,\varphi_k$, V, and $\kappa_1,\kappa_2,\ldots,\kappa_k$. Note that $\Phi_{\ell'}$ is a union of level-ℓ' CRPQs. If $\ell' > n$, $\Phi_{\ell'}$ is a UCQ.

Lemma 17. *If $\mathcal{J}$ is a $\widehat{\mathcal{B}}$-decorated interpretation, then $\mathcal{J}$ is ℓ'-consistent iff $\mathcal{J} \not\models \Phi_{\ell'}$.*

Let $K = M(n-\ell'+1)^2$. Let t be the maximal number of binary atoms in one CQ in $\Phi^{(K)}_{\ell'}$. (Note that if $\ell' > n$, the query $\Phi_{\ell'}$ is a UCQ and $\Phi^{(K)}_{\ell'}$ coincides with $\Phi_{\ell'}$.) Fix $m = t^2$ and let $\mathcal{I}'$ be an m-proper colouring of $\mathcal{I}$. On each infinite branch, select the first bag $\mathcal{M}$ such that for some bag $\mathcal{M}'$ higher on this branch, the m-neighbourhood of the target element e of the edge from the parent of $\mathcal{M}$ to $\mathcal{M}$ is isomorphic to the m-neighbourhood of the target e' of the edge from the parent of $\mathcal{M}'$ to $\mathcal{M}'$. Because the number of non-isomorphic m-neighbourhoods in a structure of bounded degree is bounded, the depth of the selected bags in the tree of bags is also bounded. The set of selected bags is finite and forms a maximal antichain. Let $\mathcal{F}$ be the interpretation obtained by taking the union of all strict ancestors of the selected bags, and for each element e as above, redirect the edge coming from the parent of $\mathcal{M}$ to e'.

Clearly, $\mathcal{F}$ is a finite level-ℓ interpretation. It is routine to check that $\mathcal{F} \models^{\mathcal{E}} \mathcal{K}$. It remains to prove that $\mathcal{F}$ is ℓ'-consistent. We know that $\mathcal{I}$ is ℓ'-consistent. By Lemma 17, $\mathcal{I} \not\models \Phi_{\ell'}$. By Lemma 16 and Fact 2, $\mathcal{I} \not\models \Phi^{(K)}_{\ell'}$. By Fact 1, $\mathcal{F} \not\models \Phi^{(K)}_{\ell'}$. By construction, $\mathcal{F}$ satisfies the assumptions of Lemma 16. Hence, by Lemma 16 and Fact 2, $\mathcal{F} \not\models \Phi_{\ell'}$. We conclude that $\mathcal{F}$ is ℓ'-consistent using Lemma 17.

Thus we have proved the converse of Lemma 14.

Lemma 18. *If there exists an ℓ'-flat (ℓ,ℓ')-model of $\mathcal{K}$ modulo $\mathcal{E}$ with bounded degree and bag size then there exists a finite (ℓ,ℓ')-model of $\mathcal{K}$ modulo $\mathcal{E}$.*

Combining Lemmas 6, 14, and 18, we get that there is a finite (ℓ,ℓ')-model of $\mathcal{K}$ modulo $\mathcal{E}$ iff there is a bounded-degree ℓ'-flat model of $\mathcal{K}$ modulo $\mathcal{E}$ whose bags are ℓ'-consistent and have bounded size. As in an ℓ'-flat model each bag is a level-ℓ' interpretation, when we restrict our search to one-bag models the problem is an instance of the (ℓ',ℓ')-model problem. Models consisting of multiple bags can be built coinductively top-down by means of a greatest fixed point algorithm (similar to type elimination), using the (ℓ',ℓ')-model problem to check if each bag exists.

Lemma 19. *The (ℓ,ℓ')-model problem for an $\mathcal{ALC}$ KB $\mathcal{K}$, a UCRPQ Φ, and an environment $\mathcal{E}$ can be solved in time*

$$2^{O(\|\mathcal{K}\|)+2^{\mathrm{poly}(\|\Phi\|)}}$$

given an oracle for the (ℓ',ℓ')-model problem (with the same UCRPQ and TBox).

7 Looking Forward (and Back)

This paper provides first positive results on finite entailment of navigational queries over DLs ontologies. The main technical contribution is an optimal automata-based 2ExpTime upper bound for finite entailment of UCRPQs in $\mathcal{ALC}$.

Let us take a look back at our journey. We devised an expansion of the semiautomaton used to represent UCRPQs to keep track of its runs that begin in all possible states, on all infixes of the input word. By making interpretations and CRPQs knowledgeable of the runs of this expansion, we are able to associate levels to them as dictated by the transitions of the expansion. To solve the entailment problem, we use a recursive method eliminating the lowest level from the query and from the interpretation, and solving then the simpler problem. In particular, we look at problem of finding (ℓ,ℓ') models, and solve it by recursively increasing ℓ and ℓ' in an alternating way, until both reach a maximum level: Section 5 and 6 respectively address the increment of the query level and of the model. We finally showed what to do when $\ell = \ell' = n+1$, which as argued, is enough to solve the original finite entailment problem.

As for future work, the first immediate step is to extend our method to deal with test atoms of the form $A?$, which are usually available in UCRPQs. For the ontology language, we believe our method can be adapted to allow inverses, nominals or counting. Regarding more expressive query languages, the natural next step is to consider *two-way* CRPQs. Our current approach relies on the fact that information only flows forward, and it is not clear whether it can be adapted to deal with queries that can go back.

Acknowledgments

This work was supported by Poland's National Science Centre grant 2018/30/E/ST6/00042. It also benefited from inspiring discussions with Charles Paperman.

References

Amarilli, A., and Benedikt, M. 2020. Finite open-world query answering with number restrictions. *ACM Trans. Comput. Log.* 21(4):27:1–27:73.

Amendola, G.; Leone, N.; and Manna, M. 2018. Finite controllability of conjunctive query answering with existential rules: Two steps forward. In *IJCAI*, 5189–5193. ijcai.org.

Baader, F.; Horrocks, I.; Lutz, C.; and Sattler, U. 2017. *An Introduction to Description Logic*. Cambridge University Press.

Baget, J.-F.; Leclère, M.; Mugnier, M.-L.; and Salvat, E. 2011. On rules with existential variables: Walking the decidability line. *Artificial Intelligence* 175(9):1620–1654.

Bárány, V.; Gottlob, G.; and Otto, M. 2014. Querying the guarded fragment. *Log. Methods Comput. Sci.* 10(2).

Bednarczyk, B., and Kieroński, E. 2022. Finite entailment of local queries in the $\mathcal{Z}$ family of description logics. In *Proc. of AAAI-2022*.

Bednarczyk, B., and Rudolph, S. 2019. Worst-case optimal querying of very expressive description logics with path expressions and succinct counting. In *IJCAI*, 1530–1536. ijcai.org.

Bienvenu, M.; Calvanese, D.; Ortiz, M.; and Simkus, M. 2014. Nested regular path queries in description logics. In *KR*.

Bienvenu, M.; Ortiz, M.; and Simkus, M. 2015. Regular path queries in lightweight description logics: Complexity and algorithms. *J. Artif. Intell. Res.* 53:315–374.

Bojańczyk, M., and Parys, P. 2011. XPath evaluation in linear time. *J. ACM* 58(4):17:1–17:33.

Calvanese, D.; Giacomo, G. D.; Lenzerini, M.; and Vardi, M. Y. 2000. Containment of conjunctive regular path queries with inverse. In *KR*, 176–185. Morgan Kaufmann.

Calvanese, D.; Eiter, T.; and Ortiz, M. 2014. Answering regular path queries in expressive description logics via alternating tree-automata. *Inf. Comput.* 237:12–55.

Chen, C.; Huang, H.; Ross, K. E.; Cowart, J. E.; Arighi, C. N.; Wu, C. H.; and Natale, D. A. 2020. Protein ontology on the semantic web for knowledge discovery. *Scientific Data* 7(1):337.

Civili, C., and Rosati, R. 2012. A broad class of first-order rewritable tuple-generating dependencies. In *Datalog*, volume 7494 of *Lecture Notes in Computer Science*, 68–80. Springer.

Cook, C. E.; Bergman, M. T.; Finn, R. D.; Cochrane, G.; Birney, E.; and Apweiler, R. 2016. The European Bioinformatics Institute in 2016: Data growth and integration. *Nucleic acids research* 44:D20–D26.

Danielski, D., and Kieronski, E. 2019. Finite satisfiability of unary negation fragment with transitivity. In *MFCS*, volume 138 of *LIPIcs*, 17:1–17:15.

Figueira, D.; Figueira, S.; and Baque, E. P. 2020. Finite controllability for ontology-mediated query answering of CRPQ. In *KR*, 381–391.

Florescu, D.; Levy, A. Y.; and Suciu, D. 1998. Query containment for conjunctive queries with regular expressions. In *PODS*, 139–148. ACM Press.

Galgonek, J.; Hurt, T.; Michlíková, V.; Onderka, P.; Schwarz, J.; and Vondrášek, J. 2016. Advanced SPARQL querying in small molecule databases. *Journal of Cheminformatics* 8(1):31.

Ginzburg, A. 1968. *Algebraic Theory of Automata*. ACM Monograph Series. Academic Press.

Glimm, B.; Lutz, C.; Horrocks, I.; and Sattler, U. 2008. Conjunctive query answering for the description logic SHIQ. *J. Artif. Intell. Res.* 31:157–204.

Gogacz, T., and Marcinkowski, J. 2013. Converging to the chase - A tool for finite controllability. In *LICS*, 540–549. IEEE Computer Society.

Gogacz, T.; Gutiérrez-Basulto, V.; Ibáñez-García, Y.; Jung, J. C.; and Murlak, F. 2019. On finite and unrestricted query entailment beyond SQ with number restrictions on transitive roles. In *IJCAI*, 1719–1725. ijcai.org.

Gogacz, T.; Gutiérrez-Basulto, V.; Gutowski, A.; Ibáñez-García, Y.; and Murlak, F. 2020. On finite entailment of non-local queries in description logics. In *KR*, 424–433.

Gogacz, T.; Ibáñez-García, Y.; and Murlak, F. 2018. Finite query answering in expressive description logics with transitive roles. In *Proc. of KR-18*.

Gottlob, G.; Manna, M.; and Pieris, A. 2018. Finite model reasoning in hybrid classes of existential rules. In *IJCAI*, 1831–1837. ijcai.org.

Gutiérrez-Basulto, V.; Ibáñez-García, Y. A.; and Jung, J. C. 2018. Answering regular path queries over SQ ontologies. In *AAAI*, 1845–1852. AAAI Press.

Hu, W.; Qiu, H.; and Dumontier, M. 2015. Link analysis of life science linked data. In *ISWC (2)*.

Ibáñez-García, Y. A.; Lutz, C.; and Schneider, T. 2014. Finite model reasoning in Horn description logics. In *KR*.

Kostylev, E. V.; Reutter, J. L.; and Vrgoc, D. 2014. XPath for DL-Lite ontologies. In *Description Logics*.

Lysenko, A.; Roznovăţ, I. A.; Saqi, M.; Mazein, A.; Rawlings, C. J.; and Auffray, C. 2016. Representing and querying disease networks using graph databases. *BioData Mining* 9(1):23.

Ortiz, M., and Simkus, M. 2014. Revisiting the hardness of query answering in expressive description logics. In *RR*, volume 8741 of *Lecture Notes in Computer Science*, 216–223. Springer.

Pratt-Hartmann, I. 2009. Data-complexity of the two-variable fragment with counting quantifiers. *Inf. Comput.* 207(8):867–888.

Rajabi, E., and Sanchez-Alonso, S. 2021. Knowledge discovery using SPARQL property path: The case of disease data set. *Journal of Information Science* 47(5):677–687.

Rosati, R. 2008. Finite model reasoning in DL-Lite. In *ESWC*.

Rudolph, S. 2016. Undecidability results for database-inspired reasoning problems in very expressive description logics. In *KR*, 247–257. AAAI Press.

Schneider, T., and Simkus, M. 2020. Ontologies and data management: A brief survey. *Künstliche Intell.* 34(3):329–353.

Stefanoni, G.; Motik, B.; Krötzsch, M.; and Rudolph, S. 2014. The complexity of answering conjunctive and navigational queries over OWL 2 EL knowledge bases. *J. Artif. Intell. Res.* 51:645–705.